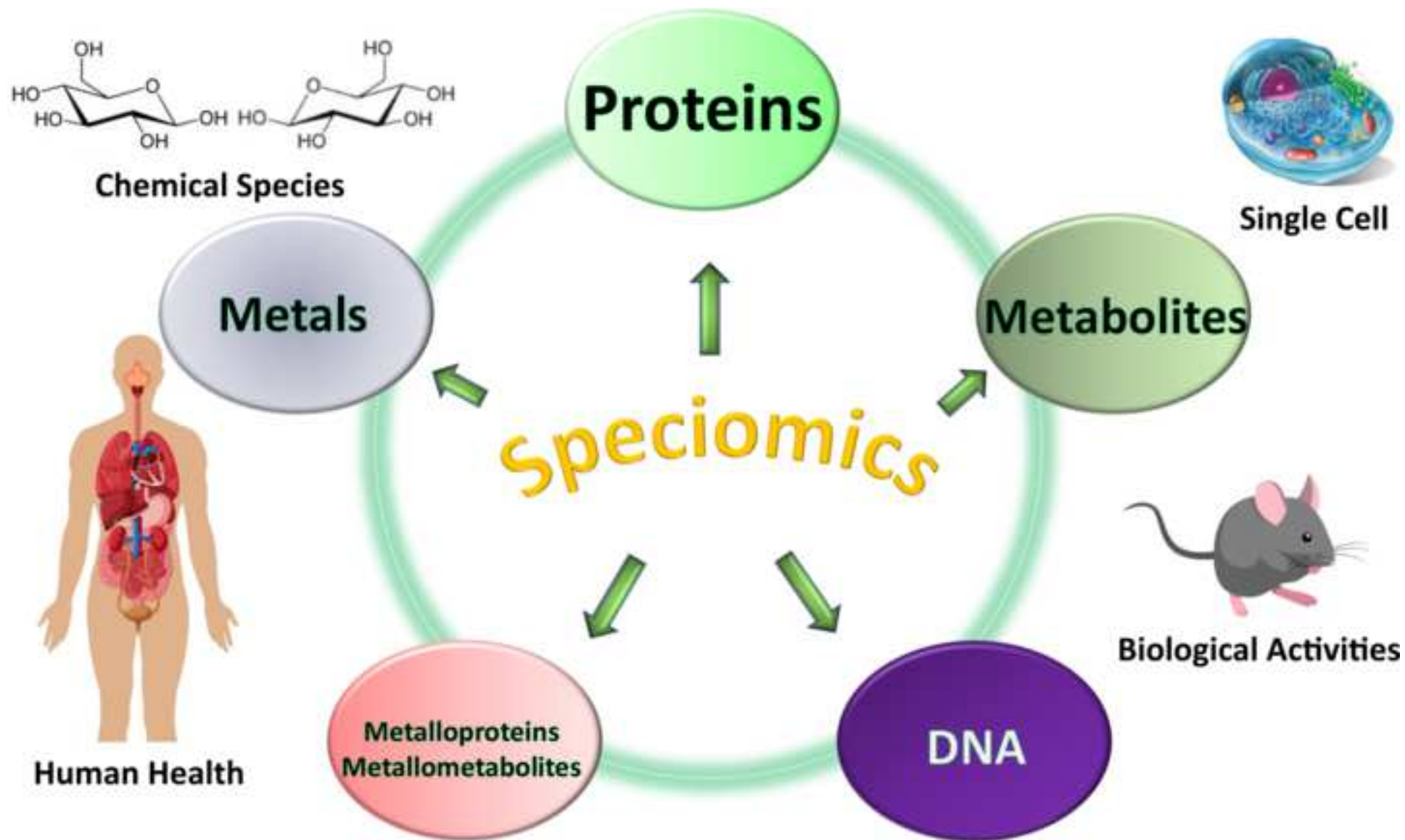


Journal of Proteomics

Speciomics as a concept involving chemical speciation and omics

--Manuscript Draft--

Manuscript Number:	JPROT-D-22-00075R1
Article Type:	Review Article
Section/Category:	Review Article
Keywords:	chemical species; proteins; metabolites; DNA; metals; Instrumentation
Corresponding Author:	Marco Arruda, PhD Unicamp Campinas, BRAZIL
First Author:	Marco Arruda, PhD
Order of Authors:	Marco Arruda, PhD Jemmyson Romário de Jesus Claudia Blindauer Alan James Stewart
Abstract:	<p>The study of chemical speciation and the refinement and expansion of omics-based methods are both consolidated and highly active research fields. Although well established, such fields are extremely dynamic and are driven by the emergence of new strategies and improvements in instrumentation. In the case of omics-based studies, subareas including lipidomics, proteomics, metallomics, metabolomics and foodomics have emerged. Here, speciomics is being proposed as an “umbrella” term, that incorporates all of these subareas, to capture studies where the evaluation of chemical species is carried out using omics approaches. This paper contextualizes both speciomics and the speciome, and reviews omics applications used for species identification through examination of proteins, metalloproteins, metabolites, and nucleic acids. In addition, some implications from such studies and a perspective for future development of this area are provided.</p>



Speciomics as a concept involving chemical speciation and omics

Marco Aurélio Zezzi Arruda^{a,b*}, Jemmyson Romário de Jesus^c, Claudia Andrea Blindauer^d, Alan James Stewart^e

^a Spectrometry, Sample Preparation and Mechanization Group, Institute of Chemistry, University of Campinas – Unicamp, P.O. Box 6154, Campinas, SP 13083-970, Brazil.

^b National Institute of Science and Technology for Bioanalytics, Institute of Chemistry, University of Campinas – Unicamp, P.O. Box 6154, Campinas, SP 13083-970, Brazil.

^c Research Laboratory in bionanomaterials, LPbio, Chemistry Department, Federal University of Viçosa, UFV, Viçosa, Minas Gerais, Brazil

^d Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom

^e School of Medicine, University of St Andrews, North Haugh, St Andrews, KY169TF, United Kingdom.

