Mass spectrometry in Genomics

MRINALINI DAMLE, ASHWINI MADGULKAR AND K. G. BOTHARA*
AISSMS COLLEGE OF PHARMACY, KENNEDY ROAD, NEAR R.T.O., PUNE-411001.

Mass spectrometry is now an indispensable tool for rapid protein and peptide structural analysis, and the widespread use of mass spectrometry is a reflection of its ability to solve structural problems that are not readily or conclusively determined with conventional techniques. High detection sensitivity and fully automated analysis of many samples in short time has paved way for current proteome research projects. The article gives overview of various mass spectrometry techniques, different methods of sample preparation along with their advantages and disadvantages. The applications based on compatible analytical techniques like liquid chromatography; electrophoresis and surface plasmon resonance coupled to mass spectrometer indicates the wide scope of mass spectrometry.

Aim of Human Genome Project is to determine complete nucleotide sequence of human DNA and to localize estimated 50 000 to 1 00 000 genes within the human genome. Once the full genome sequences are known, establishing the function of gene products is a major challenge of post genomic era. Recent advances in mass spectrometry (MS) technology together with progress in genomic sequencing and bioinformatics have dramatically changed the position of MS methods in general strategy of molecular and cell biology research. Expanding from its previous role as mining tool, for providing protein sequence information, it is becoming deeply integrated in the process of functional characterization of biologically important genes.

The ambitious role of proteomic research is comprehensive, qualitative, quantitative analysis of all proteins expressed by a genome as well as description of changes occurring at protein level under influence of biological stimuli like diseases or drug treatment. Proteomics, currently is in intense phase of progress and separation science, MS as well as bioinformatics play a major role in technological development.

Mass spectrometry:

All mass spectrometers have three essential components that are required for measuring the mass of individual molecules that have been converted to gas phase ions prior to detection. The components are an ion source, a mass analyzer and a detector. MS data is recorded as spectra, which displays ion intensity versus the m/z value. The two techniques that have become preferred methods for ionization of peptides and proteins are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), due to their effective application to a wide range of proteins and peptides. Although different combinations of ionization techniques and mass analyzers exist, MALDI usually uses a time-of-flight (TOF) tube as a mass analyzer while ESI is traditionally combined with quadrupole mass analyzers capable of tandem mass spectrometry (MS/MS). Instruments capable of MS/MS have the ability to select ions of a particular m/z ratio from a mixture of ions, to fragment selected ions by a process called collision-induced dissociation (CID) and to record the precise masses of the resulting fragment ions. If this process is applied to the analysis of the peptide ions, in principle, the amino acid sequence of the peptide can be deduced. Both ionization techniques give best results with salt and detergent free samples, although MALDI is more tolerant towards sample contaminants. Besides the improve-
ment in ionization performance, certain key events have contributed to popularize the use of ESI and MALDI.

Merits of ESI and MALDI techniques:

Micro-scale capillary reversed-phase-high performance liquid chromatography (capillary RP-HPLC) can be directly coupled to an ESI interface. Computer algorithms can correlate MS and MS/MS generated data with database information. MS software development allows better instrument control and versatility. Post-source decay (PSD), a MALDI-TOF process is developed whereby peptide sequence information can be obtained. Sample cleanup of small (1-5 ul) volumes, useful for both ESI and MALDI is possible. Development of very low ESI flow rates (10-25 nl/min), or nanoelectrospray, enables long analysis time of as little as 1 ul of sample. The long analysis time permits many experiments on the same sample, and is particularly useful for optimizing MS/MS fragmentation patterns. Liquid extraction of peptides produced by in-gel proteolytic digestion can be done.

OVERVIEW OF MS TECHNIQUES

Soft ionization techniques:

Soft ionization techniques in mass spectrometry directly yield measures of mol. wt. in the form of singly or multiply protonated or deprotonated pseudomolecular ions. Since the appearance of fast atom bombardment (FAB) in 1981, two other ionization techniques, ESI and MALDI have been developed, commercialized and now widely used for analyzing thermally labile, ionic or high molecular weight compounds. In case of FAB, L secondary ion mass spectrometry (LSIMS), ESI and MALDI, varying degrees of fragmentation is observed. MALDI usually deposits more internal energy through use of hot matrix. The actual degree of unimolecular decomposition varies with amino acid sequence in unpredictable way but use of cold matrix yields comprehensive sequence information from high-energy collisional activation. ESI is the ionization technique of choice for interfacing LC to MS. It is a sensitive technique and has ability to analyze large non-volatile chargeable molecules like proteins and nucleic acids.

