Indian Journal of Biotechnology
Vol 2, January 2003, pp 48-64

# Mass Spectrometry: An Essential Tool for Genome and Proteome Analysis

Pushpendra Kumar Gupta* and Sachin Rustgi

Molecular Biology Laboratory, Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut 250 004, India

**Mass spectrometry (MS), in its various forms, has become an essential tool for genome and proteome analysis. It involves gaseous ionization of the analyte to be examined, followed by separation of ions according to mass-to-charge (m/z) ratio and determination of molecular masses of ions from mass spectra obtained after mass spectrometry of analyte. Several methods for ionization, mainly including MALDI and ES, each coupled with a specific mass spectral analysis system (e.g. TOF-MS and quadrupole MS) are available. MS/MS is devised particularly for the determination of amino acid sequences of small peptide. The advantage of MS over other techniques is its speed, since gel electrophoresis and labeling of the analyte, needed in other techniques used for genome/proteome analysis, can be dispensed with. Applications of mass spectrometry for genome analysis include DNA sequencing and SNP detection, the latter involving PinPoint assay (minisequencing), PNA hybridization, invader cleavage, "MALDI on a chip", etc. Similarly, its applications for proteome analysis include peptide sequencing, determination of molecular weights of proteins and protein identification by database search. Protein modifications and protein–protein interactions can also be examined by coupling mass spectrometry with database search. In this manner, mass spectrometry has become an essential tool for genome and proteome analysis.**

**Keywords: mass spectrometry, genome analysis, proteome analysis, MALDI-TOF MS**

## Introduction

MS involves separation of charged atoms or molecules according to their m/z ratio and therefore helps in the determination of relative molecular masses of organic compounds and biomolecules with very high precision and sensitivity. This has led to a very wide range of applications of MS in investigations involving study of biomolecules. Although mass spectrometry had its beginning in the early years of the 20th century, it was only in 1980s and 1990s, that mass spectrometry was extensively used for research in various fields of biological sciences. Application of MS for the study of biomolecules actually took long time because it requires charged gaseous molecules for analysis, and the polymeric biomolecules, being large and polar, cannot be easily transferred into the gaseous phase and ionized. However, the availability of ionization techniques like matrix-assisted laser desorption/ionization (MALDI) and electrospray (ES) in 1980s and the major advances made in sample preparation for MS led to powerful instrumentation (sample preparation methods/protocols are out of scope of this article, while an extensive amount of information is given on PennState College of Medicine's website). This made it possible to obtain polymeric biomolecules in gaseous state and in an ionized form, so that MS has been utilized extensively for the study of biomolecules. Starting in early and mid-1990s, software algorithms also became available, which allowed study of correlations of the data collected from MS with the data available in massive databases/databanks. Thus, during the last decade of the 20th century (1990-2000), MS became an important technique for genomics and proteomics research, leading to the 2002 Nobel Prize in chemistry to J B Fenn and K Tanaka. Currently it is used mainly for characterization, identification and quality control of a variety of nucleic acid and protein molecules, which is so important for genomics and proteomics research. A vast amount of literature is available on mass spectrometry; the purpose of the present review is not to summarize all the literature but to provide researchers with an overview of recently used mass spectrometric techniques including the principles of ionization methods used in MS, the major MS instruments currently in use for the study of biomolecules, and the various applications of MS in biological sciences

## Ionization Methods

As mentioned above, MS requires that the molecules to be examined should be available in charged gaseous form. Therefore, methods were

* Author for correspondence:
Tel: 0121-2768195; Fax: 0121-2767018
E-mail: pkgupta36@vsnl.com

developed to obtain biomolecules in the form of gaseous ions to be used for MS. Following common methods are available for ionization of biomolecules.

***Matrix Assisted Laser Desorption-Ionization (MALDI)***

MALDI, developed in the year 1988 by Karas and Hillenkamp, involved co-precipitation of large excess of a matrix material (a small organic molecule) with the analyte molecule (the molecule to be analyzed). This is achieved by pipetting a submicrolitre volume of the mixture of matrix and analyte onto a metal substrate, where it is allowed to dry. The dried solid having matrix and analyte is then irradiated by nanosecond laser pulse, usually supplied by a small nitrogen laser with a wavelength of 337 nm, which is specific for the absorbance of the selected matrix material. The irradiation causes energy transfer and desorption, producing gas phase matrix ions. The non-absorbing intact analyte molecules are also desorbed into the gas phase and get ionized with the help of matrix ions (Fig. 1). The charged molecular ions of the analyte, generated during a gas-phase proton transfer reaction with the matrix molecules, are detected and analyzed by MALDI-TOF MS (see later). The matrix used with biomolecules is generally made up of one of several available substances (Table 1), which differ in the energy they impart to the biomolecules during desorption and ionization and therefore, also differ in the degree of fragmentation (unimolecular decay) that they cause. The DHB matrix, which gives highest sensitivity in MALDI, is preferred when stability of the ions for milliseconds is required as in trapping experiments. In time of flight (TOF) experiments, where stability for microseconds is required, other matrix substances like CHCA can also be used. Several methods are available for sample preparation also. For instance, matrix may be laid down in microcrystalline thin film on the substrate leading to better adherence and providing large crystalline surface from which the ions can be desorbed. Sometimes admixture of analyte with the matrix can also be beneficial. MALDI is generally used for the study of molecules with mass above 500 Daltons. However, proteins undergo fragmentation during MALDI, resulting in broad peaks and loss in sensitivity. As a result, MALDI is mostly applied to the analysis of oligonucleotides and peptides.

***Surface Enhanced Laser Desorption-Ionization (SELDI)***

The patented SELDI based protein chip (Ciphergen® Biosystems, USA)/biochips (LumiCyte,

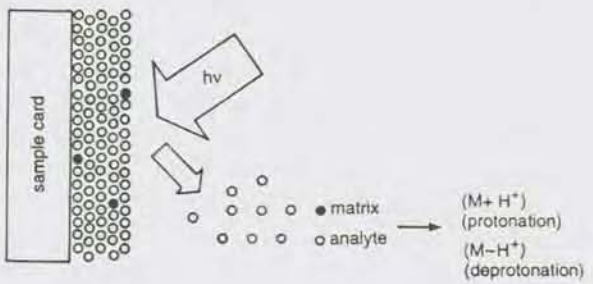


Fig. 1—Schematic diagram showing mechanism involved in matrix assisted laser desorption-ionization.

Inc., Fermont, CA) system, is a unique combination of affinity chromatography and time of flight mass spectrometry on a single platform—a combination that is designated SELDI-TOF MS (Surface Enhanced Laser Desorption/Ionization Time-of-Flight Mass Spectrometry). This system enables protein capture, purification, ionization and analysis of complex biological mixtures directly on protein chips array surface and detection of the purified proteins by Laser Desorption/Ionization Time-of-Flight Mass Spectrometric analysis (Weinberger *et al*, 2000).

SELDI system can be divided into three parts (i) protein chip arrays, (ii) the protein chip reader (LDI-TOF MS) and (iii) specialized software. The array has eight 2-mm diameter spots, each with surfaces designed to select proteins from crude extracts according to protein properties. The reader is an LDI-TOF MS (Laser/Ionization Time-of-Flight Mass Spectrometer), which utilizes pulsed nitrogen laser energy, to ionize proteins from the array and measures the mass of each protein species based on its velocity though reader's ion chamber i.e. TOF. Data is finally collected in the computer with the help of specific software. This system can detect and accurately calculate the mass of compound ranging from small molecules and peptides of less than 1 kDa to proteins of 500 kDa or more (Weinberger *et al*, 2000). Using

Table 1—Some common MALDI matrices having maximum absorbance at 337 nm and the corresponding analytes that can be studied

| Matrix | Analytes |
|---|---|
| α-Cyano-4-hydroxycinnamic acid (CHCA) | Peptides, proteins, lipids and oligonucleotides |
| 3,5-Dimethoxy-4-hydroxycinnamic acid (sinapinic acid) | Peptides, proteins and glycoproteins |
| 2,5-Dihydroxybenzoic acid (DHB) | Peptides, proteins, lipids and oligosaccharides |

SELDI-based protein chip technology, one can produce differential maps of protein expression 'phenotypic fingerprints'– directly from crude samples such as serum, urine or tissue extracts. Its general applications include disease diagnosis, toxicological screening, target-ligand interaction, peptide mapping and immunoassay development.

Recently Nelson & coworkers (2000) have combined surface plasmon resonance-biomolecular interaction analysis (SPR-BIA) with MALDI-TOF analysis (BIA-MS analysis). SPR-BIA is also a chip based technique that uses an immobilized receptor to monitor biomolecular interactions. Since the SPR detection is nondestructive, proteins may be further subjected to TOF analysis for identification. This approach has detection limits at or below 20 fmol and a mass range up to 150 kDa (for a review see Hayduk *et al*, 2002).

***Electrospray (ES)***

The technique of electrospray mass spectrometry (ESMS) was developed in late 1980s by J B Fenn, who was awarded the 2002 Nobel Prize in chemistry (Fenn *et al*, 1989; Yamashita & Fenn, 1984). In this technique, the sample is dissolved in a liquid with mobile phase (water: acetonitrile or water:methanol, 1:1) and is pumped (at the rate lower than microlitre per-minute) through a hypodermic needle at high voltage in order to either electrostatically disperse, or electrospray small droplets measuring up to a micrometre in size. These droplets rapidly evaporate and impart charge onto the analyte molecules. This ionization takes place in the atmosphere and therefore is very gentle, causing no fragmentation of analyte ions in the gas phase. It is, therefore, also described as an atmospheric pressure ionization (API) technique. A stream of nitrogen gas (the nebulizing gas) that flows through a tube co-axial to the main capillary helps in the spraying process (Fig. 2).