*Corresponding author

M.A.Z.A – zezzi@unicamp.br

Speciomics as a concept involving chemical speciation and omics

Marco Aurélio Zezzi Arruda^{a,b*}, Jemmyson Romário de Jesus^c, Claudia Andrea Blindauer^d, Alan James Stewart^e

^a Spectrometry, Sample Preparation and Mechanization Group, Institute of Chemistry, University of Campinas – Unicamp, P.O. Box 6154, Campinas, SP 13083-970, Brazil.

^b National Institute of Science and Technology for Bioanalytics, Institute of Chemistry, University of Campinas – Unicamp, P.O. Box 6154, Campinas, SP 13083-970, Brazil.

^c Research Laboratory in bionanomaterials, LPbio, Chemistry Department, Federal University of Viçosa, UFV, Viçosa, Minas Gerais, Brazil

^d Department of Chemistry, University of Warwick, Coventry, CV4 7AL, United Kingdom

^e School of Medicine, University of St Andrews, North Haugh, St Andrews, KY16 9TF, United Kingdom.

*Corresponding author

M.A.Z.A – zezzi@unicamp.br

Abstract

The study of chemical speciation and the refinement and expansion of omics-based methods are both consolidated and highly active research fields. Although well established, such fields are extremely dynamic and are driven by the emergence of new strategies and improvements in instrumentation. In the case of omics-based studies, subareas including lipidomics, proteomics, metallomics, metabolomics and foodomics have emerged. Here, speciomics is being proposed as an “umbrella” term, that incorporates all of these subareas, to capture studies where the evaluation of chemical species is carried out using omics approaches. This paper contextualizes both speciomics and the speciome, and reviews omics applications used for species identification through examination of proteins, metalloproteins, metabolites, and nucleic acids. In addition, some implications from such studies and a perspective for future development of this area are provided.

Keywords: chemical species, proteins, metabolites, metalloproteins, DNA, metals, instrumentation.

Significance

The synergic effect between chemical speciation and omics is highlighted in this work, demonstrating an emerging area of research with a multitude of possibilities in terms of applications and further developments. This work not only defines and contextualizes speciomics and individual speciomes, but also demonstrates with some examples the great potential of this new interdisciplinary area of research.

1. Introduction

In Charles Darwin’s 1859 seminal book, *On the origin of species*, he describes how living beings evolve from common ancestors, leading to the formation of new species [1]. However, the concept of species in chemistry is quite different from the biological context. While in biology, species are considered as different lifeforms created separately, to a chemist, species are defined as specific forms of an element (or molecule)

that differ based either on isotopic composition, electronic or oxidation state, and/or complexation or structure [2].

This is also emphasized in the definitions of the term “speciation” when either the biological or chemical meanings are considered. The biologist Orator F. Cook coined this term (speciation) first in 1906 for cladogenesis - the splitting of biological lineages (as opposed to anagenesis, phyletic evolution within lineages) [3]. In this context speciation means the formation of new species from a pre-existing one via a process that does not occur suddenly. In chemistry, IUPAC describes speciation as “analytical activities of identifying and/or measuring the quantities of one or more individual chemical species in a sample” [2].

To study and understand chemical speciation it has been necessary to develop techniques and approaches that can be used to evaluate different chemical species in complex mixtures. It is important to note that different approaches are required to assess different types of species, even when relating to the same element. For example, analytical strategies used to determine the total content of As in a sample are not able to evaluate either its oxidation state (III or V) or the types of complexes that may be present in a given mixture (dimethylated arseno sugar, arsenobetaine, among others). Indeed, to measure such speciation differences accurately requires distinct methods. This explains the necessity for and observed upsurge over the last few decades in the development of analytical techniques that can accurately assess speciation [4].

Recent progress made in the development of such approaches has been driven by the evolution of different separation techniques (chromatography and others), coupled with major advances in mass spectrometry (MS)-based approaches and instrumentation. With this, the study of chemical speciation has gained increased importance and become an extremely active area of research. Furthermore, the evolution of omics-based methods makes it possible to evaluate both small and large biomolecules on a large scale, greatly expanding the analytical toolkit available to examine chemical speciation [5]. Such omics-based methods aim to collectively characterize and quantify pools of biological molecules that impact on the function, structure, and dynamics of an organism (or even across different organisms) [6].

Following on from this, the “speciome”, a term adapted from biological sciences [7], can be also coined in chemistry due to the synergism of both chemical speciation and omics. Here the speciome can be defined as the entirety of the chemical species in a sample evaluated through omics-based strategies. Likewise, speciomics can then also be

defined as the study or evaluation of the chemical species present using omics approaches. It is however important to reinforce that chemical species themselves do not necessarily constitute a speciome. For example, if Cr(III)/Cr(IV) or $^{65}\text{Zn}/^{66}\text{Zn}$ are evaluated in different samples, they are considered as chemical species. However, if they are evaluated as part of a biological molecule, then all these chemical forms together could be considered as speciomes. Additionally, different protein isoforms, characterized using a specific omics approach, can be considered as speciomes.

2. Speciomics: integrating chemical species and omics

The concept of speciomics in chemistry is proposed as “umbrella” term containing all omics approaches devoted to speciation analysis. Figure 1 presents a general scheme of target analytes in biomedical and cellular studies; some examples of these will be presented in the text. Numerous examples of omics-based approaches (proteomics, metabolomics, metallomics, among others) applied to evaluate chemical species can be found in the literature. In all of these, the concept of speciation is inherent.

Here, this perspective article sets out to clarify the definition of speciomics in such contexts. Owing to recent developments involving metallopecies, many of the examples relate to metal complexes. However, it is important to reinforce that all chemical species (metallic or not) in a sample, characterized through any omics approach can be defined as a speciome.

2.1 Evaluating the Speciome

2.1.1 Proteins and Metalloproteins

MS-based proteomic techniques are commonly used to identify (and sometimes quantify) proteins in complex mixtures. Within such mixtures, there are often different species or isoforms of the same protein. Such species may be present due to alternative splicing, gene polymorphisms or post-translation modifications (PTMs). Although protein isoforms can originate from separate genes (e.g., actin isoforms) [8], single genes can code for multiple proteins due to the process of alternative (differential) mRNA splicing, which results in multiple protein forms that comprise both similar and distinct peptide sequences. Each isoform can have a distinct biological role [9-11]. Gene polymorphisms that give rise to differences in protein sequences across individuals (or between alleles) are also very common [12].

