Mass Spectrometry for Metabolite Identification

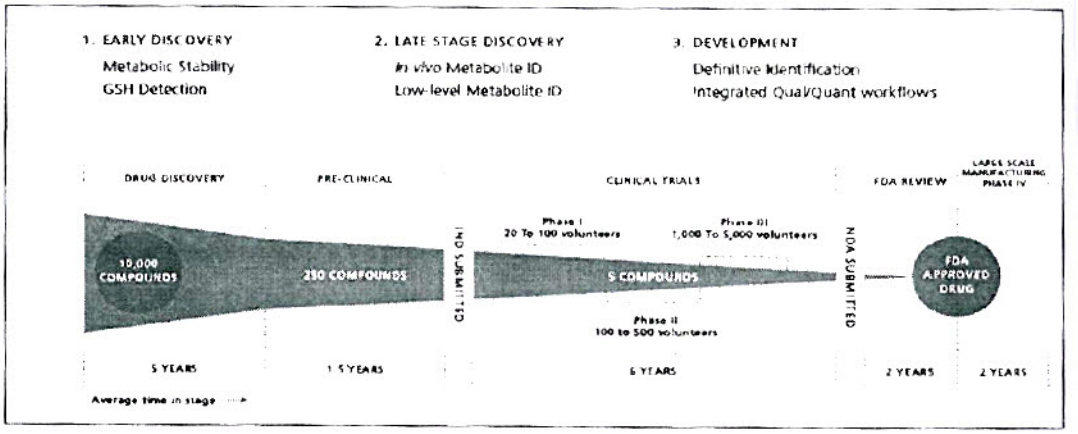
Dr. K. Ashok Kumar
Principal Scientist & Head, Fish Processing Division, ICAR-Central Institute of Fisheries Technology, Kochi - 682 029
Email: ashok1960@gmail.com

Introduction

Identification of the bio-transformation products or metabolites of drugs plays an important role in drug discovery and development. In other words, the knowledge of how a drug is absorbed, distributed, metabolized, and eliminated by the body has to be critically understood during the process of drug development. Especially, it is very essential to understand, whether any new intermediaries (metabolite) are produced during metabolism of the drug and further whether the compound so formed is toxic to the system or not. If metabolite present is toxic, it will jeopardize the entire drug development process. These intermediates can bind covalently to tissue macromolecules, such as lipids, nucleic acids, and proteins. Once a reactive metabolite has been identified, medicinal chemists can be made aware of which functional group or combination of functional groups should be avoided in the design of the next-generation of drug candidates. As a result, many pharmaceutical companies are now conducting metabolite identification studies in the early phases of drug candidate selection. Four separate but somewhat inter-related processes influence a drug's movement in the body *viz.* absorption (A), distribution (D), metabolism (M), and excretion (E). These four major components which influence a drug's level, its kinetics of exposure to tissues, and its performance as a drug are described.

Absorption is the process by which a drug molecule moves from the site of administration into the systemic circulation (bloodstream). When a drug is administered intravenously (IV), the drug is 100% absorbed (bioavailability is 100%). However, when a drug is administered *via.* other routes, such as orally, subcutaneously etc., its absorption (bioavailability) varies and is influenced by many factors, including the rate of dissolution,

metabolism before absorption and the ability to cross the gastrointestinal tract. Distribution is the process of a drug being carried *via* the bloodstream to its site of action, including extracellular fluids and/or cells of tissues and organs. Factors that affect a drug's distribution include blood flow, plasma protein binding, tissue binding, lipid solubility, pI/pKa, and membrane. Metabolism or bio-transformation is the process by which the body (human and animal) or a system (cell based or *in vitro*) breaks down and converts a drug generally via. oxidation, reduction, hydrolysis, hydration, and/or conjugation reactions into an active, inactive, or toxic chemical substance. Enzymes (e.g., cytochrome P450s) present in the liver are responsible for metabolizing many drugs. When a drug is administered intravenously (or other non oral routes such as intramuscular and sublingual), some of these metabolism pathways are avoided. Excretion/elimination is the process by which the irreversible removal (elimination) of a drug and/or its metabolites from the systemic circulation or from the site of measurement takes place. The process of elimination usually happens through the kidneys (urine) or the faeces. Unless excretion is complete, accumulation of drugs and/or metabolites can lead to adverse effects. Other elimination routes include the lungs (through exhalation), skin (through perspiration), saliva and mammary glands.



In early metabolic profiling in ADME studies, this bio-transformation is usually performed by automated metabolite identification strategies. In both metabolic profiling and metabolite identification strategies, mass spectrometer plays a prominent role.

Mass spectrometry Mass spectrometry is an analytical tool used for measuring the molecular mass of a sample. For large samples such as biomolecules, molecular masses can be measured within an accuracy of 0.01% of the total molecular mass of the sample i.e. within a 4 Daltons (Da) or atomic mass units (amu) error for a sample of

40,000 Da. This is sufficient to allow minor mass changes to be detected, e.g. the substitution of one amino acid for another or a post-translational modification. For small organic molecules the molecular mass can be measured to within an accuracy of 5 ppm or less, which is often sufficient to confirm the molecular formula of a compound.

Atomic and molecular masses are assigned relative to the mass of the carbon isotope, ^{12}C, whose atomic weight is defined as exactly 12. The actual mass of ^{12}C is 12 daltons, with one dalton is equal to $1.661\ 10^{-24}$ g. The mass of a molecule or an ion can be presented in daltons (Da) or kilodaltons (kDa). Mass spectrometry provides valuable information to a wide range of professionals: physicians, astronomers, and biologists, to name a few. More frequently mass spectrometers are used today in the following areas:

- **Biotechnology:** The analysis of proteins, peptides, oligonucleotides
- **Pharmaceutical:** Drug discovery, combinatorial chemistry, pharmacokinetics, drug metabolism
- **Clinical:** Neonatal screening, haemoglobin analysis, drug testing
- **Environmental:** PAHs, PCBs, water quality, food contamination
- **Geological:** Oil composition
- **Accurate molecular weight measurements:** Sample confirmation, determinaction of the purity of a sample, verification of amino acid substitutions, detection of post-translational modifications, calculation of the number of disulphide bridges
- **Reaction monitoring:** Monitoring enzyme reactions, chemical modification, protein digestion
- **Amino acid sequencing:** Sequence confirmation, *de novo* characterization of peptides, identification of proteins by database searching with a sequence "tag" from a proteolytic fragment
- **Oligonucleotide sequencing:** Characterization or quality control of oligonucleotides
- **Protein structure:** Protein folding monitored by H/D exchange, protein-ligand complex formation under physiological conditions, macromolecular structure determination