Detectors:

The quadrupole ion trap MS (ITMS) is now established as a compact, cost-effective and highly sensitive detector for HPLC and capillary electrophoresis. ITMS can provide both molecular weight and structural analysis of biological macromolecules. Ion trap mass spectrometers are capable of performing CID and indeed can perform (MS)n experi-

ments on a selected ion species. Acquiring a MS/MS spectrum in an ion trap consists of selecting a single ion species by ejecting the ions of a larger m/z ratio and those present of a lower m/z ratio than the target ion. The ability to perform (MS)n experiments is particularly useful for the detailed structural analysis of components; such as peptides containing post-translationally modified amino acids.

For protein identification purposes, the triple quadrupole (TQ) MS is primarily used as a device for the generation of peptide CID spectra. This type of instrument is however extremely versatile and can perform specifically specialized experiments such as neutral loss scan, precursor ion scan and thorough in source fragmentation, MS/MS/MS. These capabilities are particularly useful for the analysis of modified peptides and proteins. A consequence of the sophisticated instrument control language, as part of the instrument in the most advanced TOF-TOF instruments, is that they are characterized by a high degree of flexibility in selecting parameters and optimizing procedures. Tandem MS generally employs quadrupoles or ion traps for initial mass analysis for ions of low kinetic energy. TOF system is optimized for ions with high kinetic energy. A method reports use of MALDI TOF/TOF to obtain high-energy CID spectra that are superior to PSD spectra. The choice of matrix and collisional gas control the nature and extent of fragmentation. Nature of MALDI CID spectra depends upon amino acid composition and peptide size.

MALDI-TOF-MS peptide mapping has become important tool for identifying proteins in sequence databases. Proteins isolated by 2D-electrophoresis are digested in situ by specific protease, mostly trypsin and the resulting peptides are mass analysed. The data are compared with the expected values computed from sequence database entries according enzyme cleavage specificity. The results are scored and the rankings suggest the protein being identified or not. Enzyme cleavage specificity, no. of detected cleavage peptides and mass accuracy are the critical parameters.

Hyphenated techniques:

To meet the growing need to simultaneously monitor all types of proteins in biological systems, new separation strategies have emerged that are amenable to hyphenation to MS. High resolution, two-dimensional gel electrophoresis (TDGE) is a powerful separation method enabling fractionation of thousands of proteins within few hours. Upon separation, the proteins are usually excised, extracted from the gel and proteolytically digested. The mixture of peptide fragments is subjected to MALDI-MS or RP-HPLC-ESI-MS.
Finally for protein identification, mass spectral data is interpreted and matched to protein and DNA databases using computer algorithms. Chemical modifications of proteins have been used as a probe of protein secondary structure or active site characterization and MS is used to analyse products of such reactions. Various separation methods like RP-LC, salt promoted adsorption chromatography, affinity chromatography, multidimensional separations, capillary electrophoresis have been reviewed.

Direct tissue profiling:

MALDI TOF MS is an excellent tool for direct peptide profiling of tissues and single cells. While obtaining molecular weight information is a critical step toward peptide characterization, alone it is not sufficient to allow unambiguous peptide identification. PSD analysis takes advantage of metastable decay of precursor ions occurring in the field free region of the flight tube and mass analysis of fragment ions using reflection mode to generate sequence information, and such fragmentation can be enhanced with CID. A study reports the ability to directly sequence the peptides from biological cells using MALDI TOF MS with PSD and CID. PSD analysis is shown to be a viable technique to localize disulfide bonds, phosphorylation, or glycosylation sites in post translationally modified peptides. In contrast to the majority of the published PSD applications involving extraction and purification of large amount of material for peptide characterization, elegant work of Hsieh et al. involves use of micro column LC to fractionate peptides from a single neuron. Direct profiling of proteins in tissue sections for several organs of mouse using MALDI-MS is reported. Unique protein profiles have been generated as a function of type of tissue and location of the sample within a given tissue in the animal.

