

Methods for full protein sequence mapping

I. Sample preparation

a. Reduction/alkylation with DTT/Iodoacetamide; Protein samples need to be denatured prior to any digestion with proteases so the protease will have as much access to targeted amino acids as possible. To help achieve this, the protein is treated first with dithiothreitol (DTT) to break disulfide linkages between cysteine residues. To prevent the disulfide bonds from re-forming, iodoacetamide is used to modify the reactive cysteine -SH groups, forming s-carboxymethylated cysteines (see example <http://www.ionsource.com/Card/cmc/method.htm>)

b. Protease digestion; Endoproteinase Lys-C (Lys-C) and trypsin are proteases used for digesting proteins into a population of peptides that can be identified by the mass spectrometer. Lys-C cleaves on the c-terminal side of lysine and the resulting peptides are larger than tryptic peptides. The larger-sized peptides are compatible with the LTQ-FTMS and new ETD mass spectrometry technologies discussed below. Trypsin cleaves on the c-terminal side of lysine and arginine amino acids. Both proteases work best around pH 8. For detailed information and digestion protocols, see <http://www.roche-applied-science.com/pack-insert/1420429a.pdf> (Lys-C) or <http://www.promega.com/tbs/9piv511/9piv511.pdf> (trypsin).

Since maximal amino acid coverage of the protein is required, it is best to digest the sample with several proteases, so the resulting peptides are more likely to contain amino acid information from the entirety of the protein. Glu-C cleaves on the c-terminal side of glutamic acid and, less frequently, aspartic acid residues. Chymotrypsin specifically cleaves on the c-terminal side of tyrosine, phenylalanine and tryptophan amino acids. Less frequently, it will act on leucine, methionine, alanine, aspartic acid and glutamic acid residues. For detailed information and digestion protocols, see (Glu-C) <http://www.roche-applied-science.com/pack-insert/1420399a.pdf> and (chymotrypsin) <http://www.roche-applied-science.com/pack-insert/1418467a.pdf>. The selection of proteases depends greatly upon the amino acid sequence of the target protein. Proteases are selected which will yield peptides from 600-6000 Da and which will maximize the coverage of potential phosphorylation sites.

II. LC-MS analysis

a. Microcapillary RP-HPLC separation interfaced to electrospray ionization source Slow HPLC flow rates of ~50-100 nL/min are achieved by using "microcapillary" HPLC columns. These columns are constructed from fused silica glass capillaries (Polymicro Technologies, Phoenix, AZ), typically 75 microns inner diameter. A segment (~15cm) of fused silica is cut and a segment of polyimide coating is burned off using a small flame. A bottleneck is shaped into the glass using a laser puller and the reversed-phase packing material is loaded behind the bottleneck. The C-18 (YMC Corporation) packing material is loaded using a pressurized stainless steel bomb and Helium gas at about 500 psi to force the slurry into the capillary tube. The electrospray tip is pulled using the laser puller which forms a tip ~ 5-10 microns inner diameter.

An aliquot of proteolytic digest is loaded onto the microcapillary HPLC column, again using the pressurized stainless steel bomb. Typical sample volumes range from 1-40 μ L. Larger sample volumes (> 5 μ L) are loaded onto a “pre-column”, which is constructed from 360 x 75 μ m fused silica and 5-20 μ m C18 packing material with a licrosorb frit. The pre-column is then connected to the microcapillary column using a Teflon sleeve. The microcapillary column is connected to the HPLC solvent line. When the sample analysis begins, a gradient of increasingly organic solvent runs across the microcapillary HPLC column. At the same time, a high voltage (1.5-1.8kV) from the mass spectrometer is applied at a stainless steel union on the HPLC waste line. The result is a plume of charged droplet molecules emitting from the electrospray tip of the column at a rate of ~ 50nL/min. A typical HPLC gradient for a highly complex mixture of peptides is 0-80% organic solvent over a period of about 2 hours. The low flow rate and shallow gradient help minimize the number of peptides being analyzed by the mass spectrometer at any given time, effectively decreasing the complexity of the sample and maximizing the coverage of peptides selected for MS/MS analysis.

b. Improved Instrumentation: LTQ-FTMS We use ion trap mass spectrometry for our analyses. Recently, a Finnigan LTQ-FTMS (ThermoElectron Corp., San Jose, CA) instrument was added to the repertoire of ion trap mass spectrometers and is the primary instrument used for the full protein sequence mapping experiments. It is a hybrid instrument that combines linear ion trap and Fourier Transform ion cyclotron resonance mass analysis technologies (Syka, Marto, et al., 2004). Advantages over traditional 3D ion trap technology include: sub-femtomole limit of detection, ~50-fold increase in dynamic range, high mass resolution (100,000 at m/z 400), accurate mass determination (1-3 ppm with external calibration), and increased scan speed (5 scans per sec versus 1 scan per 2 sec). All of these features dramatically increase the amount of useful information that can be obtained from analyses of complex mixtures such as binding partner pull-downs. The sensitivity, resolution and high mass accuracy of the LTQ-FTMS enables the user to more quickly sift through MS/MS data associated with single protein pull-downs and assign sequence and post-translational modification information to spectra.