ES produces many more ions than MALDI and these ions are transferred into mass spectrometer with high efficiency for analysis. A wide range of molecules including proteins, oligonucleotides, sugars and lipids can be analyzed by ESMS. Electrospray is performed in either of the two different modes. First, is the nanoelectrospray, which is a miniaturized version and operates without pumps at a very low flow rates (few microlitres per minute), facilitated by glass capillaries with an inner diameter of one micrometre at the tip, and second, is the high-performance liquid chromatography (HPLC), which may also be combined with mass spectrometry (LC-MS), so that MS analysis of components takes place on line, as they elute from chromatographic column (Vestal, 1990).

***Fast Atom Bombardment (FAB) or Liquid Secondary Ionization (LSI)***

In this technique, the sample is first dissolved in a liquid matrix, which is a viscous, low-pressure liquid, like glycerol or 3-nitrobenzyl alcohol. The mixture is not crystallized as done in MALDI, but instead a few microlitres of this liquid are placed on a small metal target at the end of a probe, which is inserted into MS. The liquid surface is then bombarded with a beam of high kinetic energy atoms (xenon or argon) or ions (caesium) rather than laser used in MALDI. Molecules are sputtered from the surface, enter the gas phase and ionize. This method of ionization has some inborn advantages over other ionization methods viz. use of liquid matrix instead of solid surface, which if used is permanently damaged by a

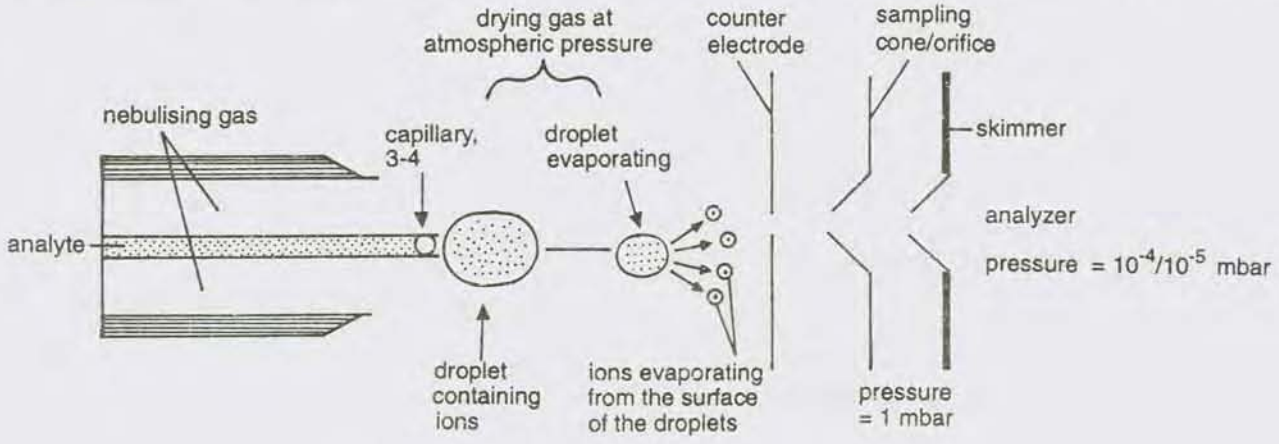


Fig. 2—Schematic diagram of standard electrospray ionization source.

beam of high energy atoms, leading to short lived samples and spectra; the mobility of liquid matrices used allows the surface to be continuously replenished and provides long life to sample and spectra. This has distinct advantages for the study of many medium to large biological molecules (Caprioli, 1989; Barber *et al*, 1981).

***Plasma Desorption Mass Spectroscopy (PDMS)***

PDMS utilizes 252Cf fission fragments to desorb large molecules from target. The target containing the analyte is made up of a thin aluminium foil, often covered with a layer of nitrocellulose. The desorption-ionization involves hydrophobic interactions, which allow salts to be washed off and chemical reaction to be carried out on the target. Alternatively, the sample can be electrosprayed directly onto Ni or Al foil. This technique is more efficient for the analysis of small peptides. PDMS has a reasonably good sensitivity for peptides and small proteins, and typically about 10 picomoles are needed for molecular mass determination of such a protein. The main virtue of this technique is its simplicity, easy usability and easily interpretable spectra, which can be easily utilized by non-experts in mass spectrometry (for review see Gordan, 2000).

***Thermospray and Particle Beam (PB)***

Thermospray ionization involves coupling of HPLC to mass spectrometry at conventional flow rate (0.5 to 1.5 ml/min). The elutriate from HPLC column is released and partially vaporized under reduced pressure by heating a stainless steel tube of 0.10 to 0.15 mm inner diameter. The resulting small droplets vaporize further due to hot gas in this low-pressure region of the ion source. When an auxiliary filament or low-current discharge device is used, complete evaporation of the solvent from the liquid droplets produces gas phase ions from ionic compounds in the sample solution (Blakeley & Vestal, 1983).

The particle beam (PB) interface, derived from the MAGIC interface (Monodisperse Aerosol Generation Interface for Chromatography), has elements in common with thermospray, but gives spectra with more fragment ions. It involves an initial step involving formation of an aerosol, followed by dispersion caused by a gas stream (usually helium), and desolvation. This leads to formation of a beam of charged particles, which makes the ion source. In the original interface, a momentum separator then separates the lighter dispersion gas and the vaporized solvent from the sample particles having higher momentum, thus facilitating entry of ions into the mass spectrometer. PB is less sensitive than thermospray or electrospray and is not best suited for the analysis of substances like ionic components, high molecular mass samples, and thermally labile compounds.

***Electron Ionization (EI) and Chemical Ionization (CI)***

In this technique, a beam of energetic electrons is generated inside a chamber by heating a metal filament to a sufficiently high temperature (approx. 2000 K). The electrons are attracted toward a plate (away from the filament), called the trap, by maintaining the trap at a relatively positive potential. A relatively volatile analyte, in gas phase, is directed into this high-energy beam traveling from filament to the trap, and an interaction with electron beam generates charged molecular ions of the analyte. Depending on the compound and the ionization energy, the molecular ions may then fragment. The spectra, usually containing many fragment-ion peaks, are useful for structural characterization and identification. Small impurities in the sample are easy to detect (McCloskey, 1990).

Chemical ionization (CI) is very similar to EI but in this case a reactive ionized reagent species is used instead of a beam of electrons and the interaction causes ionization of the analyte. Many such reagents are gases (e.g. isobutene, methane and ammonia, etc.), which are used to enhance the abundance of molecular ions. For both these ionization methods, range of molecular weight of the analyte is 50 to 800 Da. In rare cases, it is possible to analyze samples of higher molecular weight also (Harrison, 1992).

***Electron-capture Ionization***

Electron-capture, sometimes called negative ion chemical ionization (NICI), is used for molecules containing halogens, $NO_2$, CN, etc. It usually requires that the analyte be derivatized to contain highly electron-capturing moieties (e.g., fluorine atoms or nitrobenzyl groups). Such moieties are generally inserted into the target analyte after isolation and before mass spectrometric analysis. The sensitivity of NICI analysis is generally two to three orders of magnitude greater than that of positive ion chemical ionization (PCI) or EI analysis. Little fragmentation occurs during NICI, and this mode of ionization is generally employed for quantitative analysis of trace

amounts of compounds of known structure, where a heavy isotope-labeled compound is used as an internal standard.

**Mass Spectrometers and Mass Spectral Analysis**

Essentially, a mass spectrometer measures m/z ratios of analytes such as nucleic acids, oligonucleotides, proteins, peptides or peptide fragments. It has three components, ionization source, the m/z analyzer and the detector (Fig. 3). Following three principles are available for separation of molecules that differ in mass: (i) separation of ions on the basis of time of flight (TOF MS); (ii) separation of ions by quadrupole electric field generated by metal rods (quadrupole MS) and (iii) separation of ions by selective ejection of ions from a three-dimensional trapping field (ion trap or Fourier transform ion cyclotron MS). Any of the ionization methods described above, including MALDI, ESI or FAB can be coupled to any of the above three methods of ion separation. However, since MALDI produces short bursts of ions in vacuum, it is coupled with TOF MS, while electrospray, producing continuous beam of ions in atmosphere is coupled with quadrupole and ion trapping MS.

In 1900, J J Thomson introduced the first mass spectrometer, which employed fixed magnetic and electric fields (described as electric sector and magnetic sector) to separate ions of different mass and energy. He recognized that charged particles that differ in momentum behave differently in an electromagnetic field, and used this property for separation of ions with different masses. Using this instrument, he was able to prove that the noble gas Ne (Neon) was composed of two different isotopes of mass 20 Da and 22 Da (Thomson, 1913). These so called 'sector instruments' are not very sensitive in comparison with recent mass spectrometers and are now more or less of historical importance. Some advantages, although not unique, are their relatively high mass range, sensitivity and resolving power, and their compatibility with wide range of ionization techniques. The disadvantages are their size and cost compared to most other mass spectrometers developed later.

***MALDI-TOF MS***

MALDI-TOF mass spectrometers utilize MALDI for ionization and the time of flight (TOF) of the ions to reach the detector as a parameter to estimate m/z ratio. Since the velocity of ions will depend on its mass and energy, the time taken by an ion to travel a specified distance will depend on m/z ratio and is described as TOF. Several types of MALDI-TOF MS are now available (Gordan, 2000).

