

Advances In Structure Elucidation Of Small Molecules Using Mass Spectrometry

Vishalkumar Shashikant Modi

Aarti Birendra Yadav

SSR College of Pharmacy, Silvassa

Abstract: Structural elucidation of small molecules using mass spectrometry structural elucidation of small molecules plays an important role in modern life sciences and bioanalytical approaches. This review covers different soft and hard ionization techniques and figures of merit for modern mass spectrometers, such as mass resolving power, mass accuracy, isotopic abundance accuracy, accurate mass multiple-stage. The latter part discusses mass spectral data handling strategies, which includes background and noise subtraction, adduct formation and detection, charge state determination, accurate mass measurements, elemental composition determinations and complex data-dependent setups with ion maps and ion trees. The importance of mass spectral library search algorithms for tandem mass spectra and multiple-stage MS(n) mass spectra as well as mass spectral tree libraries that combine multiple-stage mass spectra are outlined. The successive chapter discusses mass spectral fragmentation pathways, biotransformation reactions and drug metabolism studies, the mass spectral simulation and generation of in silico mass spectra, expert systems for mass spectral interpretation, and the use of computational chemistry to explain gas-phase phenomena. A single chapter discusses data handling for hyphenated approaches including mass spectral deconvolution for clean mass spectra, cheminformatics approaches and structure retention relationships and retention index predictions for gas and liquid chromatography. The last section reviews the current state of electronic data sharing of mass spectra and discusses the importance of software development for the advancement of structure elucidation of small molecules.

Keywords: Structure elucidation, Mass spectrometry, Tandem mass spectra, Fragmentation prediction, Mass spectral interpretation, Mass spectral library search, Multistage tandem mass spectrometry

I. INTRODUCTION

Mass spectrometry is a standard technique for the analytical investigation of molecules and complex mixtures. It is important in determining the elemental composition of a molecule and in gaining partial structural insights using mass spectral fragmentations. The final structure confirmation of an unknown organic compound is always performed with a set of independent methods such as one- (1D) and two-dimensional (2D) nuclear magnetic resonance spectroscopy (NMR) or infrared spectroscopy and X-ray crystallography and other spectroscopic methods. The term structure elucidation usually refers to full de novo structure identification, and it results in a complete molecular connection table with correct

stereochemical assignments. Such an identification process without any assumptions or pre-knowledge is commonly the domain of nuclear magnetic resonance spectroscopy. The term dereplication often refers to the rediscovery of known natural products by means of mass spectral library search or the interpretation of known mass spectral fragmentations.

II. SCOPE OF THIS REVIEW

The review investigates theoretical and experimental structure elucidation techniques using mass spectrometry for organic molecules with a molecular mass less than 2,000 Da. The review covers newer techniques within the last 10–15

years; if none were available, then older material was included. Hyphenated separation techniques (gas chromatography coupled to mass spectrometry (GC-MS) and liquid chromatography coupled to mass spectrometry (LC-MS)) are covered due to the close relationship of those techniques with mass spectrometry. The term “small molecules,” used throughout this review, thus refers to all small molecules excluding peptides.

MASS SPECTRAL INSTRUMENTATION AND IONIZATION TECHNIQUES

The history of commercial mass spectrometry instrumentation covers more than 40 years. Brunnee covers the principles of common mass analyzers in a vibrant 1987 review. Gelpi discusses over 130 different mass spectrometers built since 1965 in a series of two reviews. Only one totally new mass spectrometer type, the Orbitrap analyzer, has been developed lately. Nevertheless, many new hybrid approaches, among them ion mobility coupled to time-of-flight (TOF) mass spectrometers have been introduced to the market recently. A series of ionization techniques and figures of merit for mass spectrometers will be discussed in the proceeding paragraphs.

SOFT AND HARD IONIZATION TECHNIQUES

Electron ionization (EI) at 70 eV is historically seen as the oldest ionization technique for small-molecule investigations. Because of the selected constant ionization energy, this technique results in consistent and fragment-rich mass spectra. These mass spectra can be easily used for a mass spectral library search. Electron ionization is commonly used for GC-MS setups. A major disadvantage of mass spectra obtained under EI conditions is the low abundant or missing molecular ion. An abundant molecular ion however is needed for the calculation of elemental compositions. Chemical ionization (CI) is a soft ionization technique mostly used in GC-MS setups to obtain molecular ion information. Supersonic molecular beam interfaces provide the ability to obtain fragment rich electron ionization spectra together with abundant molecular ions. The introduction of electrospray ionization (ESI) was a major breakthrough for the analysis of intact and large biomolecules. ESI is now the ionization method of choice for LC-MS in many laboratories worldwide. Additionally, nanoelectrospray (nanoESI) and chip based nanoelectrospray ionization have been advanced during recent years. The infusion of nanoliters of solvents using nano ESI allows for sustained analysis over long sample times with a minimum of sample material and increased sensitivity. These long infusions times are needed for structural identifications from data-dependent MS_n fragmentations obtained by ion trap mass spectrometers. The use of a new spray nozzle for each injection prevents cross-contaminations (see Fig. 1) especially when multiple compounds are infused from 396 well plates. Recently, multi-nanoelectrospray emitters (nanoESI) have been developed, which may further enhance ion production and increase the dynamic range (see Fig. 2)

