Application of MALDI Mass Spectrometry in Natural Products Analysis

Abstract

This article presents the utility of mass spectrometry with a MALDI ionization source in natural products analysis. The advantages and drawbacks of this technique for natural products analyses will be presented and discussed. In addition, the structural determination of secondary metabolites using MALDI-MS/MS will be explored, which can guide MALDI experimental methods and stimulate new research in this area. Finally, several important approaches for MALDI data processing will be discussed.

Abbreviations

- MALDI
- dereplication
- natural products
- fragmentation
- data processing
- APCI: atmospheric pressure chemical ionization
- CHCA: α-cyano-4-hydroxycinnamic acid
- CID: collision induced dissociation
- DAN: 1,5-diaminonaphthalene
- DHB: 2,5-dihydroxybenzoic acid
- DIT: dithranol
- EI: electron ionization
- FAB: fast atom bombardment
- FT: Fourier transform
- IAA: trans-3-indoleacrylic acid
- ICR: ion cyclotron resonance
- IMS: imaging mass spectrometry
- IT: ion trap
- LDI: laser desorption/ionization
- LiDHB: lithium 2,4-dihydroxybenzoate
- MALDI: matrix-assisted laser desorption/ionization
- NALDI: nanostructure-assisted laser desorption/ionization
- PSD: post source decay
- Q: quadrupole
- ReTOF: reflectron time-of-flight
- SA: sinapinic acid
- SELDI: surface-enhanced laser desorption/ ionization
- THAP: 2,4,6-trihydroxyacetophenone
- DHAP: 2,5-dihydroxyacetophenone
- 4NA: 4-nitroaniline
- 9-AA: 9-aminoacridine
- THAP: 2,4,6-trihydroxyacetophenone
- HABA: 2-(4-hydroxyphenylazo)benzoic acid
- SELEN: surface-enhanced laser desorption/ ionization

Introduction

In the late 1950s, Roland Gohlke and Fred McLafferty hyphenated an EI mass spectrometer as the detector in gas chromatography [1], which had a tremendous influence on natural products chemistry. The impact of this innovation can be measured by the number of reviews written during the 1960s about the application of mass spectrometry in natural products analysis. In the same period, another technique based on laser desorption/ionization mass spectrometry appeared and was applied to the ionization of organic molecules by irradiation with a high-intensity laser pulse. This method had a lower impact on the natural products research than electron ionization between 1950 and 1970. Research by Michael Karas and Franz Hillenkamp led to the new and alternative use of a matrix in the laser ionization process in 1987 [2]. The authors demonstrated the application of MALDI in the analysis of high-molar mass compounds by the combination of the analyte’s solubilization in an organic matrix and its excitation by a laser. In this process, the matrix must have a strong absorbance at a specific wavelength and must be easily sublimated [3]. Ionization generally occurs through acid-base reactions (protonation and deprotonation) of the analyte and the matrix in a hot plume of ablated gases, but coordination and oxidation reactions may also occur [3, 4]. Considering the chemical properties of the matrix, the initial studies were dedi-
cated to biomolecules (DNA, proteins, peptides, sugars, and others), with few applications in natural products chemistry. In 2006, Crotti and coworkers published a review that compared the growth of several ionization processes from 1992 to 2004 [5]. At that time, ESI was greatly increasing, and the total number of published studies using ESI was twice the number of those using MALDI in 1994. The use of EI was largely constant during the analysis period, which is likely due to the requirement of thermostability and volatilization of the analytes. Natural products chemistry was responsible for only 5% of all mass spectrometry publications, and the majority of them were related to ESI [5]. An update on the impact of ESI, EI, and MALDI ionization techniques on natural products chemistry is shown in Fig. 1. The bibliographic search was of the Chemical Abstracts Service (CAS) by SciFinder Scholar (March 13th, 2015) using MALDI, ESI, and EI ionization as keywords in the first step (Fig. 1A). ESI applications increased from 2004 to 2008, but the number of articles remained at approximately 5000 publications per year from 2008 to 2014. EI exhibited a slight increase compared with the previous compilation [5] but remained constant at approximately 2500 publications per year. Conversely, MALDI publications increased annually, and no stabilization was observed (Fig. 1A).

The bibliographic search refined with the term natural products in a second step (Fig. 1B). ESI is still growing, but MALDI remains constant, similar to EI. This trend likely reflects the enormous expansion of metabolomics studies, which utilizes ESI [6], and the limited knowledge or perspective of MALDI in natural products investigations. The same effect can be observed in a recent volume of Natural Products Reports that was dedicated to the mass spectrometry of natural products, coordinated by Pieter Dorrestein [7]. This report discussed the applications and advances in the instruments, the improved experimental workflows, and the important mass spectrometry tools from the perceptive of increased performance of structural elucidation and natural products functional characterization. The majority of publications were focused on new perspectives of ESI and other atmospheric pressure ionization processes in addition to the well-established use of EI. MALDI was only highlighted in tissue imaging studies applications [8–10], thus confirming the tendency of uses previously discussed and presented in Fig. 1. Outside the reduced focus on the MALDI application, various successful studies have been published on the application of MALDI-MS in natural products analysis, which will be discussed throughout this article. Therefore, significant and exciting progress in MALDI-MS natural products applications is expected in the coming years.

**MALDI-MS Analyses of Natural Products**

**MALDI technique**

MALDI is a soft ionization technique initially developed for macromolecule analyses, which is greatly expanding due to its advantages, although more research is required to understand the processes involved, primarily the reactions in the ionization steps and fragmentation, especially for methods with a high energy transfer [11–13]. Basically, the ionization processes in MALDI can be explained through three basic steps: 1) the incorporation and isolation of analytes in a matrix, 2) the excitation of the matrix producing the plume by physics desorption/ablation, and 3) the ionization of the analytes by ion–molecules reactions [11–13]. There are different models that try to explain the MALDI ionization process, but they are crude, and no method is universally accepted. One widely discussed model is the cluster model, originally called the “lucky survivors” model. One common agreement is the function of the matrix, which includes the isolation of analyte molecules, the absorption of laser energy, the co-desorption of analytes and the charge transfer. In addition, the matrix can be useful to protect the analyze molecules, reducing their in-source dissociations [3].

After laser energy deposition, photons from the laser are absorbed by the matrix molecules and a conversion of most of the energy to heat occurs. Subsequently, the matrix-analyte solid dissociates and a plume is produced, in which the secondary chemical reactions occur, promoting the ionization of the analytes by charge transfer. It can be of three types: proton, electron, and cation transfer. These ion–molecule reactions happen in the plume and they are reduced significantly with the plume expansion, thus the secondary reactions did not run to completion and many matrix ions are not neutralized, being observed in the spectra. The high-plume density is an important parameter to increase the secondary reactions and to reduce the matrix ions in the spectra. It can be targeted by a higher laser fluence (related to the concentration of primary ions) [3, 11–13]. Besides laser intensity, the secondary plume reactions can be additionally controlled by analyte concentration and matrix choice [3, 11, 12].

The matrix excitation occurs electronically by ultraviolet lasers or vibrational stages by infrared lasers. And there are different kind of used lasers in MALDI, such as nitrogen: Nd : YAG, CO₂, Er : YAG, ArF, KrF, and others. The lasers work using different wavelengths and amount of single-photon energy that influences the ionization efficiency, mainly for LDI analyses (without matrix). The most common lasers are nitrogen (337 nm, single-photon energy of 3.6 V) and frequency-tripled Nd: YAG (355 nm, single-photon energy of 3.8 V).
energy of 3.3 V). Different to the nitrogen laser, Nd:YAG has a higher laser frequency, enabling very fast data acquisition, and the Gaussian energy profile (not equally distributed) that could represent some problems of sensitivity and resolution. However, a modified Nd:YAG laser was developed without this disadvantage, showing a well-structured energy profile [13–15].

Moreover, the choice of mass analyzer is an important decision to get good results according to its specific features, which include the following: resolution, mass accuracy, mass range extension, sensitivity, dynamic range, quantification, speed, and handling. The most common analyzers applied for MALDI are IT, quadrupole (Q), Orbitrap, FT-ICR, and TOF, and there also are hybrid mass spectrometers from these analyzers. As important reviews and books have described the analyzers [6,16,17], an extensive description is not shown here, and only the special applications in the area are reported. In addition, there are mass spectrometers coupled to ion mobility spectrometry; an electrophoretic technique that separates the ions based on their mobilities in the gas phase, allowing the chemical study of conformations, separation of isomers, and isobars ions. The ion mobility spectrometry is a theme of many review articles, describing the mass spectrometers available, their advantages, and disadvantages [18, 19].