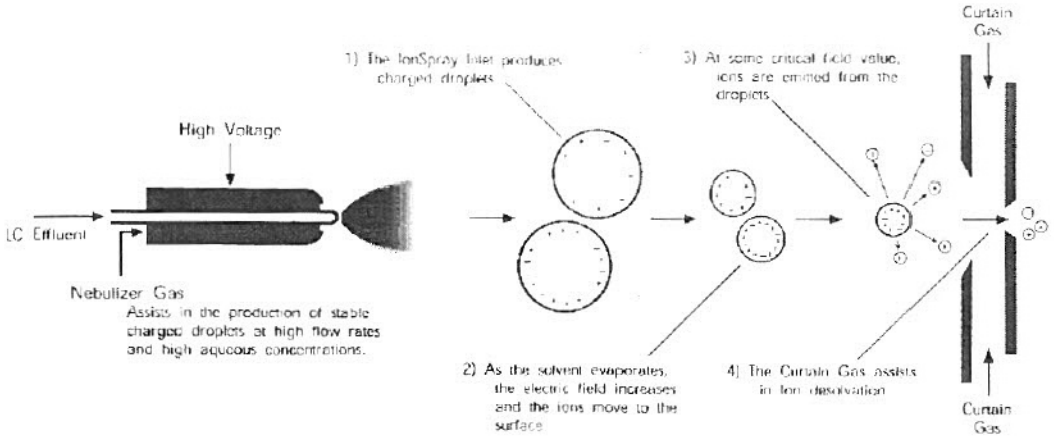
The Mass Spectrometer Mass spectrometer can be divided into three fundamental parts, namely the ionisation source, the analyser, and the detector. The sample has to be introduced into the ionisation source of the instrument. Once inside the ionisation source, the sample molecules are ionised, because ions are easier to manipulate than neutral molecules. These ions are extracted into the analyser region of the mass spectrometer where they are separated according to their mass (m)-to-charge (z) ratios (m/z). The separated ions are detected and this signal is sent to a data system where the m/z ratios are stored together with their relative abundance for presentation in the format of a m/z spectrum.

The analyser and detector of the mass spectrometer, and often the ionisation source too, are maintained under high vacuum to give the ions a reasonable chance of travelling from one end of the instrument to the other without any hindrance from air molecules. The entire operation of the mass spectrometer, and often the sample introduction process also, is under complete data system control on modern mass spectrometers.

Methods of sample ionization: Many ionisation methods are available and each has its own advantages and disadvantages. The ionisation method to be used should depend on the type of sample under investigation and the mass spectrometer available.

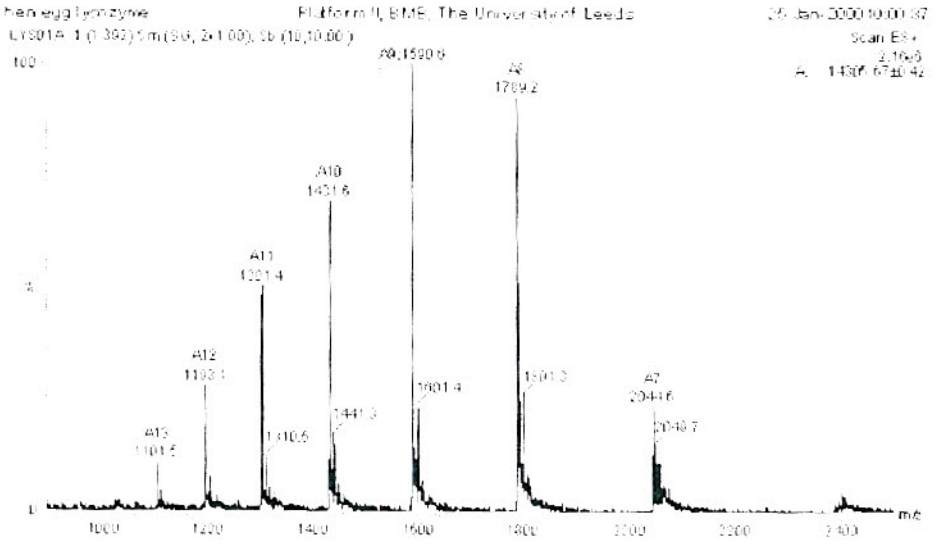
- Atmospheric Pressure Chemical Ionisation (APCI)
- Chemical Ionisation (CI)
- Electron Impact Ionisation (EI)
- Electrospray Ionisation (ESI)
- Fast Atom Bombardment (FAB)
- Field Desorption/Field Ionisation (FD/FI)
- Matrix Assisted Laser Desorption Ionisation (MALDI)
- Thermospray Ionisation (TSP)

Electrospray ionization: Electrospray Ionisation (ESI) is one of the Atmospheric Pressure Ionisation (API) techniques and is well-suited to the analysis of polar molecules ranging from less than 100 Da to more than 1,000,000 Da in molecular mass.



During standard electrospray ionisation, the sample is dissolved in a polar, volatile solvent and pumped through a narrow, stainless steel capillary (75-150 micrometers i.d.) at a flow rate of between 1 μL/min and 1 mL/min. A high voltage of 3 or 4 kV is applied to the tip of the capillary, which is situated within the ionisation source of the mass spectrometer, and as a consequence of this strong electric field, the sample emerging from the tip is dispersed into an aerosol of highly charged droplets, a process that is aided by a co-axially introduced nebulising gas flowing around the outside of the capillary. This gas, usually nitrogen, helps to direct the spray emerging from the capillary tip towards the mass spectrometer. The charged droplets diminish in size by solvent evaporation, assisted by a warm flow of nitrogen known as the drying gas which passes across the front of the ionisation source. Eventually charged sample ions, free from solvent, are released from the droplets, some of which pass through a sampling cone or orifice into an intermediate vacuum region, and from there through a small aperture into the analyser of the mass spectrometer, which is held under high vacuum. The lens voltages are optimized individually for each sample. Electrospray ionisation is a very sensitive analytical techniques but the sensitivity deteriorates with the presence of non-volatile buffers and other additives, which should be avoided as far as possible. Electrospray ionisation is known as a "soft" ionisation method as the sample is ionised by the addition or removal of a proton, with very little extra energy remaining to cause fragmentation of the sample ions.

