Trans. Natl. Acad. Sci. Tech. Philippines 28:295–299 (2006)
ISSN 0115-8848

PROTEOMICS: WEIGHING THE EVIDENCE

Leopold L. Ilag
Associate Professor, Department of Analytical Chemistry, Stockholm University and Director, Stockholm University Proteomics Facility
Stockholm, Sweden; email: leopold.ilag@anchem.su.se

Sequencing of the human genome was completed in 2001[1], a few years ahead of schedule, one of the implicit aims of which was to identify disease markers. However, to date, we still grapple with the elusive causes of many diseases and the hunt for biomarkers is still on. This is rooted in the fundamental problem that the genome is a mere blueprint for a vastly complex network of interactions orchestrated principally by proteins. The protein complement of the genome is what we refer to as the *proteome* and the variety of disciplines aiming to survey and understand the proteome fall under the umbrella of proteomics. In this talk I wish to explore the state of the field and the various applications of mass spectrometry in this actively growing area.

The complexity of studying the proteome arises from the fact that proteins encoded by genes can undergo a number of processing events that are regulated in space and time. Although a particular organism is endowed with a single genome, there are several corresponding proteomes for each cell population at a certain time and under specific environmental influences. At the molecular level, a single gene encoding a protein yields a product that is then processed (e.g. glycosylation, phosphorylation) amplifying the variants. Considering mature forms, degradation products and processed forms, numbers could easily reach 10,000. Complex biological fluids like plasma[2] not only contain resident proteins but may be invaded by extraneous proteins through tissue leakage. This can represent all the products of the genome, which can increase the 500 or so plasma proteins to about 500,000 products. Because the plasma reflects acquired immunity, the 10, 000, 000 or so immunoglobulins present further increases the complexity we have to deal with.

Given the above numbers, it seems daunting and almost impossible to make sense of the proteome. Fortunately, the methods used to address issues in proteomics simplifies the existing variation into magnitudes more manageable because the techniques offer the possibility for efficient profiling. There is no immediate need to account for all minor differences allowing a survey of relevant protein groups the levels of which change, for example as a reflection of a disease state.

The techniques involved in proteomics include those which can be used for protein separation, identification and quantification. More advanced applications allow for structural analysis, mapping modifications, defining complex interactions and even tracking protein regulation in the cellular context. One of the most common strategies include, 1D and 2D gel electrophoresis; various types of chromatography and mass spectrometry (MS). The latter is powerful as a stand-alone technique in proteomics research but maximal effectiveness is achieved when it is coupled to the above-mentioned separation techniques.

Mass spectrometry (MS) as its name implies basically measures the "weight" of molecules. It is amazing to realise how much information one can get by merely weighing things. However, when one realises that MS offers resolution enough to distinguish a mass difference of a single proton (1 Da) and in some cases even as much as resolving a mass difference in the millidalton range then it is not surprising why this technique is so powerful.

Most of the variations we try to track as indicators of abnormal physiology can be related to a mass change. When two molecules interact or are transformed, there is a corresponding change in mass (except rearrangement reactions). When peptides or proteins are modified chemically, there would also be a change in mass seen as an increase attributed to the moiety added. Because MS allows directed fragmentation, even isomers can be distinguished based on the different fragments generated as a function of structure hence products of different masses are obtained giving insight into chemical structure.

Analytes or samples to be studied by MS can be solid, liquid or gas as starting materials as long as charged analytes can be generated from them because MS depends on gas-phase mass separation and detection of ions (charged molecules). Early methods of ionization were limited to analysis of small molecules. With the introduction of fast atom bombardment (FAB)[3] it became possible to ionize intact peptides. However, it was not until soft-ionization methods namely electrospray ionization (ESI)[4,5] and matrix-assisted laser desorption ionization (MALDI)[6] were developed, that rapid and facile applications to large biomolecules was made routinely possible. These developments in MS were so important it lead to its recognition with a Nobel Prize in Chemistry in 2002[7,8].

MALDI coupled to the time of flight (ToF) analyzer has been a workhorse for the strategy called peptide mass fingerprinting (PMF)[9]. This is based on the fact that proteins having unique sequences also often have distinct patterns of cleavage sites for a given enzyme. The cleaved protein therefore yield peptides that reflect this unique distribution (distinct sequence) allowing the matching of the experimental digest with a theoretical digest stored in databases, thus, aiding the identification of a protein. However, this is not a very high-resolution method which often fails when you are looking at related proteins or isoforms. Therefore it is still important to get actual sequence information which can be obtained through *de novo* sequencing[10].