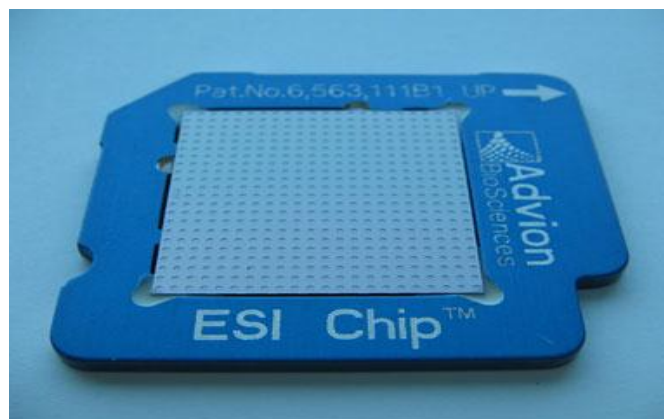


Figure 1: Chip-based nanoelectrospray allows for sensitive and contamination-free mass spectral infusions

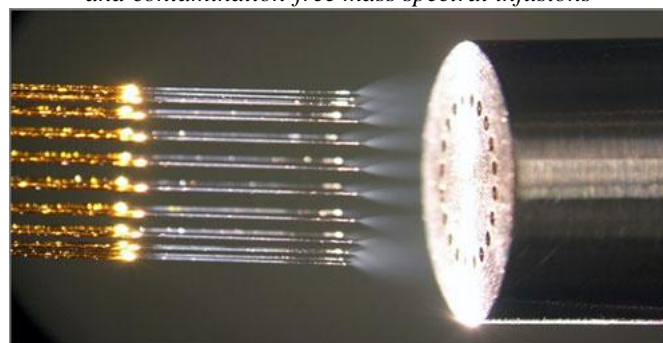


Figure 2: Nanoelectrospray emitter array for enhanced sensitivity of electrospray ionization mass spectrometry

Field desorption and field ionization are also soft ionization techniques, and both produce abundant molecular ions with few fragment ions. Direct analysis in real time (DART) is an ambient ionization technique and allows for the real time analysis of the sample. The DART source has been widely used in “open access/walk-up” laboratories together with robotic sample handling. Techniques for sampling molecules from surfaces have been extensively reviewed as well. Secondary ion mass spectrometry (SIMS) and MALDI are used for mass spectrometric imaging, a new and exciting technology to gain spatial and structural insights from tissues and organs. Several new surface-based ionization techniques including desorption electrospray ionization, desorption ionization on silicon, and nanostructure initiator mass spectrometry have been developed recently. Multi-mode or simultaneous ion sources, as well as the pulsed and parallel use of different ionization techniques, are helpful to shorten analysis time and to obtain structural information from a wide range of different substance classes (see Fig. 3). Although simultaneous positive and negative polarity switching is available within many ion source designs, the commercialization of dual- or multi-mode ion sources applying different ionization techniques is a more recent development.

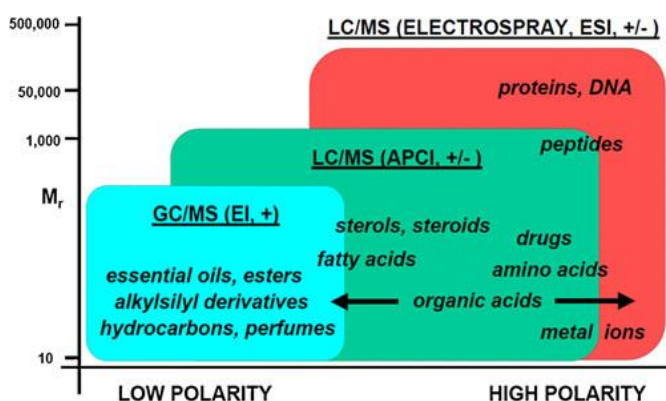


Figure 3: Coverage of molecule classes with different ionization methods

III. FIGURES OF MERIT OF MASS SPECTROMETERS

Mass spectrometers are typically designed for specific analytical aims: ion trap mass spectrometers as versatile instruments, quadrupole mass spectrometers as general work horses, triple quadrupole mass spectrometers as very sensitive instruments for targeted analysis, and Fourier transform instruments for measurements requiring high resolving power and high mass accuracy. In addition to their technical and instrumental design, mass spectrometers can be classified using specific figures of merit. These figures of merit are combinations of hardware, software, and customer experience indicators. High mass resolving power is needed to resolve overlapping interferences by mass spectrometry only (see Fig. 4). Up to one million resolving power can be achieved routinely with current commercially available Fourier transform ion cyclotron resonance (FT-ICR-MS) instruments. A series of “world records” achieved by FT-ICR-MS has been recorded. Hybrid instruments especially allow for the acquisition of high-resolution tandem mass spectra used for natural product structure elucidation. One drawback of FT-ICR-MS and Orbitrap instruments is the higher cycle time to acquire high-resolution broad band mass spectra. At one million resolving power (FT-ICR-MS), a single scan can take up to 2 s or longer. New high-field Orbitrap analyzers can now reach resolving power in excess of 350,000 at m/z 524 (full width at half maximum). Modern TOF and Q-TOF instrument are routinely capable of higher than 10,000 mass resolving power with the latest instruments reaching up to 40,000 resolving power. When coupled to ultra performance liquid chromatography and comprehensive two-dimensional GC \times GC, the data acquisition rate (scan speed) and duty cycle of the mass selective detector are very important. The chromatographic peak width can be around 2–5 s or lower, and there needs to be enough time to perform additional data-dependent tandem mass spectra (MS/MS) or MS $_n$ scans. Several new TOF and hybrid quadrupole-TOF and iontrap-TOF mass analyzers have been introduced into the market to obtain accurate masses at the MS $_n$ level at a very high data acquisition rate. High mass accuracy together with high isotopic abundance accuracy is generally important to obtain only few molecular formula candidates from an accurate mass measurement. Low machine maintenance and high robustness of the instrument operating under different temperatures and

humidity ranges in high throughput manner are additional important aspects. The software as one of the cornerstones for successful compound identification is just as important as the instrument itself. Fast software bug fixes, uncomplicated software updates, easy-to-use graphical user interfaces, and responsive soft-ware support are sometimes more important than certain instrument parameters.

