ProtTech™

Two different Methods in Protein Identification by Mass Spectrometry

There are tow major methods that are widely used for protein identification by mass spectrometry: MALDI-TOF based protein fingerprinting and LC-MS/MS based peptide sequencing. In the MALDI-TOF based protein fingerprinting method, a sample is digested with certain proteolytic enzyme (usually trypsin) and one MS spectrum is acquired which generates the massed of all peptides (or MH+), and these masses are used as a fingerprint to search proteins in a database. The hits are ranked according certain scoring method. A candidate protein that contains more proteolytic peptides, which can match measured masses, has a higher score. A Protein identified often is the protein of highest score. In contrast, protein identification by LC-MS/MS peptide sequencing is following different principle: typically, the peptide mixture from a proteolytic digestion (usually trypsin digestion) is separated a HPLC (often with 75 uM ID column with column flow rate less than 1ul/min). A Tandem mass spectrometer is on-line coupled with HPLC so peptides elute from HPLC column is fragmented by a process called collision-induced dissociation (CID) and a MS/MS spectra is acquired for each fragmented peptide (there are often several thousands MS/MS spectra from each sample). Each MS/MS spectrum (corresponding a specific peptide sequence) is used to search protein database for matched peptides. A Protein hit is identified often by multiple independently sequenced peptides from the same protein.
Protein identification based on MALDI-TOF MS fingerprinting has some significant advantages over Nano-LC-MS/MS based method. The first advantage is its speed: Acquiring a MALDI-TOF MS data often takes less than one minutes. Subsequent database search often takes less than three minutes, and analysis of search result is often rapid and straightforward. In contrast, Nano LC-MS/MS data acquisition typically takes 1-2 hours with several thousands MS/MS spectra generated, and subsequent database search may take several hours on a PC, and analyzing search result could take even longer time. The Second advantage is its easiness: operating a MALDI-TOF mass spectrometer is a much easier task than operating a Nano LC-MS/MS system.

However, there are two major factors that make the Nano LC-MS/MS a superior platform for protein identification: sensitivity and reliability. This high sensitivity of Nano LC-MS/MS is achieved mainly by to three factors: first, the use of Nano LC with 75 um ID reverse phase C18 column can effectively concentrate peptides 50-200 fold before MS detection. For example, a typical peptide peak from a 75 um column has a pick width of ~5 second, and its peak volume is about 40 nl at 500 nl/minute (or 8nl/second) column flow rate. Assuming 5 ul sample is injected into HPLC, the sample has been concentrated about 125 folds. Since mass spectrometry detection is concentration-dependent, such concentration increase is effectively translated into the increase of MS signal. Consequently, it can dramatically increase real sensitivity in identifying proteins. Second, protein identification by LC-MS/MS is based on independent sequencing of peptides. It is often possible to confidently identify a protein based on MS/MS sequencing of only one peptide, in contrast to 10-20 peptides required for MALDI-TOF fingerprinting. Since number of peptides observed from a protease digestion is often related to amount of sample for a giving protein, the ability to make a positive protein identification based on much less peptides is translated into the real sensitivity gain. Third, during MALDI based MS analysis, only very small percentage (typically 1/1000 to 1/10000) of sample loaded on a MALDI plate are utilized, while almost all the sample are utilized during electrospray process in LC-MS/MS. So in the situation that concentration of analyte is the limiting factor for detection, the ability to utilize more analyte will result in higher sensitivity.

Another important advantage of Nano LC-MS in protein identification is its reliability. First, the outcome from a MALDI-TOF fingerprinting is a list of candidates each with a ranking score. In many cases scores of many protein candidates are so close that hit is picked up based on a range of constrains such as MWt, pI, and species. Although computer database search with some MALDI-TOF MS data can always generate a list of "hits", it is often very difficult evaluate the fidelity of these hits. In contrast,

identification of a protein based on one peptide sequenced by MS/MS with 70%-80% sequence coverage (a typical case) is at about 90% confident level. The confident level is about 99% if the protein identification is based on MS/MS sequencing of two independent peptides. It is usually 100% certain is three peptides are sequenced. Majority of the protein identification carried out at ProtTech is based on large than 10 peptides sequenced. Second, if a gel band contains two or more proteins, although database search will still generate a hit list, the reliability of protein identification based on MALDI-TOF fingerprinting will become a very serious issue. Unfortunately, roughly 80% of all the samples we have analyzed contain two or more proteins. In contrast, nano LC-MS/MS is able to analyze very complex mixture since each peptide is independently sequenced. Theoretically a LC-MS/MS can identify a thousand protein mixtures with the same reliability as that in identifying one protein, but such capacity is limited by one-dimensional HPLC, which is often difficult to separate a very large number of peptides.

Based on above technical consideration, ProtTech carries out all its protein identification service by using Nano LC-MS/MS technology.