SAMPLE PREPARATION TECHNIQUES FOR MALDI-MS ANALYSIS

MALDI technique's popularity for the characterization of biological polymers is due to high tolerance of inorganic salts and buffers. But many other agents (salts, chaotropic agents, surfactants, preservatives, solubilising agents) employed by the biologists can cause problems with the subsequent MS characterization. Such interfering agents present in most biological samples necessitate some sort of cleanup procedure prior to MS analysis. A published method reports effects of various buffers, surfactants and organic additives on molecular weight determination of peptides by MALDI-TOF-MS. They observed that S/N ratio decreases with increase in buffer concentration. Ionization of organic additive was buffer dependent. Ionization of neutral surfactants was suppressed in ammonium acetate without impacting peptide spectra. When methanol, tetrabutylammonium or polyvinyl alcohol was used with ammonium acetate or sodium phosphate, there was suppression of peptide ionization.

Several sample preparation techniques have been specifically designed for MALDI-MS analysis of contaminated biological samples. One strategy relies upon thin film of matrix deposited in acetone. The peptide sample is applied to the surface of matrix film and desalted using aqueous organic acids. Recrystallization method involves incorporation of more peptide molecules than the salt contaminants. The excluded salts are then selectively washed away from the adhering crystals. In the technique based on the use of polymeric membranes, samples are deposited onto polymeric membranes such as porous polyethylene (PE) and poly(vinylidene difluoride) (PVDF). The salts are washed away from the sample that itself binds to the membrane by hydrophobic interactions. It has been demonstrated that the MALDI target can be covalently modified so that it behaves like a medium for solid phase extraction. These modified targets are capable of binding the molecules of interest but not contaminants when the sample solutions are placed on them. The target can simply be washed to remove the contaminants prior to MALDI-MS analysis. The technique of adsorption on micro beads involves concentration of proteins from aqueous solution by adsorption on RP polymeric micro beads. The beads are then washed extensively to remove contaminants and then bound proteins are digested with Trypsin. Analysis is performed by MALDI-TOF-MS by direct deposition of beads on MALDI target along with the matrix solution. For membrane proteins, which are generally insoluble in aqueous buffers, MALDI unfriendly reagents like detergents have to be used for sample preparation. A method reports use of mixed solvent systems like methanol-water, acetonitrile-water, 2-propanol-water for solubilising the proteins, which also accelerated proteolytic digestion by denaturation. The amino acid coverage for the analysed proteins also increased. A method suitable for enriching membrane proteins from complex protein mixture without prior isolation of membrane fractions is reported. The proteins were separated by two-dimensional electrophoresis by immobilized pH gradient IEF on IPG strips followed by SDS electrophoresis. Peptides obtained by trypsin digestion were analyzed by MALDI-TOF-MS.

SAMPLE PREPARATION TECHNIQUES FOR ESI-MS ANALYSIS

May - June 2004 Indian Journal of Pharmaceutical Sciences 267
ESI-MS has a low tolerance for salts, detergents, non-volatile buffers and other contaminants, because they interfere with the ionization process and may obscure relevant signals in the mass spectrometer. Removal of these contaminants from peptide samples is therefore essential and can efficiently be performed using reversed-phase chromatography resin. Peptides remain bound to the resin while contaminants are washed away. This process can be performed on-line with the MS, LC-MS or off-line. The coupling on-line of micro scale RP-HPLC columns and ESI-MS has been a major contribution to the success of ESI-MS. In addition to sample clean-up LC-MS has the following advantages: i) peptides are concentrated during separation, thus improving the level of sensitivity; ii) peptides in peptide mixtures are separated. Therefore peptides in individual fractions are analyzed sequentially by MS and MS/MS; iii) post column flow-splitting allows for the recovery of a large fraction of the purified peptide sample for further analysis, without significant reduction in the detection sensitivity in the ESI-MS instrument.