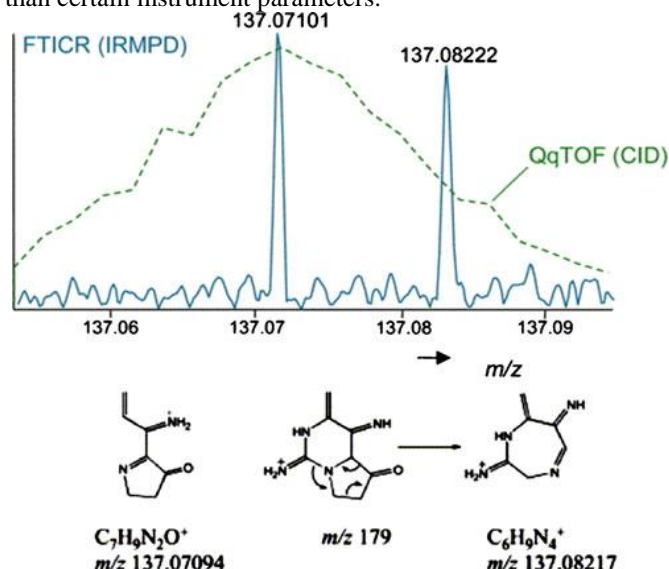


Figure 4: The importance of mass resolving power showing a high-resolution FT-ICR-MS spectrum with low resolution Q-TOF mass spectrum. Only the high resolution instrument can resolve peaks with 0.0112 Da difference

IV. TANDEM MASS SPECTROMETERS AND MODES OF OPERATION

Ion trapping instruments such as quadrupole ion traps and FT-ICR mass spectrometers can be used to create tandem mass spectra, and multiple-stage MS $_n$ experiments can be performed without instrument modification or couplings of different mass analyzers. Hybrid mass spectrometers are favorable for structural elucidation because they either increase the total peak resolution or they introduce another separation dimension. The different modes of operation, which include precursor ion scans, product ion scans, neutral loss scans and selected reaction monitoring, are discussed in De Hoffmann. Multiple precursor ions can be manually selected or the software can acquire tandem mass spectra when a certain peak abundance or signal/noise ratio is exceeded. For example, electrospray ionization with ion mobility mass spectrometry coupled to time-of-flight mass spectrometry (ESI-IMMS-TOF-MS) was used for metabolic profiling of *Escherichia coli* metabolites, phospholipid, and drug analysis.

V. ION ACTIVATION MODES

Collision-induced dissociation (CID), or collisionally activated dissociation, is the most common technique to obtain tandem mass spectra. Precursor ion stability and internal energy under CID have been previously discussed. A series of new fragmentation modes are aimed at improved protein and

peptide identification rates by creating more specific fragmentations. These modes include electron capture dissociation (ECD), electron transfer dissociation, and infrared multiphoton dissociation. They are not fully exploited yet for small-molecule applications outside proteomics.

VI. TWO-DIMENSIONAL, THREE-DIMENSIONAL, HYBRID, AND ORTHOGONAL CHROMATOGRAPHIC APPROACHES

On the chromatography side, the usual aim is directed at increasing the peak resolution, which therefore provides a better separation of overlapping compound peaks. The peak capacity can be increased by using different selective chromatographic phases in a two-dimensional or multicolumn setup. These approaches are known for liquid chromatography and prominently used for protein identification by coupling an ion exchange column together with a reversed phase column, which coined the term multidimensional protein identification technology. The detector must be a fast scanning detector with a high acquisition rate and an example of this is a time-of-flight mass analyzer. Sampling rates are usually between 100 and 200 spectra per second for GC×GC-TOF-MS instruments. Miniaturization and the introduction of chip-based liquid chromatography play a major role in high throughput methods.

VII. MASS SPECTRAL DATA HANDLING

The following section discusses basic steps that have to be performed to obtain clean and background free mass spectra. Charge state deconvolution, accurate mass measurements, and software algorithms for elemental composition calculations are reviewed. Certain hardware specific setups are discussed when required.

VIII. BACKGROUND AND NOISE SUBTRACTION

Automatic background and noise subtraction are standard techniques to obtain clean and interference free mass spectra. The Biller–Biemann algorithm or similar algorithms by Dromey et al. have been in use for more than 30 years. It is generally advisable to perform blank or solvent runs to obtain possible noise or contamination data. These infusion mass spectra or complete LCMS and GC-MS runs must be subtracted from the real sample data. Most modern mass spectrometry software tools have inbuilt algorithms to perform these tasks. Many of the mentioned algorithms have been developed for EI (70 eV) mass spectra. Several approaches have been introduced with the CODA algorithm of Windig et al. for ESI and LC-MS data, and similar methods have been applied in drug discovery studies. A more detailed discussion about automated approaches is covered in the mass spectral deconvolution and biotransformation sections. Adduct formation and detection Ionization techniques such as CI, MALDI, ESI, or APCI show not only single adduct ions but also sets of multiple adducts. The process of adduct formation

can be studied using heuristic and computational methods. Solvent and buffer constitution, pKa, pH, substance proton donor and acceptor properties, and gas-phase acidities influence the formation of adducts. Different adducts also can result in different fragmentation pathways. The correct adduct ion must be detected in order to obtain the accurate mass of the neutral molecule. One possible solution is to increase the concentration of specific ions in the liquid phase to obtain preferably those adducts. When analyzing lipids, lithium is used as modifier to obtain characteristic $[M+Li]^+$ ions. An extended list of common electrospray adducts, including $[M+H]^+$, $[M+NH_4]^+$, $[M+Na]^+$ and $[M-H]^-$, has been prepared. In case of MALDI, metal cation adducts $[M+Na]^+$ and $[M+K]^+$ are often observed. Software tools such as CAMERA and IntelliXtract, and tools for infusion spectra can help detect adduct ions in mass spectra automatically. Currently, no software exists that can predict adduct probabilities based on a given compound structure for a specified ionization mode (CI, ESI, APCI, and APPI).