*(a) Linear-TOF MS.* In linear spectrometers using MALDI and TOF, samples are deposited on a metal substrate, capable of holding several hundred analyte spots. Irradiation of these spots by a laser pulse generates a short burst of ions, which are accelerated to a fixed amount of kinetic energy in order to travel down a flight tube (Fig. 4). The small ions have a higher velocity than the larger ones, thus producing a TOF spectrum.

*(b) DE MALDI-TOF MS.* Linear MALDI-MS described above lacks resolution and sensitivity to analyze complex mixtures like those of Sanger's sequencing reactions. This is because the initial velocities and therefore also the range of flight times of ions of the same mass extracted in conventional MALDI-TOF MS involving continuous ion extraction may differ due to a variety of reasons thus reducing the resolution of the technique. Improvements have been made by varying the time between formation and extraction of ions from the source. The ion source may be field free during ionization and only after a specified time delay, a high amplitude pulse of irradiation extracts ions. This results in the arrival of ions of the same mass simultaneously at the detector and helps in correlating the velocity differences in ions with their masses thus improving the resolution of a time of flight (TOF) mass analyzer. The

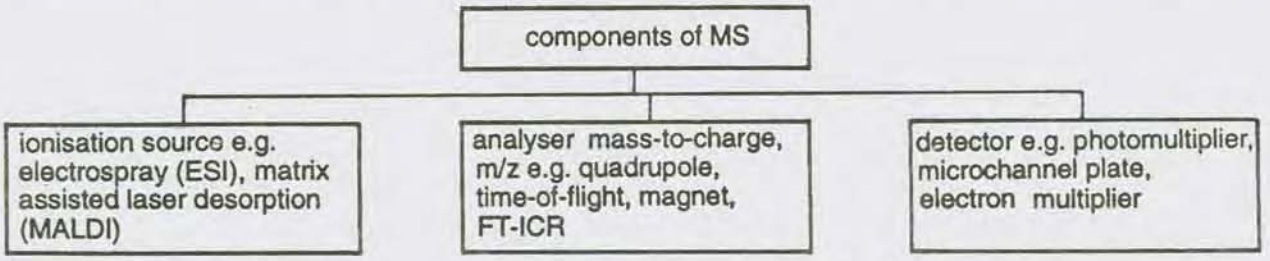


Fig. 3—Simplified schematic diagram of a mass spectrometer (MS).

technique is described as delayed extraction MALDI (DE MALDI) and has been used for DNA sequencing.

*(c) Reflectron TOF (RETOF)-MS.* In this instrument, a single stage or a dual stage reflectron is used at the end of the flight tube, which leads to improvement in mass resolution. The reflectron, located at the end of the flight tube, is used to compensate for the difference in flight times of ions with same m/z, but different kinetic energies (Fig. 5).

### *Quadrupole Mass Filter Instruments*

A quadrupole is a mass filter consisting of four rods, to which an oscillating electric field is applied and which lets only a certain mass pass through, the other masses being on unstable trajectories, do not reach the detector. By scanning the amplitude of the electric field and recording the ions at the detector, one obtains a mass spectrum. Quadrupole MS is capable of resolving mass accuracy of 0.1 to 1 Da and thus excels in quantitative measurements. Most experiments are performed on triple quadrupole instruments, consisting of three sections: two mass separating quadrupole sections separated by a central quadrupole (or a higher multipole) section, which contains the ions during fragmentation.

Quadrupole MS can sometimes be combined with TOF, so that the dissociated ions enter the TOF analyzer for getting the mass spectrum. For instance in a quadrupole TOF instrument, the first quadrupole (q0) focuses the ions, second quadrupole (q1) separates the ions and the third quadrupole (q2) dissociates the ions to enter the TOF analyzer, where the ions are pulse accelerated into the reflector and then to the detector (Fig. 6). Quadrupole TOF analyzer may sometimes be used with MALDI ion source as done in MALDI-quadrupole TOF (MALDI-q TOF) (Fischer, 1959).

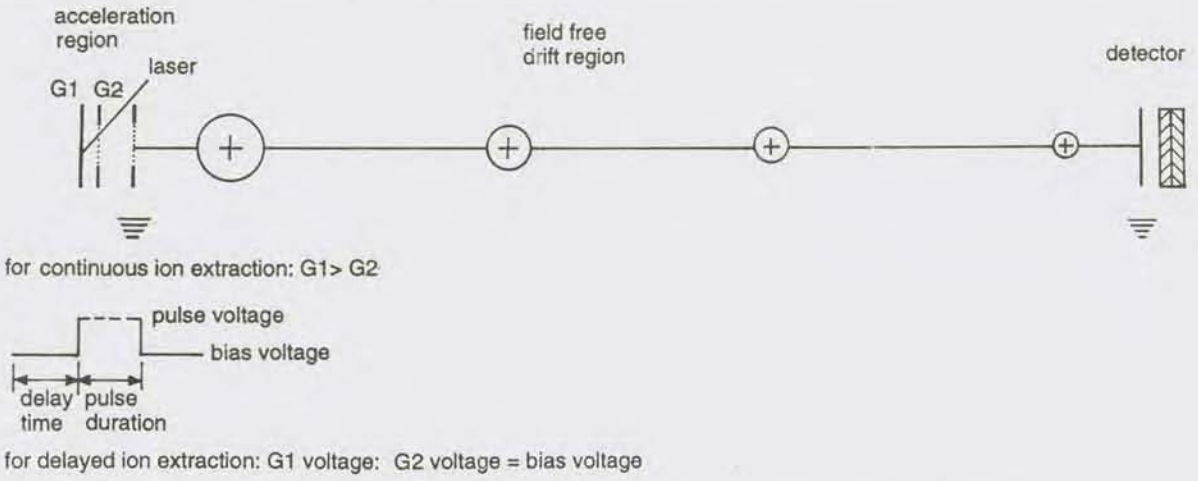


Fig. 4—Schematic representation of basic components of a linear TOF mass spectrometer.

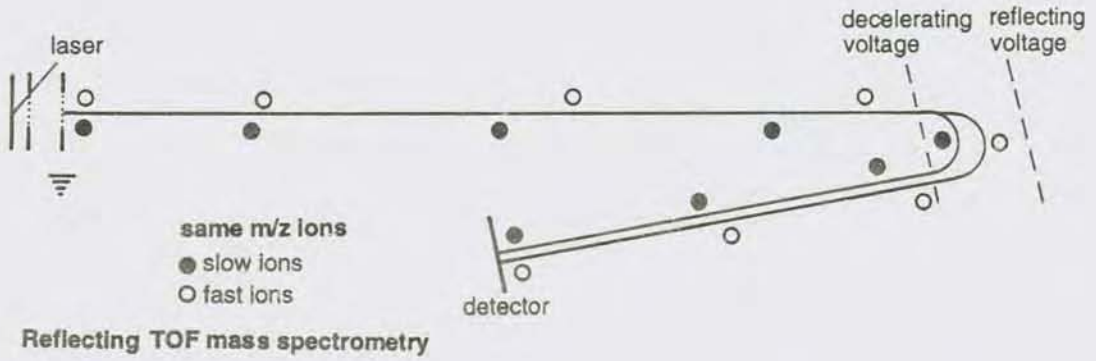


Fig. 5—Schematic representation of basic components of a reflecting TOF mass spectrometer.

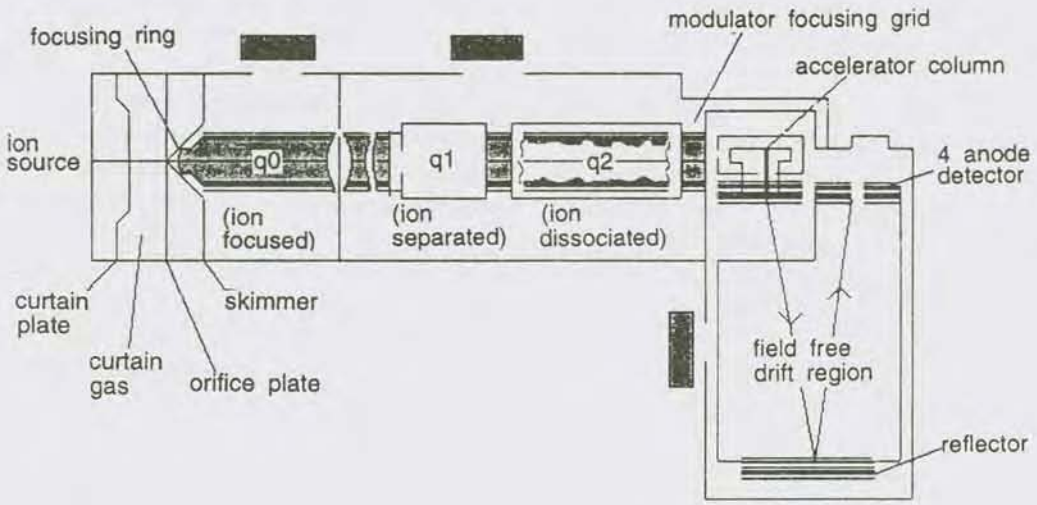


Fig. 6—Schematic diagram of quadrupole TOF instrument, representing different sections of quadrupole (modified from Mann *et al*, 2001).

### *Trapping Instruments*

In these instruments, ions in a continuous beam are trapped in three-dimensional electric fields. The trapped ions are ejected one after another from the trap by application of additional electric field and are analyzed to produce a mass spectrum. Ion trap instruments are compact, versatile and highly automated, but in practice the accuracy and resolution achieved by these instruments do not reach the level of quadrupole instruments.

