PIONEER
ACADEMIC PUBLISHING LIMITED

Journal of Progress in Engineering
and Physical Science
ISSN 2709-4006
www.pioneerpublisher.com/jpeps
Volume 3 Number 1 March 2024

A Review of Mass Spectrometry as an Important Analytical Tool for Structural Elucidation of Biological Products

Oluwaseun Olanike Ajayi[1]

[1] Department of Chemistry, University of Georgia, Athens
Correspondence: Oluwaseun Olanike Ajayi, Department of Chemistry, University of Georgia, Athens.

doi:10.56397/JPEPS.2024.03.05

Abstract

Mass spectrometry (MS) is one of the most versatile analytical measurement tools available to scientists today and finds application in virtually most aspects of chemistry research. MS is an efficient method to elucidate the structure and chemical composition of different samples or molecule. Recently, it has been used to classify biological products, in particular proteins and protein complexes, in a number of species. The success of mass spectrometry is driven both by innovative instrumentation designs, especially those operating on the time-of-flight or ion-trapping principles, and by large-scale biochemical strategies, which use mass spectrometry to detect the isolated proteins. Due to MS capability to distinguish between substances, it is used to identify unknown substances and its isotopes. When compared with other analytical tools or orthogonal methods used in analytical laboratories, mass spectrometers are often used as it has less background interference since it is performed in a vacuum. The increase in sensitivity and resolution of the MS has opened new dimensions for structural elucidation of biological products. In this review, the recent research application of MS techniques for the structural elucidation of some biological products was presented.

Keywords: mass spectrometry, structural elucidation and biological products

1. Introduction

Mass spectrometry is a key technique in analytical chemistry and plays a vital role in many aspect of day-to-day life. Mass spectrometry (MS) involves the measurement of mass-to-charge ratios of ions in order to determine their molecular weight. In other words, mass spectrometry is an analytical technique that measures the mass-to-charge ration of molecular ions or their fragments. Samples are injected into the mass spectrometer, ionized, fragmented and detected according to molecular mass and signal intensity. Mass spectrometer is a versatile analytical tool due to the fact that it can give qualitative and quantitative information on the elemental, isotopic, and molecular composition of organic and inorganic samples. Depending upon the type of mass spectrometer, these measurements can often be used to determine the exact molecular weight of the sample components and to identify unknown compounds.

All mass spectrometers contain at least three major components namely an ion source, a mass analyzer, and an ion collection/detection system. The first component is some means by which molecules or atoms from sample can be ionized. Neutral species cannot be steered by electric fields used in mass spectrometers; therefore it is necessary to produce ions. However, there are many different means by which this can be done, and they are collectively referred to as ion sources. The second component of all mass spectrometers is the mass analyzer itself. There are several different means by which the m/z ratio of ions can be measured. Time-of-flight (ToF), magnetic sector and quadrupole mass analyzers are the most common, each with its own set of strengths and weaknesses. The third component common to all mass spectrometer systems is a means of detecting or counting the number of ions of a specific m/z value. These devices are called detectors and they too come in several different forms with the most common being electron multipliers, Faraday cups, channel irons and channel plates. The instrument must be connected to a computer system to process and record the data and a vacuum pump to control the pressure within the mass spectrometer (Figures 1a & 1b).

