Trends in Analytical Chemistry 69 (2015) 52–61

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

# Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics

Arpana Vaniya [a,b], Oliver Fiehn [b,c,*]

[a] Department of Chemistry, University of California Davis, One Shields Avenue, Davis, CA 95616, USA
[b] West Coast Metabolomics Center, Genome Center, University of California Davis, 451 Health Sciences Drive, Davis, CA 95616, USA
[c] Biochemistry Department, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

*Keywords:*
Fragmentation tree
Ion tree
Mass spectral tree
Mass spectrometry
Metabolite identification
Metabolomics
$MS^n$
Multi-stage analysis
Tandem mass spectrometry
Unknown compound

ABSTRACT

Identification of unknown metabolites is the bottleneck in advancing metabolomics, leaving interpretation of metabolomics results ambiguous. The chemical diversity of metabolism is vast, making structure identification arduous and time consuming. Currently, comprehensive analysis of mass spectra in metabolomics is limited to library matching, but tandem mass spectral libraries are small compared to the large number of compounds found in the biosphere, including xenobiotics. Resolving this bottleneck requires richer data acquisition and better computational tools. Multi-stage mass spectrometry ($MS^n$) trees show promise to aid in this regard. Fragmentation trees explore the fragmentation process, generate fragmentation rules and aid in sub-structure identification, while mass spectral trees delineate the dependencies in multi-stage MS of collision-induced dissociations. This review covers advancements over the past 10 years as a tool for metabolite identification, including algorithms, software and databases used to build and to implement fragmentation trees and mass spectral annotations.



**Contents**



## 1. Introduction

Mass spectrometry (MS) is the dominant analytical technique in metabolomics. The elemental composition and structural information of a molecule can be readily determined by information provided by MS, such as accurate mass-to-charge ratio ($m/z$), isotope abundance [1] and fragmentation patterns [2]. The Metabolomics Standards Initiative (MSI) categorizes structure elucidation into four different levels: identification, annotation, characterization and classification [3,4]. These levels establish a thorough standard for the validation of metabolites that are identified across non-targeted metabolomic studies [4]. However, MSI does not provide a scoring schema to rank identified compounds within the identified and annotated categories, a caveat that was recently highlighted by metabolomics investigators [5]. Identification of metabolites refers to complete identification of the structure, including molecular connections and stereochemical assignments [6]. The identification process of small molecules in metabolomics is similar to that in other fields, such as toxicology and proteomics. All fields use accurate mass analysis, databases or libraries, and mass spectral fragmentations, such as LC-MS/MS. Some major differences between metabolomics and proteomics are the presence of multiply-charged ions from

* Corresponding author. Tel.: +1 530 754 8258.
*E-mail address:* ofiehn@ucdavis.edu (O. Fiehn).

http://dx.doi.org/10.1016/j.trac.2015.04.002
0165-9936/

peptides and the much larger chemical diversity in metabolomics and exposome analyses [7–9]. Synthesizing reference standards for confirmation of putative identifications is limited, time consuming, and uneconomical. According to MSI, annotation is putative compound identification in which the assignment of structures is highly likely, but not validated through chemical-reference standards [4]. Structure annotations are often ambiguous due to the large number of possible isomers, data inaccuracies, limited amounts of corroborating information, and human errors, including misclassification of sub-structures. However, annotation can also be viewed as a strategy to reduce the need for isolation of compounds and *de-novo* elucidation. The idea is to annotate mass spectra using the most probable elemental compositions found in public databases and to add additional orthogonal filters to decrease the number of structure hits [10].

Computer-assisted structural elucidation (CASE) encompasses structural dereplication using various analytical techniques from tandem MS ($MS^2$) and multi-stage MS ($MS^n$) to ultraviolet-visible (UV), infrared (IR) and nuclear magnetic resonance (NMR) spectroscopies. CASE first reduces chemical and spectral properties of an unknown compound, second generates candidate structures compatible with spectral features, and then ranks these candidates [11–13]. CASE can be used when manual interpretation of data is impractical and outcomes are unreliable using certain techniques, such as artificial intelligence, pattern recognition, library search, and spectral simulation [12,14]. Conversely, structural dereplication is performed by comparing experimental data to well-known databases that have standard reference data. Essentially, dereplication is a process to identify "known unknowns", which are compounds that are unknown at the time of detection and with further investigation are then found to be known compounds [15]. For example, the National Institute of Standards and Technology (NIST) database can be used to identify unknown compounds in gas chromatography-MS (GC-MS) studies [16]. Both structural dereplication and CASE are not considered *de-novo* identification because they rely on database searches with pre-existing known metabolites or reference standards [17]. Full *de-novo* identification by MS alone can hardly be achieved because isomers are difficult to distinguish by MS [10]. Mass spectral data inform about elemental compositions by combining accurate mass and isotopic information [1]. Collision-induced fragmentation data on the $MS^2$ or $MS^n$ levels are used to find structural information from unique fragmentation patterns to test for the presence and the absence of functional groups. Interpretation of data in CASE may subsequently yield a partial structure or a sub-structure [12] (e.g., by using graphs that represent $MS^n$ fragmentation-tree spectra in a hierarchical and data-dependent format). In CASE, rules, such as the calculation of "Rings plus double-bond equivalents" (RDBE), the nitrogen rule and the "even-electron rule", are applied when interpreting MS data to identify the formation of fragment ions and neutral species [18].