At least 400 types of PTM are currently known [13], ranging from small chemical modifications (e.g., phosphorylation and acetylation) to the addition of complete proteins (e.g., ubiquitination). However, not all such modifications have been extensively characterized. PTMs that are well understood include phosphorylation, N-, O- or C-linked glycosylation, acetylation, ubiquitination, methylation, SUMOylation, hydroxylation, carboxylation, palmitoylation, sulfation, and nitrosylation. These PTMs are known to occur at >1,000 unique modification sites, with each modified form constituting a different species. Stastna & Van Eyk have discussed developments in methodology and instrumentation for the enhanced detection of protein isoforms as well as examples of their biological importance [14]. Besides others, these include the study of fibronectin, a component of the extracellular matrix, which exists in many alternatively spliced isoforms [15]. They examined the phenotypes of mouse strains with different targeted mutations in the FN1 gene, linking the functions of alternatively spliced isoforms of fibronectin to human pathologies.

Metalloproteins are proteins that contain bound metal ions or metal complexes, with such interactions often being important for their functioning [5]. In such proteins, *apo*- and *holo*-forms can be simultaneously present in biological systems. Such examples include carbonic anhydrase (Zn), calmodulin (Ca), myoglobin (Fe) and ferritin (Fe). Some metalloproteins such as superoxide dismutase (Cu, Zn), simultaneously present with different metals in their structures. In addition, there are several superoxide dismutases that require different metal cofactors (Mn, Fe, Ni). These have been highlighted as prime examples of “cambialism” [16], a term that indicates that prediction of the specificity of a protein for binding a particular metal from sequence data alone is not possible. Hence this emphasizes the importance of (and requirement for) determining the native speciation of metalloproteins in biological samples.

Techniques used for the examination of metalloprotein speciation are most commonly based on liquid chromatography (LC), coupled with inductively-coupled plasma-MS (ICP-MS). Other separation techniques involving gel electrophoresis (GE) or capillary electrophoresis (CE) can also be used, with each presenting its own advantages and limitations. The best approach for a given application depends upon the matrix complexity, sample characteristics, or species information desired. With LC-ICP-MS, frequently size exclusion chromatography (SEC) is employed, with the column linked to a flow system with UV-Vis detection to detect both organic (through UV-Vis) and inorganic (through ICP-MS) species in different fractions [5,17]. To increase the

resolution of chromatographic separations for refinement of speciation detection, an additional column may be used - thus performing a 2D separation. This can be done on-line by having the sample pass through the columns sequentially, or off-line, where collected fractions are separated through the first column, and then subsequently, one or more fraction(s) are injected onto another column allowing a second separation (LC-LC-ICP-MS) [18]. Some examples of LC-LC-ICP-MS can be found in the literature [19-21]. For example, González-Fernández *et al.* [19] used this strategy for the identification of Cu-binding proteins in different mouse brains, where they used 2D separation (SEC followed by reverse phase). The use of these dual separation strategies can aid the examination of speciation owing to enhanced separation of the metalloproteins as they are based on use of two distinct properties of the molecules (orthogonal separation mechanisms). In the case above, separation was based on the hydrodynamic radii of the proteins followed by differences in polarity or charge. However, it is also possible to further improve resolution by adding more dimensions to the separation. These include using 2D chromatographic separations combined with non-chromatographic ones (such as SDS-PAGE). Such a “3D” approach has been used to identify metalloproteins in soybeans [22].

Besides these strategies, examination of metalloprotein speciation using Travelling Wave Ion Mobility Spectrometry coupled to MS (TWIMS-MS) is an elegant alternative. Even without chromatographic separation, this approach increases the selectivity in elemental speciation at atmospheric pressure. As a proof-of-concept, we identified different metal-bound forms of bovine carbonic anhydrase. The protein was found to bind not only Zn^{2+} , which is natively bound, but also Pb^{2+} , Cd^{2+} , Cu^{2+} , Se^{4+}/Se^{6+} and Cr^{3+}/Cr^{6+} . This work demonstrates the broad utility of TWIMS-MS for study of interactions between metals (presenting with different oxidation states) and biomolecules [23,24].

In addition to (metallo-)protein identification, quantification is also important. In this regard, the term “heteroatom-tagged proteomics” was first coined by Sanz-Medel [25], and used for quantitative evaluation of protein species. This was enabled due to the exceptional abilities of ICP-MS to track these heteroatoms (any bioelement of interest, e.g., Se, S, P, I, except C, H, O or N). Through use of LC-ICP-MS, and knowledge of the stoichiometry between the protein and the heteroatom of interest, the quantification of the protein is feasible. This includes proteins presenting with post-translational modifications (PTMs) [26]. In addition, if the element is not naturally present in the species, it can be

introduced chemically with bifunctional reagents containing a heteroatom (detected by ICP-MS) and a biomolecule recognition moiety (e.g., an antibody). In fact, the evaluation of heteroatoms through ICP-MS can simplify the complexity of a natural protein or peptide mixture (if a protein or group of proteins possess specificity toward that heteroatom), overcoming many analytical hurdles derived from the complexity of the mixture [27].

2.1.2 Metabolites and Metallometabolites

The metabolome of an organism reflects its diet, lifestyle, and genetic makeup [28], and many disorders are associated with metabolic dysregulation [28,29]. Metabolites are small molecules (molecular weight <1 kDa) that are formed through physiological chemical reactions [26]. Many complex diseases, such as cancer, diabetes, cardiovascular and neurological diseases are linked to multiple metabolic processes [29]. In this sense, the identification and quantification of metabolites associated with other chemical species can lead to a better understanding of the cause, prevention, and treatment of human disease [28].