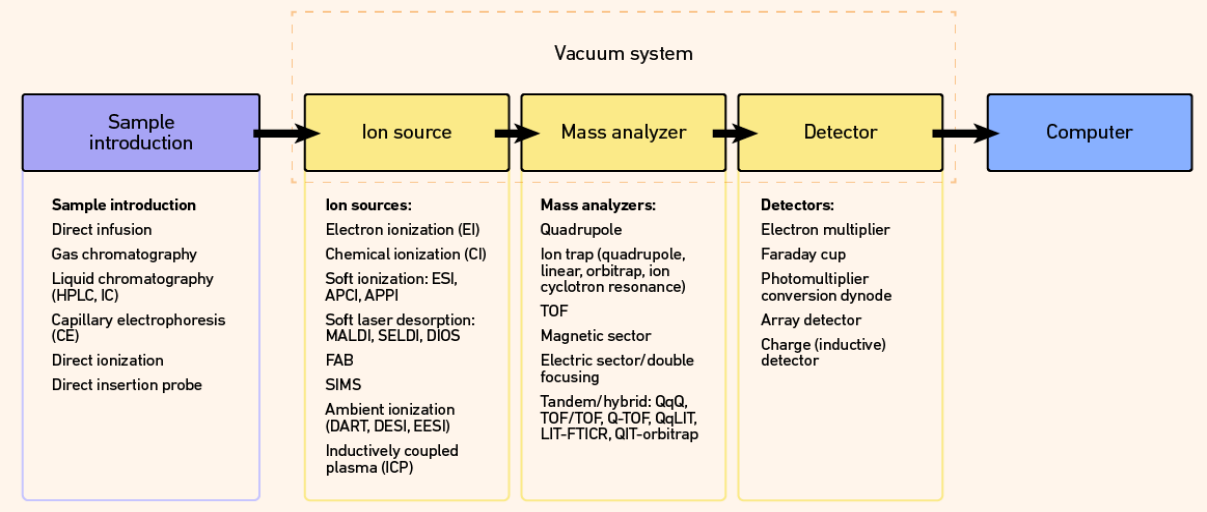

Figure 1a. Main Steps of MS and Common Variants available at each Step (TN, 2023)

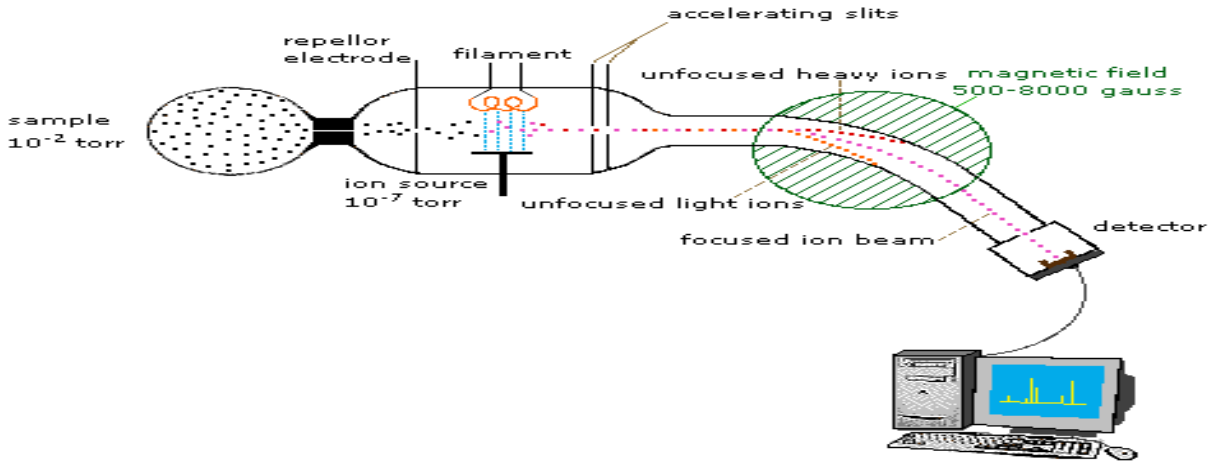

Figure 1b. Basic Components of a Mass Spectrometer

Mass Spectrometry (MS) has become an essential analytical tool in structural elucidation of biological products such as proteins and protein complexes because in obtaining molecular weight from the molecular or deprotonated ions, the intact ions can be monitored to undergo fragmentation processes so that key fragment ions can be produced,

detected, and used to reconstruct the unknown molecule. Hence, MS is a pivotal technique in the process of structural elucidation. In other words, Mass spectrometry is a standard technique for the analytical investigation of molecules and complex mixtures, since it is important in ascertaining the elemental composition of a molecule and in gaining partial structural insight using mass spectral fragmentations. The final structural composition of an unknown compound is usually done with a set of independent methods such as one and two-dimensional nuclear magnetic resonance spectroscopy (NMR) or infrared spectroscopy and X-ray crystallography and other spectroscopic methods. MS has played a central role for the structure elucidation of biological products because it can used to determine the molecular weights and elemental compositions of substances with a wide range of chemical and physical properties. Structure elucidation refers to full de novo structure identification, and it results in a complete molecular connection table with correct stereochemical assignments. Structural elucidation is the process of determining the chemical structure of a compound. Structural elucidation describes the fundamentals of chemical structure and symmetry, and the use of molecular orbital theory and quantum mechanics to understand molecular properties and chemical spectroscopy. In this review, biological products are refers to as biologics. According to Food and Drug Administration (2018), biological products are composed of sugars, proteins, nucleic acids, or complex combinations of these substances. Impurities or unknown substances may be difficult to identify. Structural elucidation of biological products is necessary to identify or confirm the structural identity of a chemical compound during chemical research and product development in industrial sectors such as agrichemicals and pharmaceutical. Various studies on the use of mass spectrometry techniques to determine molecular weight and apply tandem MS/MS approaches to provide valuable data in the identification of an unknown molecule in biological products are reviewed.