c. MS/MS analysis of peptide ions; Software provided with the mass spectrometer enables the acquisition of mass spectra in a non-redundant manner so that as many peptides as possible are surveyed from a complex mixture. This type of data acquisition is known as "data-dependent MS/MS analysis". During this process, an experimental method is set up in which one full range MS scan (300-2000 m/z) is acquired followed by the acquisition of ten MS/MS scans.

The top ten most abundant ions present above a user-specified signal threshold in the full MS spectra are chosen for subsequent collision with helium atoms. These collisions fragment the peptide, typically along the amide backbone, producing a series of fragment ions which is represented as the MS/MS spectrum. The experimental method is designed to place those ten most abundant peptide ions on a dynamic exclusion list for a user-determined interval. The cycle is repeated and another full MS scan is taken, followed by MS/MS of the ten most abundant ions (when the ten peptide ions from the

first cycle are on the exclusion list, the next ten most abundant ions are chosen for MS/MS). The cycle continues for the duration of data acquisition.

III. Peptide sequence identification using SEQUEST

SEQUEST compares experimental MS/MS spectra to theoretical MS/MS spectra from database proteins and calculates a correlation value, which can be examined in order to determine how well a particular experimental spectrum matches the corresponding theoretical spectrum ([Eng et al., 1994](#)). Since false negative hits are not uncommon, database hits obtained from SEQUEST searches are often examined by the researcher to determine the validity of the hit. This examination can be very cursory if the correlation value is high and multiple peptides hit for the same protein. Alternatively, hits can be compared to individual MS/MS spectra if correlation values are mediocre but hit on a biologically relevant protein. The extent of post-acquisition data analysis depends upon the goal of the experiment. For full protein sequence mapping, the MS/MS spectra are searched against a database of the target protein sequence. For this step, it is essential that the full amino acid sequence of the target protein is accurate. The plasmid itself, used for transfection of target DNA into host cell, should be sequenced to insure confidence in the amino acid sequence that is ultimately translated. Once the single protein database is established, a search for phosphorylation sites is done by dictating that SEQUEST search for a possible addition of 80amu to the amino acids Serine, Threonine and Tyrosine. All SEQUEST hits above scoring criteria are examined manually to validate the peptide/phosphopeptide sequence.

IV. IMAC enrichment of phosphopeptides

Immobilized Metal Affinity Column (IMAC) chromatography is used to select out phosphopeptides from the mixture of modified and unmodified peptides in the target protein sample. The protocol calls for the use of ferric chloride to bind to the metal chelating packing material of the IMAC column. Phosphorylated peptides will bind to the immobilized iron atoms when loaded onto the column and can be eluted under acidic conditions (reduction of Fe^{+3} to Fe^{+2} will bump the phosphopeptides off the column). Since acidic amino acid residues will also bind to the Fe^{+3} on the IMAC column, non-specific binding to the column can be a hindrance. To prevent this, the sample is treated with 2 M methanolic hydrochloride prior to loading onto the IMAC column. This process converts all acidic amino acid residues, as well as the C-terminus of the peptide, from carboxylic acid groups to methyl esters. The phosphate groups remain unmodified, allowing for selective binding onto the IMAC column ([Ficarro et al., 2002](#), [Ficarro et al., 2003](#)). The phosphorylated peptides that bind to the column are eluted off with ascorbic acid onto a microcapillary C-18 column. The sample is analyzed as described above and searched against a protein database using the SEQUEST search tool, taking into account the methyl ester and phosphorylation modifications.

Web Resources

General Mass Spectrometry resources

<http://i-mass.com/guide/protocol.html> - General proteomics site containing information about protocols and resources

<http://us.expasy.org/> - Proteomics tools and resources

<http://prospector.ucsf.edu/> - Mass spec data analysis

http://depts.washington.edu/~yeastrc/ms_home.htm - General tandem mass spec information

Mass Spectrometry Protocols and Procedures

<http://www.ionsource.com/Card/ester/esters.htm> - Making methyl esters

<http://www.ionsource.com/Card/cmc/why.htm> - Reduction of proteins

<http://www.ionsource.com/Card/cmc/intro.htm> - Reduction/alkylation of proteins

<http://prowl.rockefeller.edu/recipes/sproteas.htm> - Proteolysis

Text References

Eng JK, McCormack AL, Yates JR III. 1994. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **5**:976. PubMed

Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. 2002. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature Biotechnology.* **20**(3): 301. PubMed