Different conventional matrices have been used for several purposes, such as DHB, CHCA, DHAP, SA, 4NA, THAP, nicotinic acid, picolinic acid, ferulic acid, and others, but few matrices were well characterized and many points are still unclear, for example, a deep explanation about their efficiencies, the chemical species produced in the plume for each matrix, and the information about the physics and chemistry properties of plume expansion and reactions [12, 20]. The good features for matrices are linked to their solubility, absorbivity, reactivity, volatility, and desorption and a considerable number of reports provide details on preparation methods of different matrices, which includes dried droplet, crushed crystal, fast evaporation, overlay method, spin coating, and electrospray [20]. Some matrices have known applications, such as for oligonucleotides, proteins, lipids, polymers, and carbohydrates (Table 1), while indication guides for specific natural product classes have not been reported [11] that stimulates more studies in this area, since there are innumerable advantages of MALDI (described on next item). The main analyses of natural products by MALDI, including the applied matrices, are summarized in Tables 1 and 2. Another important point for choice matrix is variable amounts of energy are transferred to analytes, which depends on the matrix and can represent different effects on ion fragmentation due to changes of their internal energy [21], but these effects are underexplored, mainly for natural products.

LDI analyses without a matrix have also been described, mainly for conjugated compounds, which is possible for ionization without a matrix since they can absorb the laser radiation. Some natural products were already analyzed by LDI, such as flavonoids and carotenoids, but the role of the matrix is fundamental in reducing the in-source dissociation as well as increasing the efficiency of ionization and thereby increasing sensitivity [22,23], as confirmed in the investigation of ionization and in-source dissociation of twenty-six flavonoids [22] and others studies (section Application – for more examples and details).

**Drawbacks and advantages**

The main disadvantages reported for MALDI analyses include the background matrix ions, which are observed in the mass range used for small compounds (<1000 Da), the possibility of sample photodegradation, the in-source dissociations of analytes, and the difficulty in working on line with liquid chromatography, which is not yet commonly applied but yields additional information about isomers [6,11]. Some alternatives have been described to reduce this background, for example, carbon nanotubes, ionic liquids, graphene, surfactants with traditional matrices, and a low proportion of matrix (matrix: analyte) [24–26]. In addition, there have been great efforts to reduce the fragmentation of analytes in the source [22]. This information is relevant for improving MALDI results and applications in many study areas, such as metabolomics, dereplication, quantification, biological fingerprinting analyses, and IMS, stimulating its utility in natural products research. The low reproducibility is another drawback already reported, but several approaches have been described to improve it, many of which are related to crystal homogeneity in different matrices, sample preparation methods, and/or increasing the number of runs per sample [24,26–30]. The methods to improve the cocrystallization of the analyte-matrix mixture include procedures such as fast evaporation preparation, electrospray sample deposition, and others [6,26,31,32]. So MALDI has also been successfully applied to quantitative studies for alkaloids [33–35], anthocyanins [36–38], flavonoids [39], acetogenins [40], spirolides [41], curcuminoids [42], and rotenoids [43].

From new technologies overcoming the bottlenecks of MALDI analyses, the scientific community can have access to all its benefits and the application of it can be enlarged in the natural products area, mainly because the MALDI technique exhibits several advantages. They include the ability to analyze complex mixtures with less ion suppression compared to ESI, high sensitivity, high tolerance of salts and contaminants, low sample consumption, high throughput, simple and rapid sample preparation, low time consumption to obtain the spectra (≈ 60 s), and the production of singly charged species [6,27].

### Applications

MALDI-MS has been applied to determine the molecular weight of some natural products, to identify the chemical structures, as well as for metabolomic studies, quantification and others, however, recently it has also been used to establish tissue distribution of the metabolites by MALDI imaging [6,25,27,31] (Fig. 2). Although there are several applications of MALDI in natural products chemistry, the higher number of published articles is related to its use for molecular weight determination by low or high resolution. Several glycosylated and non-glycosylated secondary metabolites have been analyzed, such as hydrolysable and condensed tannins, anthocyanins, alkaloids, flavonoids, saponins, rotenoids, carotenoids, xanthophylls, glycosylated triterpenes, theaflavins, thearubigins, phenolics, sesquiterpene lactones, steroids, diterpenes, sesterterpenes, cyanogenic glycosides, and others (Fig. 3, Table 2) [22,23,36,37,44–49], demonstrating its huge applicability and ability to analyze secondary metabolites. Among glycosides analyzed by MALDI-MS from the period 1999 to 2010 [44–49], saponins, steroids, triterpenes, and flavonoids are the glycosylated metabolites predominantly described (Fig. 4).

New studies have shown successful applications for MALDI for metabolic profiling in comparison to other techniques [27,50,51], for example, the analyses of fractions from *Psoralea corylifolia* L. (Fabaceae). They were analyzed by different analytical methods, such as LC-DAD, LC-APCI-MS, and MALDI-TOF MS, with the use of oxidized carbon nanotubes as the matrix. The LC-APCI-MS increased the number of detected substances compared to
the UV detector because some peaks were observed only by APCI-MS and coeluting peaks could be identified. MALDI-TOF MS showed a high ability to detect the substances, even without previous chromatographic separation, and with sensitivity for low mass substances. In this study, a total of 188 components were identified from enriched fractions using all the techniques, 65% of all identified components could be identified by MALDI, and almost 50% of them were identified only by MALDI analyses [50]. This superiority of MALDI for metabolic profiling studies compared with LC-ESI-MS was also verified from analyses of Balanites aegyptiaca (L.) Del. (Zygophyllaceae) extracts and the identification of additional saponins was only observed by MALDI analyses [51]. Thus, the clear superiority for the detection and identification of compounds in a complex sample has been confirmed for it. Other interesting MALDI applications include structural identification, chemical screening, single plant cell analyses, molecular interactions with target molecules, and the determination of the identity of medicinal plants.

Although the technique shows some drawbacks for quantification studies, Wang and collaborators [36] compared the quantification by HPLC-DAD and MALDI-TOF MS of glycosylated anthocyanins (Fig. 5A) and demonstrated the substantial potential of high-throughput MALDI quantification. The MALDI technique proved to be faster at accurately identifying and quantifying the anthocyanins; therefore, MALDI-TOF is a faster alternative to HPLC analyses. In addition, the quantification of glycosylated flavonols (Fig. 5B) by MALDI-TOF MS was also highly correlated with the results obtained by HPLC-UV, which indicates that MALDI-TOF is a valid system for these evaluations [39]. The important tools for quantification by MALDI are the high repetition rate laser, which improves precision and sensitivity, and the use of an internal standard to average out variations in instrument response. Ideally, the internal standards must be chemically similar to the target substances, and the best approach is the use of the same target substance that is isotopically labeled [31,52,53].

### Table 1 Typical commercial MALDI matrices and their applications and characteristics.

<table>
<thead>
<tr>
<th>Matrix</th>
<th>MW (Da)</th>
<th>Excitation wavelength</th>
<th>Applications</th>
<th>Natural product classes analyzed</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-Aminoacridine (9-AA)</td>
<td>194</td>
<td>337</td>
<td>Peptides, sugars, amino acids, sulfated sugars, phospholipids</td>
<td>Flavonoid, naphthydianthrene, phloroglucinol, glucosinolate</td>
</tr>
<tr>
<td>α-Cyano-4-hydroxycinnamic acid (CHCA)</td>
<td>189</td>
<td>337, 355</td>
<td>Peptides, lipids, nucleotides, small proteins, oligosaccharides, sugars</td>
<td>Flavonoid, isoflavonoid, aflatoxin, alkaloid, anthocyanin, tannin, curcuminoid, rotenoid, phenylpropanoid, saponin, spiridole, theaflavin, thearubigin</td>
</tr>
<tr>
<td>2,5-Dihydroxybenzoic acid (DHB)</td>
<td>154</td>
<td>266, 337, 355</td>
<td>Peptides, small proteins, carbohydrates, glycols, glycopeptides, glycoproteins, sugars, lipids, nucleotides, oligonucleotides, oligosaccharides</td>
<td>Acetogenin, alkaloid, anthocyanin, carotenoid, tannin, curcuminoid, flavonoid, limonoid, phospholipid, phenylpropanoid, saponin, spiridole</td>
</tr>
<tr>
<td>Dithranol (DIT)</td>
<td>226</td>
<td>337, 355</td>
<td>Polymers, Lipids</td>
<td>Tannins</td>
</tr>
<tr>
<td>2-(4-Hydroxyphenylazo) benzoic acid (HABA)</td>
<td>242</td>
<td>266, 337</td>
<td>Protein mixtures, negatively charged proteins, glycoproteins, polymers, oligosaccharides</td>
<td>Flavonoid, isoflavonoid</td>
</tr>
<tr>
<td>3-Hydroxy picolinic acid (3-HPA)</td>
<td>139</td>
<td>266</td>
<td>Oligonucleotides</td>
<td>Alkaloid, flavonoid, isoflavonoid</td>
</tr>
<tr>
<td>4-Nitroaniline (4NA)</td>
<td>138</td>
<td>337</td>
<td>Peptides, lipids, oligosaccharides, phosphatidylycholines</td>
<td>Flavonoids</td>
</tr>
<tr>
<td>Sinapinic acid (SA)</td>
<td>224</td>
<td>266, 337, 355</td>
<td>Proteins (MW higher than 10 kDa), glycoproteins, hydrophobic proteins</td>
<td>Alkaloid, tannin, flavonoid, spiridole</td>
</tr>
<tr>
<td>2,4,6-Trihydroxy-acetophenone (THAP)</td>
<td>168</td>
<td>337, 355</td>
<td>Nucleic acids, proteins contaminated with salts, acidic peptides, dendrimers, acidic glycans</td>
<td>Alkaloid, anthocyanin, curcinimod, flavonol, isoflavonoid, glycoalkaloid, tannin, theaflavin, thearubigin</td>
</tr>
</tbody>
</table>