2. Biological Products

In this review, biological products are referred to as biologics. Biological products are a diverse category of products and are generally large, complex molecules. Biological products could be made of sugars, proteins or nucleic acids or complex combinations of these substances, or may be living entities such as cells and tissues. In other words, biological products encompass blood, blood components, gene therapy, somatic cells, tissues, recombinant proteins, and vaccines (such as Shingles vaccine and flu vaccines). Biological products are typically derived from natural sources microorganisms, plant, animal, or human cells. These products may be produced through biotechnology in a living system. Biological products are used to either treat or cure diseases and medical conditions, prevent diseases, or diagnose diseases.

3. Structural Elucidation of Proteins with Mass Spectrometry

Proteins are nitrogenous organic compound that are polymers of large numbers of amino acid units joined together by peptide bonds. Proteins are essential part of living systems as they are involved in structure and function of cell. Proteins can have different structure depending upon the type of amino acids from which they are formed. To understand protein function and mechanism of action, it is essential to determine protein complex assembly and structure. The main structures of protein are primary, secondary, tertiary, and quaternary structures (Figure 2)

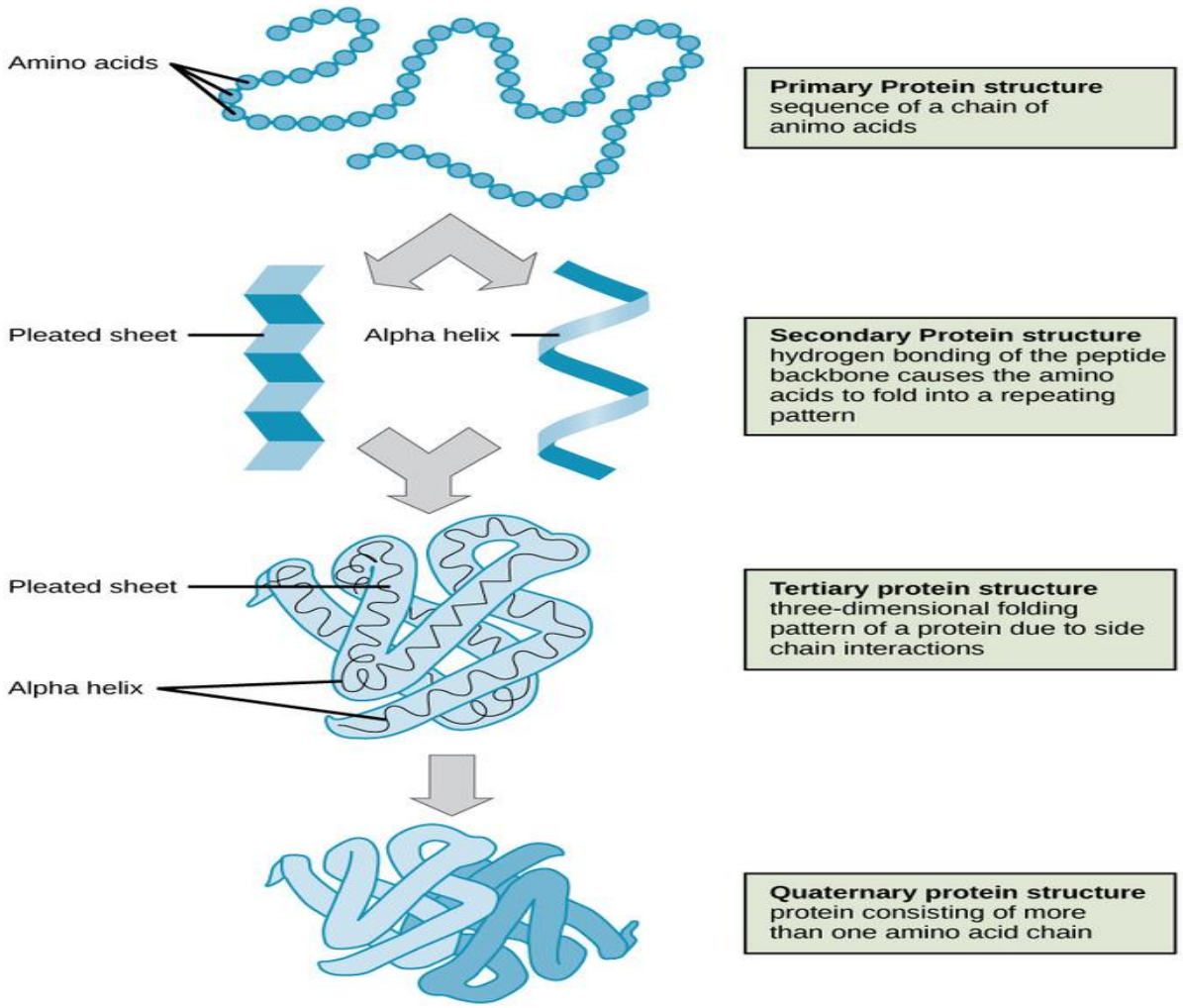

Figure 2. Main structures of Protein (CDER, 2018)