Samples (M) with molecular weights greater than ca. 1200 Da give rise to multiple charged molecular-related ions such as $(M+nH)^{n+}$ in positive ionisation mode and $(M-nH)^{n-}$ in negative ionisation mode. Proteins have many suitable sites for protonation as all of the backbone amide nitrogen atoms could be protonated theoretically, as well as certain amino acid side chains such as lysine and arginine which contain primary amine functionalities. An example of multiple charging, which is practically unique to electrospray ionisation, is presented in the positive ionisation m/z spectrum of albumin.



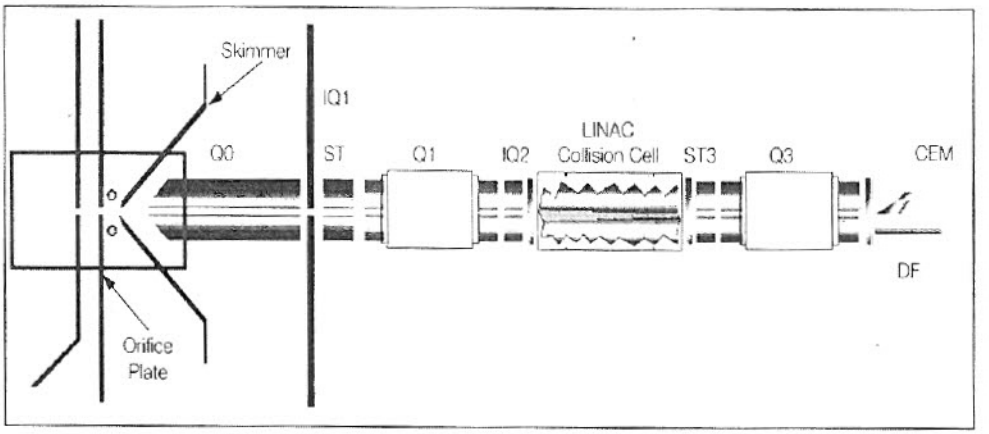
Positive ESI-MS m/z spectrum of albumin

Analysis and Separation of Sample Ions: The main function of the mass analyser is to separate, or resolve, the ions formed in the ionisation source of the mass spectrometer according to their mass-to-charge (m/z) ratios. There are a number of mass analysers currently available, the better known of which include quadrupoles, time-of-flight (TOF) analysers, magnetic sectors and both Fourier transform and quadrupole ion traps .

These mass analysers have different features, including the m/z range that can be covered, the mass accuracy, and the achievable resolution. The compatibility of different analysers with different ionisation methods varies. For example, all of the analysers listed above can be used in conjunction with electrospray ionisation, whereas MALDI is not usually coupled to a quadrupole analyser.

Tandem (MS-MS) mass spectrometers are instruments that have more than one analyser and so can be used for structural and sequencing studies. Two, three and four analysers have all been incorporated into commercially available tandem instruments, and the analysers do not necessarily have to be of the same type, in which case the instrument is a hybrid one. More popular tandem mass spectrometers include those of the quadrupole-quadrupole, magnetic sector-quadrupole, and more recently, the quadrupole-time-of-flight geometries. A tandem mass spectrometer is a mass spectrometer that has more than one analyser, in practice usually two. The two analysers are separated by a collision cell into which an inert gas (e.g. argon, xenon) is admitted to collide with the selected sample ions and bring about their fragmentation. The analysers can be of the same or of different types, the most common combinations being:

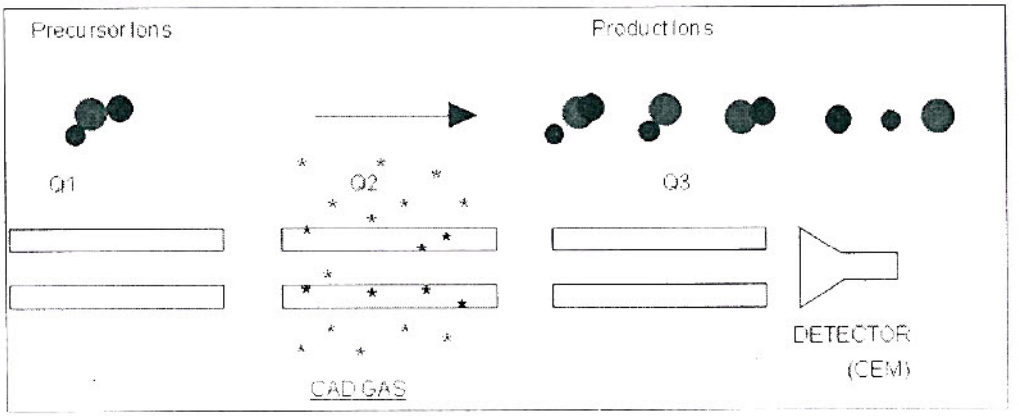
- Quadrupole - quadrupole
- Magnetic sector - quadrupole
- Magnetic sector - magnetic sector
- Quadrupole - time-of-flight.



Schematic diagram of a triple quadrupole mass spectrometer (MSMS)

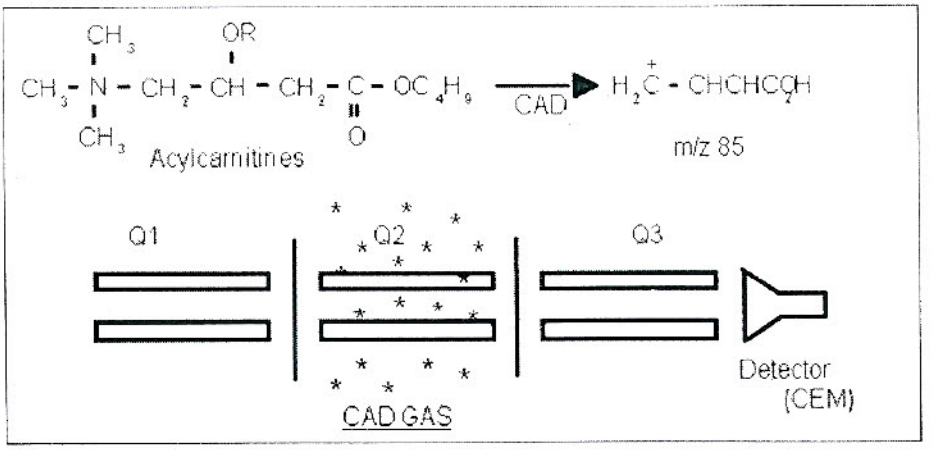
Fragmentation experiments can also be performed on certain single analyser mass spectrometers such as ion trap and time-of-flight instruments, the latter type using a post-source decay experiment to effect the fragmentation of sample ions. Matrix-assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS) system provides reliable mass information with higher sensitivity and throughput; and need for lesser individual assay optimization for molecular identification and characterization. In addition, the MALDI-TOF MS system has the added potential of detecting sub-microscopic deletions within chromosomal DNA.