**MW**: molecular weight
Table 2  MALDI MS applications for examination of natural products.

<table>
<thead>
<tr>
<th>Classes of natural products</th>
<th>Source</th>
<th>Matrix</th>
<th>Study goal</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetogenins</td>
<td>Annona muricata</td>
<td>DHB</td>
<td>Quantification</td>
<td>[40]</td>
</tr>
<tr>
<td>Aflatoxins</td>
<td>Peanuts</td>
<td>CHCA</td>
<td>Screening</td>
<td>[63]</td>
</tr>
<tr>
<td>Alkaloids</td>
<td>Urine</td>
<td>CHCA, DHB</td>
<td>Quantification</td>
<td>[35]</td>
</tr>
<tr>
<td>Alkaloids</td>
<td>Corydalis yanhusuo, Coptis chinensis and Aconitum Carmichaeli</td>
<td>CHCA and DHB</td>
<td>Metabolite profiling and identification</td>
<td>[56]</td>
</tr>
<tr>
<td>Alkaloids</td>
<td>Aconitum carmichaeli</td>
<td>SA, DHB and CHCA</td>
<td>Metabolite profiling and quantification</td>
<td>[57]</td>
</tr>
<tr>
<td>Alkaloids</td>
<td>Strychnos nux-vomica</td>
<td>SA, THAP, 3-aminoquinoline (3-AMQ), 3-hydroxypicolinic acid (3-HPA), CHCA and DHB</td>
<td>Metabolite profiling</td>
<td>[58]</td>
</tr>
<tr>
<td>Alkaloids</td>
<td>Sinomenium acutum</td>
<td>CHCA, SA, and DHB</td>
<td>Identification and differentiation of herbal samples</td>
<td>[60]</td>
</tr>
<tr>
<td>Amino acid</td>
<td>Berberis baradana</td>
<td>Matrix-free</td>
<td>Identification</td>
<td>[80]</td>
</tr>
<tr>
<td>Anthocyanins</td>
<td>Solanum melongena</td>
<td>DHB</td>
<td>Spatial analysis*</td>
<td>[109]</td>
</tr>
<tr>
<td>Anthocyanins</td>
<td>Vaccinium corymbosum</td>
<td>THAP</td>
<td>Identification and quantification (comparison between HPLC and MALDI)</td>
<td></td>
</tr>
<tr>
<td>Anthocyanins</td>
<td>Red wine, fruit juice, and syrup</td>
<td>THAP</td>
<td>Identification and quantification</td>
<td>[37]</td>
</tr>
<tr>
<td>Anthocyanins</td>
<td>Arabidopsis thaliana</td>
<td>Caffeic acid (CAF), ferulic acid (FER), DHB, CHCA and THAP</td>
<td>Quantification</td>
<td>[38]</td>
</tr>
<tr>
<td>Anthocyanins</td>
<td>Oryza sativa</td>
<td>DHB</td>
<td>Identification and spatial analysis*</td>
<td>[168]</td>
</tr>
<tr>
<td>Carotenoids</td>
<td>Lycopersicon esculentum, Arabidopsis thaliana and Capsicum species</td>
<td>DHB</td>
<td>Metabolite profiling and identification</td>
<td>[64]</td>
</tr>
<tr>
<td>Carotenoids</td>
<td>Citrus reticulata and Citrus sinensis</td>
<td>DHB</td>
<td>Profiling and identification*</td>
<td>[136]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Pinus pinea and Pinus radiata</td>
<td>DHB</td>
<td>Profiling and identification</td>
<td>[65]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Salix alba, Picea abies, Tilia cordata and Fagus sylvatica</td>
<td>CHCA, SA, DIT, 3-p-indole acetic acid (IAA) and DHB</td>
<td>Metabolite profiling and identification*</td>
<td>[66]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Quebracho wood, mimosa, and cacao</td>
<td>DHB</td>
<td>Fragmentation study*</td>
<td>[128]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Theobroma cacao</td>
<td>DHB</td>
<td>Identification and fragmentation study*</td>
<td>[153]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Cinnamonum zeylanicum</td>
<td>DHB</td>
<td>Identification*</td>
<td>[154]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Eugenia dysenterica</td>
<td>DHB</td>
<td>Identification*</td>
<td>[155]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Anadenanthera colubrina, Commiphora leptophloeos and Myrocarpoum urundueva</td>
<td>DHB</td>
<td>Identification*</td>
<td>[156]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Pityrocampa moniliformis</td>
<td>DHB</td>
<td>Identification*</td>
<td>[157]</td>
</tr>
<tr>
<td>Condensed tannins</td>
<td>Prunus dulcis</td>
<td>DHB</td>
<td>Qualitative profiling and identification*</td>
<td>[158]</td>
</tr>
<tr>
<td>Curcurminoids (diarylethananoid)</td>
<td>Curcuma longa</td>
<td>CHCA, DHB and THAP</td>
<td>Detection and quantification</td>
<td>[42]</td>
</tr>
<tr>
<td>Diverse classes</td>
<td>Psoralea corylifolia</td>
<td>oxidized carbon nanotubes</td>
<td>Identification</td>
<td>[50]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Onion and green tea</td>
<td>THAP, HABA, DHB, and CHCA</td>
<td>Targeted analysis and identification*</td>
<td>[68]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Lyciphopora species</td>
<td>Matrix-free</td>
<td>Identification and spatial analysis*</td>
<td>[74]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Apple (delicious golden)</td>
<td>CHCA and DHB</td>
<td>Spatial analysis</td>
<td>[96]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Purified compounds</td>
<td>CHCA and DHB</td>
<td>Fragmentation studies*</td>
<td>[122]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Isolated standards</td>
<td>DHB</td>
<td>Ion Cluster formation studies*</td>
<td>[124]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Isolated standards</td>
<td>CHCA, DHB, 4N, vanillin (VAN), nicotinic acid (NI), SA, SAM, LiDHB and NALDI</td>
<td>Ionization and in-source dissociation mechanisms*</td>
<td>[22]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Isolated standards</td>
<td>Norhman</td>
<td>Fragmentation and in-source dissociation studies*</td>
<td>[125]</td>
</tr>
<tr>
<td>Flavonoids, fatty acids, and others</td>
<td>Arabidopsis thaliana, Drosophila melanogaster, Acrysthophus pisum</td>
<td>1,8-bis(dimethylamino)naphthalene (DMAN)</td>
<td>Targeted analysis and identification</td>
<td>[72]</td>
</tr>
<tr>
<td>Flavonoids, naphthodianthrones and phloroglucinols</td>
<td>Arabidopsis thaliana and Hypericum species</td>
<td>9- AA</td>
<td>Spatial analysis</td>
<td>[107]</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>Prunus dulcis</td>
<td>THAP</td>
<td>Identification and quantification</td>
<td>[39]</td>
</tr>
<tr>
<td>Glucosinolates</td>
<td>Arabidopsis thaliana</td>
<td>9- AA</td>
<td>Spatial analysis</td>
<td>[106]</td>
</tr>
<tr>
<td>Glycoalkaloids</td>
<td>Solanum tuberosum</td>
<td>THAP</td>
<td>Quantification</td>
<td>[34]</td>
</tr>
<tr>
<td>Hydrolysable tannins</td>
<td>Chinese gall</td>
<td>DHB, THAP, and CHCA</td>
<td>Identification*</td>
<td>[59]</td>
</tr>
<tr>
<td>Hydrolysable tannins</td>
<td>Castanea sativa, Caesalpina spinosa, Quercus infectoria (galls)</td>
<td>DHB</td>
<td>Fragmentation study*</td>
<td>[137]</td>
</tr>
<tr>
<td>Hydrolysable tannins</td>
<td>Rosa chinensis</td>
<td>Matrix-free</td>
<td>Identification*</td>
<td>[159]</td>
</tr>
<tr>
<td>Hydrolysable tannins</td>
<td>Mangifera indica</td>
<td>DHB</td>
<td>Identification*</td>
<td>[160]</td>
</tr>
<tr>
<td>Hydrolysable tannins</td>
<td>Purified compounds</td>
<td>DHB</td>
<td>Identification and evaluation of cationization reagents*</td>
<td>[161]</td>
</tr>
<tr>
<td>Hydrolysable tannins</td>
<td>Astronium urundueva and Astronium graveolens</td>
<td>DHB</td>
<td>Identification*</td>
<td>[162]</td>
</tr>
<tr>
<td>Isoflavonoids</td>
<td>Soy beans</td>
<td>HCCA, THAP, DHB, HABA</td>
<td>Targeted analysis and identification*</td>
<td>[135]</td>
</tr>
<tr>
<td>Limonoids</td>
<td>Azadirachta indica and Melia azedarach</td>
<td>DHB</td>
<td>Detection and identification</td>
<td>[207]</td>
</tr>
<tr>
<td>Lipids</td>
<td>Saccharomyces cerevisiae</td>
<td>DHB</td>
<td>Targeted analysis and identification</td>
<td>[67]</td>
</tr>
</tbody>
</table>

continued next page
In addition, MALDI-MS has also been applied to identify and classify microorganisms based on proteomic fingerprints [54]. This strategy was demonstrated by Ernst and collaborators [27] upon examination of low mass (< 1200 Da) metabolites, but using plant extracts. In this paper, the first protocol for creating a metabolic fingerprint of plants by MALDI-TOF MS was proposed. Three different MALDI matrices and subsequent multivariate data analysis by in-house algorithms implemented in the R environment were employed to taxonomically classify plants from different genera.