There are several methods used for structural elucidation of proteins. However, the primary focus of this study is Mass Spectrometry (MS). This technique is based upon the principle of ionization of the biomolecules and the arrangement of these molecules on bases of their mass to charge ratio. Tandem mass spectrometry is an advanced technique of structural elucidation of proteins which based upon fragmentation method of structural identification. Technology developments in MS have given rise to several applications of structural biology, both at the single protein and protein complex level. The ability to perform experiments at proteome scale, to analyze protein in their native biological state, and to reduce the minimum required sample size are the main advantages of MS-based techniques. Mass spectrometry (MS) is commonly used to determine both the primary and higher-order structures of proteins. New advances in MS technologies, combined with chemical modification and proteolysis strategies, allow the study of both single proteins and protein complexes as well as further exploration of protein structure and structural dynamics. Mass spectrometry has now become a crucial technique for almost all proteomics (Proteomics refers to the analysis of complete protein content in a living system, including co-and post translational modified proteins and alternatively spliced variants) experiments. It allows precise determination of the molecular mass of peptides as well as their sequences. This information can be used for protein identification, de novo sequencing, and identification of post-translational modifications.

The advent of soft-ionization mass spectrometry for biomolecules has opened up new possibilities for structural analysis of proteins. Some MS-based structural elucidation techniques can be carried out on proteins in their natural environments such as lipid, cells,

membranes, organelles and tissues. Petrotchenko and Borchers (2022) concluded that, mass spectrometric analysis of protein structures can supplement standard structural biology methods, such as X-ray crystallography, nuclear magnetic resonance (NMR), and recently, cryo-electron microscopy (cryo-EM), especially in cases where the application of traditional methods is complicated by sample heterogeneity, sample complexity, or impurities.

Some new advances in Mass Spectrometry technologies are thermo scientific orbitrap MS solutions, liquid chromatography mass spectrometry (LC-MS), Hydrogen deuterium exchange mass spectrometry (HDX-MS) and cross-linking mass spectrometry (XL-MS). The thermo scientific orbitrap MS solutions enable highly specific and sensitive workflows that allow highly specific and sensitive workflow that allow you to analyze samples with increasing analytical depth, delivering information to accelerate the path from structure to function. Proteins are enzymatically digested to their peptide components and analyzes by LC-MS (Figure 3). The resulting sequence data is used to determine the original protein components of the sample. Information on posttranslational modifications (PTMs) and stoichiometry can also be obtained with this approach.

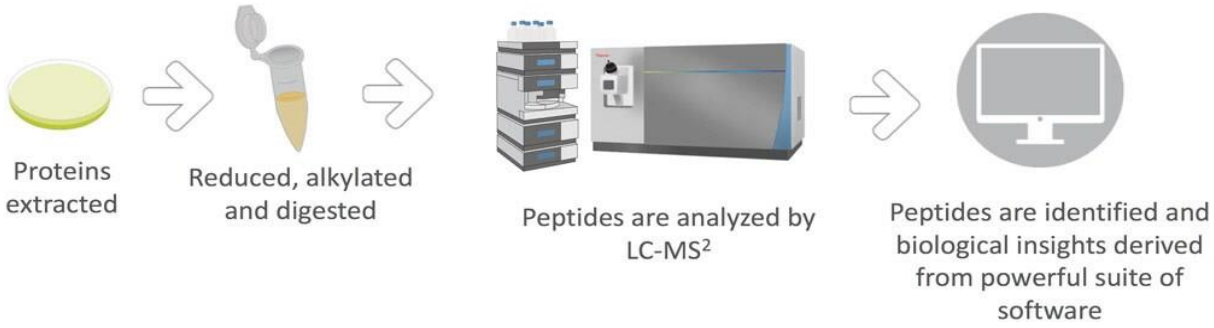

Figure 3. Liquid Chromatography Mass Spectrometry (Thermofisher, 2020)

Hydrogen deuterium exchange mass spectrometry (HDX-MS) takes advantage of the labile nature of protons present on protein backbone amides, and is powerful tool in the elucidation of protein structure. When dissolved in solution, proteins exchange these protons with hydrogen groups present in a deuterated buffer, and protons from the protein are exchanged with deuterium. Only the protons present on the backbone amides are measured. The rate of hydrogen to deuterium exchange provides solvent accessibility data, which can be used to infer information on protein structure and conformation. MS can be used to measure the rate of deuterium uptake. HDX-MS can be used to obtain information on structure, protein-protein or protein-ligand interaction sites, allosteric effects, intrinsic disorder, and conformational changes induced by posttranslational modifications (PTMs). HDX-MS has the advantage of not being limited by the size of proteins or protein complexes, and it is highly sensitive, able to detect coexisting protein conformations. Protein structure data can be obtained independently or simultaneously as part of protein interaction studies using cross linking mass spectrometry (XL-MS). In such analysis, XL-MS is often combined with high-resolution techniques. Such technologies help determine protein region distance constraints via 3-D structure information or topology, XL-MS can also be used to identify protein complexes.