IX. CHARGE STATE DECONVOLUTION

Charge state determinations play an important role in proteomics but are less frequently required in small-molecule investigations. Many small organic molecules are usually singly charged. Certain molecule classes, such as cardiolipins, may occur as singly and doubly charged ions. The occurrence of multiply or doubly charged ions can be influenced by buffer concentration, analytes concentration, amount of organic modifier, or flow rate. Open-source software tools, such as Decon2LS, exist (see Fig. 5), which can automatically determine charge states. Most vendor mass spectrometry software has charge state determinations included.

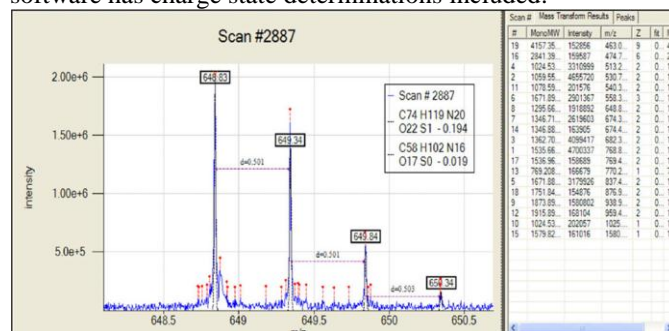


Figure 5: Change state deconvolution with the freely available software Decon2LS

X. ACCURATE MASS MEASUREMENTS

Accurate masses and isotope abundances are reported in an IUPAC report. The statistical evaluation of measured mass accuracies should include the proper terminology and basic statistic tests. An inter comparison study from 45 laboratories showed that FT-MS and magnetic sector field instruments in peak matching mode routinely achieved less than 1 ppm mass accuracy. Quadrupole-TOF, TOF, and magnetic sector field instruments in magnet scan mode achieved between 5 and 10 ppm. Newer publications reported that time-of-flight instruments can reach around 1 ppm or even sub-ppm mass

accuracies. Orbitrap technology in hybrid mode can routinely reach sub-ppm mass accuracy and in non-hybrid mode less than 2 ppm. The mass error introduced by this calculational error can be as high as 5 ppm at 100m/z (see Fig. 6). The red line marks 300 ppb, which can be obtained from broadband FT-ICR-MS experiments. The current accurate electron mass is reported as: $m(e^-) = 0.00054857990924$ u. The reported mass error on a LTQ-Orbitrap dropped from $\pm 1-2$ ppm to an absolute median error of 0.21 ppm. The process of selected ion monitoring (SIM) stitching was investigated. The authors concluded that an average mass error of 0.18 ppm could be obtained routinely on a high-resolution FT-ICR mass spectrometer. If instruments are uncalibrated or out of tune, then an automated post-calibration routine can be used to remove systematic precursor mass errors. The authors' reason that in case of sample overload, the automatic gain control system (AGC) is not able to control the optimal number of ions to inject into the Orbitrap cell, which finally results in space charge effects causing noticeable systematic mass errors.

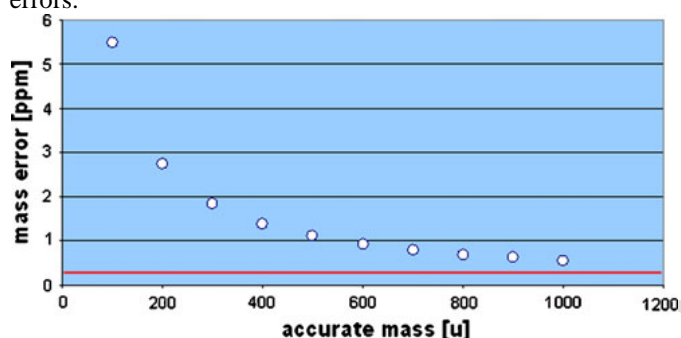


Figure 6: A mass error upto 5ppm is the penalty if the electron mass is not accurately included in accurate mass calculations. The lower red line marks 0.3 ppm mass accuracy, which can be reach by FT-ICR-MS

These algorithms use an internal calibrant that is later used for post-calibration of mass accuracy errors. If data are obtained in centroid mode or stick mode, then no such post-correction can be performed. A correction for spectral accuracy can also be performed with high-resolution data. Artificial neural network calibration in conjunction with AGC and better peak centroiding can improve the mass accuracy on FT-MS instruments to reach 100 ppb for certain experiments. Several unit mass resolution instruments, including ion traps and triple quadrupole instruments, allow a hardware-based high-resolution or an ultra-zoom scan. The resolving power usually can be increased by one order of magnitude, or from 1,000 resolving power to 10,000 resolving power. However the m/z scan range is usually very limited, and the duty cycle is high for enhanced resolution scans.