Another version of the trapping principle is employed in Fourier transform ion cyclotron MS (FTMS), which involves trapping of ions formed by EI, CI, MALDI, and ESI and uses Fourier transform detection, permitting simultaneous detection of a wide-range of masses. The method has several advantages over other methods, such as highest resolving power, a high upper mass limit of analyte that can be examined, high sensitivity, nondestructive detection, and high accuracy for mass measurement.

### *Tandem Mass Spectrometers*

Tandem mass spectrometry (MS/MS) is utilized, where a mixture of peptides is used for getting amino acid sequences of peptides one at a time. The first part of tandem spectrometer selects one peptide on the basis of mass, which is then dissociated or fragmented by collision with an inert gas like argon or nitrogen. In the second part of tandem mass spectrometer, the resulting fragments are separated and tandem mass spectra or MS/MS spectra are obtained (Fig. 7). The fragments differ in size depending upon the position of amide bond fragmentation and can be C-terminal or N-terminal, the former being predominant. Most peptides thus examined are tryptic peptides (obtained due to digestion with trypsin enzyme) having arginine or lysine residues at their C-terminus (Busch *et al*, 1988).

Tandem mass spectra are usually interpreted with computer assistance or are matched against databases, while interpreting the data of C-terminal or N-terminal fragments. One will notice that C-terminal fragments start with masses 147 (lys) or 175 (arg) and the next fragment peak differs by mass of a single second amino acid, which can be identified by mass difference. Consecutive amino acids will be identified using this principle thus giving the amino acid sequence of a small peptide. Similar will be the case with the N-terminal ion series.

## Other Mass Spectrometers

In addition to four basic types of mass spectrometers described above, following are other instruments, which are relatively less important.

### *Gas Chromatography-Mass Spectrometry (GC-MS)*

It permits separation of a complex mixture into its components, before ionization and mass analysis. This is particularly useful when analyzing relatively low levels of target compounds derived from complex biological mixtures. The target analyte must be relatively volatile or must be susceptible to conversion into a volatile derivative to permit GC separation. In general, the derivatized analyte should

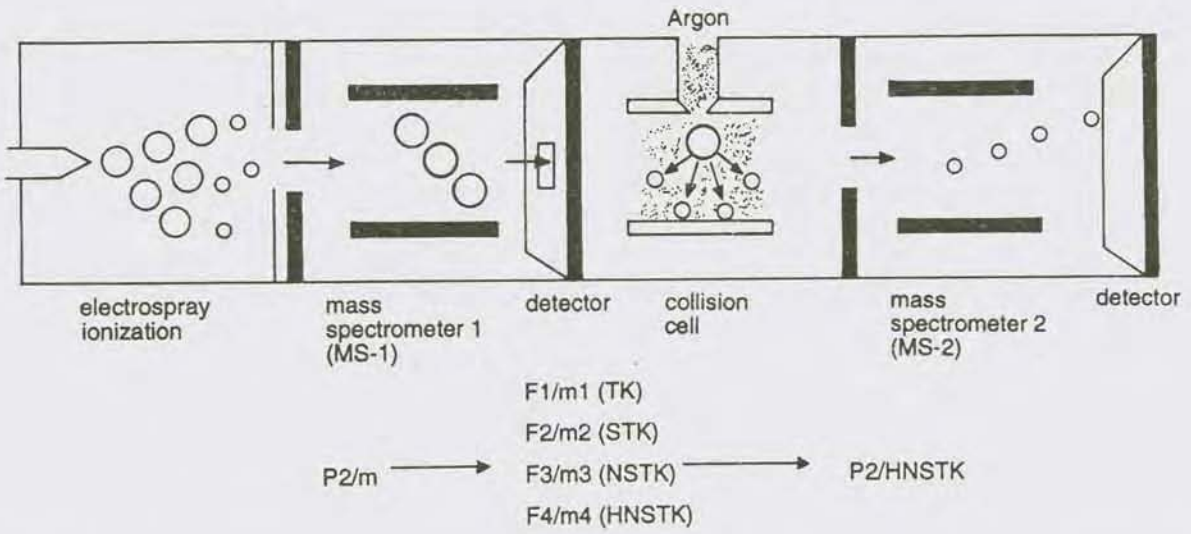


Fig. 7—Schematic representation of tandem mass spectrometry for determination of a peptide sequence (P1 to P3, different peptides; F1 to F5 peptide fragments obtained from peptide P2 and m1-m4 corresponding masses determined by MS2; H, N, S, T and K = different amino acids).

have a MW of less than 1000 Da to be used for GC-MS. In special cases, derivatized analytes with MW up to 1000-2000 Da can also be analyzed. EI and CI in positive and negative modes are employed for ionization.

### *Liquid Chromatography and Tandem Mass Spectrometry (LC-MS)*

Liquid chromatography (LC) coupled with tandem mass spectrometry, called LC-MS/MS or popularly known as LC-MS, is a powerful technique for analysis of proteins and peptides. In a typical LC-MS/MS experiment, the analyte is eluted by reversed-phase high performance liquid chromatography (HPLC) as applied elsewhere in protein chemistry; the only difference is that the dimension and flow rates are much smaller, that are achieved by nano-LC columns of internal diameter of 50-100 μm. The sensitivity of this technique depends upon column diameter, because the lower the flow-rate for the chromatographic separation, the higher will be the sensitivity. The detection limit of few femtomoles (1-10 fmol) in this technique makes it compatible for detection of proteins/peptides at low copy number per cell (Vestal, 1990; Caprioli, 1994).

Complicated mixtures containing hundreds of proteins can be analyzed directly by LC-MS even when concentrations of different proteins differ. LC-MS/MS can be used alone or in combination with 1-D or 2-D electrophoresis, immunoprecipitation, or protein purification techniques. In this manner, in a single LC-MS/MS run, a large number of peptides can be sequenced even if they have the same molecular weight but differ in hydrophobicity.

### *Isotope Ratio Mass Spectrometry (IRMS)*

Isotope ratio MS, which can be combined with GC-MS is capable of very precise determination of 13C/12C ratios. It is exploited principally in examining trace enrichment of 13C in small molecular-weight analytes (e.g. protein-derived amino acids) after biosynthetic incorporation of a 13C-labeled precursor. Applications of IRMS include the study of substrate disposition in humans after infusion of 13C-labeled precursors. Other elemental ratios (e.g. 2H/1H, and 18O/16O) can also be measured although it is mainly used for detection of 13C/12C and 15N/14N ratios that are useful in proteomics research.

## Applications of Mass Spectrometry for Genome/Proteome Analysis

Mass spectrometry in the form of MALDI-TOF has been extensively used for the analysis of nucleic acids, proteins and peptides, although initially the technique was more often used for analysis of peptides (Hillenkamp *et al*, 1991; Fitzgerald & Smith, 1995; Nordhoff *et al*, 1996; Yates III, 1998; Li *et al*, 2000). For biotechnology applications, we need to study both nucleic acids and proteins and therefore use of MS for the study of both nucleic acids and proteins has been briefly discussed.

***Genome Analysis using MS***

Several methods that require neither the gel electrophoresis nor labeling, utilize MALDI-TOF MS for genome analysis. For instance, a study of DNA polymorphism using MS, depends largely on the determination of relative masses of individual DNA fragments with sufficient accuracy, so that even a single base replacement (transition or transversion) can be identified due to unique mass of each base. However, the methods have been applied only to relatively short oligonucleotides, because the resolution and sensitivity of MALDI-MS generally falls off dramatically with increase in size of oligonucleotides. Analysis of nucleic acid by MALDI-TOF MS has several advantages including the following: (i) The ionization, separation by size, and detection takes milliseconds to complete, so that the technology is fast, in contrast, electrophoretic separation and detection takes hours, sometime days; (ii) The results are more accurate than the electrophoresis-based and hybridization-based methods, which are influenced by secondary structures formed by nucleic acids; and (iii) Complete automation involving both sample preparation and acquisition/processing of data is possible (Leushner, 2001).

*(a) DNA sequencing by DE-MALDI-TOF MS.* As we know, 'Sanger's dideoxy termination' method of DNA sequencing depends on resolution to allow sequential distribution of the termination fragments differing by single nucleotides. These termination products of Sanger's termination reactions can be examined by delayed extraction MALDI-TOF MS. In this approach, mass spectra of each of the four specific dideoxy termination reactions generated from Sanger's chemistry are overlaid and each sized product correlated to one of the four base termination reactions, since mass of each base is known. A major advantage of this approach is the fast acquisition of data by MALDI-TOF MS, which needs only minutes per sample. An additional advantage is that there is no need for either the gel electrophoresis or for any radioactive or fluorescence labeling. Short sequences of oligonucleotides were distinguished from each other by this technique (Smith, 1993; Fitzgerald *et al*, 1993; Williams, 1994). But the actual analysis of reaction products of Sanger's sequencing reactions proved difficult firstly, due to availability of only small amount of each fragment in the mixture and secondly, due to interference by buffer salts and other components of the reaction. Delayed extraction (DE-MALDI) allowed further improvement in the sensitivity and resolution of the technology of MALDI-MS (Colby *et al*, 1994; Vestal *et al*, 1995). It involved improvement of the resolution of time of flight (TOF) mass analyzer by correcting for the velocity distribution of ions (by altering the time between formation and extraction of ions) formed in the ion source. DE-MALDI is coupled with high yield cycle sequencing protocol to make it suitable for analysis of sequencing mixtures of oligonucleotides, 40-50 bases long (Roskey *et al*, 1996) (Fig. 8). Since fragments larger than 100 bases cannot be utilized for sequencing (Nordhoff *et al*, 1992; Fitzgerald *et al*, 1993; Mouradian *et al*, 1996; Koster *et al*, 1996; Monforte & Becker, 1997; Taranenko *et al*, 1997; 1998; Fu *et al*, 1998; Kirpekar *et al*, 1998), the technique can not be employed for genome sequencing, but can certainly be used for minisequencing required for single nucleotide polymorphism (SNP) genotyping (Ross *et al*, 1998; Timothy & Smith, 2000).