4. Structural Elucidation of Carbohydrates with Mass Spectrometry

Carbohydrates (also called sugars, oligosaccharides, and glycans) are defined as polyhydroxyaldehydes, polyhydroxyketones and their simple derivatives, or larger compounds that can be hydrolyzed into such units. Carbohydrates are the most abundant organic compounds in nature, widely distributed in animals, plants and microorganisms and its plays a vital role in metabolism and structural properties of biological systems (Devlamynck, 2019). In Carbohydrates, building blocks (compositional monosaccharide), glycosidic linkages and anomeric configuration, sequence of sugar residues, molecular mass or degree of polymerization and substitutions are critical aspects to analyze in other to understand their

structure properties. Elucidating the structure of an unknown carbohydrate compound is like solving a complicated puzzle. Carbohydrates may vary in monomer composition, e.g., glycose, fructose, and galactose. These building block/s are connected via glycosidic linkages, in an alpha or beta configuration, forming linear or branched chains of disaccharides, oligosaccharides, or polysaccharides, or be covalently linked as glycans to proteins, or lipids.

Moreover, many plant metabolites constitute glycosylated non-carbohydrate compounds. Mixtures of oligosaccharides such as mammalian milk are also found in nature. Therefore, understanding detailed structural composition of such mixtures is often important. Mass spectrometry was first applied to carbohydrate analysis in 1958 (Finan, Reed & Snedden, 1958) with electron impact (EI) ionization (Bleakney, 1929). EI is said to be "hard" ionization method since it induces extensive fragmentation and the molecular ions are typically not observed. Since EI is only suitable for volatile organic molecules, hydroxyl groups in non-volatile carbohydrates had to be protected by permethylation or peracetylation to increase the volatility prior to analysis (Ciucanu & Kerek, 1984). Since then, many researches for structural analysis of glycans using mass spectrometry have been reported with the improvement of analytical techniques, mass spectrometric instrumentation, as well as development of "soft" ionization methods which enable native carbohydrates to be ionized with molecular ion information to be obtained.

The unique functions of glycans are directly related to their structures. To fully elucidate the structure of an unknown glycan, the sugar unit identity, anomeric configuration, linkage types, sequence, and branching location are the main levels of structural information needed. Kudzai and Rune (2020) demonstrated the structural characterization of carbohydrates based on their molecular mass, as well as the mass of their respective fragment ions using mass spectrometry (MS). However, one of the challenges limiting the advance of glycoscience is the lack of sensitive tools for detailed glycan structural analysis. However, studies by Konda (2019), Maki and Renkonen (2021) revealed that, Mass spectrometry (MS/MS) is a powerful tool for structural elucidation of carbohydrate with its high sensitivity and selectivity. For instance, glycan structural determination typically stops at the level of topology (relative location of sugar units) using other analytical tools.

The elucidation of carbohydrate structures remains extremely difficult and unsolved. However, Konda (2019) developed a multiple stage tandem mass spectrometry approach toward a sensitive and full level structural analysis of linear oligosaccharides by obtaining information of individual sugar unit identity, anomeric configuration, linkage types, and sequence. The Mass spectrometry (MS^3) approach roots in the discovery of diagnostic ions, which are sub-structures of disaccharides and the fragmentation pattern of which can be linked to a specific level of structural information. An MS^3 screening method was developed and a new diagnostic ion for linkage and structural determination was established for the first time. In another study, Juan, Ace, Mathew, Eshani, Garret, Nikita, and Carlito (2021) developed *de novo* mass spectrometry-based workflow to isolate and structurally elucidate oligosaccharides derived from plants to provide sequence, monosaccharide compositions, and glycosidic linkage positions. The approach developed by Juan et al (2021) employed liquid chromatography-tandem mass spectrometry (LC-MS/MS) based methods in a 3-dimensional concepts which includes one high performance liquid chromatography-quadrupole time-of-fight mass spectrometry (HPLC-QTOF MS) analysis for oligosaccharide sequencing and two ultra high performance liquid chromatography-triple quadrupole mass spectrometry (UHPLC-QqQ MS) analyses on fractionated oligosaccharides to determine their monosaccharide's and linkage compositions.