XI. ISOTOPE ABUNDANCE MEASUREMENTS AND ISOTOPIC PATTERN CALCULATIONS

The isotopic abundances of common monoisotopic (F, Na, P, and I) or polyisotopic (H, C, N, O, S, Cl, and Br) elements are listed. Isotopic abundances are measured and have been utilized in mass spectrometric measurements since the beginning of mass spectrometry. Sensitive and accurate method for isotopic abundance measurements is accelerator

mass spectrometry, and this method is used for age determination, forensics, and food monitoring. Its precision is around 0.05% for the measurement of the $^{13}\text{C}/^{12}\text{C}$ ratio requiring total combustion of the sample. The availability of commodity mass spectrometers delivering isotopic abundance errors less than $\pm 5\%$ was utilized for LCMS-based screening approaches and environmental screening applications. To filter or match elemental compositions according to their experimental isotopic abundances, the high- or low resolution isotopic envelopes of molecular formulas must be calculated. Several algorithms have been proposed to calculate the isotopic fine structures and allow the modeling of Gaussian peak shapes according to the selected resolving power of the instrument. Several of the algorithms implement either polynomial-based methods or Fourier transform-based methods (IsoDalton, MWTWIN, Mercury, Isotope Calculator, IsoPro, emass/ qmass, lib mercury++, ISOMABS, and Decon2Ls).

XII. ELEMENTAL COMPOSITION DETERMINATION

The determination of the molecular formula or elemental composition requires a clean mass spectrum with no interfering noise or coeluting compounds. The algorithm includes a decision making step for proton and alkali metal adducts, automated determination of charge states and overlapping peaks, and an isotopic pattern matching. It was validated with 220 pharmaceutical compounds and yielded a success rate of 90%. Isotope-enriched metabolites can be investigated using a method that includes spectral correlation methods along with mass accuracy and isotope ratio filters. The freely available software SIRIUS (Sum formula Identification by Ranking Isotope patterns Using mass Spectrometry) has a user-friendly graphical interface and can be used on LINUX, MAC, and Windows platforms. The newer implementation "SIRIUS Starburst" also includes features such as peak intensity, number of hetero atoms in the molecular formula, neutral losses, and tandem mass spectral information. The Seven Golden Rules are a set of heuristic rules for elemental composition calculations, including the Senior and Lewis rules, element ratio rules, and an isotopic abundance matching filter. The rules were developed with a set of 68,237 existing elemental compositions and validated with 6,000 molecular formulae by means of an internal database of 432,968 existing elemental compositions. The freely available software was used to calculate the molecular formula space (elements CHNSOP; <2,000 u) covering more than two billion elemental compositions, and it was deduced that only 623 million elemental compositions are highly probable (see Fig. 7). The automated correction of isotope pattern abundance errors using peak shaping and correction algorithms resulted in better identification rates of the molecular formulas. An algorithm for isotopic pattern calculation that includes stable isotope markers (^{13}C and ^{15}N labeled) was developed. Recently, an approach was developed that uses elemental formula calculations with database lookup and a subsequent in silico generation of CID mass spectra from the obtained isomer structures. The obtained in silico

tandem mass spectra (calculated by MassFrontier) were then compared with experimental CID spectra.

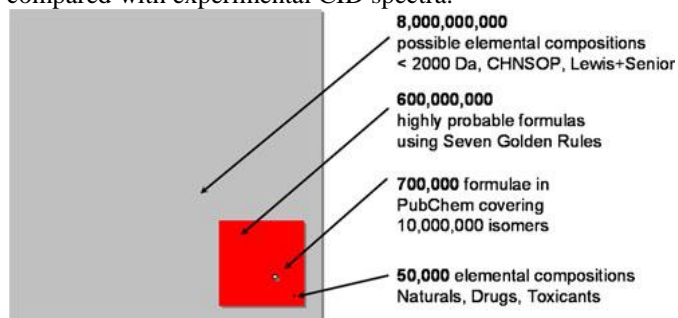


Figure 7: The molecular space between 2,000 Da (elements CHNSOP) covers more than eight billion elemental compositions and can be reduced to 600 million highly probable molecular formulas using the Seven Golden Rules

XIII. ALGORITHMS FOR FORMULA CALCULATION FROM HIGH-RESOLUTION MS/MS DATA

If the mass spectrometer is capable of obtaining accurate mass multistage product ions (MS_n), then this information should be utilized during the elemental composition determination. The possible elemental formulae for single peaks should be shown, and the algorithm should analyze if the elemental composition of the product ion could be combined to generate feasible elemental compositions of the complete molecule. Bruker (Billerica, MA, USA) developed the Smart Formula three-dimensional (3D) algorithm that includes this information by using a recursive algorithm to exclude unfeasible molecular formulae from lower mass fragments. Tandem mass spectra obtained under EI can be used together with isotope abundance analysis to obtain correct elemental compositions. Polynomial expansion algorithms to calculate the isotope patterns for precursor ion, neutral loss, and MS_n product ion tandem mass spectra have been discussed in Ramaley and Herrera, and Rockwood et al. Another approach used accurate masses from MS/MS product ions during the investigation of fragmentation processes of some natural products. Sirius Starburst is a freely available software that combines MS/MS fragment and element ratio information with elemental composition determinations. A useful hardware based approach, the acquisition of exact masses at high and low ionization energy MSE, can lead to more accurate elemental formula determinations.

XIV. COMPLEX DATA-DEPENDENT SETUPS INCLUDING ION MAPS AND ION TREES

Data-dependent acquisition methods are used in most of today's tandem mass spectrometers. The mass spectrometry software triggers MS/MS or MS_n product ion scans based on specific events. The trigger can be set on specific events such as the highest abundant peaks, manually selected masses, specific neutral losses, or specific isotopic pattern.

Specific data-dependent setups such as total molecule ion maps are very powerful features for molecule fragmentation studies. The process to create ion maps has been known since

more than 20 years; however, it has not reached its full potential, mainly due to missing data handling options. Ion maps contain product ion mass spectra over the mass range of all precursor ions from 20 Da increasingly up to the molecular mass of the compound. These ion maps can be obtained by a longer direct infusion process with autosamplers or better by nanoESI using Nanomate (Advion Inc.) robotic injections to allow long-enough scan times. The method should not be confused with spatial ion maps obtained from secondary ion mass spectrometry TOF-SIMS or mass spectrometric imaging or ion maps that refer to retention time– m/z visualizations (LC-MS ion maps). The total ion map is a function of precursor m/z value versus product ion m/z value and intensity, and it can be represented in two- or three-dimensional space.