Product or daughter ion scanning The first analyser is used to select user-specified sample ions arising from a particular component; usually the molecular-related i.e. $(M+H)^+$ or $(M-H)^-$ ions. These chosen ions passing into the collision cell are bombarded by the gas molecules which cause fragment ions to be formed, and these fragment ions are analysed i.e. separated according to their mass to charge ratios, by the second analyser. All the fragment ions arise directly from the precursor ions specified in the experiment, and thus produce a fingerprint pattern specific to the compound under investigation. This type of experiment is particularly useful for providing structural information concerning small organic molecules and for generating peptide sequence information.



Precursor or parent ion scanning

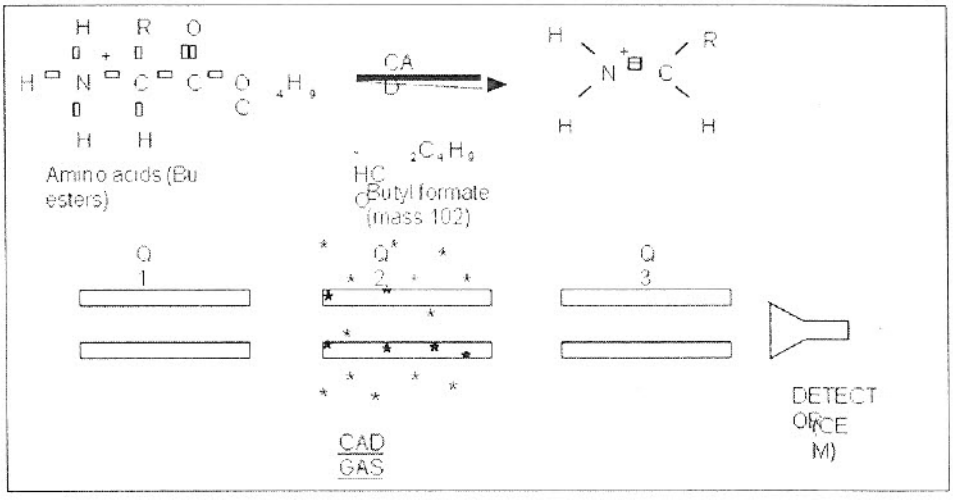
The first analyser allows the transmission of all sample ions, whilst the second analyser is set to monitor specific fragment ions, which are generated by bombardment of



the sample ions with the collision gas in the collision cell. This type of experiment is particularly useful for monitoring groups of compounds contained within a mixture which fragment to produce common fragment ions, e.g. glycosylated peptides in a tryptic digest mixture, aliphatic hydrocarbons in an oil sample, or glucuronide conjugates in urine.

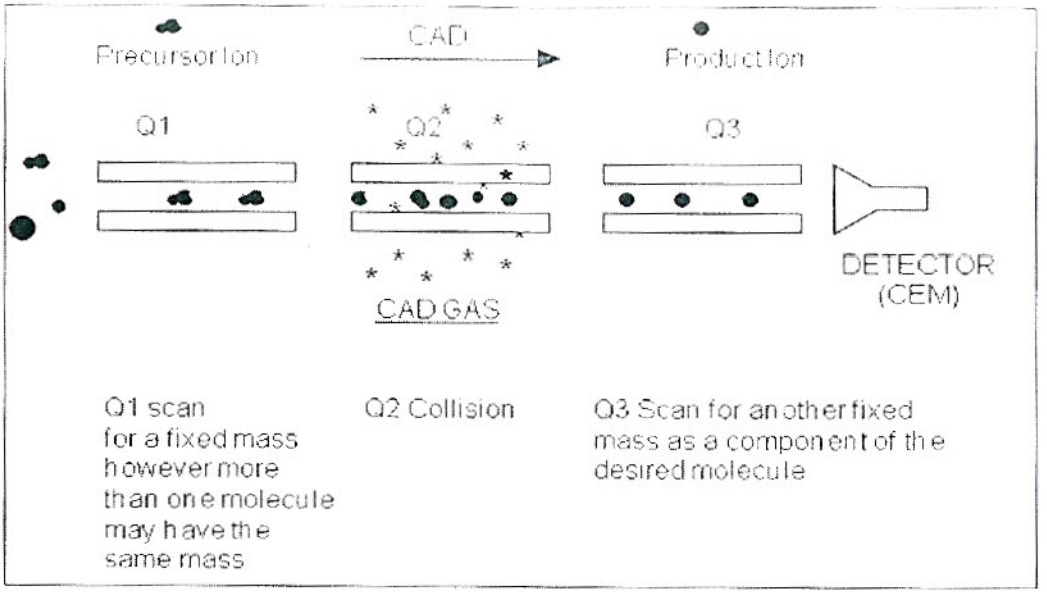
Constant neutral loss scanning

This involves both analysers scanning, or collecting data, across the whole m/z range, but the two are off-set so that the second analyser allows only those ions which differ by a certain number of mass units (equivalent to a neutral fragment) from the ions transmitted through the first analyser. This type of experiment could be used to monitor all of the carboxylic acids in a mixture. Carboxylic acids tend to fragment by losing a (neutral) molecule of Carbon dioxide, CO_2, which is equivalent to a loss of 44 Da or atomic mass units. All ions pass through the first analyser into the collision cell. The ions detected from the collision cell are those from which 44 Da have been lost.



Selected/multiple reaction monitoring

Both of the analysers are static in this case as user-selected specific ions are transmitted through the first analyser and user-selected specific fragments arising from these ions are measured by the second analyser. The compound under scrutiny must be known and have been well-characterised previously before this type of experiment is undertaken. This methodology is used to confirm unambiguously the presence of a compound in a matrix e.g. drug testing with blood or urine samples. It is not only a highly specific method, but also has very high sensitivity.