*(b) SNP detection using MS.* A variety of methods utilizing MS have been developed for analysis of point mutations that are due to SNP (Gupta *et al*, 2001).

(i) *PinPoint assay for SNP detection using MALDI-TOF MS of PCR products*—One of the methods that was described as PinPoint assay (Taranenko *et al*, 1996; Tsuneyoshi *et al*, 1997; Paracchini, 2002) is based on using a primer directly upstream of a known point mutation or SNP, so that in PCR, the primer is extended by a single ddNTP. Such primers may sometime anneal several bases downstream of SNP and extend more than one bases before incorporating a ddNTP, which terminates the reaction. The PCR extended primers are 'solid–phase purified and detected by m/z value specific for the nucleotide added in the extension reaction. The technique is described as minisequencing, since it involves sequencing of a segment few bases long. SNP in several genes like p53 gene, apolipoprotein E gene and RET proto-oncogene have been identified using this approach. Multiplex SNP analysis has also been possible using MALDI-TOF MS, as shown for PCR amplification from BRCA1 exon 13 locus (Haff & Smirnov, 1997) having five SNPs. This is facilitated by using mass-tagged minisequencing primers (extra non- complementary dT nucleotide at the 5' end) for each SNP position, so that all the five primers could be extended in a single reaction and PCR products could be simultaneously analyzed by MALDI-TOF

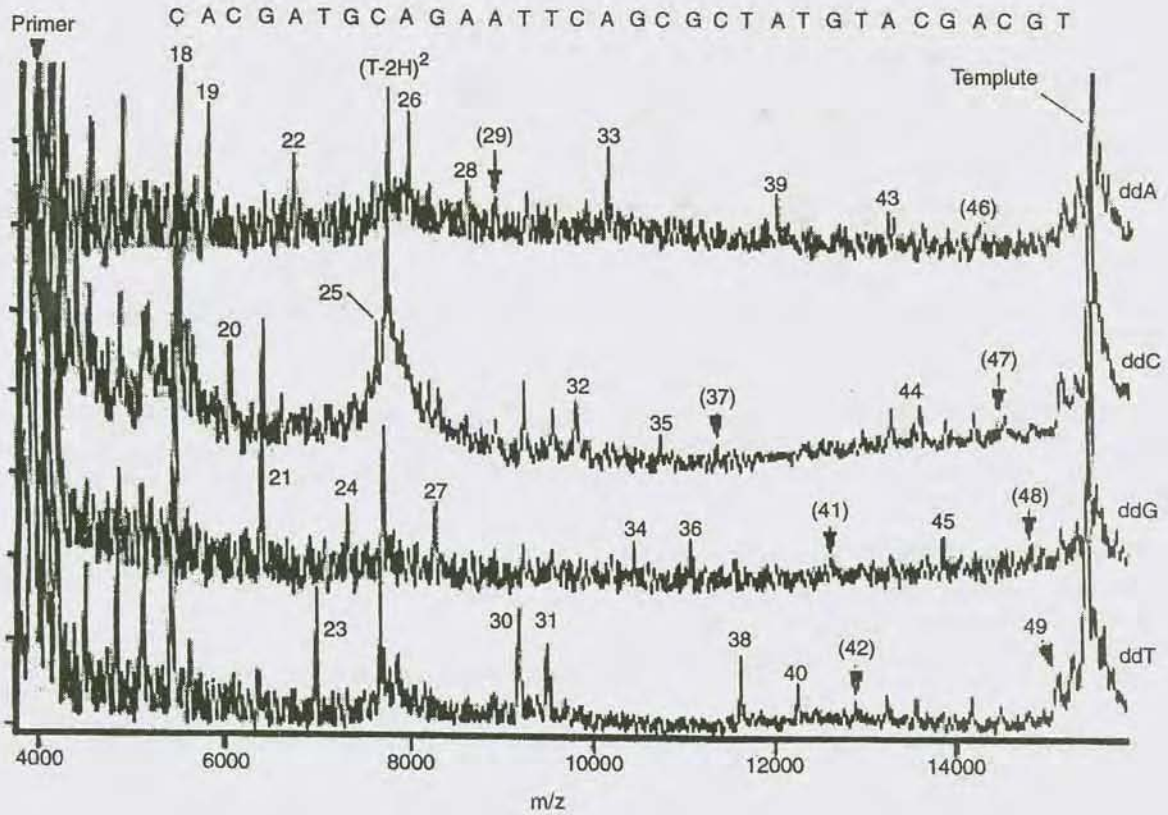


Fig. 8—Diagram showing negative ion mass spectra of four dideoxy reactions generated by primer extension on a 50-base template. Numbers indicate peaks that correspond to expected termination products numbered according to the total DNA fragment length including primers; sequence deduced is given at the top of the figure (modified from Roskey *et al*, 1996).

MS. Mass tagged ddNTPs have also been used to increase the mass difference between PCR products, so that this mass difference can be resolved by MALDI-TOF MS (Fei *et al*, 1998). In another study, multiplex genotyping of 12 different SNPs (selected from human genomic SNPs) was possible using minisequencing (Ross *et al*, 1998). The products can be analyzed by Voyager DE-MALDI-TOF platform. Mass spectrometry has been used both for the discovery of SNPs and for genotyping SNPs at the known position (Nordhoff *et al*, 1992; Fitzgerald *et al*, 1993; Mouradian *et al*, 1996; Koster *et al*, 1996; Monforte & Becker, 1997; Taranenko *et al*, 1997; 1998; Fu *et al*, 1998; Kirpekar *et al*, 1998). For instance, PCR products from p53 tumour suppressor gene have been sequenced and analyzed by MALDI-TOF MS (Fu *et al*, 1998).

(ii) *SNP detection using PNA probes*—Peptide nucleic acids (PNAs) are being used for a variety of purposes, including their use as hybridization probes. They are actually preferred over DNA/RNA probes, due to (i) their ability to hybridize under low ionic strength, (ii) increased hybridization specificity for complementary DNA, and (iii) the increased thermal stability they provide to the hybrid duplexes formed. PNAs can also be more easily analyzed by MALDI-TOF MS, since peptide backbone does not fragment during MALDI. For SNP analysis, two allele specific PNA probes, one for each SNP allele are annealed to biotinylated single-stranded PCR products that are immobilized on streptavidin coated magnetic beads. After annealing, the beads are stringently washed so that only perfectly matched probes remain hybridized. The entire bead solution is then spotted on MALDI probe tip and acidic matrix solution is added to the beads, which dissociates the PNA probe from immobilized DNA. The mixture is then irradiated so that PNA is desorbed and ionized and detected by its mass; each unequally mass-labeled PNA probe detected corresponds to a specific allele (Fig. 9). The PNA probe may be mass labeled either by (i) incorporating a variable number of 8-amino-3,6-dioxaoctanoic acid residues on the N-terminal end of the PNA probe, as done for the detection of SNPs in

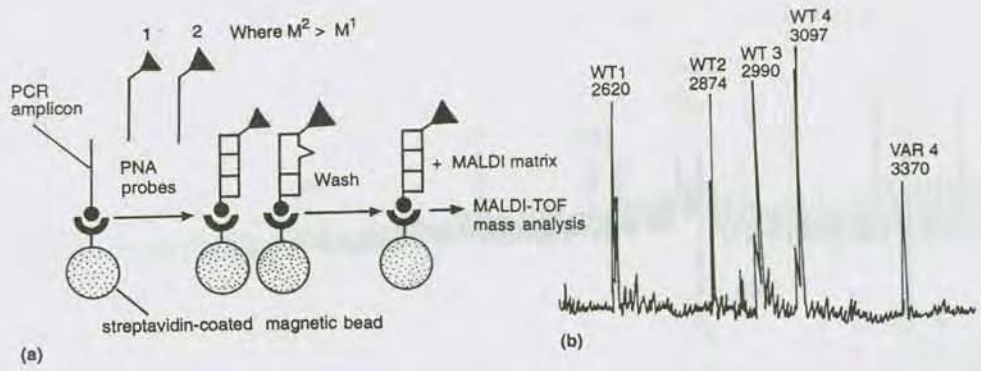


Fig. 9—(a) Schematic diagram of steps involved in the mass-spectrometric analysis of peptide nucleic-acid hybridization probes. (b) Representation of composite mass spectra representing four hypothetical wild types (WT – 1-4) and a variant (VAR) allele for a unique single-nucleotide polymorphism. In this case sample was heterozygous for WT 4 allele and homozygous for other three WT alleles (modified from Griffin *et al*, 1997).

human tyrosinase gene (Griffin *et al*, 1997) or (ii) by adding extra non-complementary dT to the 3' end of PNA oligomer, as done for detection of SNPs at human leucocyte antigen DQ α locus (Jiang-Bucom *et al*, 1997).

(iii) *SNP detection using invader cleavage*—Invader cleavage is a technique that allows SNP detection without PCR and has been discussed elsewhere (Taranenko *et al*, 1996; Gupta *et al*, 2001).

(iv) *SNP detection involving minisequencing using 'MALDI on a chip' technology*—In this technology, nanolitre amount of sample from minisequencing SNP-analysis reactions are piezoelectrically pipetted onto a silicon chip. This silicon chip is directly inserted into mass spectrometer (MS) and each separate sample spot is automatically analysed using MALDI-TOF MS approach. High quality mass spectrum from 100 separate spots (each containing only 8 femtomoles of an oligonucleotide) could be obtained successfully using this approach, within six minutes (Little *et al*, 1997).