### Table 2 MALDI MS applications for examination of natural products.  

<table>
<thead>
<tr>
<th>Classes of natural products</th>
<th>Source</th>
<th>Matrix</th>
<th>Study goal</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipids</td>
<td>Isolated standards</td>
<td>DHB</td>
<td>Fragmentation study*</td>
<td>[138]</td>
</tr>
<tr>
<td>Phenylpropenoic and flavonoids</td>
<td><em>Scutellaria baicalensis</em></td>
<td>CHCA, DHB, graphene and graphene oxide</td>
<td>Metabolite profiling and ionization studies*</td>
<td>[25]</td>
</tr>
<tr>
<td>Phenolics</td>
<td>Different Lichen species</td>
<td>Matrix-free</td>
<td>Dereplication studies</td>
<td>[208]</td>
</tr>
<tr>
<td>Phospholipids</td>
<td>Egg yolk</td>
<td>DHB</td>
<td>Identification</td>
<td>[77]</td>
</tr>
<tr>
<td>Polyphenols</td>
<td>cranberry, grape, sorghum, and pomegranate</td>
<td>IAA</td>
<td>Identification*</td>
<td>[163]</td>
</tr>
<tr>
<td>Quaternary alkaloids</td>
<td><em>Corydalis yanhusuo</em> (rhizoma)</td>
<td>DHB</td>
<td>Identification and quantification*</td>
<td>[33]</td>
</tr>
<tr>
<td>Rotenoids</td>
<td><em>Boraginaceae</em></td>
<td>CHCA</td>
<td>Quantification*</td>
<td>[43]</td>
</tr>
<tr>
<td>Saponins</td>
<td><em>Echinacea angustifolia</em></td>
<td>DHB</td>
<td>Metabolite profiling</td>
<td>[51]</td>
</tr>
<tr>
<td>Saponins</td>
<td>Panax ginseng and Panax quinquefolius</td>
<td>CHCA, SA and DHB</td>
<td>Identification and differentiation of the species*</td>
<td>[55]</td>
</tr>
<tr>
<td>Saponins</td>
<td>Bacopa monnieri</td>
<td>CHCA</td>
<td>Comparison of methods and identification</td>
<td>[134]</td>
</tr>
<tr>
<td>Saponins</td>
<td><em>Quillaja saponaria</em></td>
<td>DHB</td>
<td>Targeted analysis*</td>
<td>[146]</td>
</tr>
<tr>
<td>Saponins</td>
<td>Holothuria forskali</td>
<td>CHCA</td>
<td>Identification and spatial analysis*</td>
<td>[147]</td>
</tr>
<tr>
<td>Saponins</td>
<td>Holothuria lessoni</td>
<td>CHCA</td>
<td>Structural elucidation*</td>
<td>[150, 151]</td>
</tr>
<tr>
<td>Saponins and triterpenes</td>
<td><em>Centella asiatica</em></td>
<td>CHCA</td>
<td>Identification</td>
<td>[78]</td>
</tr>
<tr>
<td>Spiroloides</td>
<td>Phytoplankton</td>
<td>CHCA, DHB, and SA</td>
<td>Identification and quantification*</td>
<td>[41]</td>
</tr>
<tr>
<td>Steroidal lactones and alkaloid</td>
<td><em>Withania somnifera</em> and <em>Nicotiana tabacum</em></td>
<td>Matrix free</td>
<td>Screening*</td>
<td>[62]</td>
</tr>
<tr>
<td>Steroids and lipopeptides</td>
<td>Standards</td>
<td>Coumarins</td>
<td>Evaluation of matrix efficiency</td>
<td>[209]</td>
</tr>
<tr>
<td>Theaflavin, thearubigins</td>
<td>Yunnan black tea</td>
<td>THAP, CHCA</td>
<td>Identification</td>
<td>[169]</td>
</tr>
<tr>
<td>Theaflavin and flavan-3-ol derivatives</td>
<td>Black tea leaves</td>
<td>DHAP</td>
<td>Structural elucidation</td>
<td>[170]</td>
</tr>
<tr>
<td>Undermined</td>
<td><em>Echinacea species</em></td>
<td>CHCA, SA</td>
<td>Metabolite profiles, differentiation of species</td>
<td>[61]</td>
</tr>
</tbody>
</table>

* Based on MS/MS data

**Fig. 2** Some applications of MALDI in natural products. (Color figure available online only.)

**Fig. 3** Some examples of natural products analyzed by MALDI: cyanogenic glycosides (1, 2), triterpenoid saponins (3, 4), aflatoxins (5, 6), carotenoids (7), flavones (12, 13), typical linear condensed tannins possessing B-type linkage (8, 9), rotenoids (14, 15), sesterterpenes (16, 17), and theaflavins (18, 19).
families, and orders. Initially, analyses without or with only one matrix were performed, but the results did not provide sufficient chemical information to correctly classify the plants. Then, several matrices were evaluated to select the best matrices that yield a higher number of ionized compounds and have less dissociation in the source in both ion modes. 4NA and CHCA matrices were selected for negative and positive ion modes, respectively. However, the nonpolar compounds, such as some triterpenes and diterpenes, were not ionized using these common matrices, and they could only be ionized with the LiDBH matrix, a synthesized matrix. Thus, the chemical information from plant extracts was enlarged with the nonpolar compounds that, using careful algorithms and parameter selections, allowed a close taxonomic classification with 92% similarity to the taxonomic classifications found in the literature [27].

Recently, another relevant MALDI-MS application was reported to identify different plants. It was also successfully used to differentiate the herbs Panax ginseng C.A. Meyer and Panax quinquefolium L. (Araliaceae) because they have similar chemical and physical properties, including a problematic botanic differentiation, but with substantially different therapeutic effects that highlight the importance of a correct identification. The methodology allowed the unambiguous differentiation between the two species, required a small quantity of material, and was fast, robust, and simple [55]. Consequently, MALDI-MS can potentially characterize adulterants within the plants and perform rapid dereplication and quantification studies. The characterization of medicinal plants by MALDI-MS has been described for other plants, such as the species Aconitum carmichaeli Debx. (Ranunculaceae), Corydalis yanhusuo W.T.Wang (Papaveraceae), Echinacea species (Asteraceae), and others [56–61]. MALDI can also be applied directly to the analysis of the powdered plant material [62], representing an important tool for quality control, since secondary metabolites have been successfully screened by MALDI-MS, such as aflatoxins [63], saponins [51], anthocyanins [36,37], carotenoids [64], tannins [59,65,66], lipids and phospholipids [67], flavonoids [68], and others (Fig. 4). In general, metabolomics studies can be obtained by MALDI-MS directly from extracts, tissue, or single cells [6,69–72], which highlights the potential to define the tissue distribution of the metabolites by MALDI imaging [73–75]. This application is extremely useful in metabolomics, representing a promising future in studies such as metabolic compartmentalization [6,69,70] (see MALDI imaging section below).