An even more powerful method to investigate mass spectral fragmentations and fragmentation pathways of molecules are ion tree experiments. The methodology has been available for many years, and in principle, any mass analyzer capable of MS_n fragmentation can make use of it. Different ionization voltages and adduct-dependent fragmentations, as well as the use of high-resolution measurements and accurate mass MS_n spectra from hybrid instruments, can reveal additional fragmentation pathways. However, these complex multidimensional setups were rarely used in the past due to data handling and software issues. Application examples include fragment studies of polyphenols, lipids, glycans, and carbohydrates.

XV. MASS SPECTRAL LIBRARY SEARCH

Mass spectral library search is the first step in any mass spectral interpretation and therefore will be discussed in deeper detail. Mass spectral search can be performed with unit mass and high-resolution mass spectra of all stages (MS to MS_n). The aim of a library search is either to obtain a correct structure hit of compounds already in the library or to obtain partial structural insights from compounds that nearly match. For that purpose, an experimental mass spectrum is searched against a large collection of already recorded mass spectra that are stored in a database. A general review of mass spectral libraries and mass spectral search algorithms has been written.

XVI. MS AND MS/MS AND MS_n LIBRARIES AND SEARCH ALGORITHMS

Search algorithms for electron ionization spectra were developed first, and these include the INCOS algorithm, probability-based matching (PBM), and dot-product algorithm. Currently, the NIST08 MS/MS collection is a large commercially available database with 14,802 MS/MS spectra from 5,308 precursor ions. There are a variety of commercial libraries that have been generated for certain instrument types and settings. The publicly available Massbank and ReSpec database (RIKEN) are databases currently covering 24,772 mass spectra and tandem mass spectra from 13,200 compounds. An electrospray tandem mass spectrometry library (ESI-MS/MS) for forensic applications covered 5,600

spectra of 1,253 compounds acquired at different ionization voltages using a hybrid tandem mass spectrometer coupled to a linear ion trap. Smaller but specialized libraries are in use for toxicological screening and drug analysis. An in-house library of MS/MS spectra from 1,200 natural products with the majority of entries having $[M+H]^+$ adducts and 95% of those compounds being able to ionize in positive mode. Tandem mass spectra are not as reproducible as electron ionization spectra when obtained from different instruments. However, the creation of reproducible and transferable MS/MS spectral libraries for use on multiple instrument types is possible. A fragmentation energy index was proposed for LC-MS to normalize collision energies and create reproducible spectra comparable to 70-eV electron ionization spectra. Another study compared tandem mass spectra obtained from quadrupole–quadrupole–time of flight, quadrupole–quadrupole–linear ion trap, quadrupole–quadrupole–quadrupole, and linear ion trap–Fourier transform ion cyclotron resonance mass spectrometer and came to the conclusion that platform independent MS/MS spectra can be obtained with multiple fragmentation voltage settings. Peptide mass spectra usually show specific fragmentations, and a series of specialized search algorithms were developed for these purposes. MS/MS spectra can be searched according to spectral similarity, probability match (PBM), or dot-product algorithm search. If the MS/MS spectra were obtained in data-dependent mode and precursor mass information is available, this precursor mass can be used as a powerful first filter for all subsequent MS/MS matches. The precursor m/z search window can be selected according to the experimental mass accuracy of the instrument. Well-calibrated unit mass resolution instruments can reach a mass accuracy of ± 0.5 Da (or better with post-calibration methods). In this case, a precursor search window of ± 0.5 Da can be set for MS/MS search. The subsequent MS/MS match uses a product ion window search tolerance that is slightly higher due to possible hydrogen shifts. Well-established dot product, PBM, and reverse search algorithms are used to match the filtered MS/MS spectra. The accuracy, recall, precision, true, and false discovery rate of the selected algorithm and all other statistical parameters are best obtained from test sets with known spectra and decoy mass spectral datasets as seen from the proteomics community. Moreover, NIST MS Search can handle and search molecular structures together with their associated mass spectra, which is an obligatory prerequisite for any advanced library search program.

XVII. MASS SPECTRAL TREES COMBINE MULTIPLE-STAGE MASS SPECTRA

Ion traps and hybrid mass spectrometers can be used to create multiple-stage mass spectra (MS_n) by consecutively fragmenting precursor and all product ions. Usually, the abundance of the obtained product ions decreases, which sets a practical limit at MS_6 to MS_{10} . The feasibility of using MS_n data for the investigation of drugs, monosaccharides, oligosaccharides, and other molecules has been shown. The authors show that similar building blocks will have similar product ion mass spectra, and therefore, the utilization of MS_n

spectra of all stages can aid in structure elucidation of the core molecule structures. For example, if a set of molecules would have different substitutions or side chains, then an accurate mass precursor search could not identify these molecules. If the side chain is cleaved off or lost in a dissociation step, then the remaining core molecules would generate similar product ion spectra and therefore could be matched among this set of similar compounds. The representation of a spectral tree of compound mass spectra and their associated structures was obtained from MassFrontier (HighChem Ltd).