ESI-MS detection of proteins and peptides is concentration dependent. Accordingly, a post-column flow-splitter is frequently used, which directs a small fraction of the eluting analyte directly into the MS, while the remainder of the fraction can be collected for further analysis. In contrast to microbore columns, post-column flow-splitting from capillary columns is technically difficult. Hunt and co-workers have demonstrated, however, that even from capillary columns significant fractions of the separated analytes can be recovered. For further information, Yates and co-workers have written an excellent guide to preparation and operation of capillary LC columns coupled with ESI-MS. A method has been reported for identification of proteins in complexes by solid phase micro extraction-multistep elution-capillary electrophoresis and tandem MS. A sheath less liquid–metal junction interface is used to interface CE and ESI-MS/MS. Solid phase extraction serves as semi separation dimension using organic phase step elution gradient in combination with second separation dimension for increased resolving power. This improves concentration detection limit for CE and allows more proteins in the complex mixture to be identified. CE offers high separation efficiency, short separation time and lower sample consumption. Atomol level detection limits have been reported for peptides. By using preconcentration on hydrophobic media, buffers, salts and other contaminants can be removed.

**ADVANCED APPLICATIONS USING COMPATIBLE ANALYTICAL TECHNIQUES**

2D-PAGE is a well suited technique to study protein expression in biological systems. But some proteins are notably difficult to separate by 2D-PAGE. They are membrane, low copy number, large (> 150 kDa) and highly basic proteins. Hence alternative separation methods have been tried. In most cases, different selections of proteins are to be identified.

**Surface plasmon resonance bimolecular interaction analysis (SPR-BIA)-MALDI-TOF-MS:**

A study reports multiple detection and protein identification at low femtomole and subfemtomole levels. Epitome tagged tryptic peptides were captured via affinity interactions with either chelated Ni²⁺ or monoclonal antibodies and detected using surface plasmon resonance bimolecular interaction analysis (SPR-BIA). The tagged peptides were either eluted from the biosensor chips for MS analysis or analysed directly from the chip using MALDI-TOF. These two techniques performed for different analytical purposes are complimentary and can be interfaced. SPR-BIA is used to investigate interaction between surface immobilized receptor and the solution borne ligands giving information about association-dissociation kinetics. Here it is used as biosensor to quantitatively detect presence of specific molecules. Alternate use of MALDI-TOF extends to detection and quantification of target analytes selectively retrieved from the biological carriers via immunoaffinity isolation. SPR-BIA and MALDI-TOF are highly compatible. Since optical detection of SPR is non-destructive, a secondary analysis of the species retained during SPR-BIA is possible using MALDI-TOF. The analysis is then capable of distinguishing number and nature of species bound to sensor chip and relative contribution of each to the composite binding curve.

**Capillary isoelectric focusing (CIEF)-FTICR-MS:**

A published method reports capillary isoelectric focusing (CIEF) hyphenated to FTICR-MS. Protein complexes were separated using CIEF under native conditions. The molecular masses of the intact complexes were measured as they emerged from capillary using FTICR-MS. It was shown that either the intact non-covalent complexes or their constituent protein subunits could be detected by variation of the sheath liquid added at electropray MS interface. Thus two successive experiments permitted fast and efficient characterization of intact complex stoichiometry.

**LC-ESI-MS:**

A study reports characterization of cysteine residues and disulfide bonds in proteins by liquid chromatography
ESI-MS\textsuperscript{41}. Free cysteine residues in proteins were labeled with PEO - maleimide-biotin followed by denaturation with 8M urea and digestion with trypsin or chymotrypsin. The resulting peptide mixture was analysed by LC-ESI-MS to identify free cysteine residues. A portion of the peptide mixture was subsequently reduced with dithioeritrol, alkylated with iodoacetamide and analyzed by LC-ESI-MS, to identify cysteine residues involved in disulfide bonds.