De novo sequencing can be achieved when powerful mass analyzers for example the quadrupole time of flight (Q-ToF) is used (Figure 1). The Q-ToF is

endowed with the capability of performing tandem MS. This refers to the possibility of using the quadrupole as a mass filter allowing the isolation of ions of a desired mass/charge ratio (m/z) and subsequently subjecting these to collisions with inert gas molecules to cause fragmentation. It is because of this that structural information via MS could be obtained from peptides. The weakest bonds happen to be along the peptide bond making the reading of the sequence from the mass differences of the various fragments rather straightforward (Figure 2). Fragment sizes differ by a discernable mass difference corresponding to known amino acids from which the sequence can be pieced together. An additional feature of this strategy is that it allows detection of post-translational modifications when you detect mass differences not corresponding to any of the essential amino acids. Typical modifications may include a mass increase of 80Da for phosphorylation or addition of 16Da for oxidation etc. When the molecules exceed the size of typical peptides such as intact proteins, the energies absorbed by ions in the collision cell are not enough to cause fragmentation but only simple dissociation. Although this cannot give sequence information, it provides useful data in terms of structural organization (subunits that dissociate first may indicate surface binding or a weaker interaction) and stoichiometry [11,12]. Given a macromolecular complex, through dissociation it is possible to detect heterogeneous composition[13,14]. Finally it could

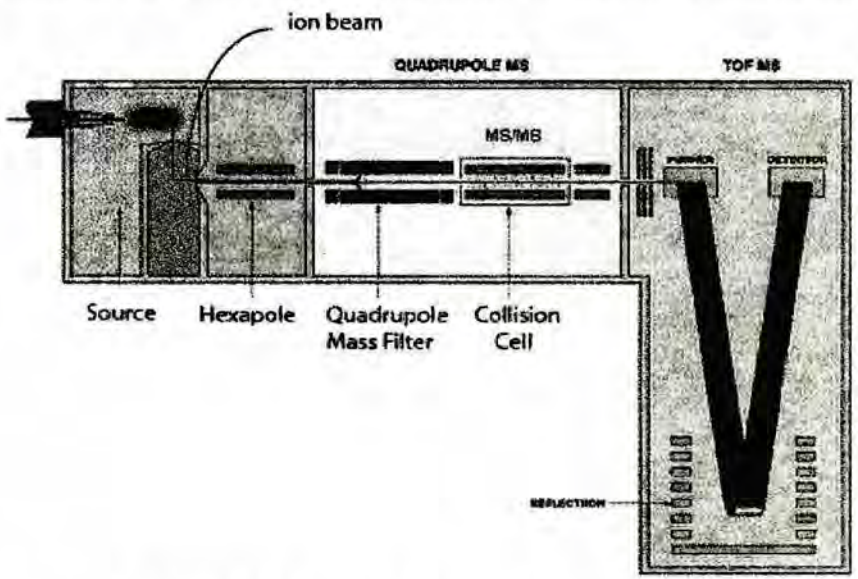

Figure 1. The Q-ToF is an integrated quadrupole and time-of-flight (ToF) mass spectrometer. The ion beam typically generated from an Electrospray or a MALDI source is guided through to the mass filter where specific ions can be isolated and subjected to fragmentation or dissociation in the collision cell (filled with inert gas usually Argon). The product ions are then analysed by the ToF analyzer.

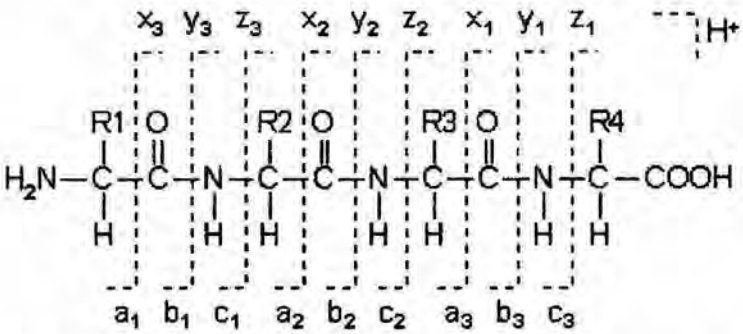

Figure 2. Most common fragments produced by MS/MS are those along the peptide backbone as shown above. To be visible in the mass spectrometer, the fragments must carry at least a single charge. Fragment ions are referred to as a,b,c or their complement x,y,z ions according to convention. The mass differences between fragments should correspond to a particular aminoacid defined by its unique R group or its modified form allowing for sequence information to be derived. (Adapted with permission from Matrix Science)
http://www.matrixscience.com/help/fragmentation_help.html