XVIII. MASS SPECTRAL INTERPRETATION

Many of the developments in mass spectral interpretation are deeply rooted in the era of electron ionization mass spectrometry from the 1970s and 1980s. Hence, mass spectral fragmentation interpretation rules are best developed for EI mass spectrometry. The red book entitled “Interpretation of mass spectra” written by Turecek and McLafferty, the book entitled “Introduction to Mass Spectrometry” by Watson and Sparkman, and “Understanding mass spectra: a basic approach” by Smith are standard sources for mass spectrometrists investigating electron ionization spectra. These books contain very detailed explanations of reactions and fragmentation pathways, including rearrangement reactions, homolytic or heterolytic bond cleavages, hydrogen rearrangements, electron shifts, resonance reactions, and aromatic stabilizations. Any de novo interpretation without any pre-knowledge is still challenging, if not totally impossible, due to the high molecular diversity and many similar compound structures. The even electron rule states that usually neutral molecule fragmentations are observed from molecular ions, but radical loss can also occur in case of aromatic and nitroaromatic compounds. Under positive electrospray (ESI), most fragment ions were reported even electron, whereas the formation of odd electron under EI was significantly higher. The Stevenson rule states that ions with low ionization energy are more stable and will gain high peak abundance in the mass spectrum. The nitrogen rule should in principle only be used for unit resolution mass spectra because high-resolution and high accuracy mass spectrometry can always calculate the correct number of nitrogen atoms. The Rings Plus Double Bonds Equivalent (RDBE) should not be used with elements that allow multiple valence counts (such as phosphorus and sulfur) as otherwise only possible RDBE ranges can be obtained instead of unique solutions. Mass spectral visualization techniques such as van Krevelen or Kendrick plots, and spectral mappings using dimension reduction methods with principal component analysis are helpful for the investigation of unresolved and complex organic matter (petroleum, coal, sediments, and fulvic acids).

XIX. DETERMINATION OF STEREOCHEMISTRY USING MASS SPECTROMETRY

The determination of stereochemical (absolute) configuration usually requires a separation technique such as GC, CE, or LC with chiral columns. ESI-MS was used to

determine the binding affinities of ion-molecule reactions by performing CID experiments of host-guest complexes. It is possible to determine the chirality of molecules without pre-separation using chiral selector agents and ESIMS/MS. Additionally, traveling wave ion mobility

spectrometry can be used to determine stereochemistry. The book titled "Applications of Mass Spectrometry to Organic Stereochemistry" discusses practical approaches for stereochemical investigations of molecules. Mainly, protein folding and dynamics have been studied in recent years. The determination of the conformational changes of small molecules is possible using ion mobility mass spectrometers or hybrids thereof. This approach requires the experimental determination of cross sections from known molecules and the use of such data for theoretical models.

XX. APPROACHES FOR HYPHENATED TECHNIQUES (GC-MS AND LC-MS)

Mass spectral deconvolution for clean mass spectra Mass spectral deconvolution refers to the process of creating background- and noise-free mass spectra from GC-MS or LC-MS data. Traditionally, chromatographers would use a simple chromatographic peak detection method and would manually select a detected peak to obtain the related mass spectrum. This manual process is error prone and time consuming, and requires manual background subtraction in front and in the back of the chromatographic peak. With an automated deconvolution, routine peaks can be detected under the baseline total ion chromatogram or overlapping peaks can be resolved (see Fig. 8). Additionally, if the chromatographic resolution is not sufficient, then the process is also able to separate (deconvolute) overlapping compound mass spectra. The automated deconvolution process itself is now standard in many GC-MS investigations and is mostly known from the freely available AMDIS. The AMDIS process includes four sequential steps:

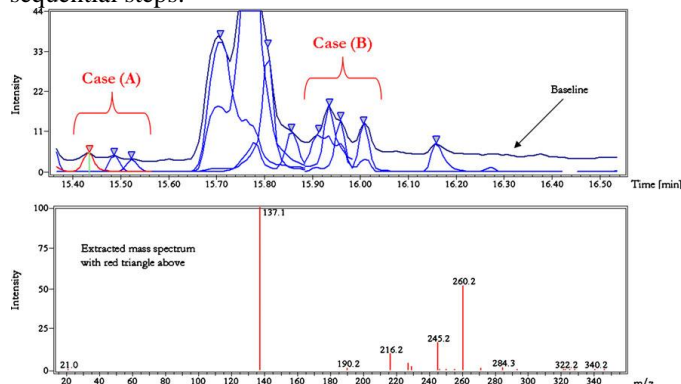


Figure 8: Peak picking and mass spectral deconvolution. The program can automatically detect peaks under the baseline (case A). Overlapping (non-resolved) peaks can be detected, and clean mass spectra are extracted (case B)

(1) noise analysis, (2) component perception, (3) spectral deconvolution, and (4) compound identification. AMDIS was recently adapted to monitor air quality and identify toxic gases on board of the International Space Station. Multiple other software solutions for the analysis of GC-MS and LC-MS exist. That includes LECO ChromaTOF, SpectralWorks

AnalyzerPro, Ion Signature Quantitative Deconvolution Software, HighChem MassFrontier, and

TargetSearch. The use of peak picking and peak detection algorithms for LC-MS data is still an active field of research due to high noise ratios, broader chromatographic peaks, and mass spectra that show less fragments than electron ionization spectra. The deconvolution process itself usually performs best if it is optimized for a specific scan rate; otherwise, false-positive and false-negative peak detections may occur. The detection of these deconvolution errors is best solved by using reference compound mixes with a known number of analytes and a subsequent optimization process to detect the correct number of compounds. Deconvoluted compound mass spectra are subsequently submitted to a mass spectral database search. If additional MS/MS spectra were extracted, then a tandem mass spectral search can be performed.