In general, metabolomic workflows to characterize the metabolite profiles of an organism can be divided into two important areas: (i) targeted approaches; and (ii) untargeted approaches. The first strategy aims to quantify specific metabolites from the organism of interest, while untargeted approaches focus on a more comprehensive metabolomic assessment of metabolites derived from the specific organism, potentially allowing the discovery of biomarkers. Both strategies have their own advantages and limitations and are often used in combination to discover and quantify metabolites [30]. However, one of the main challenges involving targeted and untargeted approaches is associated with the biotransformation of metabolites, rendering their identification, quantification, and the understanding of their mechanisms of actions in the organism difficult [30]. For example, resveratrol (3,5,4'-trihydroxy-trans-stilbene, RSV) is one of the most widely studied polyphenols because of its potential health benefits [31]. Despite this, its mechanisms of action in the body are still unclear, even though several *in vitro* and *in vivo* studies have been performed. Its analysis is complicated by its pharmacokinetic profile and biotransformation into other metabolites, including isomeric and isobaric species, such as *trans*-resveratrol-3-*O*-sulfate, *cis*-resveratrol-3-*O*-sulfate, *trans*-resveratrol-3-*O*-glucuronide, *trans*-resveratrol-4'-*O*-glucuronide, *trans*-

resveratrol-3,4'-*O*-disulfate, *trans*-resveratrol-glucuronide-sulfate, among others (Figure 2) [31].

Furthermore, due to chemical similarity, isomeric or isobaric species also pose important challenges in studies of identification and quantification of metabolites. Molecules such as hexoses and related sugars (glucose, fructose, galactose, sorbose, mannose and inositol) are examples of metabolite species with important identification challenges [32]. The metabolites with the highest isomer counts are hexose-phosphate ($C_6H_{13}O_9P$) with 22 isomers, hexoses ($C_6H_{12}O_6$) with 17 isomers, hexose-lactones ($C_6H_{10}O_6$) with 13 isomers [32,33]. Lipids are similarly challenging, with leukotriene B4 ($C_{20}H_{32}O_4$) with 13 isomers. This challenge is also encountered in identifying some examples of metabolites in central carbon metabolism, such as isocytic and citric acid, isoleucine and leucine, as well as aminocaproic acid, β -leucine and norleucine [33,34]. In this sense, to ensure proper interpretation of data and understanding of the mechanism of action of particular metabolites, a reliable design of experiments is essential in metabolomics [35,36].

Another subclass of metabolites of interest in metabolomics studies relevant to human disease are metallometabolites [37,38]. Metallometabolites can generally form in different parts of the body due to their complexation capacity combined with the inter- and intracellular availability of metals, either during normal physiological processes or as the result of environmental exposure (toxicity) [39-41]. As an example, citrate and malate act as ligands for several metals, typically transition metals, while sulfur-rich metabolites tend to bind different elements, such as Se and Hg among others. Transition metals, such as Cu and Fe, as well as the d10 metal Zn, play important roles in the physiology of life [42,43]. Zn^{2+} , the most abundant d-block metal ion in cells, plays vital roles in the functioning of more than 300 enzymes and in DNA stabilization and gene expression [43]. However, the imbalance of metals in their natural cellular ratio – occasionally seen with small metal species – can result in detrimental effects such as the generation of reactive oxygen species (ROS) [41]. On the other hand, thiol peptides play important roles in the coordination of different heavy metals [42]. Intracellular thiol peptides possess sulfur containing groups (thiol or sulfhydryl group: -SH) that can serve as ligands for metal coordination. Such peptides include glutathione (GSH) and phytochelatins. In plant studies, thiol peptides complexed with Cd(II), Pb(II), As(III)/As(V), and Hg(II) have been investigated [40,43]. The fundamental difference

between these individual complexes is that the As- and Hg-thiol complexes are most stable thermodynamically due to the covalent character of the As-S and Hg-S bonds.

Spectroscopic (nuclear magnetic resonance, NMR), spectrometric (MS), and separation techniques (LC, GC, CE) are widely employed to identify and quantify species of metabolites using both targeted and untargeted approaches. The choice for a given technique is influenced by the sample matrix, the concentration and properties of the metabolites, and the amount of sample available. Furthermore, high precision analytical tools are required to investigate metallometabolites, [44,45]. For such an approach to be useful, it must allow identification of all distinct species of a particular element present in a sample, or at least identify some and clearly explain the unidentified number (in other words, produce a quantitative speciation model) [46,47]. MS offers a diversity of advantages for the identification of metabolites at sub-nanomolar levels [47], becoming one of the preferred analytical tools for detecting and identifying (metallo-)metabolites [40]. This has been facilitated by the development of "softer" electrospray ionization (ESI) techniques, which better enable complexes to remain intact during ionization. Since ICP-MS does not provide structural information, the combination of ICP-MS with ESI-MS can be deployed in a unique LC system for simultaneous structural identification and quantitation of complexes formed [44,45,48-52]. Additionally, advances in hyphenated techniques of CE (CE-ICP-MS and CE-ESI-MS) also show great promise in the characterization of complex metallometabolite species [47]. Table 1 summarizes a comparison of the main analytical techniques applied in studies for metabolomics, highlighting the uses and limitations of each one.

In approaches that involve separation techniques, the choice of chromatographic approach essentially depends on the characteristics of the metabolites of interest. In some situations, orthogonal configurations are necessary for reliable identification [30,32]. Recently, in a study to quantify metallometabolites of Cr in *Taraxacum officinale*, Markovic et al. developed a new analytical procedure based on LC-ICP-MS using strong anion-exchange chromatography for separation of the Cr-metabolites [37]. In this study, Cr-oxalate, Cr-malate, Cr-citrate, Cr-aconitate, Cr-quininate, and Cr-nitrate were efficiently separated and quantified. In another study, Ramírez-Acosta et al. developed a method using column switching and species-unspecific isotopic dilution to quantify selenometabolites and selenoproteins in human hepatic cells [38]. In this study, the authors observed that Se supplementation increased the levels of selenometabolites and selenoproteins in cells.

2.1.3 Metal-nucleic acid complexes

Interactions between metal ions and nucleic acids (DNA and RNA) readily occur. This is not surprising considering that the structure of nucleic acids includes a phosphate backbone with a net-negative charge, a (deoxy-)ribose sugar with O atoms, and purine and pyrimidine bases that contain O and N atoms [53]. Metal ions (such as K^+ , Na^+ , and Mg^{2+}) bind DNA and RNA and are essential to their stability as they neutralize the net negative charge arising from the phosphate backbone [54]. In fact, metal ions can bind to nucleic acids directly or indirectly through hydrogen bonding to the coordinating water molecules surrounding the metal ions. Metal binding to the bases usually disrupts base pair hydrogen bonding and destabilizes the double helix [55].