5. Structural Elucidation of Lipids with Mass Spectrometry

Lipids are small but complex biomolecules due to their immense structural and functional diversity that are intricately linked. Most lipids contain aliphatic chains, which cause the structural variability of molecular species within a lipid class. The LIPID MAPS classification system organizes lipids into eight categories based on their chemical structures namely fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), saccharolipids (SL), polyketides (PK), prenol lipids (PR), and sterol lipids (ST) (figure 4). The chains can be straight or branched, saturated or

unsaturated, and modified by various functional groups. The multitude of biological functions is reflected in a tremendous structural diversity and coexistence of multiple isomers. Therefore, it is important to completely elucidate the structure and spatial mapping of individual species since its biological function depends on structure. However, Biehn and Lindert (2021) ascertained that mass spectrometry is an effective method for elucidating lipids structure when compared to other analytical methods. Although, common mass spectrometry methods usually do not reveal much about the structure of the lipid chains especially the distinction between frequently occurring isomers but technological advancements in Mass Spectrometry (MS) based techniques such as coupling of MS with laser light (Kirschbaum & Pagel, 2022) and atmospheric-pressure ionization MS-based technique (Cvacka, Vrkoslav & Strnad, 2023) have tremendously advanced the possibilities of structural elucidation of lipids. The recent advances in lipidomics applications employ laser-induced ultraviolet (UV) and infrared (IR) photo dissociation, derivatization of lipids and advanced ion activation and imaging techniques to further elucidate the structure of the chains.