Another promising methodology is TLC combined directly with MALDI-MS to identify compounds. The TLC is an easy and fast technique to separate compound mixtures and it is widely used in natural products laboratories, facilitating its implementation for this purpose [76]. TLC-MALDI MS has been applied to analyze phospholipids from chicken eggs [77], centellosides from Centella asiatica (L.) Urb. (Apiaceae) [78], siderophores from microbial samples [79], and alkaloids from Berberis barandana S. Vid. (Berberidaceae) [80]. Despite the advantages of TLC-MALDI MS, such as low cost for chromatographic separation, low time consumed, and direct analysis (without the extraction from TLC), its use has been restricted to natural products chemistry due to the yet low expansion of the technique [76]. The TLC-LDI-MS (analysis without the matrix) was applied to the analysis of quaternary protoberberine alkaloids from B. barandana (Fig. 6A). The compounds suffered in-source dissociations [80], which were likely intensified due to the matrix absence; therefore, more studies are required to understand the matrix role for secondary metabolites analyses and their chemical reactions in the source, such as the evaluations performed with aromatic carboxylic acids and flavonoids [22,81,82]. The in-source dissociations of twenty-six flavonoids (glycosylated and non-glycosylated), including flavones, flavonols and others (Fig. 6B–D), were evaluated without a matrix and with different matrices. The in-source dissociations, the influence of laser intensity, and the applied matrix type were investigated. The flavonoid O-glycosides eliminated the sugar in-source, even in the presence of the...
matrix and in both ion modes, producing radical ions ([M – H – sugar]+) in the negative ion mode, similar to ESI [83]. The MALDI conventional matrix application reduced the in-source fragmentation of C- and O-glycosylated flavonoids, but they were not eliminated. In addition, the methyl radical losses from methoxylated flavonoids were completely eliminated only with the LiDHB matrix; in addition, the methyl radical losses from methoxylated flavonoids were completely eliminated only with the LiDHB matrix; besides this, all the retro Diels–Alder fragmentations in the source were also eliminated [22]. Thus, the use of a specific matrix may be required, depending on the drawn objectives and the type of molecular target.

Moreover, the screening and selection of bioactive compounds from extracts can be performed by MALDI-MS, including the search for specific enzymatic inhibitors and the establishment of the enzyme reaction kinetics constants [84, 85]. A quantitative method by MALDI-FT MS was developed and used to analyze the product of the acetylcholinesterase enzymatic reaction and the potential inhibitors from Rhizoma Coptidis extracts. The assay interferents (acetylcholinesterase and cholinesterase) were also eliminated [22]. Thus, the use of a specific matrix may be required, depending on the drawn objectives and the type of molecular target.

Tination of the ion source represent the main disadvantages. However, the oxidation and chemical functionalization of carbon nanotubes has improved some of these problems, facilitating the sample preparation, since the water solubility is higher, and improving the efficiency of desorption/ionization analytes and reproducibility [87–89]. Finally, the adduction profiles of quinone-thioether metabolites by cytochrome C have also been established by MALDI-TOF [90], as well as the identification of protein interactions, such as for hydrolysable and condensed tannins with bovine serum albumin [91, 92]. Thus, the biological fingerprinting analyses by MALDI can be used for screening targets and assisting with studies of absorption, distribution, metabolism, elimination, and toxicity [85, 86, 93].

MALDI imaging
MALDI imaging, another important technique in natural products chemistry, was introduced by Caprioli in 1997 and was initially applied to proteins and peptides [94]. The technique directly analyzes tissue using MS, producing maps of the ion tissue distribution. The highlighted advantage is not a necessity of the previous extraction of the compounds for analyses, avoiding the losses of information about spatial distribution [95]. This method was first used to analyze animal tissue, but currently includes other tissue analyses, such as for plants, microorganisms, and insects [9, 95–98]. There are some protocols of sample preparation from animal tissue for protein and peptide analyses; however, no universal methods of sample preparation and data processing for others tissues have been reported, such as for plants [9, 94]. In addition, plant tissue analyses by MALDI imaging are far from being
a routine technique, since there are many difficulties in preparing the samples and the requirements to adapt protocols for specific samples.

IMS has been successfully applied to diverse studies due to its high sensitivity (compounds of very low concentration can be analyzed; attomole to low femtomole range), selectivity (similar compounds can be differentiated), and ability to identify structurally the metabolites, since mass spectrometric data give chemical information useful to identify them [9, 69, 73, 75, 99]. The high mass accuracy obtained by high-resolution analyzers can be used to differentiate similar compounds that show differences only in their exact masses, as well as a reduction of the interference of matrix ions in the images. In addition, the images of tissue distributions of metabolites produced from MS/MS data have a higher selectivity, since the isomers (compounds with the same molecular formula) with unlike fragmentation pathways can be distinguished, producing images with high reliability [9, 73, 100]. Although IMS has been performed by other ionization methods, such as desorption electrospray ionization (DESI), laser ablation electrospray ionization (LAESI), and secondary ion mass spectrometry (SIMS), most studies applied MALDI and LDI techniques that are explainable because of their advantages, such as good spatial resolutions (around 20 µm) and a high speed of data acquisition (lasers with high frequency) [9, 75]. The huge challenge of MALDI analysis is still the background matrix ions in the low mass range, the size crystals of the matrix, since it is directly related to spatial resolution, as well as the scarce information about the ionization and in-source dissociation of secondary metabolites to improve the data and result qualities.

The tissue distribution of specific metabolites can be established by MALDI imaging, unlike classical histochemical methods (un-specific) applied for plant tissue, making it possible to distinguish between individual compounds, which can be confirmed by MS/MS data, and it has a high accuracy, which reduces the problems related to isobaric matrix ions and creates more reliable image data. However, most studies did not use MS/MS data and they only operate in full scan mode [9]. Some articles, including relevant reviews, have shown the application of MALDI imaging for plant tissues, food, and microorganisms (for example, to elucidate the interaction between them). They described diverse details about sample preparation, spectra acquisition, post-acquisition analysis, and data evaluation related to biosynthesis, spatial dynamic, ecology, physiology, and morphology [9, 73, 75, 101–105]. So, new issues are reported here to extend the information not yet addressed in other review articles.

MALDI and LDI imaging have been used for analyses of fruits, leaves, stems, sepals, seeds, roots, tubers, flowers, pollen grains, and rhizomes [9, 73, 75], such as from the leaves of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae – glucosinolates) [106], *Lycnophora* species (Asteraceae) [74], *Hypericum perforatum* L. and *Hypericum reflexum* L.f. (Hypericaceae) [107], apples [108], eggplant [109], and others. The details of MALDI imaging of plant tissue are not described here because excellent review articles have already been published in this area [9, 73, 75, 101, 102, 110, 111].

Recently, Bjarnholt and collaborators performed an extensive revision (up to 2013) about IMS from plant tissue, and almost 65% applied MALDI and LDI for image acquisition using mainly CHCA and DHB as the matrix and cryosectioning to slice the materials [9]. The published articles, applying MALDI imaging from plants, between 2014 and 2015 are summarized in Table 3 and some special review articles have been published from analytical strategies for data obtainment, including higher spatial resolution and sample preparation [112–116]. In addition, a recent published review summarized the advances of IMS for lipidomics, showing the insights related to the spatial compartmentalization of lipids and their metabolism in plants, and the matrices DHB and 9-AA are predominantly used for these analyses [117].

A critical step in MALDI imaging experiments is the sample preparation, which requires some special care to avoid the degradation or metabolism of the compounds. In addition, the vacuum in the source (not for atmospheric pressure MALDI) makes the analyses of transversal plant sections (*in natura*) difficult because of water losses in vacuum due to the tissue contraction, which makes the correlation between the MALDI imaging and anatomical data difficult, as well as the flatness of the tissue that may interfere in the accuracy. Another important point is about the spatial resolution of images, since the higher resolution (few microns) only has been presented by prototype machines produced in specialized research laboratories. The images at resolutions of 10–20 µm are possible for commercial equipment, but the focus laser beam should be considered an experimental parameter and the size of the matrix crystals are also relevant [9, 73]. Nowadays, the use of MALDI imaging from plants is still limited, but it can give valuable information about surface metabolite and tissue distribution maps, which can assist in understanding the pathway biosynthetic, metabolite translocations, defense of plants, and others [9, 73, 118]. Although there are huge challenges for small compound analyses and plant tissues, MALDI imaging is a technique with many advantages and gives new perspectives in the natural products area, helping to explore and understand diverse issues, such as ecological and physiological.

MALDI-MS/MS to identify natural products

MALDI has been coupled to different analyzers such as quadrupole (Q), IT, orbitrap, TOF, and FT-ICR. The mass accuracy, resolution, m/z range, sensitivity, speed, and other characteristics are different for each analyzer and should be selected to best fit the specific requirements of each experiment [6, 16, 102, 119, 120]. MS/MS is performed in two stages: the first includes ion precursor isolation and its activation, which subsequently leads to the fragment ions being separated and detected. The activation step involves increasing the internal energies of ions, resulting in the rupture of chemical bonds by homolytic and/or heterolytic fissions. Different ion activation methods are available such as low- and high-energy CID, electron capture dissociation (ECD), blackbody infrared dissociation (BIRD), surface-induced dissociation (SID), ultraviolet photodissociation (UVPD), electron transfer dissociation (ETD), electron-induced dissociation (EID), infrared multiphoton dissociation (IRMPD), and PSD [121]; however, not all of the methods are available for the MALDI system, and CID is most often applied.

Although MALDI-MS/MS is widely used to identify and characterize peptides, there is little information on its use for small compound natural products, especially on the in-source fragmentation and the gas reactions for ion activation methods with a high energy transfer. In addition, the influence of the amount of energy transferred in the ionization processes by MALDI and the subsequent fragmentation is poorly understood. Therefore, this review explored the use of MALDI-MS/MS for natural products identification and the principal considerations reported, which stimulates future studies to expand its applications.