also set the stage for comparing relative binding strengths of complexes opening avenues for screening of drug binding for example[15,16]. One should bear in mind however that in any MS analysis, the interactions being studied are gas-phase interactions. Though interactions existing in solution may still be reflected in the gas phase, one should not assume that one is still looking at native solution structures or interactions.

In recent times, the application of MS has moved into the realm of MS-imaging[17]. This principally uses MALDI-ToF instrumentation wherein the laser is used to raster over a whole tissue sample (usually mounted on conductive glass). Each spectrum collected over time not only encodes the chemical entities present but also the spatial distribution of those entities over the tissue sample. Through a simple algorithm it becomes possible to display the distribution of specific biomolecules e.g. peptides (at a particular m/z) of choice thus monitoring fluctuations that may be correlated with disease or any abnormal physiology. This is an excellent approach for obtaining global difference maps without previous knowledge of what exact chemical species is of interest.

Mass spectrometry indeed is a powerful technique not limited to mere mass measurements for small molecules but could be powerfully employed for the identification, characterization and structural analysis of any biomolecule particularly proteins. By weighing molecules and their fragments we could weigh the evidences for potential biomarkers related to disease and understanding molecular and cellular networks.

References

1. Schuler, G.D. et al (1996) A gene map of the human genome. Science 274:540-6.
2.Anderson, N.L., and Anderson. N.G. (1998) Proteome and proteomics: new technologies, new concepts and new words. Electrophoresis (1998) 19(11):1853-61.
3. Greer, F.M. and Morris, H.R. (1997) Fast atom bombardment and electrospray mass spectrometry of peptides, proteins and glycoproteins. Methods Mol Biol 64:147-63.
4. Mann, M., and Wilm, M. (1995) Electrospray mass spectrometry for protein characterization. Trends Biochem Sci 20(6):219-24.
5. Fenn, J.B., Mann, M, Meng, C.K., Wong, S.F. and Whitehouse, C.M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246: 64-71.
6. Hillenkamp, F. and Karas, M. (1990) Mass spectrometry of peptides and proteins by matrix-assisted ultraviolet laser desorption/ionization. Methods Enzymol. 193(280-95).
7. Tanaka, K. (2002) The origin of macromolecule ionization by laser irradiation. Nobel Lecture. Stockholm, Sweden
8. Fenn, J.B. (2002) Electrospray wings for Molecular Elephants. Nobel Lecture, Stockholm,Sweden
9. Sommerer, N., Centeno, D. And Rossignol, M. (2007) Peptide mass fingerprinting: identification of proteins by MALDI-ToF. Methods Mol Biol 355:219-34.
10. Domon, B. and Aebersold, R. (2006) Mass spectrometry and protein analysis. Science 312: 212-7.
11. Benesch, J.L. and Robinson, C.V. (2006) Mass spectrometry of macromolecular assemblies: preservation and dissociation. Curr Opin Struct Biol 16(2):245-51.
12. Hernandez, H. et al (2006) Subunit architecture of multimeric complexes isolated directly from cells. EMBO Rep 7(6): 605-10.
13. McKay, A.R. et al (2006) J Am Chem Soc 128(35):11433-42.
14. Ilag, L.L. et al (2004) Mass spectrometry of *Escherichia coli* RNA polymerase:interactions of the core enzyme with sigma 70 and Rsd protein. Structure 12(2):269-75.
15. McCammon, M.G. and Robinson, C.V. (2004) Structural change in response to ligand binding. Curr Opin Chem Biol 8(1):60-5.
16. Ilag, L.L. et al (2004) Drug binding revealed by tandem mass spectrometry of a protein-micelle complex. J Am Chem Soc 126(44):14362-3.
17. Chaurand, P. et al. (2006) New developments in profiling and imaging of proteins from tissue sections by MALDI mass spectrometry. J Proteome Res 5(11): 2889-900.