The scope of this review is to discuss advancements in techniques used by MS for structure elucidation, specifically the use of $MS^n$ ion trees for small organic molecules with molecular weights less than 2 kDa.

## 2. Limitations of tandem mass spectrometry

While collision-induced dissociation (CID) MS/MS today is the dominant technique for library matching and interpreting fragment patterns to find structural information [6], using MS/MS alone falls short because product ions found in the MS/MS spectrum may be derived from intermediary ions instead of being produced directly from the molecular adduct precursor ion. For example, although epinine (deoxyepinephrine) conjugates in urine can be determined by MS/MS via precursor ion and neutral loss scans [19], MS/MS is unable to distinguish between positional isomers of such catecholamines. In addition, many fragment ions in MS/MS cannot be explained through fragmentation pathways even when structures are known [19]. Isomeric flavonoid O-diglycosides may yield different product-ion ratios in MS/MS fragmentation spectra [20]. However, such fragment-ion ratios cannot be used to infer interglycosidic linkages or glycan sequences in structural annotations of unknowns (Fig. 1) even though the authors successfully constructed a decision tree to differentiate these O-diglycosyl flavonoids [20].

Similarly, the annotation of positional sub-structures of taxanes in *Taxus* could not be achieved by MS/MS alone but only by using additional analytical methods [21]. Taken together, MS/MS certainly does not provide full structural information to elucidate an unknown compound completely. MS/MS fails to yield specific positional information of sub-structures, and many fragment ions remain unannotated with respect to presence of sub-structures or detailing fragmentation pathways.

## 3. Fragmentation trees and mass spectral trees

Trees are data structures defined by graph theory to organize and store data (e.g., the fragmentation process of an analyte of interest, or $MS^n$ spectra generated by an ion-trap mass spectrometer). A tree is generated by nodes that are linked by edges (Fig. 2). Typically, the graphs are called fragmentation trees [23], family trees [24] or identification trees [25], if these trees show the fragmentation pathway of a molecule (Fig. 2A). Fragmentation trees are generated computationally to predict the fragmentation pathway of a molecule [23]. An implication of the fragmentation relationship between precursor ions and product ions is made before acquiring $MS^n$ data. Conversely, ion trees or mass spectral trees refer to the sequential stages and relationships of mass spectral acquisition in $MS^n$ processes, representing precursor and product ions as nodes and neutral losses as edges [26,27] (Fig. 2B). $MS^n$ trees can therefore link ion-fragmentation pathways with (sub)structure relationships in a hierarchical order. An important aspect of $MS^n$ trees is that they reveal both the dependency of precursor/product ion and product ion/product ion within the same $MS^n$ stage or between different $MS^n$ stages. This idea is rooted in the concept that any two $MS^n$ spectra can ideally be treated as virtual MS/MS data: an ion has no memory. Hence, organizing large $MS^n$ libraries will yield a tremendous expansion of publicly available MS/MS spectra, as long as each mass spectrum (Fig. 2B) is associated with a defined structure (Fig. 2A). For both fragmentation and mass spectral trees, computational methods are required to organize dependencies and extract specific information.