XXI. CONCLUSIONS

Structure elucidation using mass spectrometry is a challenging field of research with many success stories. Mass spectrometry itself is seldom used for the de novo structure elucidation of small molecules but serves as an important building block together with NMR, IR, X-ray crystallography, and other spectroscopic techniques. Together with hyphenated chromatographic techniques, (GC and LC) mass spectrometry serves as a powerful tool for the elucidation of drugs, pesticides, metabolites, and complex chemical mixtures. Mass spectrometry hardware is currently in a very advanced stage with many technologies not fully exploited yet. More data centric approaches have to be taken in the future. This includes the electronic publishing of investigated structures and their associated multiple-stage mass spectra with open data licenses. The ultimate success of structure elucidation of small molecules lies in better software programs and the development of sophisticated tools for data evaluation of high-resolution and accurate mass multiple-stage (MSⁿ) mass spectral data.

REFERENCES

- [1] Gordin A, Fialkov AB, Amirav A (2008) Classical electron ionization mass spectra in gas chromatography/mass spectrometry with supersonic molecular beams. *Rapid Commun Mass Spectrom* 22(17):2660–2666
- [2] Fenn JB (2003) Electrospray wings for molecular elephants (Nobel Lecture)*. *Angew Chem Int Ed* 42:3871–3894
- [3] Gaskell SJ (1997) Electrospray: principles and practice. *J Mass Spectrom* 32(7):677–688
- [4] Cech NB, Enke CG (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* 20(6):362–387
- [5] Schmidt A, Karas M, Dulcks T (2003) Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: when does ESI turn into nano-ESI? *J Am Soc Mass Spectrom* 14(5):492–500

- [6] Lazar IM, Grym J, Foret F (2006) Microfabricated devices: a new sample introduction approach to mass spectrometry. *Mass Spectrom Rev* 25(4):573–594
- [7] Pereira-Medrano AG, Sterling A, Snijders APL, Reardon KF, Wright PC (2007) A systematic evaluation of chip-based nanoelectrospray parameters for rapid identification of proteins from a complex mixture. *J Am Soc Mass Spectrom* 18(9):1714–1725
- [8] Almeida R, Mosoarca C, Chirita M, Udrescu V, Dinca N, Vukeli Ž, Allen M, Zamfir AD (2008) Coupling of fully automated chip-based electrospray ionization to high-capacity ion trap mass spectrometer for ganglioside analysis. *Anal Biochem* 378(1):43–52
- [9] Lydic TA, Busik JV, Esselman WJ, Reid GE (2009) Complementary precursor ion and neutral loss scan mode tandem mass spectrometry for the analysis of glycerophosphatidylethanolamine lipids from whole rat retina. *Anal Bioanal Chem* 394 (1):267–275
- [10] Wickremsinhe ER, Singh G, Ackermann BL, Gillespie TA, Chaudhary AK (2006) A review of nanoelectrospray ionization applications for drug metabolism and pharmacokinetics. *Curr Drug Metab* 7(8):913–928
- [11] Koulman A, Cao M, Faville M, Lane G, Mace W, Rasmussen S (2009) Semi-quantitative and structural metabolic phenotyping by direct infusion ion trap mass spectrometry and its application in genetical metabolomics. *Rapid Commun Mass Spectrom: RCM* 23(15):2253–2263
- [12] Hobby K, Gallagher RT, Caldwell P, Wilson ID (2009) A new approach to aid the characterisation and identification of metabolites of a model drug; partial isotope enrichment combined with novel formula elucidation software. *Rapid Commun Mass Spectrom* 23(2):219–227
- [13] Alon T, Amirav A (2006) Isotope abundance analysis methods and software for improved sample identification with supersonic gas chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* 20(17):2579–2588
- [14] Rogers S, Scheltema RA, Girolami M, Breitling R (2009) Probabilistic assignment of formulas to mass peaks in metabolomics experiments. *Bioinformatics* 25(4):512–518
- [15] Werner E, Heilier J-F, Ducruix C, Ezan E, Junot C, Tabet J-C (2008) Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. *J Chromatogr B* 871(2):143–163
- Advances in structure elucidation of small molecules 49
- [16] Reemtsma T (2009) Determination of molecular formulas of natural organic matter molecules by (ultra-) high-resolution mass spectrometry. Status and needs. *J Chromatogr A* 1216(18):3687–3701
- [17] Koch BP, Ludwiczowski KU, Kattner G, Dittmar T, Witt M (2008) Advanced characterization of marine dissolved organic matter by combining reversed-phase liquid chromatography and FT-ICR-MS. *Mar Chem* 111(3–4):233–241
- [18] Bocker S, Letzel MC, Liptak Z, Pervukhin A (2009) SIRIUS: decomposing isotope patterns for metabolite identification. *Bioinformatics* 25(2):218
- [19] Bocker S, Rasche F (2008) Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics* 24(16):i49
- [20] Lee S, Mandrake L, Bornstein B, Bue B (2009) Quantification of trace chemicals using vehicle cabin atmosphere monitor. *Aerospace Conference 2009 IEEE*
- [21] Mandrake L, Lee S, Bornstein B, Bue B (2009) Adapting AMDIS for autonomous spectral identification of hazardous compounds for ISS monitoring. *Aerospace Conference 2009 IEEE*
- [22] Renard BY, Kirchner M, Steen H, Steen JAJ, Hamprecht FA (2008) NITPICK: peak identification for mass spectrometry data. *BMC Bioinformatics* 9:355
- [23] Lu H, Liang Y, Dunn WB, Shen H, Kell DB (2008) Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *TrAC, Trends Anal Chem* 27(3):215–227
- [24] Luedemann A, Strassburg K, Erban A, Kopka J (2008) Tag-Finder for the quantitative analysis of gas chromatography–mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics* 24(5):732
- [25] Fiehn O (2008) Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Anal Chem* 27(3):261–269
- [26] Ebada SS, Edrada RA, Lin W, Proksch P (2008) Methods for isolation, purification and structural elucidation of bioactive secondary metabolites from marine invertebrates. *Nat Protocols* 3(12):1820–1831