The fragmentation of nine flavonoids (two aglycone flavones, one isoflavone, three aglycone flavonols, two O-glycosylated flavonols, and one flavanone) were compared by ESI-QTOF, MALDI-
cursor ions produced by MALDI are reaccelerated to 20000 eV ($E_{cm} = 150–310$ eV). Thus, MALDI TOF ReTOF works at a higher energy than ESI-QTOF (accelerated to 30 eV, $E_{cm} = 1–4$ eV) and MALDI-QIT ($E_{cm} = 52$ eV) when compared to a specific ion [122]. The rare loss of a hydrogen radical from typical flavonoid fragments (for example, from ions produced by retro-Diels-Alder cleavages) was only observed in MALDI TOF ReTOF at high center-of-mass energies and included other radical ions, and no competing fragmentation processes at low center-of-mass energies have been observed [121, 122]. The high center-of-mass energies have been described to induce charge-remote fragmentation and diverse fragmentation pathways, which may be useful in dereplication studies because they offer additional information for structural elucidation.

Through previous fragmentation studies of standards, compounds such as ferulic acid, wogonin (a flavone), and scutellaran (a glycoside flavone) could be identified in extracts of traditional Chinese medicine herbs by MALDI-MS/MS with graphene and graphene oxide matrices, which improved the limit of detection and reduced the in-source fragmentation. The graphene belongs to the multidimensional carbon nanomaterial family, which is composed of two dimensional layers of sp² bonded carbon and recently it has been applied as a MALDI matrix for small compound analyses, since a low background interference of matrix ions is also observed in the spectra. In addition, the graphene oxide is easily produced by oxidizing graphite, showing hydroxyl and epoxide groups on the base of the structural plane of graphene oxide sheets that confer strong hydrophilic properties and help water dispersion and swelling. The excellent results could be related to their properties such as thermal, electronic, and mechanical [25, 123]. Specific matrices can reduce the in-source dissociation and significantly influence the final results. However, there is little information about ionization processes, in-source dissociations, and fragmentation by MALDI with high-energy CID, PSD, and LIFT. Fragmentation studies using high-energy CID, PSD, and LIFT are in demand because, as reported by March and coworkers [122], only fragmentation data obtained by low-energy CID are similar to ESI and MALDI for flavonoids, and such data are widely available for ESI.

Silva and Lopes [22] evaluated the in-source dissociations of various glycosylated and non-glycosylated flavonoids (flavanones, flavones, flavonols) without a matrix and with different matrices in addition to the influence of the laser intensity on these in-source reactions. The flavonoid-matrix cluster ions, which depend on the structure of the flavonoid, were elucidated by MALDI-MS/MS, confirming the formation of cluster ions involving fragments produced by retro-Diels-Alder fragmentation [124]. The formation of radical fragments in-source and MS/MS experiments are evident [22, 81, 122, 125], and increasing the understanding of these reactions will enlarge the application of MALDI as a tool in high-throughput chemical analysis and the identification of natural products [25, 27, 51, 55, 68, 126].

Although the O-glycosylated flavonoids easily lose the sugar in the source, Wang and Sporros [68] thoroughly studied the fragmentation pathway of glycosylated flavonoids with up to three glycosides. The best matrix reported by the authors was THAP, because it had good repeatability (spot-to-spot), produced ions in both positive and negative ion modes, and exhibited a high affinity for alkali metals in the ionization process. The main fragment ions were generated from sugar losses, as observed in ESI [127]. However, the fragment ions produced from the aglycone were not described [68], which is important for its elucidation.

PSD for TOF analyzers and the LIFT system have been employed in MS/MS experiments, and both can be associated with CID to increase the internal energy of the ions, thus increasing fragment ions [128, 129]. Metastable fragments are produced from ions with excess internal energy during ion acceleration from the source in the free field region, and the reflector voltage is set to detect these fragments in the PSD technique. However, PSD is time consuming, and difficulties exist in mass calibration and the detection of fragments with a low $m/z (< 150)$ [129]. In LIFT experiments, a low voltage of 8 kV is applied to a precursor ion for acceleration: the ion is then isolated and raised to a higher potential in the LIFT cell. Subsequently, the fragment ions are reaccelerated toward the detector, and changes in the reflector voltage are not required [129]. PSD and LIFT were used to analyze clusters of the CHCA matrix, demonstrating the many advantages of the LIFT technique, with a better detection of ions with a low abundance and mass, and a relatively fast acquisition time [130]. Initially, PSD was applied in peptides [131, 132] and carbohydrate sequencing [133], and currently, there are some applications in natural products studies. It was used to identify saponins [134], flavonoids [122], isoflavones [135], carotenoids [136], and hydrolysable and condensed tannins [66, 137] as well as to examine triacylglycerols and phosphatidylethanolamines from plants, algae, and animal tissue [138-140].

Different carotenoids (polyenes without oxygen) and xanthophylls (polyenes with oxygen) were analyzed by MALDI and produced radical ions (molecular ions) through the removal ($M^+$) of one electron (Fig. 7) [136], which is similar to ESI and FAB. The increase of a conjugation extension is a relevant factor for electron loss and radical stability, and produces spectra complexes from sequential in-source homolytic cleavages, as observed in the MALDI source [4, 23, 141-143]. The fragmentation patterns of carotenoids and xanthophylls were proposed based on MALDI-PDS data [23, 136]. The loss of one molecule of water ($M-18$) confirms the presence of a hydroxyl group (such as for luteolin), and [M-92]$^+$ and [M-106]$^+$ ions are diagnostic of tolune and xylene losses, respectively. The intensity relation between these fragment ions ($I_{M-92}/I_{M-106}$) is important to define the presence of double bonds, and when this relation is $> 10$, the carotenoids contain nine double bonds. The loss of fatty acids ($[M-RCOOH]^+$) is common for carotenoid fatty acid esters (such as for $\beta$-cryptoxanthin palmitate) and characterizes their presence in the carotenoid structures, while the epoxy species can be confirmed by the fragment ion ($M-80$). Other diagnostic carotenoid fragment ions have been described in detail and are useful for elucidating their chemical structures [23, 136].

The O-glycosylated isoflavones from soy [Glycine max L. (Fabaceae)] showed fragments with structural importance by PSD, including the main fragment that occurs through the loss of sugars, which is similar to ESI. In addition, the DHB and THAP matrices were evaluated, and DHB showed better ionization results and a higher number of fragments in MS/MS spectra [135]. PSD was also applied in the study of peracylated isoflavone glycosides from their protonated and cationized different metal ions to understand the gas phase molecule cluster complexes. A reduc-
tion in the number of fragments was detected in the order of \( \text{Li}^+ > \text{Na}^+ > \text{Ag}^+ > \text{Cu}^+ > \text{H}^+ > \text{K}^+ > \text{Rb}^+ = \text{Cs}^+ \) [144]. Fragmentation was lower for \( \text{Ag}_3^+ \) clusters than \( \text{Ag}^+ \), demonstrating a stronger gas phase interaction with \( \text{Ag}_3^+ \) [145].

Saponins have been studied using different fragmentation techniques and, in some cases, the fragmentation was applied in dereplication studies. MALDI has limited applications in this area and is sometimes used together with LC-ESI-MS/MS data because more saponins can be identified and detected by MALDI [51, 146, 147]. ESI-IT, MALDI-IT, MALDI-IT/TOF, and MALDI-TOF/TOF, including high-energy CID, were applied for dammarane-type triterpenoid saponins from Bacopa monnieri (L.) Wettst. (Plantaginaceae) [134]. In low energy techniques, the spectra were similar for ESI and MALDI ionization, and their product ions are applied to elucidate the sequence and branching of the sugar moieties [134, 147]. The useful fragment ions from the aglycones and the glycosidic moiety were obtained from high-energy CID, which yielded the same fragments observed by PSD [134] and showed a similar charge-remote fragmentation described for high-energy CID for conjugated steroids and others [148, 149]. These diagnostic fragment ions were also observed by the LIFT system, which assisted in the structural identification of saponins from the sea cucumber Holothuria forskali [150, 151].

Hydrolysable and condensed tannins are another group of secondary metabolites analyzed by MALDI due to its advantages, such as sensitivity, the formation of singly charged molecular ions, and lower ion suppression for complex mixture analysis [66, 128, 137, 152]. Several articles have reported the potential to identify such compounds by MALDI-MS, which contains different combinations and connectivity between the monomers and produces tannins of the same molecular weight. However, unambiguous identification can only be completed with MS/MS data, establishing the linkage between the specific monomers. Monomer losses involving fragment ions by hydrogen transfer and typical retro-Diels-Alder fragmentations are observed from each oligomer, followed by water elimination, but the radical fragment ions are less visible in the spectra [153–157]. These observations were applied to identify condensed tannins of A-type and B-type linkages from different species described from Theobroma cacao L. (Sterculiaceae) [153], Salix alba L. (Salicaceae), Picea abies (L.) H. Karst (Pinaceae), Fagus sylvatica L. (Fagaceae), Tilia cordata Mill. (Malvaceae) [66], Cinnamomum zeylanicum L. (Lauraceae) [154], Eugenia dysenterica DC. (Myrtaceae) [155], quebracho wood [128], Commiphora leptophloeos (Mart.), J.B. Gillett (Burseraceae), Anadenanthera colubrina (Vell.) Bremann var. colubrina (Fabaceae), Myracrodruon urundeuva Allemão (Anacardiaceae) [156], Pityrocarpa moniliformis (Benth.) Luckow & R.W. Jobson (Fabaceae) [157], almond [158], and others. In addition, it allowed for the identification of a rare polymeric series of up to eight flavan-3-ol units with pentose and hexose sugars by MALDI-TOF/TOF LIFT [153].