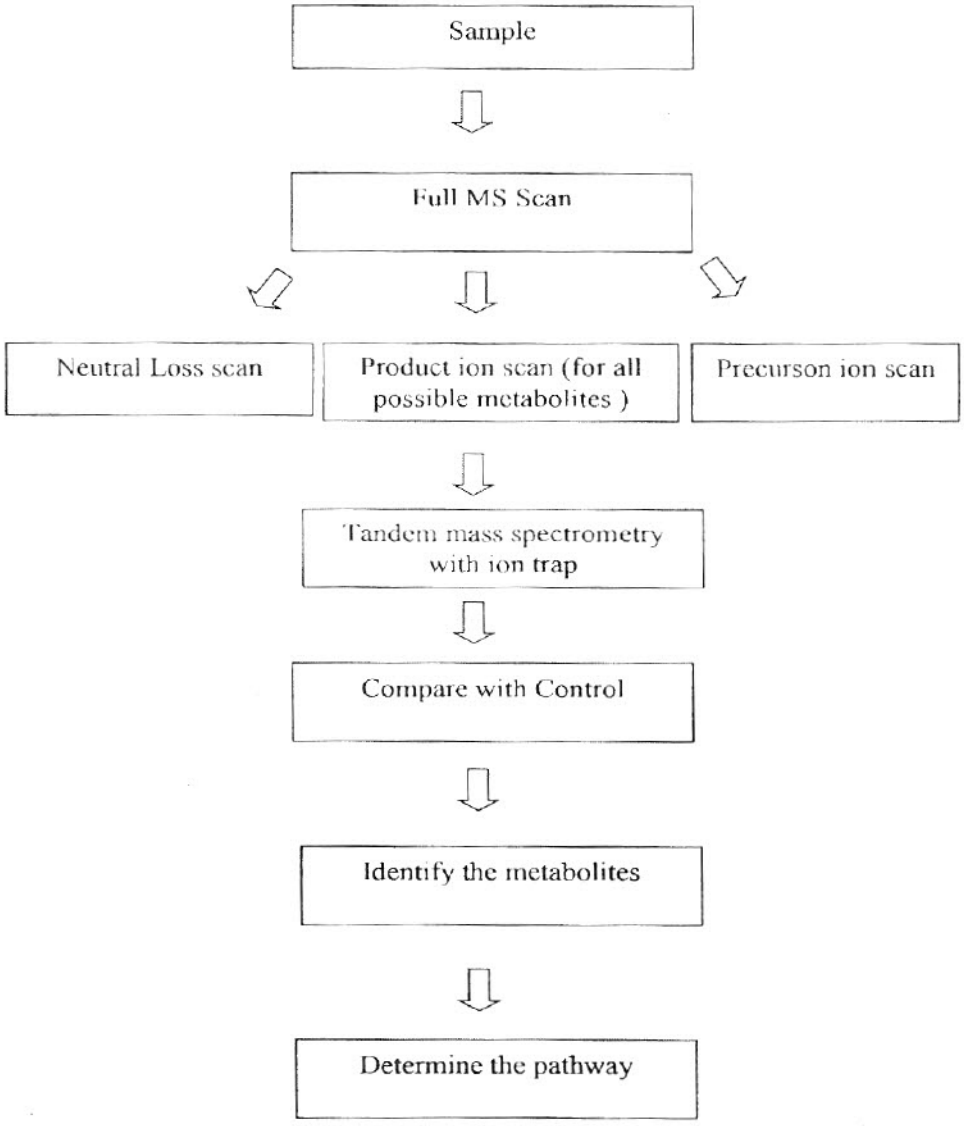
Mass Spectrometer for metabolite identification

The mass spectrometer was long considered a useful and challenging analytical tool but largely limited to the specialist users. The steady movement from specialist use to general use gained considerable speed in the 1990s, particularly due to the development of practical, sensitive liquid chromatography–mass spectrometry (LC–MS) interfaces and advances in front -end softwares. The rapid proliferation of quadrupole ion trap, linear ion trap, orbi trap, quadrupole mass filter, time-of-flight, and other types of mass spectrometers has impacted the industry from the earliest stages of disease determination through the final stages of clinical testing. Today, mass spectrometry has a profound influence on the direction and speed of drug discovery and development, especially in the area of drug metabolism (DM) and pharmacokinetics (PK).

Mass spectrometry is currently employed in all phases of drug discovery. Early in the discovery, target proteins are identified and characterized by MS followed by LC or two-dimensional gel electrophoresis separation. The make-up of an isolated protein is then determined by enzymatically digesting the protein and then analyzing the peptides by mass spectrometry. In almost all pharmaceutical companies, open-access MS laboratories have been set up to allow medicinal chemists to confirm and assess the purity of their synthesis or isolated products. Once the compounds or compound series are confirmed, high-throughput screening (HTS) assays are used to weed out compounds that do not show any activity towards a host (protein, ribonucleic acid (RNA),

deoxyribonuclete acid (DNA), etc.). MS has emerged as an ideal technique for the identification of such structurally diverse metabolites. When coupled with on-line HPLC, the technique is extremely robust, rapid, sensitive, and easily automated. Not surprisingly, LC/MS and LC/MS/MS have become the methods of choice for pharmacokinetic studies, yielding concentration versus time data for drug compounds from *in vivo* samples such as plasma.

Strategy for metabolite identification



Nevertheless, identifying metabolites remains a time-consuming process because a range of instrumental techniques and software applications are needed to obtain the appropriate data. The analyst must be quite experienced to handle these various instruments and software packages. Furthermore, just as in a police investigation, the data is rarely obvious or completely conclusive, but rather requires previous experience to unravel it. For all these reasons, interpreting the data is typically the largest bottleneck in metabolite identification.

The basic approach depicted above assumes that it is possible to predict numerous common alterations to the drug such as oxidation, glucorination, hydration, oxidative conjugation etc. The common alterations are given in the Table 1. Most of the times, the expected metabolites will be structurally very similar because a prospective lead compound's structure is 'fine-tuned' for better selectivity and potency toward the receptor of interest. Therefore, the scientist can easily understand the most common metabolic alterations to the parent structure and any novel modifications. This experience and information allows for a guided analysis with targeted searches for expected metabolites.