Metal-bound nucleic acids can be considered as separate chemical species, particularly in cases where nucleic acid-metal interactions are sequence-specific. Furthermore, metals are indirectly important for forming nucleic acid-protein complexes (e.g. Mg^{2+} frequently serves as an enzyme cofactor in DNA and RNA polymerases) and regulatory proteins (e.g. zinc fingers are amongst the most abundant DNA-binding domains in eukaryotes).

Additionally, Tanaka *et al.* made a synthetic DNA where the natural bases were replaced with artificial ones [56]. Each of those substituted bases possessed a high affinity for either Cu^{2+} or Hg^{2+} . The modified DNA was only found to form a double helix if the opposing bases preferred (and bound to) the same metal ion. Such artificial DNAs can be employed in organic chemistry as catalysts for enantioselective reactions, which are important for the synthesis of certain biologically active compounds.

While DNA speciation is relevant in the context of metal binding, integrating the concept of chemical species in the context of genes, and genomics, is a hard task, particularly as genes are rarely considered as chemical species. Previously, Ascone *et al.* proposed the concept of metallogenomics, involving the study of metalloproteins in the context of structural genomics programs [57], where the 3D structure of every protein encoded by a particular set of genes is systematically characterized. Examples include the work of the Structural Genomics Consortium, who aim to obtain representative structures across a broad range of families of human proteins and make their coordinates publicly accessible [58,59]. Ascone *et al.* suggested that X-ray absorption spectroscopy data could be incorporated within such platforms to provide structural data pertaining to metal sites in proteins.

3. Implications and Perspectives

The concept of speciomics has important implications for capturing the diversity of studies involving examination or measurement of chemical “species”. A better understanding of these can lead to paradigm shifts in a range of life science disciplines, allowing a more complete understanding of the diversity of biological, chemical, medical and environmental studies. For example, nucleic acid-metal interactions encompass a large area of research that ranges from the most fundamental characterization of metal ion binding to the role of nucleic acid-bound metal ions in health and disease. Alternative DNA base-pairing mediated by metal binding is currently being investigated and manipulated for applications in nanotechnology, molecular machines, and logic gates [49].

Additionally, with respect to instrumentation, it is necessary to think in terms of multi-speciation-omics, with multimodal strategies, linking inorganic (metal) with organic (protein or other biomolecule) species-focused instrumentation. Current work is aimed at improving multi-dimensional separation, potentially by combining multi-dimensional with multi-modal separations to observe a target species in a complex mixture [44]. Figure 3 summarizes strategies employing spectrometry across different wavelengths and times. However, mass spectrometry-based approaches are also possible. These include MALDI-IMS, in which the sample, often a thin tissue section, is moved in two dimensions while the mass spectrum is recorded [60], TWIMS-MS (already exemplified in this text, see section 2.1.1) [23,24], or Orbitrap-ICP-MS, allowing the unambiguous observation of isotopic patterns and identification of compounds with high precision [61].

All these examples provide a means for multi-speciomic evaluation. Many of these studies can generate very large data files furthering the need for use of artificial intelligence and techniques that facilitate big data analysis. All of these factors can be accurately observed in Figure 4, where through a bibliometric evaluation using the VOSviewer program, employing “chemical speciation” and “omics” as search terms, a diversity of further linked terms are found. These incorporate terms relating to chemical species (“chemical composition” and “heavy metals”), organism (“plants” and “gut microbiota”), instrumentation (“spectroscopy” and “mass spectrometry”), environment (“plasma”, “sediments” and “sea water”) and disease processes (“oxidative stress” and “insulin resistance”). Additionally, when focusing the search to articles between 2012 and 2020, thematic areas such as “artificial intelligence”, “big data”, “validation”,

“microbiome”, “gut microbiota”, and “precision medicine” emerge, providing insight into areas where speciomics studies will be highly useful going forward.

4. Conclusions

It is easy to rationalize that the complexity involved in performing speciomics studies is enormous. The quantity of the data produced may be so large that artificial intelligence is employed for data analysis, as demonstrated through the bibliometric analysis performed. Speciomics allows the discovery of new species and, with them, a more complete biochemical understanding of environmental, physiological and disease processes. In the latter case it may be used to aid diagnosis or disease management.

We propose that speciomics represents a jigsaw puzzle, with each chemical species representing the individual pieces. Only by discovering the right piece and putting it in the right place, will the picture appear in its fullness.

Authors statement

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *J. Proteomics*.

Conflict of interest

There is no conflict of interest regarding this manuscript

Acknowledgments

The authors thank the FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) [grant numbers 2014/50867-3, 2017/50085-3, 2018/25207-0, 2019/00018-3, 2019/24445-8, and 2020/08543-7], the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brasília, Brazil) [grant number 303231/2020-3], and also the Biotechnology and Biological Sciences Research Council (BBSRC, United Kingdom, BB/V014684/1)

Abbreviation

CE- capillary electrophoresis

ESI - electrospray ionization

GE - gel electrophoresis

ICP-MS - inductively-coupled plasma mass spectrometry

LC - liquid chromatography

MALDI-IMS - Matrix-Assisted Laser Desorption/Ionization Imaging Mass Spectrometry

NMR - nuclear magnetic resonance

PTMs - post-translation modifications

SEC - size exclusion chromatography

SH - sulfhydryl group

TWIMS-MS - travelling wave ion mobility spectrometry coupled to MS

5. References

1. C.R. Darwin, *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*, London: John Murray, 1859.
2. D.M. Templeton, F. Ariese, R. Cornelis, L. G. Danielsson, H. Muntau, H. P. Van Leeuwen, R. Łobinski, Guidelines for terms related to chemical speciation and fractionation of elements. Definitions, structural aspects and methodological approaches (IUPAC recommendations 2000), *Pure Appl. Chem.* 72 (2000)1453-1470.
3. O.F. Cook, Factors of species-formation, *Science.* 23 (1906) 506–507.
4. R. Cornelis, *Handbook of elemental speciation, Techniques and Methodology*, Wiley, 2003.
5. J.R. Jesus, L.F. Costa, E.L. Lehmann, R.M. Galazzi, K.C. Madrid, M.A.Z. Arruda, *Chemical Speciation and Metallomics*, in: M.A.Z. Arruda (Ed), *Metallomics: The science of biometals*, Springer, 2018, pp. 183-211.
6. AltTox.org. <https://alttox.org/mapp/emerging-technologies/omics-bioinformatics-computational-biology/>, 2021 (Accessed 26 July 2021).
7. Omics.org. https://omics.org/Main_page, 2021 (Accessed 26 July 2021).
8. B.J. Perrin, J.M. Ervasti, The actin gene family: function follows isoform, *Cytoskeleton.* 67 (2010) 630–634.
9. P.J. Grabowski, D.L. Black. Alternative RNA splicing in the nervous system. *Prog Neurobiol.* 65 (2001) 289–308.
10. D.L. Black, Mechanisms of alternative pre-messenger RNA splicing, *Annu Rev Biochem.* 72 (2003) 291–336