MS/MS data obtained by MALDI QIT-TOF, PSD, and CID have been applied to determine the chemical structure of hydrolysable tannins from Rosa chinensis Jacq. (Rosaceae) [159], Chinese galls [59], tara, Turkey gall, chestnut woods [137], Mangifera indica L. (Anacardiaceae) [160], and others, which are preferentially acquired in the positive ion mode due to the visualization of fewer signal peaks in the negative ion mode for gallotannins [59]. However, an extensive application of different matrices must be evaluated to confirm this statement. The hydrolysable tannins were also evaluated by MALDI-TOF/TOF (positive ion mode) with two cationizing agents, sodium and cesium, but significant differences in the fragmentation pathway were not observed [160, 161]. First, the galloyltannins lost one or more galloyl moieties (152 Da), accompanied by the loss of one or more molecules of H\(_2\)O. The molecules of gallic acid (170 Da) or ellagic acid, when present, can be lost with the hydrogen transfer, and a double bond in the glucose is formed [137]. Many important fragment ions were not observed by ESI-MS/MS, which could complicate their structural identification [162–165]. A complex fragmentation pattern obtained from < 300 Da is observed, which complements the information to identify them [59, 137, 160]. The fragment ions from sugar core tannins can be observed, occurring by the internal cleavage of the glycosidic ring, with a similar fragmentation already reported for carbohydrates by MALDI-CID [166, 167].

MALDI-MS is an advantageous technique for macromolecule analyses due to the characteristics described previously, and currently, its application to small molecules is just beginning [24, 26, 28]. Although more research in this area is required, MALDI-MS has demonstrated valuable contributions to secondary metabolites over other analytical techniques [50]. Many publications described the use of MALDI-MS for establishing metabolic fingerprints [6, 27, 168, 169]. A chemical profile of the compounds and the use of MALDI-MS/MS to identify the chemical structure of natural products are recent and not extensively applied in the identification of tannins (hydrolysable and condensed) and saponins. In addition, few studies by MALDI-MS/MS have been described for a restricted number of natural compounds, such as anthocyanins, carotenoids, xanthophylls, betaines, glycosylated isoflavones, phenolic acids, flavonoids, theaflavins and thearubigins, limonoids, alkaloids, and others, which are sometimes combined with LC-ESI-MS/MS data [23, 25, 68, 74, 122, 125, 126, 136, 143, 153, 159, 168–171].

All the statements appointed here confirm the relevant MALDI applications in the natural products area, stimulating its applications in wide studies such as the research of biomarkers, hierarchical clustering, taxonomy, imaging from tissues, and others. However, the data processing is fundamental to improve the quality of results and to avoid mistaken conclusions. So, the data analysis can be divided into the steps of preprocessing and statistical analysis. The preprocessing will be described here, since it is extremely important and restrictedly described in the literature.

**Data Processing**

In the previous section, we have shown a broad range of MALDI ionization applications to natural products. Due to the mass spectrometer hardware and the advances in applications, many authors consider that data analysis is going to play an important role in most of these applications and will become a bottleneck for many inquiry fields [172, 173].

In many cases, specific techniques are required to analyze a sample and to answer a scientific question, for example, by MALDI imaging or a MALDI-MS/MS data set. Each spectrum in an experimental data set, usually composed of many samples and replicates, should be processed to improve the results and thus guide the study to correct conclusions [174]. For this reason, a common modular analysis flow diagram was proposed here, and some applications to MALDI-MS data were described to exemplify the addressed points. Following this flux analysis, here the similarities with MS data obtained by other ionization methods and more generally for other large-scale data sampling techniques are highlighted. The optimization of data analysis steps described be-
low is much more critical to correctly associate patterns in data to the studied biological changes.

In the following, each step presented in Fig. 8 will be briefly explored, highlighting its importance to an unbiased analysis, reviewing applications for MALDI-MS data and software availability. Some steps, such as smoothing and baseline removal, may switch their locations in the analysis flux [175]. In Fig. 8, the position of visualization is presented as a parallel procedure to all analysis steps, highlighting the importance to plot spectra summaries before and after preprocessing. Morris and collaborators [175] showed how simple heat maps allowed the discovery of unexpected data patterns.

The intensity of a given m/z value is proportional to the relative abundance of an ion represented by this m/z value [174]. Given a set of S spectra containing information from a total of n small molecules, the goal of our analysis flux is to generate an S × m matrix, whose rows correspond to the individual spectrum, and the columns contain some (relative) quantification of a molecule (or mass signal associated with a molecule) [175]. This matrix, which represents a multidimensional data set, is generated through the preprocessing described in Fig. 7, and the final goal is to associate changes or patterns in this data set with our studied problem. According to Ge and Wong [176], a mass spectrum can theoretically be decomposed into three components: the baseline value, the true signal, and the noise. Therefore, baseline and noise can be potentially associated with inter- and intra-sample variability, and data preprocessing is required to reduce the baseline and noise in the raw data before any multivariate analysis can occur [177, 178].

Morris and collaborators [175] showed illustrative examples that ignore the basic assumption of experimental planning, namely, blocking and randomization, and how it can completely compromise an experiment. The authors also showed how a change in the biological sample collection protocol after the first 20 subjects had been collected resulted in changes between the subtypes of cancer studied that could not be attributed to the biological question, instead it could be easily attributed to systematic and reproducible changes due to the experimental procedure. Oberg and Vitek [179] discuss how randomization, replication, and blocking helped to avoid systematic biases due to the experimental procedure and assisted in optimizing its ability to detect true quantitative changes between groups in mass spectrometry-based proteomic experiments. The importance of experimental design in mass spectrometry has been reviewed [180, 181]. The experimental design is also commonly used to test the combination of two or more factors of interest. Zhang et al. [182] obtained optimum automated data acquisition settings that yielded the highest reproducibility of replicate mass spectra based on experimental design statistical analysis of intact cell MALDI-MS spectra of Pseudomonas aeruginosa.

According to Xiang and Prado [183], the calibration methods are usually divided into two classes for high-resolution analyzers (mostly TOF analyzers): external and internal. The authors recorded a standard sample in a separate spectrum, in the external calibration process, and the calibration parameters were drawn from this standard sample and used to calibrate samples in other spectra. In the internal calibration process, standard molecules were mixed with the sample, a spectrum of the mixture acquired, and standard molecules peaks were identified and used to calibrate the entire spectrum. Fraser and coworkers [64] have used a single point external calibration and achieved mass accuracies of 10–70 ppm in MALDI-TOF analyses. When the same authors used an internal calibration, using two carotenoid calibrants flanking the m/z of interest, they achieved mass accuracies below 10 ppm. Saideman and coworkers [184], using internal calibration, have determined peptide mass accuracies ranging from 0.124 to 5.17 ppm (n = 3).

The baseline is associated with a series of matrix materials, impurities, ionization by-products, electronic signal noise, and sample preparation contamination [177, 185, 186]. Baseline correction methods were reviewed by Hilario and collaborators [187].
The MALDIquant package [188] implements different approaches to adjust the baseline. The package’s default is the Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP) algorithm proposed by Ryan and collaborators [189] that yields a smooth baseline and leads to positive corrected intensities. Systematic shifts can appear in repeated experiments, and the spectra for two identical metabolites can have different m/z values. The spectral alignment consists of aligning corresponding peaks across samples to address this problem [190]. For a given m/z value selected in different spectra, the peak position can differ slightly, and generally, this effect cannot be corrected by the instrument calibration [174].

The spectral alignment is a theme of constant research. A simple approach for spectral alignment is presented by Pham and Jimenez [172], in which peaks in individual spectra are matched against the closest peak in a mean spectrum. He and coworkers [191] compared several algorithms currently in use for mass spectrometry and showed that their correlations-based algorithm performed better in a SELDI-MS dataset. The alignment method described by He and coworkers [191] is also implemented in the MALDIquant package. Bloemberg and coworkers [192] wrote a comprehensive tutorial explaining the working principles of several methods and provided implementations in R [193] and Matlab (The Mathworks Inc.) languages.