Production and Neutral loss analysis

In drug discovery, each new compound normally arises from a small alteration to an initial parent molecule. The final compound may bear very little similarity to that

Table 1. Common metabolic reaction and the shift in molecular weights of the parent drug

Metabolic reaction	Formula	Affect on molecular weight of parent drug (M)
Reduction	$+H_2$	M + 2
Methylation	$+CH_2$	M + 14
Hydroxylation	$+O$	M + 16
Hydrolysis of a ring	$+H_2O$	M + 18
Acetylation	$+C_2H_2O$	M + 42
Glycine conjugation	$+C_2H_3NO$	M + 57
Sulphate conjugation	$+SO_3$	M + 80
Taurine conjugation	$+C_2H_5NO_2S$	M + 107
S-Cysteine conjugation	$+C_3H_5NO_2S$	M + 119
Glucuronide conjugation	$+C_6H_8O_6$	M + 176
S-Glutathione conjugation	$+C_{10}H_{15}N_3O_6S$	M + 305
Alcohol to ketone	$-H_2$	M − 2
Demethylation	$-CH_2$	M − 14
Defluorination	$-F+H$	M − 17
Dechlorination	$-Cl+H$	M − 34

initial lead, but it has a lineage that stretches back to the initial compound. Common metabolic alterations can be predicted, and a list of expected metabolites can be compiled, on the basis of a previously analyzed series. By combining this list with a list of suspected metabolites identified by precursor and neutral-loss scan data, a series of ions can be targeted for product ion analysis. Any type of mass spectrometer capable of product ion scanning can be used at this point, including a triple quadrupole. Thus, all of the experiments discussed so far can be completed using a single instrument. However, to localize alterations within a molecule to a specific site, additional stages of tandem MS (MS^n) require an ion trap mass spectrometer.

Determining sites of modification

Multiple stages of MS can provide large amounts of structural information regarding each analyte, thereby allowing for a more detailed characterization of the metabolites. Completing MS^n experiments requires a mass spectrometer that can capture and store ions. While the ions are stored, they can be subjected to excitation and collisional fragmentation. The trapping instrument can then capture the resultant fragment ions, which can then be forced to undergo further fragmentation. The second-generation mass spectrum will now give structural information regarding the isolated fragment, allowing easier characterization of that ion. Because this procedure can be applied to each of the initial parent ion fragments, detailed structural information can be acquired rapidly.

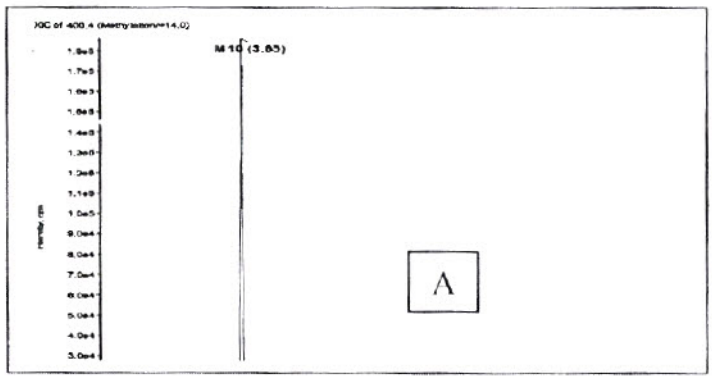
Although MS^n experiments can often pinpoint the site of modification very accurately, occasionally a metabolite fragments in a manner that does not provide the required information to identify the type of modification. This is typically the case when an expected fragment ion appears as a modified entity within the tandem MS/MS spectrum, but the modification cannot be explained by a common metabolic alteration (e.g., M+4 to a dimethyl benzene ring). In these cases, the answer can sometimes be determined by using accurate mass measurement, which in turn, allows the calculation of an empirical formula for the fragment.

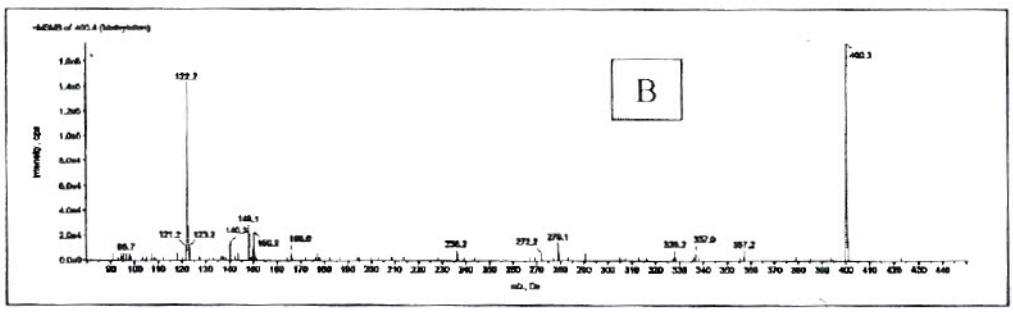
Identification of a dosed compound

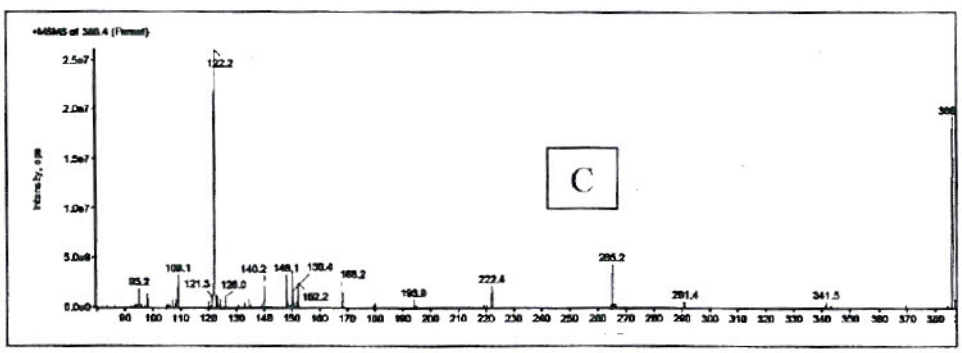
With the basics of instrumentation and techniques described, a full metabolite identification study of a compound dosed *in vivo* can be examined. Drugs to be tested have to be dosed intravenously to any test animals @ 2 mg/kg body weight and orally at 10 mg/kg. Urine and bile are collected over 24 h at regular time intervals. A control is also collected at 0 hour. No further sample preparation is performed. Minimal sample

cleanup is used because the nature, number, and concentrations of metabolites present are unknown, and it is therefore impossible to determine if any will be lost during a sample preparation procedure. Neat bile and urine samples were simply centrifuged to remove any particulates, and the supernatant is injected directly onto an analytical LC column, minimizing metabolite loss and decreasing sample preparation time.