11. B.J. Blencowe, Alternative splicing: new insights from global analyses, *Cell*. 126 (2006) 37–47
12. S. Ismail, M. Essawi, Genetic polymorphism studies in humans, *Middle East Journal of Medical Genetics*. 1 (2012) 57–63
13. P. Minguéz, L. Parca, F. Diella, D. R. Mende, R. Kumar, M. Helmer-Citterich, A.C. Gavin, V. van Noort, P. Bork, Deciphering a global network of functionally associated post-translational modifications, *Mol Syst Biol*. 8 (2012) 599–613.
14. M. Stastna, J. E. Van Eyk, Analysis of proteins isoform: can we do it better? *Proteomics*. 12 (2012) 2937–294.
15. E.S. White, F.E. Baralle, A.F. Muro, New insights into form and function of fibronectin splice variants, *J Pathol*. 216 (2008) 1–14.
16. K.A. Frye, K.M. Sendra, K.J. Waldron, T.E. Kehl-Fie, Old dogs, new tricks: New insights into the iron/manganese superoxide dismutase family. *J Inorg Biochem*. 4,220(2022)111748. doi: 10.1016/j.jinorgbio.2022.111748. Epub ahead of print. PMID: 35151099.
17. J.P.C. Coverdale, J.P. Barnett, A.H. Adamu, E.J. Griffiths, A.J. Stewart, C.A. Blindauer, A metalloproteomic analysis of interactions between plasma proteins and zinc: elevated fatty acid levels affect zinc distribution, *Metallomics*. 11 (2019) 1805–1819.
18. M. A. O. da Silva, A. Sussulini, M. Z. Arruda, Metalloproteomics as an interdisciplinary area involving proteins and metals, *Expert Rev. Proteomics*. 7 (2010) 387–400.
19. M. González-Fernández, T. García-Barrera, A. Arias-Borrego, J.J.C. Pueyo, J.L. Barea, J.L. Gómez-Ariza, Metallomics integrated with proteomics in deciphering metal-related environmental issues, *Biochimie*. 91 (2009) 1311–1317.
20. I. Neverova, J.E.V. Eyk, Role of chromatographic techniques in proteomic analysis, *J. Chromatogr. B*. 815 (2005) 51–63.
21. A. Motoyama, J.R. Yates III, Multidimensional LC separations in Shotgun Proteomics, *Anal Chem*. 80 (2008) 7187–7193.
22. L.R.V. Mataveli, M.A.Z. Arruda, Expanding resolution of metalloprotein separations from soybean seeds using 2D-HPLC-ICP-MS and SDS-PAGE as a third dimension, *J. Proteomics*. 104 (2014) 94–103.
23. G.S. Pessôa, E.J. Pilau, F.C. Gozzo, M. A. Z. Arruda, Ion mobility mass spectrometry: an elegant alternative focusing on speciation studies, *J. Anal. At. Spectrom*. 26 (2011) 201–206.
24. G.S. Pessôa, E.J. Pilau, F.C. Gozzo, M.A.Z. Arruda, Ion mobility spectrometry focusing on speciation analysis of metals/metalloids bound to carbonic anhydrase, *Anal. Bioanal. Chem*. 405 (2013) 7653–7660.
25. A. Sanz-Medel, From metalloproteomics to heteroatom-tagged proteomics, *Anal. Bioanal. Chem*. 381 (2005) 1–2.
26. A. Sanz-Medel, Heteroatom(isotope)-tagged proteomics via ICP-MS: screening and quantification of proteins and their post-translational modifications, *Anal. Bioanal. Chem*. 391 (2008) 885–894.

27. A. Sanz-Medel, "Heteroatom-tagged" quantification of proteins via ICP-MS, *Anal. Bioanal. Chem.* 408 (2016) 5393-5395.
28. M.M. Rinschen, J. Ivanisevic, M. Giera, G. Siuzdak, Identification of bioactive metabolites using activity metabolomics, *Nat Rev Mol Cell Biol.* 20 (2019) 353-367.
29. K. Mi, Y. Jiang, J. Chen, D. Lv, Z. Qian, H. Sun, D. Shang, Construction and analysis of human disease and metabolites network, *Front. Bioeng. Biotechnol.* 8 (2020) Article 398.
30. V. Lelli, A. Belardo, A.M. Timperio From Targeted Quantification to Untargeted Metabolomics. In: Zhan X, editor. *Metabolomics – Methodology and Applications in Medical Science and Life Science*, Vol 1. London: IntechOpen Limited; 2021. Doi:10.5772/intechopen.96852.
31. M. Springer, S. Moco, Resveratrol and its human metabolites – Effects on metabolic health and obesity, *Nutrients.* 11 (2019)1-17. Doi:10.3390/nu11010143.
32. H. Oberacher, Applying Tandem Mass Spectral Libraries for Solving the Critical Assessment of Small Molecule Identification (CASMI) LC/MS Challenge 2012, *Metabolites.* 3 (2013) 312-324.
33. S.L. Ramírez-Garza, E.P. Laveriano-Santos, M. Marhuenda-Muñoz, C.E. Storniolo, A. Tresserra-Rimbau, A. Vallverdú-Queralt, R.M. Lamuela-Raventós, Health Effects of Resveratrol: Results from Human Intervention Trials, *Nutrients.* 10 (2018) 1892. doi:10.3390/nu10121892.
34. T. Opialla, S. Kempa, M. Pietzke, Towards a More Reliable Identification of Isomeric Metabolites Using Pattern Guided Retention Validation, *Metabolites.* 10 (2020) 457. Doi:103390/metabo10110457.
35. K. Segers, S. Doclerck, D. Mangelings, Y.V. Heyden, A.V. Eeckhaut. Analytical techniques for metabolomics studies: a review, *Bioanalysis.* 11 (2019) 2297-2318.
36. A. Alonso, S. Marsal, A. Julià. Analytical methods in untargeted metabolomics: state of the art in 2015, *Front. Bioeng. Biotechnol.* 3 (2015) 1-20.
37. S. Markovic, L. Levstek, D. Zigon, J. Scancar, R. Milacic. Speciation and bio-imaging of chromium in *Taraxacum officinale* using HPLC post-column ID-ICP-MS, high resolution MS and laser ablation ICP-MS techniques. *Front. Chem.*
38. S. Ramírez-Acosta, R.Uhlířova, F. Navarro, J.L. Gomez-Ariza, Tamara García Barrera. Antagonistic interaction of selenium and cádmium in human hepatic cells through selenoproteins, *Front. Chem.* (2022).
39. J.R. de Jesus, M.A.Z. Arruda, Human disease biomarkers: challenges, advances, and trends in their validation, *JIOMICS.* 11 (2021) 1-28.
40. M. Montes-Bayón, M. Sharar, M. Corte-Rodriguez, Trends on (elemental and molecular) mass spectrometry based strategies for speciation and metallomics, *TrAC - Trends in Analytical Chemistry.* 104 (2018) 4–10.
41. S.C. Wilschefski, M.R. Baxter, Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects, *Clin. Biochem. Rev.* 40 (2019) 115-133.