Many spectral smoothing methods generate an intensity value by averaging the values within a span of data points. Smoothing is directly proportional to the span size. Therefore, care must be taken when choosing the span size, as a large span size may lead to information loss. Smoothing is important to correctly estimate the noise and to improve the peak picking process [187]. Several algorithmic procedures for informative peak identification have been proposed, named as peak detection or peak picking, that correspond to true signals. A comparative review of some of these algorithms, their principles, implementations, and performance is presented by Yang and collaborators [194]. The author concluded that continuous wavelet transform (CWT)-based algorithms provided the best performance. The peak picking process is important in extracting the relevant signals from a mass spectrum and for its significant reduction of the data set dimensionality. Different implementations are available in the R package MALDIquant, as well as on XCMS [195], which can be straightforwardly adapted for MALDI-MS data.

In repeated experiments, it is common to observe differences arising from sample preparation and instrumental measure variation, which cannot be attributed to differences in the biological samples. Normalization is the transformation of sample-wise mass intensities to the same scale, which enables the comparison of different samples [196]. The impact of normalization was addressed by Ejigu and collaborators [197] in an LC-MS study. The authors state that normalization should be regarded more as a remedial measure, not to correct for all sources of variability introduced by different sources of biases. The authors also recommend data-driven (assuming that a large amount of the metabolites stay constant) normalization methods over model-driven (based on internal standards or intermediate quality control runs) normalization methods. Deininger and coworkers [198] have shown the importance of normalization in MALDI imaging experiments, where the incorrect normalization could completely change the potential biomarker spatial distribution. According to the same authors, median and noise level normalizations were significantly more robust than TIC normalization.

The success of multivariate data analysis depends heavily on the correct execution of the previous analysis steps. For example, incorrect peak picking can leave out peaks from important molecules, incorrect alignment can merge peaks from different molecules, and incorrect normalization can show intensity differences caused by the experimental procedure in place of biological differences. These mistakes impact the conclusions drawn from multivariate methods.

The multivariate analyses are commonly classified into two groups: unsupervised and supervised methods. Unsupervised methods are used on exploratory data analysis and make use of the data latent structure, without any previous assumption of sample classes. Supervised methods use information from the data structure and sample classes to estimate the model parameters [199].

From a large dataset, it is possible to predict the desired outcome only by chance. There are good confidence measures available to check the consistency of this analysis. For the most common unsupervised analysis, the hierarchical clustering, resampling, and regrouping techniques provide a consistency check [200]. For classification models (supervised analysis), care must be taken. A very popular model in the natural products community is PLS (partial least squares), which is prone to overfitting [201]. To avoid overfitting, an important technique in the classification models is the K-fold cross validation (CV) [175]. Many hybrid exploration/classification approaches have been proposed on MALDI. Alexandrov and coworkers [202] developed a spatially aware clustering approach for MALDI imaging. Briefly, the spectra are grouped by a measure of similarity taking into account its position, and then all pixels are pseudo-color coded according to the classification assignment. Pham and Jimenez [172] employed a grid search with exponential spacing to find the optimal values for support vector machine model selection. The authors report a final accuracy of 83% on a separate validation set (Breast Cancer Study with MALDI-MS Data) of 78 samples.

Tong and coworkers [177] argue that there is a lack of guidelines for data processing, and inexperienced practitioners can add variation when handling complex models. The authors also state that the functionalities of many tools designed for SELDI-MS or LC-ESI-MS are not transferable to MALDI-MS data analysis. As there is constant development of new computational statistical methods, many scientific communities are fostering the development of modular software in such a way that the modules can be combined to achieve different goals [203]. There is also an increasing realization of data standards, data storage, and exchange [204,205]. These good experimental practices will ultimately allow improvement in experimental reproducibility among different research groups and promote faster improvements in the groups that successfully adopt these practices.

He and coworkers [185] have a similar opinion regarding the difficulties of experimentalists in using software toolkits. The authors recommend the use of flexible software platforms that enable new users to construct new fit-to-purpose workflows. The authors cite case-based reasoning (CBR), a process of solving new problems based on the solutions of similar past problems [206].

There is also an agreement that most of the complex experiments should be performed by multidisciplinary groups and that data analysis in particular could benefit from the participation of statisticians, computer scientists, or more generally quantitative scientists. Their participation would be effective if they could be
present in all phases from study design to interpretation of information extracted from experimental data [175, 177, 185].

Conclusions

MALDI-MS is an underexplored technique in natural products chemistry, but the advantages encourage its use in the field. Its advantages include decreased ion suppression in complex mixtures, increased speed, higher tolerance to impurities, increased sensitivity, and low sample consumption compared to ESI. MALDI techniques can be applied to various fields of study, including studies of structural identification, tissue imaging, metabolomics, molecular interactions with target molecules, medicinal plant identification, quantification, and chemical screening.

Although several points are unknown in secondary metabolite analysis, MALDI has been successfully applied, thus demonstrating its potential and usefulness for the analyzed metabolite classes listed in this review. Several types of metabolites can be analyzed by MALDI, such as hydrolysable and condensed tannins, alkaloids, saponins, flavonoids, anthocyanins, sesquiterpene lactones, carotenoids, steroids, diterpenes, and lipids, including glycosylated and non-glycosylated compounds. The development of chemical analysis guidelines for different metabolite classes, as in this review and others in the future, will assist in the planning of the experimental design, interpretation, and data processing. In addition, the MALDI-MS/MS data obtained by low energy transfer methods can be easily compared to ESI fragmentation, facilitating its use in the structural identification of compounds because several ESI-MS/MS fragmentation pathways have already been reported. Additionally, fragment ions are often different due to charge-remote fragmentation in the high energy transfer methods of MS/MS that are commonly applied in MALDI analyses. These fragment ions can give additional and useful information for structural identification, as observed, for example, in carotenoids, saponins, and tannins. However, more studies on the fragmentation mechanisms for different metabolite classes are required.

With the increasing availability of MALDI experimental applications, dedicated software is needed; however, for most analytical needs, numerous powerful algorithms and free software implementations are available. The scientist’s main task is to understand the data structure and assumptions of an analytical method and correctly apply it to their own data sets. Most readers may not have experience with data analysis, and the availability of dedicated software is important, but this review sought to recognize data analysis as an important point in the experimental process and emphasize the importance of data analysts in multidisciplinary groups.

Acknowledgments

The authors thank FAPESP (Process: 2009/54098-6, 2012/18031-7, 2014/20302-4 and 2014/01884-2), CNPq, INCT_IF, CAPES, and FUNDECT.

Conflict of Interest

The authors declare that no conflicts of interest exist.

References

44 Ivanova B, Spitteler M. Simultaneous quantitation of naturally occurring insecticides, acaricides, and piscicides in rapeseed oil by UV–MALDI mass spectrometry. J Food Meas Charact 2014; 8: 15–28
53 van Kampen JFA, Burgers PC, de Groot R, Luider TM. Qualitative and quantitative analysis of pharmaceutical compounds by MALDI-TOF mass spectrometry. Anal Chem 2006; 78: 5403–5411

70 Kueger S, Steinhauser D, Willmitzer L, Giavalisco P. High-resolution plant metabolomics: from mass spectral features to metabolites and from whole-cell analysis to subcellular metabolic distributions. Plant J 2012; 70: 39–50

71 Hegeman AD. Plant metabolomics—meeting the analytical challenges of comprehensive metabolite analysis. Brief Funct Genomics 2010; 9: 139–148


75 Lee VJ, Perdian DC, Song Z, Yeung ES, Nikkau BJ. Use of mass spectrometry for imaging metabolites in plants. Plant J 2012; 70: 81–95


94 Goodwin RJ. Sample preparation for mass spectrometry imaging: small mistakes can lead to big consequences. J Proteomics 2012; 75: 8993–8911


115 Heeren RMA. Getting the picture: The coming of age of imaging MS. Int J Mass Spectrom 2015; 377: 672–680


124 Kotter RJ. MALDI-TOF mass spectrometry in lipid research. Prog Lipid Res 2010; 49: 450–475


138 Trentin DS, Silva DB, Amaral MW, Zimmer KR, Silva MV, Lopes NP, Giordani RB, Macedo AJ. Tannins possessing bacteriostatic effect im...
158 Pérez-Jiménez J, Torres JL. Analysis of proanthocyanidins in almond blanch water by HPLC–ESI-QqQ/MS/MS and MALDI–TOF/TOF MS. Food Res Int 2012; 49: 798–806
163 Reed JD, Krueger CG, Vestling MM. MALDI-TOF mass spectrometry of oligomeric food polyphenols. Phytochemistry 2005; 66: 2248–2263
177 Tong DL, Boocock DJ, Coveney C, Snijf J, Gomez SG, Querol S, Rees R, Ball GR. A simpler method of preprocessing MALDI-TOF MS data for differ-
Pogam PL, Schinkovitz A, Legouin B, Lamer ACL, Boustie L, Richomme P.

208

Annangudi SP, Myung K, Adame CA, Gilbert JR.

213


214


215


216


217


218


219


220


221


222


223


224


225


226


227


228


229


230


231


232