MALDI-MS is used generally to achieve high throughput and automation in typing SNPs from a large number of genotypes in a fraction of time that is needed in other approaches. Krebs *et al* (2001) reported its utility in typing of short tandem repeats (STRs or SSRs) also. The approach of typing SSRs using MALDI-MS has many advantages over the electrophoretic discrimination of alleles, which depends largely upon relative mobility of fragments during electrophoresis; in this case, allele sizing is relative and generally prone to error because of artefacts. These assays are also time consuming, have poor resolution and are non amenable to automation (Braun *et al*, 1997; Ross & Belgrader, 1997; Ross *et al*, 1998).

### *Proteome Analysis using MS*

In the post-genomic era, proteome analysis will be crucial for understanding different biological processes. To achieve this objective, mass spectrometry has been and will be extensively utilized for (i) peptide sequencing, (ii) identification of proteins, (ii) study of protein expression in different tissues and under different conditions, (iii) identification of post-translational modifications (phosphorylation, glycosylation, etc.) of proteins in response to different stimuli and for (iv) characterization of protein interactions that include protein-ligand, protein-protein and protein-DNA interactions. Recent improvements in MS have significantly improved its applications in the study of protein structure and function (Loo *et al*, 1995; Yates, 1998). Much of the work with proteins falls into two categories; (i) analysis of intact proteins and (ii) analysis of peptides derived from chemically or enzymatically cleaved proteins.

To-date, majority of systems used for protein identification from complex mixtures generally involves proteolytic digestion followed by some form of chromatographic separation of tryptic peptides combined with MS analysis. This approach of protein identification is referred to as the 'bottom-up approach' (Chalmers & Gaskell, 2000; Pandey & Mann, 2000; Mann *et al*, 2001). In contrast, 'top-down approach' initially described by Mortz *et al* (1996) and recently by Stephenson *et al* (2002) involves starting with intact proteins and obtaining

sequence information directly from a mass spectrometric experiment without resorting first to proteolytic digestion (Kelleher *et al*, 1999; Stephenson *et al*, 1999; Cargile *et al*, 2001; Reids *et al*, 2001; Wells *et al*, 2001; Forbes *et al*, 2001).

*(a) Peptide sequencing.* The sequence of a peptide can be determined by interpreting mass spectrometry data resulting from either of the following three techniques: (i) tandem mass spectrometry of peptide fragments ionized using electrospray from a mixture (Li & Assmann, 2000); (ii) TOF-MS of a mixture of fragments resulting from chemical degradation of a peptide from N-terminus or C-terminus, a technique described as ladder sequencing (Chait *et al*, 1993; Patterson *et al*, 1995), or (iii) RETOF-MS of the post-source decay (PSD) fragmentation of metastable ions of a peptide produced with MALDI (Spengler *et al*, 1991). The basic principle involved in the approach, where MS/MS is used for peptide sequencing has been described earlier (Fig. 7).

*(b) Determination of molecular weight of intact proteins.* Mass spectra of intact protein fragments allow detection of precise molecular weights of major and minor proteins, although larger proteins, being heterogeneous, make determination of their precise molecular weights difficult. However, high-resolution Fourier transform ion cyclotron resonance mass spectrometry (FTMS) has been used for accurate determination of mass of small intact proteins (Horn *et al*, 2000 a b c; Kelleher *et al*, 1998; Kelleher *et al*, 1999; Stephenson *et al*, 1996). But the molecular weight alone may not be enough for protein identification, and peptide sequencing by tandem mass spectrometry, as described above, may be needed.

*(c) Protein identification using MS followed by database search.* Mass spectrometry has been combined with database search to create a valuable and automatic protein identification tool. Mainly three types of databases (genomic DNA, cDNA/EST and nonredundant protein database) are searched by mass spectrometric data. In this application, intact proteins are degraded into pools of peptides whose masses are determined by mass spectrometry and then searched against genomic DNA, cDNA/EST database entries. Nonredundant protein database contains the known set of full-length protein sequences, devoid of duplicates. These databases can also be searched both by mass fingerprint and tandem mass spectrometric data. Recently, Weiller *et al* (2001) developed a specialized database for the analysis of MALDI-TOF MS data derived from tryptic peptides of *Sinorhizobium meliloti* proteins. This database contains amino acid sequence data of proteins predicted from the complete chromosome. It allows accessing and comparing mass spectrometric data with already available protein profiles in the database. Matches can quickly be annotated via links to annotated databases such as Swiss-Prot, etc. EST databases, such as dbEST at the National Centre for Biotechnology Information (NCBI), and EMEST at European Bioinformatics Institute (EBI), contain millions of short single-pass sequences from random sequencing of cDNA libraries. These can be searched usually by translating into the six reading frames (tBLASTn). However, EST databases are highly error prone and each covers only a part of the gene, but it is experienced that, all proteins encountered in a proteomics project can be correlated to their respective ESTs based on minimal mass spectrometric information (Mann *et al*, 2001). Finally, genome databases can also be searched with mass spectrometric data. Surprisingly it has been found that peptides can match a raw genomic sequence, without any information about the reading frame, coding region and without translating it into amino acid sequence. The advantage of searching databases of completely sequenced genomes are that often the mass spectrometric data can help to define the structure of the gene, with the help of start codon, stop codon and intron-exon junctions.

Protein modifications do not present an obstacle to identification because, for a typical 50 kD protein generally 50 peptides are obtained after tryptic digestion and only a few are modified. Only a small number of peptides are required for unique matching to a database entry, especially in case of data from tandem mass spectrometry; therefore very extensive modification only marginally increases the difficulty of protein identification. There are three main programmes used for database mining for protein identification using MS data (Table 2).

(i) *Searching with peptide mass fingerprints*—In this method, first a mixture of proteins is subjected to 2-D gel electrophoresis. A specific protein is then subjected to in-gel digestion with a sequence-specific protease such as trypsin and a mass spectrum described as 'mass fingerprint' is obtained by MALDI-TOF. The mass data is compared to the theoretically expected tryptic peptide masses for each entry in the database. The proteins in the databases may differ according to the number of peptide

Table 2—Three main programs for protein identification using database mining

| Programme | Web site | Based on |
|---|---|---|
| Mascot | http://www.matrixscience.com | It incorporates a probability based implementation of the MOWSE* algorithm. |
| ProFound | http://prowl.rockefeller.edu | It calculates the probability that a protein in a database is the protein analyzed here and it uses Z-score as an indicator of the quality of the search result. |
| Protein Prospector (MS-Fit) | http://prospector.ucsf.edu | It uses MOWSE score to evaluate a hit in protein identification. |

*The MOlecular Weight SEarch algorithm (MOWSE) capitalizes on information obtained through mass spectrometric (MS) technique, using both the molecular weights of intact proteins and those resulting from digestion of the same proteins with specific protease. Calculations are based on information contained in the nonredundant (nr) protein databases. The score value indicates how often the molecular weight of a fragment occurs in proteins within a certain range of molecular weight.

matches. More sophisticated scoring algorithms take the mass accuracy and the proportion of protein covered into account and attempts to calculate a level of confidence for the match. When high mass accuracy is needed, as a rule at least 5 peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification. After a match has been found, a second-pass search is performed to correlate remaining peptides with the database sequence of the match, taking into account possible modifications.

(*ii*) *Searching with tandem mass spectrometric data*—Database can also be examined for similarities with tandem mass spectrometric data obtained on peptides from the protein of interest. Because tandem mass spectra contain structural information related to the sequence of peptide, rather than only its mass, these searches are more reliable and specific. Several approaches exist; one of the most important approaches is the peptide sequence tag method. It makes use of the fact that nearly every tandem mass spectrum contains at least a short run of fragment ions that unambiguously specify a short amino acid sequence. Such a peptide sequence tag will then be used to retrieve the corresponding sequence from the database. The procedure can be automated and is highly specific, especially when performed using instrument, such as quadrupole TOF (Little *el al*, 1997).

(*iii*) *Searching for protein modifications*—Protein modifications do not interfere in protein identification by database searching. For instance, a phosphopeptide can be correlated to a peptide sequence in the database with an additional mass increment due to the phosphogroup (80 Da). The algorithm used for peptide sequence tag allows detection of mass difference on either side of tag sequence. For instance, the sequence between the tag and C-terminus of the peptide could agree with the database entry, but the mass for the sequence from tag to the N-terminus could be larger by 80 Da. This would suggest that there is modification due to a single phosphogroup yielding a mass difference of 80 Da between the N-terminus of peptide and the start of the sequence tag.

(*d*) *Analysis of post-translational modifications.* Mass spectrometry is also used for the study of post-translational modifications in proteins. For instance, protein phosphorylation, which is so common with in the cell, can be detected by mass spectrometry. However, it requires more time and additional preparative steps. When tryptic peptides are usually analyzed in mass spectrometer, the phosphorylated peptides may need to be separated from nonphosphorylated peptides, the presence of which may obscure the signal from the phosphorylated peptides. Since, detection of the exact phosphorylated residue may still be difficult, it might be worthwhile to identify the relevant peptides that are phosphorylated by treatment with phosphatase and observe the mass difference of the peptides before and after treatment.

Other modifications most commonly found in proteins during routine proteome analysis include acetylation, glycosylation and the N- or C- terminal processing by peptidases. Acetylated peptides are quite useful because they help in determining the amino terminus of mature proteins that are formed after post-translational processing. Generally, all modifications that lead to a mass change can be analyzed by MS (for review see Krishna & Wold, 1993).