42. E. Rathahao-Paris, S. Alves, C. Junot, J.C. Tabet. High resolution mass spectrometry for structural identification of metabolites in metabolomics, *Metabolomics*, 10 (2016) 1–15. DOI: 10.1007/s11306-015-0882-8
43. J. López-Barea, J.L. Gómez-Ariza. Environmental proteomics and metallomics, *Proteomics*, 6 (2006) 51-62.
44. M. Tsednee, Y.C. Huang, Y.R. Chen, K.C. Yeh, Identification of metal species by ESI-MS/MS through release of free metals from the corresponding metal-ligand complexes. *Scientific Reports*. 6 (2016) 1-13.
45. J.R. Jesus, M. Arruda, Unravelling neurological disorders through metallomics-based approaches, *Metallomics*. 12 (2020) 1878-1896.
46. D. Wesenberg, G.J. Krauss, D. Schaumlöffel, Metallo-thiolomics: Investigation of thiol peptide regulated metal homeostasis in plants and fungi by liquid chromatography-mass spectrometry, *International Journal of Mass Spectrometry*, 307 (2011) 46–54.
47. Y.F. Li, C. Chen, Y. Qu, Y. Gao, B. Li, Y. Zhao, Z. Chai, *Metallomics, elementomics, and analytical techniques*. Pure and Applied Chemistry. 80 (2008) 2577–2594.
48. H. Haraguchi, Metallomics as Integrated biometal science, *J. Anal. At. Spectrom*, 19 (2004) 5–14.
49. S. Mounicou, J. Szpunar, R. Lobinski, *Metallomics: the concept and methodology*. Chem. Soc. Rev. 38 (2009) 1119–1138.
50. R. Ge, X. Sun, Q. He, Overview of the Metallometabolomic methodology for metal-based drug metabolism, *Curr. Drug. Metab.* 12 (2011) 287–299.
51. Y. Ogra, T. Hirata, *Metallomics: Recent analytical techniques and applications*, Springer, 2017.
52. J.R de Jesus, G.S Pessôa, A. Sussulini, J.L.C. Martínez, M.A.Z. Arruda, Proteomics strategies for bipolar disorder evaluation: From sample preparation to validation, *J. Proteomics*, 145 (2016) 187-196.
53. D.L. Morris Jr., DNA-bound metal ions: recent developments, *Biomol Concepts*. 5 (2014) 397–407.
54. N.V. Hud, A.E. Engelhart, Sequence-specific DNA-metal ion interactions, in: N.V. Hud (Ed.), *Nucleic acid-metal ion interactions*, Cambridge: Royal Society of Chemistry, 2009, pp. 75–117.
55. J. Anastassopoulou, Metal-DNA interactions, *J. Mol. Structure*. 651-653 (2003) 19-26.
56. K. Tanaka, G.H. Clever, Y. Takezawa, Y. Yamada, C. Kaul, M. Shionoya, T. Carell, Programmable self-assembly of metal ions inside artificial DNA duplexes, *Nature Nanotechnology*. 1 (2006) 190–194.
57. I. Ascone, R. Fourme, S. Hasnainc, K. J. Hodgson, Metallogenomics and biological X-ray absorption spectroscopy, *J. Synchrotron Radiat*. 12 (2005) 1-3.
58. A.R. Williamson, Creating a structural genomics consortium, *Nat. Struct. Biol.* 7 (2000) 953.

59. I. Bertini, A. Rosato, Bioinorganic chemistry in the postgenomic era, PNAS 100 (2003) 3601-3604.
60. M. Aichler, A. Walch, MALDI Imaging mass spectrometry: current frontiers and perspectives in pathology research and practice, Laboratory Investigation 95 (2015) 422-431.
61. P. Flis, L. Ouerdane, L. Grillet, C. Curie. S. Mari, R. Lobinski, Inventory of metal complexes circulating in plant fluids: a reliable method based on HPLC coupled with dual elemental and high-resolution molecular mass spectrometric detection, New Phytologist 211 (2016) 1129-1141.

Tables

Table 1- Comparison of the analytical techniques applied in targeted and untargeted metabolomics. Symbols (from – to +++) indicate the level of property.