Step 1. Collecting precursor ion scan and neutral-loss data: Before further analysis, tuning of MS with a pure sample of drug has to be performed. This is followed by



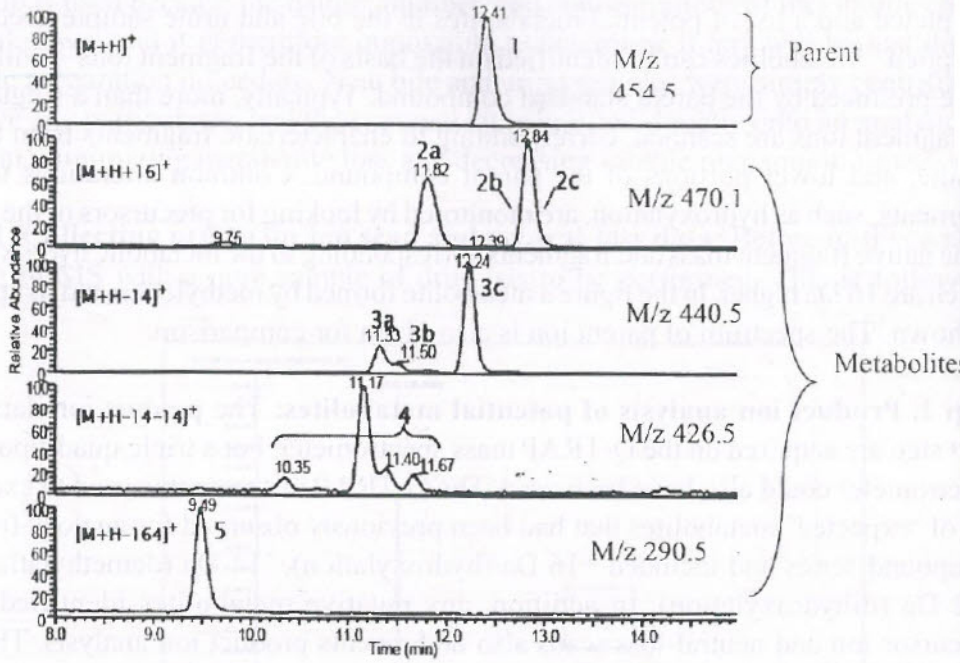




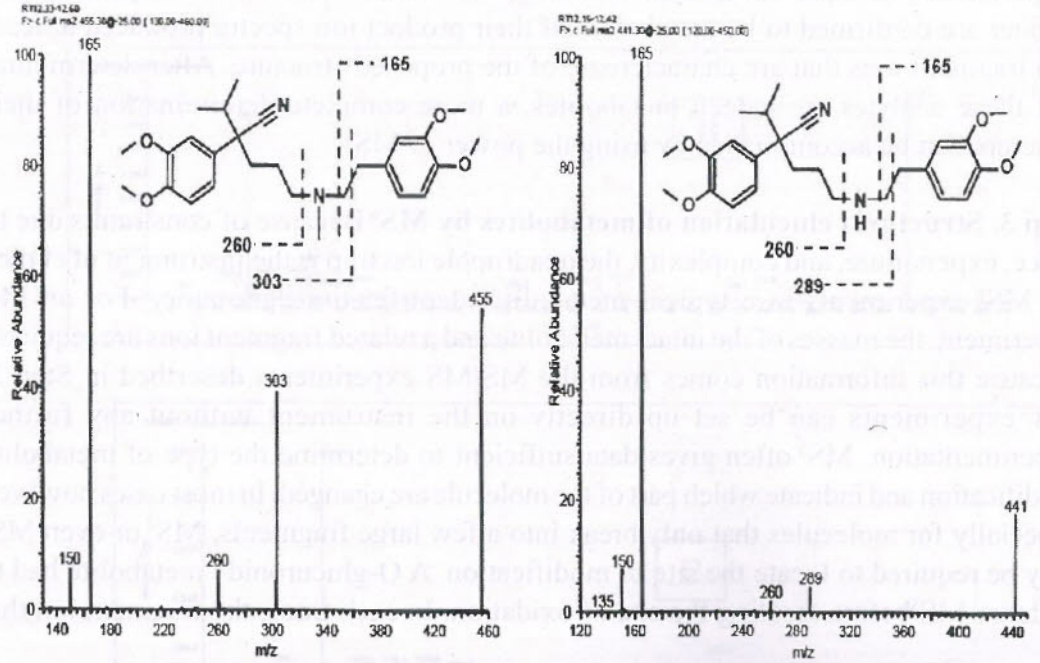
analysis of serial analysis of sample. Both precursor and neutral loss analysis has to be completed and a list of potential metabolites in the bile and urine sample needs to be compiled. Metabolites can be identified on the basis of the fragment ions' similarity to those produced by the parent standard compound. Typically, more than a single series of fragment ions are scanned, corresponding to characteristic fragments from the top, middle, and lower portions of the parent compound. Common alterations to these fragments, such as hydroxylation, are monitored by looking for precursors of the analyte at the native fragment mass and fragments corresponding to the metabolic hydroxylation, which are 16 Da higher. In the figure a metabolite formed by methylation and its spectrum is shown. The spectrum of parent ion is also given for comparison.

Step 2. Product ion analysis of potential metabolites: The product ion data in the next step are acquired on the Q-TRAP mass spectrometer, but a triple quadrupole mass spectrometer could also have been used. The Q-TRAP instrument is used to examine a list of 'expected' metabolites that had been previously observed for analogs from this compound series and included +16 Da (hydroxylation), -14 Da (demethylation), and +32 Da (dihydroxylation). In addition, any putative metabolites identified by the precursor ion and neutral-loss scans also underwents product ion analysis. The rapid scanning abilities of the Q-TRAP instrument allow several product ion experiments to be performed in rapid succession in a single LC run. Compounds analyzed in this manner are confirmed to be metabolites if their product ion spectra produced at least two fragment ions that are characteristic of the proposed structure. After determining that these analytes are indeed metabolites, a more complete determination of their structure can be accomplished by using the power of MS^n.