(*e*) *Quantitative mass spectrometry.* While the transcriptome experiments (cDNA or oligonucleotide array on chips) provide a comprehensive picture of the expression of individual genes in a given cell type or cell state, there are additional levels of controls during maturation of gene products. This is why sometimes there is a loose correlation between mRNA and protein expression levels. This bottleneck has been removed due to recent advances in MS techniques and instrumentation (Shevchenko *et al*, 1996; Oda *et al*, 1999). Now 'differential display proteomics' that makes use of MS is used to resolve many problems in proteomics, viz. knowledge of quantitative changes of protein expression in a cell, tissue, or body fluid, etc (Mann, 1999). Previously these changes used to be examined through straining intensities of proteins on gels, a labour-intensive method that is prone to error. Recently, stable isotope methods have been introduced and used successfully to access quantitative changes in protein expression (Han *et al*, 2001). The principle involved in this method involves incorporation of a stable isotope derivative in one of the two states to be compared. Stable isotope incorporation shifts the mass of the peptides by a predictable amount.

Mass spectrometry is not only used in genome and proteome analysis but also used routinely to trace the flow of organic compounds at the molecular level. It is also used for detecting metabolic pathways and to identify specific populations of microorganisms with the help of lipid biomarkers (specific compounds), characteristic of those populations. MS can also used for a study of diversity among microorganisms. For these purposes GC-MS is generally used (Zhang *et al*, 2002).

## Genomics/Proteomics Research in India Using Mass Spectrometry

There are several centres in India, where mass spectrometry is being used for the study of macromolecules. Keeping in view the importance of this tool for research, Indian Society for Mass Spectrometry (ISMAS) was established in 1978, with its headquarters at Bhabha Atomic Research Centre (BARC), Mumbai. This society, which has as many as 600 members, conducts a workshop on mass spectrometry every year. The society has also compiled a directory on mass spectrometry, which provides useful information about the types of mass spectrometers available in various laboratories in India and their various applications along with the names of scientists engaged in research.

In India, facilities for MALDI-TOF MS have been created at several centres including Centre for Cellular and Molecular Biology (CCMB), Hyderabad; Indian Institute of Science (IISc), Bangalore; Bhabha Atomic Research Centre (BARC), Mumbai; Rajeev Gandhi Centre for Biotechnology (RGCB), Thiruvananthapuram, etc. At CCMB, efforts are being made to characterize (i) differentially expressed proteins in cancerous cells, and (ii) antibacterial peptides from the skin of *Rana tigerina*. Similarly, IISc is working on maltose binding protein and the proteomics of renal disorders (Ganesh *et al*, 1999; Kumar *et al*, 2002).

## Conclusion

Mass spectrometry, and in particular MALDI-TOF MS, has become an essential tool for genomics and proteomics research. It provides for accuracy and high throughput, often avoiding the time-consuming gel electrophoresis. PCR amplification, which is so pervasive in molecular biology, now can also be dispensed with in some applications of mass spectrometry. The third generation molecular marker system i.e. SNP, which is becoming the molecular marker of choice in genomics research, and the precise identification of proteins and protein-protein interactions in proteomics research will make MS indispensable for future research involving genome and proteome analysis. In India also, facilities for mass spectrometry that are being developed and used in some laboratories, will be increasingly used in future for genomics and proteomics research.

## Acknowledgement

This review was written during senior author's tenure as CSIR Emeritus Scientist. Financial assistance provided by CSIR and the facilities provided by Head, Department of Agricultural Botany, CCS University are duly acknowledged.

## References

Barber M, Bordoli R S & Sedgewick R D, 1981. Fast atom bombardment of solids (F.A.B.): A new ion source for mass spectrometry. *J Chem Soc Chem Commun*, 325-327.

Blakeley C R & M L Vestal, 1983. Thermospray interface for liquid chromatography/mass spectrometry. *Anal Chem*, **55**, 750-754.

Braun A, Little D P, Reuter D, Muller-Mysok B & Koster H, 1997. Improved analysis of microsatellites using mass spectrometry. *Genomics*, **46**, 18-23.

Busch K L, Glish G L & McLuckey S A, 1988. Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry. VCH Publishers, Inc., New York.

Caprioli R M, 1989. Continuous-Flow Fast Atom Bombardment Mass Spectrometry. *in* Active Biological Molecules, edited by U Schlunegger. Springer-Verlag, Berlin. Pp 59-77.

Caprioli R M, 1994. On-Line LC/MS and CE/MS. *in* Biological Mass Spectrometry: Present and Future, edited by Matsuo T, Caprioli R M, Gross M, and Seyama Y. John Wiley and Sons, New York. Pp 75-100.

Cargile B J, McLuckey S A & Stephenson J L, 2001. Identification of bacteriophage MS2 coat protein from *E. coli* lysates via ion trap collisional of intact protein ions. *Anal Chem*, **73**, 1277-1285.

Chait B T, Wang R, Beavis R C & Kent S B H, 1993. Protein Ladder Sequencing. *Science*, **262**, 89-92.

Chalmers M J & Gaskell S J, 2000. Advances in mass spectrometry for proteome analysis. *Curr Opin Biotechnol*, **11**, 384-390.

Colby S M, King T B & Reilly J P, 1994. Improving the resolution of matrix-assisted laser desorption /ionization time-of-flight mass spectrometry by exploiting the correlation between ion position and velocity. *Rapid Commun Mass Spectrom*, **8**, 865-868.

Fei Z, Ono T & Smith L M, 1998. MALDI-TOF mass spectrometric typing of single nucleotide polymorphisms with mass-tagged ddNTPs. *Nucleic Acids Res*, **26**, 2827-2828.

Fenn J B, Mann M, Meng C K, Wong S F & Whitehouse C, 1989. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, **246**, 64-71.

Fischer E, 1959. Three-dimensional stabilization of charge carriers in a quadrupole field. *Z Phys*, **156**, 1-26.

Fitzgerald M C & Smith L M, 1995. Four common mutations of the cystathionine -synthase gene detected by multiplex PCR and Matrix-assisted laser desorption/ionization Time-of-Flight Mass Spectrometry. *Annu Rev Biophys Struct*, **24**, 117-140.

Fitzgerald M C, Zhu L & Smith L M, 1993. The analysis of mock DNA sequencing reactions using matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom*, **7**, 895-897.

Forbes A J, Mazur M T, Patel H M , Walsh C T & Kelleher N L, 2001. Toward efficient analysis of > 70 kDa protein with 100% sequence coverage. *Proteomics*, **1**, 927-933.

Fu D J, Tang K, Braun A, Reuter D, Darnhofer-Demar B, Little D P, O'Donnell M J, Cantor C R & Koster H, 1998. Sequencing exons 5 to 8 of the p53 gene by MALDI-TOF mass spectrometry. *Nature Biotechnol* **16**, 381-384.

Ganesh C, Banerjee A, Shah A & Varadarajen, 1999. Disordered N-terminal residues eject the folding thermodynamics and kinetics of maltose binding protein. *FEBS Lett*, **454**, 307-311.

Gordan D B, 2000. Mass Spectrometric Techniques. *in* Practical Biochemistry Principles and Techniques, edited by K Wilson and J Walker, 5th edn. Cambridge low price edition, U K 527-579.

Griffin T J, Tang W & Smith L M, 1997. Genetic analysis by peptide nucleic acid affinity MALDI-TOF mass spectrometry. *Nature Biotechnol*, **15**, 1368-1372.

Gupta P K, Roy J K & Prasad M, 2001. Single nucleotide polymorphism: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr Sci*, **80**, 524-535.

Haff L A & Smirnov I P, 1997. Multiplex genotyping of PCR products with MassTag-labeled primers. *Nucleic Acids Res*, **25**, 3749-3750.

Han D K, Eng J, Zhou H & Aebersold R, 2001. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nature Biotechnol*, **19**, 946-951.

Harrison A G, 1992. Chemical Ionization Mass Spectrometry, 2nd edn, CRC Press, Boca Raton, FL.

Hayduk E J, Choe L H & Lee K H, 2002. Proteomics tools in discovery-driven science. *Curr Sci*, **83**, 840-844.

Hillenkamp F, Karas M, Beavis R C & Chait B T, 1991. Matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem*, **63**, 1193A-1203A.

Horn D M, Zubarev R A & McLafferty E W, 2000a. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J Am Soc Mass Spectrom*, **11**, 330-332.

Horn D M, Zubarev R A & McLafferty E W, 2000b. Automated *de novo* sequencing of protein by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci USA*, **97**, 10313-10317.

Horn D M, Ge Y & McLafferty E W, 2000c. Activated ion electron capture dissociation for mass spectral sequencing of large (42 kDa) proteins. *Anal Chem*, **72**, 4778-4784.

Jiang-Baucom P, Butler J & Belgrader P, 1997. DNA Typing of human leukocyte antigen sequence polymorphisms by peptide nucleic acid probes and MALDI-TOF MS. *Anal Chem*, **69**, 4894-4898.

Karas M & Hillenkamp F, 1988. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*, **60**, 2299-2303.

Kelleher N L, Taylor S V, Grannis D, Knisland C, Chiu H J, Begley T P & McLafferty F W, 1998. Efficient sequence analysis of six genes products (7-74 kDa) from the *Escherichia coli* thiamine biosynthesis operon by tandem high-resolution mass spectrometry. *Protein Sci*, **7**, 4778-4784.

Kelleher N L, Lin H Y, Valaskovic G A, Aaserud D J, Fridriksson E K & McLafferty F W, 1999. Top-down versus bottom-up protein characterization by tandem high-resolution mass spectrometry. *J Am Chem Soc*, **121**, 806-812.