### 3.1. $MS^n$ ion tree for fragmentation analysis in natural products research

$MS^n$ multistage analysis provides means to link all product ions to specific precursor ions, hence enabling recursive reconstruction of fragmentation pathways that link specific sub-structures to complete molecular structures [28]. Oligosaccharides and sugar nucleotides were annotated using $MS^4$ ion trees with Mass Frontier 2.0 software [29], but the ion trap used lacked accurate mass capabilities to associate fragmentation rules unambiguously with potential fragmentation pathways to identify unknown metabolites detected in plant-phloem samples. Fabre et al. [30] successfully used $MS^n$ to characterize structurally fragment ions and fragmentation mechanisms of flavonoid aglycones in negative-ion mode. $MS^3$ data supported fragmentation mechanisms, helped distinguish common neutral losses for specific sub-structures, and gave sufficient information to propose reasonable structures for fragments using both experimental and computational MS. However, for some

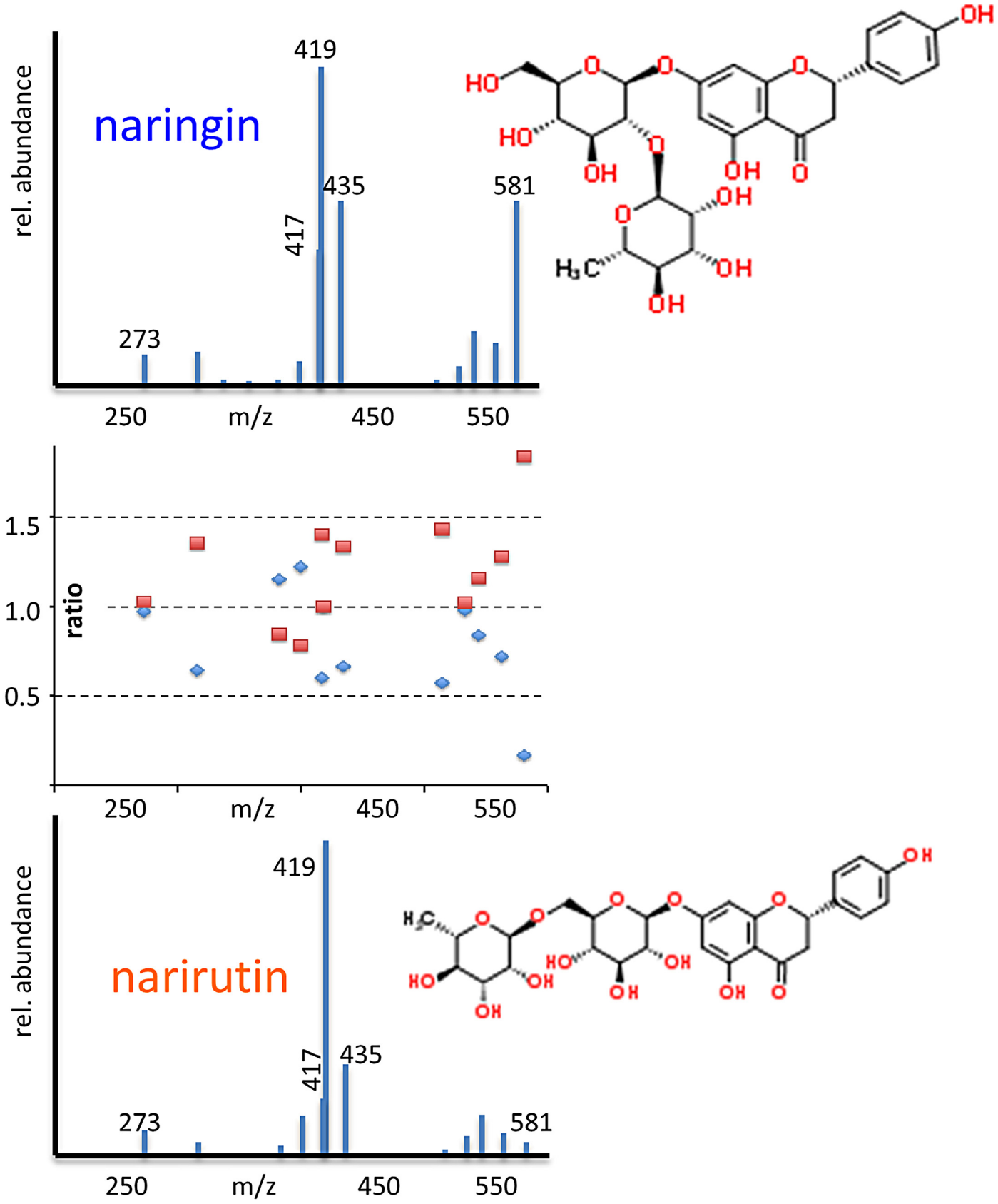


**Fig. 1.** Ion trap MS/MS spectra of (a) naringin (blue), (b) narirutin (red), acquired at 20% CID. While *m/z* values of MS/MS product ions are identical, normalized ion ratios (mid panel) can distinguish these isomeric flavonoids. Ion ratios cannot be used to determine glycan sequences on the aglycone backbone or specific interglycosidic linkages. Mass spectra adapted from [20].