Microcapillary HPLC-nanoelectrospray-FTICR-MS:

Subfemtomole peptide sequence analysis is published using microcapillary HPLC columns with integrated nanoelectrospray emitters coupled directly to FTICR-MS\textsuperscript{42}. Simple variable flow HPLC apparatus provides for tandem MS analysis of tryptic peptides at 400 amole level.

Monolithic capillary column HPLC-ESI-MS:

A team of researchers has reported use of monolithic capillary columns for proteomic studies\textsuperscript{43}. Tetrahydrofuran/decanol were used as porogens for fabrication of micropellicular poly (styrene/divinylbenzene) monoliths. It leads to rapid, highly efficient separation of peptides and proteins by RP-HPLC. The loadability of monoliths, both for small peptides and large proteins, was within 0.4 to 0.9 pmole range for 60x0.2 mm capillary column. Low concentration of trifluoroacetic acid was used in the eluent for successful coupling to ESI-MS. Detection limits for the proteins were in low femtomole range. The lack of micropores enabled RP chromatographic separation of peptides and proteins ranging in molecular mass from few hundred to more than 55,000 with high efficiency. Because of robustness of fabricated capillary columns, it could be on-line hyphenated to ESI-MS.

Two dimensional preparative liquid phase electrophoresis (2D-LPE):

In this technique the analytes remain in solution throughout the process. The first step fractionates proteins in defined pH ranges by liquid phase IEF. In the second purification step the proteins are isolated on the basis of their size differences in the liquid phase by SDS-PAGE and eluted in liquid fractions\textsuperscript{44}. This allows high protein loads and large volumes, thus yielding sufficient amounts of low abundance proteins for further characterization by MS.

Hybrid electrophoretic – chromatographic approach:

Recently, a combination of preparative IEF and NP-RP-HPLC has been described for the separation of proteins in 5–70 kDa range from HEL cell line lysate\textsuperscript{45}. About 700 proteins in the pI range of 3.2 to 9.5 were resolved and 38 were identified through peptide mapping.

BASIC INSTRUMENT MODIFICATIONS

Electrodynamic ion funnel:

A study reports introduction and optimization of newly designed electrodynamic ion funnel in ESI interface improving ion beam characteristics in quadrupole electrostatic ion guide and modification of the guide\textsuperscript{46}. This provided detection limit of ~30 zeptomole for the proteins with molecular weight of 8 to 20 kDa. The sensitivity is found to depend upon the overall ion transmission which combines probability of ionization, transmission efficiency and ion trapping in FTICR cell.

MALDI-Qq-TOF-MS:

MALDI has generally been applied in TOF-MS where introduction of delayed extraction increases accuracy of mass measurements. But the problem faced is decreased efficiency of protein digestion at subpicomole level. Recently it has also become apparent that MALDI peptide mapping alone may not allow identification of proteins in genomic sequence databases because of insufficient accuracy of exon-intron prediction algorithms\textsuperscript{47}. MALDI mapping is efficient only when statistically significant number of peaks are identified. Recently MALDI ion source has been interfaced to orthogonal injection TOF-MS and then to Qq TOF to form MALDI-Qq-TOF instrument. In these devices, collisional damping interface cools the ions produced in MALDI source before they enter analytical quadrupole Q and fragmented in collisional cell q and then TOF. Since TOF is almost decoupled from MALDI ion production process, it provides high mass accuracy for both the precursor and product ions, high resolution, precise tuning of collisional energies and simplified calibration procedure.

Activated ion electron capture dissociation:

A technique of colliding the ions with the background gas while subjecting them to electron capture has been reported. Presumably this ion activation breaks intramolecular non-covalent bonds of the ions' secondary and tertiary structure. In comparison to collisionally activated dissociation, this activated ion ECD provides more extensive and complimentary sequence information.

Nanoflow gradient generator:

HPLC and uLC coupled to ESI-MS/MS dominates the field of protein identification by tandem mass spectrometry. Analysis of protein digests by HPLC or LC using columns
with 50–100 μm diameters, requires delivery of solvent gradients at low to mid nanoliter per minute flow rates. Daniel Figeys and co-workers report alternative system for delivery of nanoliter per minute gradients48.