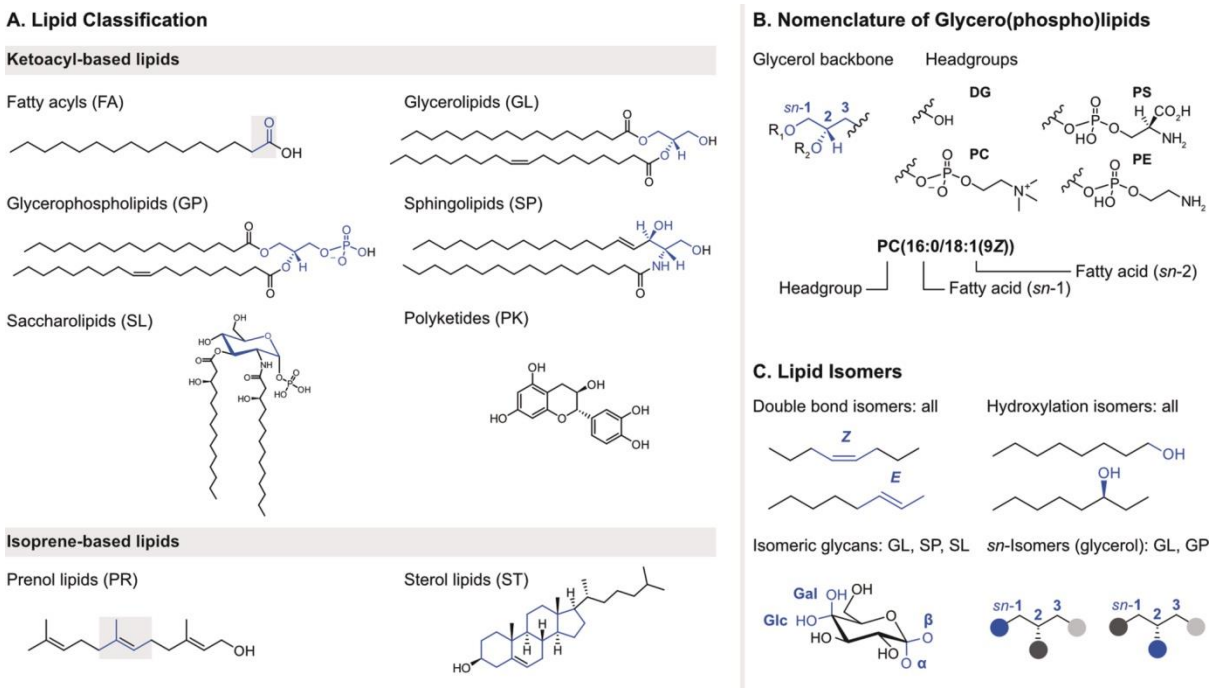

Figure 4. Lipids Categories based on their Chemical Structures

6. Structural Elucidation of Nucleic Acids with Mass Spectrometry

Nucleic acids are the third class of biopolymers (polysaccharides and proteins being the others). Deoxyribonucleic acid (DNA is the carrier of genetic information) and ribonucleic acid (RNA is an intermediate in the expression of genetic information and other diverse roles). The monomeric units for nucleic acids are nucleotides (nucleotides are made up of three structural subunits namely sugar, heterocyclic base and phosphate) (Figure 5). Nucleic acids complexes are at the crossroads between chemistry and biology. On the one hand, nucleic acids are essential players in the central dogma of molecular biology. On the other hand, nucleic acids and derivatives can be chemically synthesized, and form structures that can adopt a variety of functions (molecular recognition, catalysis, therapeutics and so on). Nucleic acids play key roles in the storage and processing of genetic information and in the regulation of processes. Nucleic acids represent attractive targets for drugs against gene-related diseases and have also found useful as chemotherapeutic agents targeting cellular DNA and RNA. The development of effective nucleic acid-based chemotherapeutic approaches requires effective analytical strategies capable of providing adequate information regarding the nucleotide patterns, structural modifications, formation of higher-order structures and the interaction of nucleic acids with other cellular components

and chemotherapeutic agents.

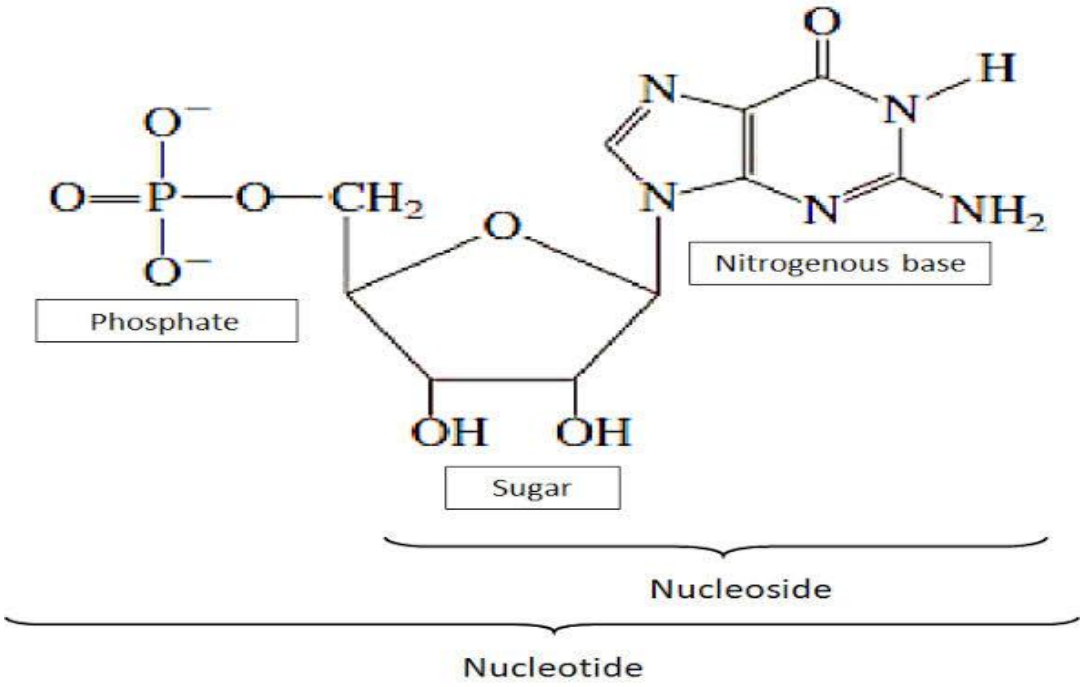

Figure 5. Nucleotide

To elucidate non-covalent complex nucleic acids, mass spectrometry has the unique advantage of providing an assumption-free determination of the stoichiometries present in the sample, while the peak intensities inform on the relative abundance of each species in solution. In addition, advanced mass spectrometry techniques can be exploited to further elucidate the structural complexes in the gas phase (such as kinetic stability, reactivity, shape or spectroscopic properties). Tandem mass spectrometry (MS) has evolved to one of the most powerful analytical tools to elucidate nucleic acids due to the recent impressive technical and methodological developments (Schurch, 2014). Structural elucidation of nucleic acids with MS has their specific features as compared with the mass spectrometry of protein molecules. First and foremost, this refers to the matrices used in MALDI mass spectrometry. A typical matrix for MALDI is an aromatic acid efficiently absorbing the energy of a laser.