Property	Analytical technique					
	LC-ESI-MS	GC-MS	CE-ESI-MS	NMR	LC-ICP-MS	CE-ICP-MS
Sensibility	+++	+++	+	-	+++	++
Reproducibility	high	high	high	high	high	lower than LC-ICP-MS
Quantification	+++	+++	+++	+	+++	+++
Sample volume for injection	2-20 μ L	1-20L	2-20 μ L	100-500 μ L	2-20 μ L	2-20 μ L
Identifications of organic species	Yes (requires standard)	Yes (spectral libraries)	Yes (required standard)	Yes	No	No
Elemental identification	No	No	No	Yes (depend on the probe)	Yes	Yes
Untargeted approach	++ (difficult identification)	+++ (easy identification using spectra libraries)	++ (difficult identification)	+++ (easy identification)	++	++
Targeted approach	+++ (requires standard)	+++	+++ (requires standard)	+++ (easy identification)	+++ (requires standard)	+++ (requires standard)

Figure captions

Figure 1. Types of molecules examined in speciomimetic studies. Analysis of these chemical species is important for the evaluation of human health, single cell functioning and other biological activities. The size of each cluster gives an approximate idea of the number of related applications in the literature.

Figure 2. Resveratrol species in the human metabolism. (1) *trans*-resveratrol, (2) dihydroresveratrol, (3) 3,4'-*O*-dihydroxy-*trans*-stilbene, (4) lunularin, (5) *trans*-resveratrol-3-*O*-sulfate, (6) *trans*-resveratrol-4'-*O*-sulfate, (7) *trans*-resveratrol-3,4'-*O*-disulfate.

Figure 3. Multimodal instrumentation using for speciomimetic studies based on spectroscopy techniques. (A) Schematic of spectroscopy instrumentation for analysis of chemical species (metals and biomolecules). (B) Acquisition and analysis of spectra to produce 2D and 3D images. Resultant images present the distribution of the analytes at each wavelength (left) and at each analysis time (right).

Figure 4. Mapping of keyword co-occurrence obtained using the VOSviewer program. The search comprised articles published between 2012-2020.

Figure 1

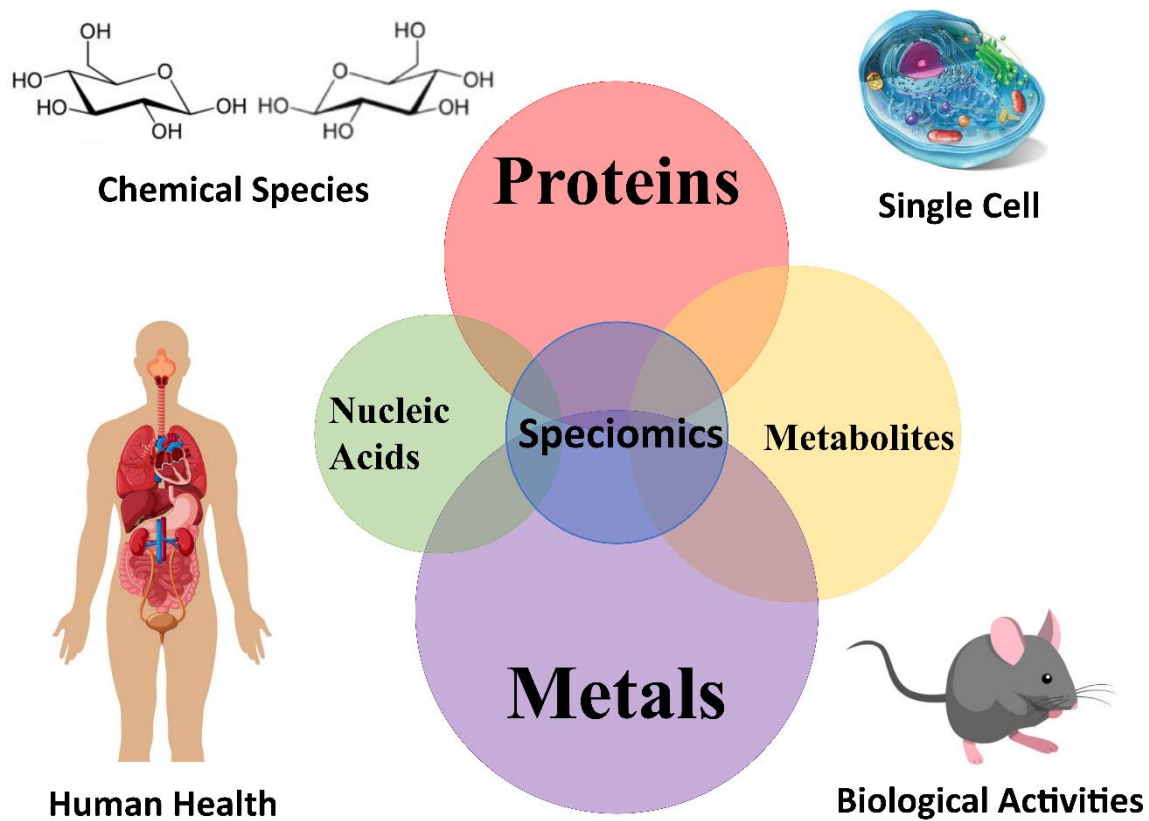
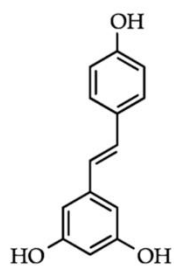
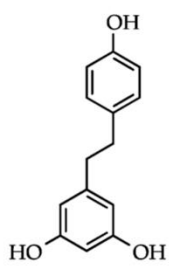


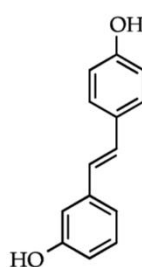
Figure 2



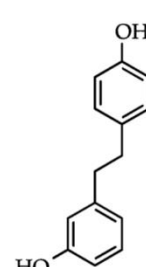
1



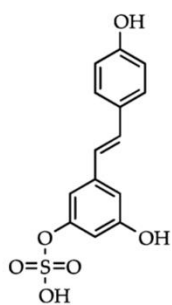
2



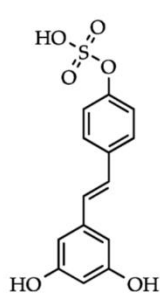
3



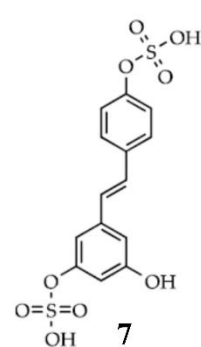
4



5



6



7

Figure 3