ADVANCED TECHNIQUES FOR PROTEIN ANALYSIS

Constrained database searching using accurate mass of cysteine containing peptide:

Since CID spectra reflect amino acid sequence of the peptides analysed, the constraints provided by CID of selected peptides in mass spectrometer are highly discriminating. Peptides eluting from separation systems are analysed using 1st stage MS that also selects peptide ions for CID via data dependent procedures followed by fragment ion analysis in 2nd stage MS. The accurately measured masses of peptides in protein digest represent a type of constraint for database searching. Their use for protein identification is referred to as peptide mass mapping or fingerprinting. Protein identification using single accurate peptide mass would combine the advantages of peptide mapping and tandem MS approaches. Inclusion of additional constraints like estimated molecular weight of parent protein, cleavage specificity of protease used, presence of relatively rare amino acid like cysteine, methionine or tryptophan in peptide sequence would further enhance stringency of database search. Scientists report a method using mass measurement of single cysteine containing peptide49. Mass spectra are acquired in a single stage using ESI-FTICR-MS at mass accuracy of 1 ppm. Cysteine containing peptides were identified in a mixture by selective alkylation with an isotope distribution encoded tag (IDENT).

Site specific mass tagging with stable isotopes:

In proteolytic peptide mass mapping, specificity can be increased by incorporation of selected amino acid residues enriched with stable isotopes into protein sequence. Selected amino acids are labeled with 13C, 15N, 3H and incorporated into proteins in sequence specific manner during cell culturing. Each of the labeled amino acid carries a defined mass change encoded in its monoisotopic distribution pattern. Peptides with mass tags can readily be distinguished in mass spectra. A reported method states that stable isotope labeling in proteins can generate internal mass signatures with characteristic mass shifts in their isotope distribution patterns without affecting their chemical and structural properties50. Magnitude of mass shift will also reflect content of particular amino acid in that peptide. This technique increases data search efficiency, specificity and accuracy of peptide sequencing and protein identification.

Accurate mass tag (AMT):

A method proposes protein identification based on the mass of a single peptide51. The technique largely depends upon the mass measurement accuracy that can be achieved. The results showed that low ppm level measurements have practical utility for analysis of small proteomes. Sub-ppm mass measurement accuracy levels are attainable using FTICR-MS. Additional information like sequence constraints enable study of more complex proteomes. They suggest that if a molecular mass of single peptide could be measured with high enough mass measurement accuracy, such that its mass was unique among all the possible peptides predicted from a genome, it could be used as an accurate mass tag (AMT). The generation of such AMTs would allow the fragments generated from entire proteome to be analyzed with greater speed and sensitivity. Thus proteome studies would progress in two phases, the first in which proteins are identified, AMTs generated and validated and the second in which subsequent measurements occur rapidly.

Isotope tagged cross-linking reagents for protein interaction analysis:

An important aspect of protein function is the interaction with the binding partners. At molecular level, the interaction analysis thus comprises of identification of interacting proteins and characterization of interaction sites. The most promising approach to tackle both simultaneously is analysis of covalently cross linked complexes by modern MS. Muller and co-workers incorporated cross-links with distinct isotope pattern into the microtubule destabilizing protein Op18 and into complexes formed by Op18 with tubulin52. The deuterium labeled cross-linking reagents were prepared together with their undeuterated counterparts and applied as 1:1 mixture of the respective d0 and d4 isotoformers. The resulting d0/d4 isotope tags allowed straightforward mass spectrometric detection of peptides carrying the linker even in complex enzymatic protein hydrolysates. The tag also facilitated the assignment of linked amino acids.

Elastase digestion and neutral loss tandem MS for phosphorylation analysis:

Loss of phosphoric acid is the most effective fragmentation reaction of pSer and pThr containing phosphoepitopes of small size in low energy CID mode. Lehmann et al. report a method for analysis of protein phosphorylation53. They performed digestion using elastase to generate smaller peptide fragments. The neutral loss scanning resulted in observation of set of partially overlapping
phosphopeptides with high abundance, providing a complete coverage of PKA phosphorylation sites. This scan mode provided highest specificity for detection of singly charged phosphopeptides (neutral loss of 98).

REFERENCES