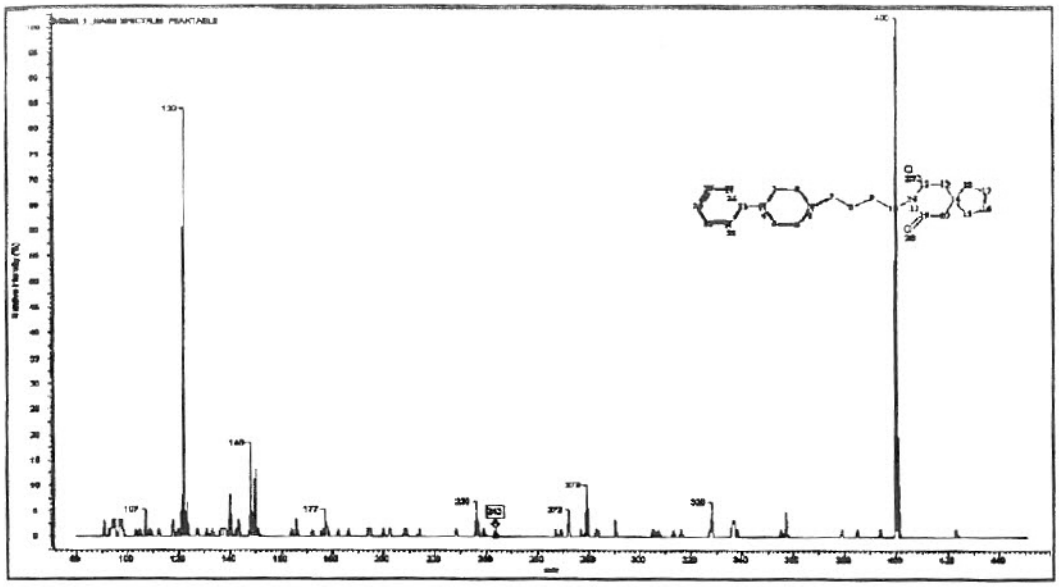
Step 3. Structural elucidation of metabolites by MS^n Because of constraints due to space, expenditure, and complexity, the quadrupole ion trap is the instrument of choice for MS^n experiments in a typical metabolite identification laboratory. For an MS experiment, the masses of the intact metabolite and a related fragment ions are required. Because this information comes from the MS/MS experiments described in Step 2, MS^3 experiments can be set up directly on the instrument without any further experimentation. MS^3 often gives data sufficient to determine the type of metabolite modification and indicate which part of the molecule are changed. In most cases however, especially for molecules that only break into a few large fragments, MS^4 or even MS^5 may be required to locate the site of modification. A O-glucuronide metabolite had to undergo MS^5before locating the site of oxidation. In each trace, the fragment ion that



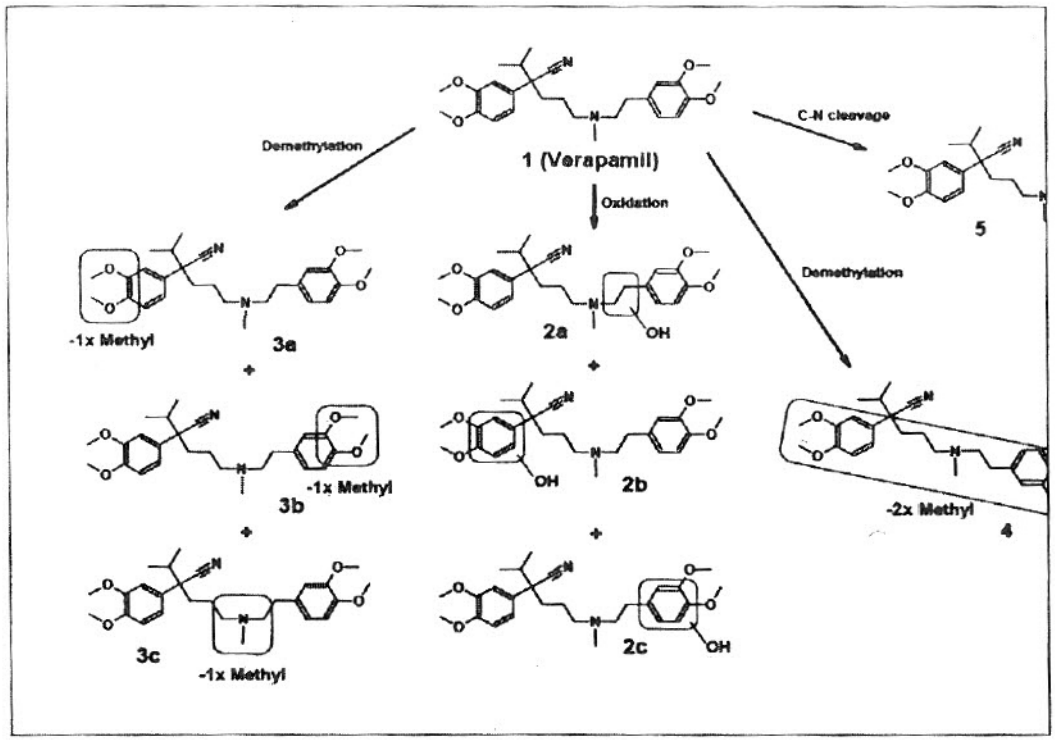
Selected Ion Monitoring of antibootic Verapamil (I) and its metabolites (2-5



Mass spectra of the antibiotic Verapamil and one of metabolites



With complete metabolite identified, entire metabolic pathway can be mapped as given ir the figure. These can used to investigate the toxicity and other details.



will undergo additional fragmentation is highlighted. Thus, this technique quickly pinpoints a very small area that has been altered.

Despite the obvious versatility and utility of MS^n to provide detailed structural information, the exact structure of the metabolite is still impossible to determine in certain cases. MS^n will often provide only enough information to narrow down the choices to two or three possible structures. When MS analysis fails to provide absolute identification of analytes or their isomers, LC/NMR can be used. Despite improvements made by these approaches, physical interpretation of the data is still a major bottleneck in the metabolite identification process, because it is so time consuming. These new approaches to data acquisition have led to many more samples being run in the same time period, which in turn, increase the amount of data. Software that can reduce operator workload by using a series of criteria to analyze data and report apparent metabolites will significantly improve throughput in metabolite identification.

Instrument manufacturers are creating software packages that help the metabolite identification process. Although these packages are simple at present, the impetus is growing rapidly, and software such as 'Metabolynx' (Micromass), 'Light Sight' (AB Sciex), and the new 'Metabolite Data Browser' (Thermo Finnigan) are helping reduce the workload of analysts. These packages automatically perform functions that researchers presently complete by hand such as the background subtraction of a control data file or the application of an isotope cluster analysis for chlorine- and bromine-containing compounds. At a minimum, the software searches for a list of expected metabolites from full-scan data and returns a list of possible hits. This allows numerous 'extraneous noise' responses to be automatically discarded.

As these packages evolve, they will handle correlation analysis of MS/MS data and data-dependent MS/MS acquisition of potential metabolites. All of these factors aid in reducing the data set that the analyst has to examine, thereby increasing throughput. The new direction for metabolite identification is the integration of automated data interpretation with LC/MS/MS in combination with the protocols described in this chapter. This approach will significantly accelerate metabolite identification.