Kirpekar F, Nordhoff E, Larsen L K, Hillenkamp F, Kristiansen K & Roepstorff P, 1998. DNA sequence analysis by MALDI mass spectrometry. *Nucleic Acids Res*, **26**, 2554-2559.

Koster H, Tang K, Fu D J, Braun A, Van den Bloom D, Smith C L, Cotter R J & Cantor C R, 1996. A strategy for rapid and efficient DNA sequencing by mass spectrometry. *Nature Biotechnol*, **14**, 1123-1128.

Krebs S, Seichter D & Forster M, 2001. Genotyping of dinucleotide tandem repeats by MALDI mass spectrometry of ribozyme-cleaved RNA transcripts. *Nature Biotechnol*, **19**, 877-880.

Krishna R G & Wold F, 1993. Post-Translational Modifications of Proteins. *in* Methods in Protein Sequence Analysis, edited by K Imahori and F Sakiyama. Plenum Press, New York. Pp 167-171.

Kumar Y, Venkata Uppuluri N R, Babu K, Phadke K, Kumar P, Ballal S & Tatu U, 2002. Proteomics of renal disorders: Urinary proteome analysis by two-dimensional gel electrophoresis and MALDI-TOF mass spectrometry. *Curr Sci*, **82**, 655-663.

Leushner J, 2001. MALDI TOF: Mass spectrometry--An emerging platform for genomics and diagnostics. *Expert Rev Mol Diagn*, **1**, 11-18.

Li J & Assmann S M, 2000. Mass spectrometry: An essential tool in proteome analysis. *Plant Physiol*, **123**, 807-809.

Li J, Wang X-Q, Watson M B & Assmann S M, 2000. Regulation of abscisic acid-induced stomatal closure and anion channels by guard cell AAPK kinase. *Science*, **287**, 300-303.

Little D P, Cornish T J, O'Donnell M J, Braun A, Cotter R J & Koster H, 1997. MALDI on a chip: Analysis of arrays of low-femtomole to subfemtomole quantities of synthetic oligonucleotides and DNA diagnostic products dispensed by a piezoelectric pipet. *Anal Chem*, **69**, 4540-4546.

Loo J A, 1995. Bioanalytical Mass spectrometry: Many flavors to choose. *Bioconjugate Chem*, **6**, 644-665.

Mann M, 1999. Quantitative proteomics? *Nature Biotechnol*, **17**, 954-955.

Mann M, Hendrickson R C & Pandey A, 2001. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem*, **70**, 437-473.

McCloskey J A, 1990. Electron Ionization Mass Spectra of Trimethylsilyl Derivatives of Nucleosides. *in* Methods in Enzymology: Mass Spectrometry, edited by J A McCloskey. Academic Press, Inc., San Diego, CA, 193, 825-842.

Monforte J A & Becker C H, 1997. High-throughput DNA analysis by time-of-flight mass spectrometry. *Nature Med*, **3**, 360-362.

Mortz E, O'Connor P B, Roepstorff P, Kelleher N L, Wood T D, McLafferty F W & Mann M, 1996. Sequence tag identification of intact proteins by matching tandem mass spectral data against sequence databases. *Proc Natl Acad Sci USA*, **93**, 8264-8267.

Mouradian S, Rank D R & Smith L M, 1996. Analyzing sequencing reactions from bacteriophage M13 by MALDI mass spectrometry. *Rapid Commun Mass Spectrom*, **10**, 1475-1478.

Nelson R W, Nedelkov D & Tubbs K A, 2000. Biosensor chip mass spectrometry: A chip-based proteomics approach. *Electorphoresis*, **21**, 1155-1163.

Nordhoff E, Ingendoh A, Cramer R, Overberg A, Stahl B, Karas M, Hillenkamp F & Crain P F, 1992. Matrix-assisted laser desorption/ionization mass spectrometry of nucleic acids with wavelengths in the ultraviolet and infrared. *Rapid Commun Mass Spectrom*, **6**, 771-776.

Nordhoff E, Kirpekar F, Roepstorff P, Hahner S & Hillenkamp F, 1996. Matrix-assisted laser desorption/ionization mass spectrometry as a new method for the characterization of nucleic acids. *Trends Anal Chem*, **15**, 240-250.

Oda Y, Huang K, Cross F R, Cowburn D & Chait B T, 1999. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA*, **96**, 6591-6596.

Pandey A & Mann M, 2000. Proteomics to study genes and genome. *Nature (Lond)*, **405**, 837-846.

Paracchini S, Arredi B, Chalk R & Tayler-Smith C, 2002, Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res*, **30**, e27.

Patterson D H, Tarr G E, Regnier F E & Martin S A, 1995. C-terminal ladder sequencing via matrix-assisted laser desorption mass spectrometry coupled with carboxypeptidase Y time-dependent and concentration-dependent digestions. *Anal Chem*, **67**, 3971.

Reids G E, Wu J, Chrisman P A, Wells J M & McLuckey S A, 2001. Charge state-dependent sequence analysis of protonated ubiquitin ions via ion trap tandem mass spectrometry. *Anal Chem*, **73**, 3274-3281.

Roskey M T, P Juhasz *et al*, 1996. DNA sequencing by delaye laser dextraction-matrix-assisted desorption/ ionization time of flight mass spectrometry. *Proc Natl Acad Sci USA*, **93**, 4724-4729.

Ross P L & Belgrader P, 1997. Analysis of short tandem repeat polymorphism in human DNA by matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem*, **69**, 3966-3972.

Ross P, Hall L, Smirnov I & Haff L, 1998. High level multiplex genotyping by MALDI-TOF mass spectrometry. *Nature Biotechnol*, **16**, 1347-1351.

Shevchenko A, Jensen O N, Podtelejnikov A V, Sagliocco F, Wilm M, Vorm O, Mortensen P, Boucherie H & Mann M, 1996. Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proc Natl Acal Sci USA*, **93**, 14440-14445.

Smith L M, 1993. The future of DNA sequencing. *Science*, **262**, 530-532.

Spengler B, Kirsch D & Kaufmann R, 1991. Metastable decay of peptides and proteins in matrix-assisted laser desorption mass spectrometry. *Rapid Commun Mass Spectrom*, **5**, 198.

Stephenson J L, Cargile B J & McLuckey S A, 1999. Ion trap collisional activity of disulfide linkage intact and reduced multiply protonated polypeptides. *Rapid Commun Mass Spectrom*, **13**, 2040-2048.

Stephenson J L & McLuckey S A, 1996. Ion/ion proton transfer reaction for protein mixture analysis. *Anal Chem*, **68**, 4026-4032.

Stephenson Jr J L, McLuckey S A, Reid G E, Wells J M & Bundy J L, 2002. Ion/ion chemistry as a top-down approach for protein analysis. *Curr Opin Biotechnol*, **13**, 57-64.

Taranenko N I *et al*, 1996. Laser desorption mass spectrometry for point mutation detection. *Genet Anal Biomol Eng*, **13**, 87-94.

Taranenko N I, Chung C N, Zhu Y F, Allman S L, Golovlev V V, Isola N R, Martin S A, Haff L A & Chen C H, 1997. Matrix-assisted laser desorption/ionization for sequencing single-stranded and double-stranded DNA. *Rapid Commun Mass Spectrom*, **11**, 386-392.

Taranenko N I, Allman S L, Golovlev V V, Taranenko N V, Isola N R, & Chen C H, 1998. Mass spectrometry sequencing of DNAs longer than 100 nucleotides. *Nucleic Acids Res*, **26**, 2488-2490.

Thomson J J, 1913. Rays of positive electricity and their application to chemical analysis. Longmans, Green and Co., London.

Timothy J G & Smith L M, 2000. Single nucleotide polymorphism analysis by MALDI-TOF mass spectrometry. *Trends Biotechnol*, **18**, 77-83.

Tsuneyoshi T, Ishikawa K, Koga Y, Naito Y, Baba S, Terunuma H, Arakawa R & Prockop D J, 1997. Mass spectrometric gene diagnosis of one base substitution from PCR amplified human DNA. *Rapid Commun Mass Spectrom*, **11**, 719-722.

Vestal M L, 1990. Liquid Chromatography-Mass Spectrometry. *in* Methods in Enzymology, edited by J A McCloskey. Academic Press, Inc., San Diego, CA. Pp 107-130.

Vestal M L, Juhasz P & Martin S A, 1995. Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom*, **9**, 1044-1050.

Weiller G F, Djordjevic M J, Caraux G, Chen H & Weinman J J, 2001. A specialized proteomics database for comparing matrix-assisted desorption/ionization-time-of-flight mass spectrometry data of tryptic peptides with corresponding sequence database segments. *Proteomics*, **12**, 1489-1494.

Weinberger S R *et al*, 2000. Recent trends in protein biochip technology. *Pharmacogenomics*, **1**, 395-416.

Wells J M, Reid G E, Engel B J, Pan P & McLuckey S A, 2001. Dissociation reaction of gaseous ferro-, ferri-, and apo-cytochrome c ions. *J Am Soc Mass Spectrom*, **12**, 873-876.

Williams P, 1994. Time-of-flight mass spectrometry of DNA laser-ablated from frozen aqueous solutions: applications to the human genome project. *Int J Mass Spectrom Ion Processes*, **131**, 335-344.

Yamashita M & Fenn J B, 1984. Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem*, **88**, 4451-4459.

Yates J R III, 1998. Mass spectrometry and the age of the proteome. *J Mass Spectrom*, **33**, 1-19.

Zhang C L, 2002. Stable carbon isotopes of lipid biomarkers: Analysis of metabolites and metabolic fates of environmental microorganisms. *Curr Opin Biotechnol*, **13**, 25-30.