The appropriate selection of matrix components is the particular factor that determines the sensitivity and accuracy of analysis. The decisive event in the history of MALDI mass spectrometry of oligonucleotides was when picolinic acid derivatives 3-aminopicolinic acid (Tanaka, 2018) and 3-hydroxypicolinic acid (Wu, Steding & Becker, 2019) were used as a matrix. Nucleic acids are prone to fragmentation in the ionization process of mass spectrometry (MS). Following the introduction of soft-ionization techniques of electrospray ionization and matrix-assisted laser desorption ionization (MALDI) at the end of the 1980s, structural elucidation of nucleic acids with mass spectrometry has gained broader applicability. Advance developments of the associated instrumentation and optimized preparation and purification methods, a dramatic enhancement of the mass range, detection sensitivity, and resolution became possible in routine analysis of oligonucleotides, up to a length of 50 nucleotides. Mass spectrometry based techniques is a rapid, accurate and sensitive tool ideally suited for structural elucidation of nucleic acids (Chowdhury & Guengerich, 2021). Mass spectrometry offers unique advantages due to its capability to distinguish each stoichiometry present in a mixture. Likewise, the recent MS techniques (reactive probing, fragmentation techniques, ion mobility spectrometry and ion spectroscopy) provide more detailed information on the structure of nucleic acids.

7. Conclusion

This review can be of interest for a wide range of audiences who may not have extensive

experience in Mass Spectrometry based techniques. It is evident from this review that MS-based techniques have proven its unique capabilities in terms of analyzing complex mixtures of biological product with high mass accuracy (the mass can be measured accurately to a 4 decimal places. As a result, sum formulas can be determined for a mass signal), high-resolution (due to MS high resolution, the isotope pattern is specific to an elemental composition and the distinctive isotope ratios are clearly visible), fragment analysis and high-sensitivity. The structural elucidations of some biological products are still a challenge due to their complexities. However, the developments of high-resolution mass spectrometry technologies made it possible to rapidly or further elucidate most biological products structures through accurate mass, instrumentation and diagnostic fragments. Mass spectroscopy provides rich elemental information, which is an important asset to interpret complex mixture components. Thus, it is an important tool for structure elucidation of biological products.

References

Bleakney, W. (1929). Carbohydrates Analysis. *Physical Science Research, 34*, 157-160.

Chowdhury, G. & Guengerich, F.P. (2021). Mass spectrometry of nucleic acids. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, 3*(2), 13-19.

Ciucanu, I. & Kerek, F., (1984). Carbohydrate. *Research Journal of Learning*, 131, 209-217.

Cvacka, J., Vrkoslav, V., & Strnad, S. (2023). Structural characterization of lipids using advanced mass spectrometry approaches. *Wiley Online Library, 7*(3), 23-28.

Devlamynck, T. (2019). Trans-glucosylation of stevioside by the mutant glucansucrase enzyme: improve its taste profile. *Journal of Food Chemistry, 30*(1), 653-662.

Finan, P. A., Reed, R. I., & Snedden, W. (1985). Glycoscience and Mass Spectrometry. *Chemistry and Industries (London),* 36, 1172-1192.

Food and Drug Administration. (2018). Biological and interchangeable products. Retrieved on January 4, 2024 from www.fda.gov/drug/biosimilars/biosimilar-and-interchangeable-products

Juan, J.C., Ace, G.G., Mathew, J., Eshani, N., Garret, C., Nikita, B., Ye, C., & Carlito, B.L. (2021). A multidimensional mass spectrometry-based workflow for De Novo Structural elucidation of Oligosaccharides from Polysaccharides. *Journal of American Society of Mass Spectrometry, 32*, 2175-2185.

Kirschbaum, C. & Pagel, K. (2022). Lipid analysis by mass spectrometry coupled with laser light. *European Chemical Societies Publishing, 2*(2), 23-29.

Kudzai, E.M. & Rune, M. (2020). Analysis of carbohydrates by mass spectrometry. *Methods in molecular biology*, 289-301.

Nier, A.O. (1947). Non-volatile carbohydrates structures. *Research in Science and Instrumentation*, 18, 398-411.

Petrotchenko, E.V., & Borchers, C.H. (2022). Protein chemistry combined with Mass Spectrometry for protein structure determination. *Chemical Review, 122,* 7488–7499.

Schurch, S. (2014). Characterization of nucleic acids by tandem mass spectrometry- The second decade (2004-2013): From DNA to RNA and modified sequences. *Wiley Analytical Science, 35*(4), 54-59.

Tanaka, K. (2018). MALDI Mass spectrometry of oligonucleotides was when picolinic acid derivatives 3-aminopicolinic acid. *Rapid Communication and Mass Spectrom*, *1*(2), 151-153.

Wu, K.J., Steding, A., and Becker, C.H. (2019). Mass spectrometry and 3-hydroxypicolinic. *Rapid Communication and Mass Spectrom*, *1*(7), 142–146.