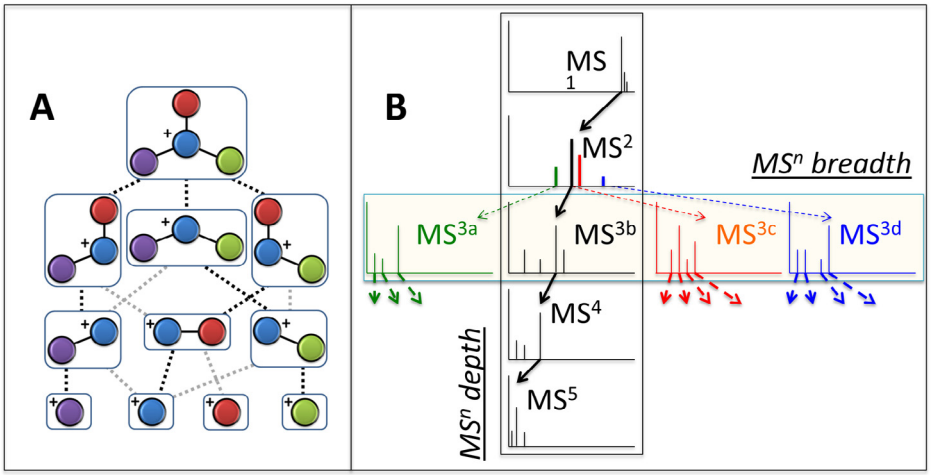


**Fig. 2.** Left panel: A fragmentation tree and graph (A) represents structures or chemical formulas as nodes, here symbolized by rectangles with color coded 'sub-structures'. Grey edges represent a fragmentation graph and black edges show the fragmentation process and mechanism used to generate the fragmentation tree. {Figure adapted from [22]}. Right panel: A mass spectral tree (B) shows nodes as individual mass spectra. Mass spectral trees are characterized by depth ($MS^n$ level) and breadth (the number of ions from each mass spectra that are selected for subsequent $MS^n$-level fragmentations).

flavonoid aglycones, $MS^3$ experiments did not provide sufficient data to deduce fragmentation mechanisms conclusively [30].

Importantly, $MS^n$ can be used for dereplication of natural products, specifically to differentiate C-glycosidic flavonoid isomers [31]. $MS^2$ yielded insufficient data to distinguish 6-C and 8-C-glycosidic flavonoids because no specific diagnostic ions were present to differentiate such isomers, even though ion-intensity ratios were different (Fig. 3). In order to establish rules about how ion ratios could distinguish isomers, many more natural-product $MS^2$ spectra would need to be acquired and computationally analyzed. However, on the $MS^3$ stage, clear diagnostic ions were present to distinguish vitexin and isovitexin (Fig. 3) because mechanisms of C-ring cleavages were very different for these two isomers. Moreover, $MS^4$ yielded data to prove the exact position of C-glycosylation on vitexin-2-O"-rhamnoside [31]. These data formed a decision tree for dereplication of flavonoids in the analysis of complex mixtures [31]. $MS^n$ data therefore provide more information to identify natural products unambiguously. Twenty-five citrus flavonoid O-diglycosides were identified by comparing experimental $MS^3$ spectra to $MS^3$ spectra of reference compounds isolated from *F. aurantii* [32], as MS/MS spectra proved to be insufficient for high-confidence identifications.

## 4. $MS^n$ data-acquisition methods

Data-dependent ion-tree experiments (dd-ITe) are generally used to collect $MS^n$ data on ion trap mass spectrometers such as the linear quadrupole ion trap or Orbitrap instruments. Usually, spectra are acquired under electrospray ionization (ESI) either using direct infusion or flow injection [33]. Direct infusion-$MS^n$ data acquisition provides the necessary time to populate the ion trap with ions of sufficient signal intensity and to acquire exhaustive mass spectral trees (Fig. 2). As an example, structure analysis of lipid A in *Francisella tularensis* subspecies *novicida* [34] was acquired in this manner on a hybrid linear ion-trap Fourier transform (FT) ion-cyclotron resonance mass spectrometer. In comparison, flow injection, while being compatible with liquid chromatography (LC) systems, often does not yield sufficient time during the elution profile of the flow-injection peak to yield strong enough signals for mass spectral tree acquisitions [35]. An alternative might be posed by the Orbitrap Fusion mass spectrometer that is equipped with a quadrupole, ion trap and Orbitrap mass analyzer. Reportedly, this instrument may be compatible to fit the timescale of ultra-high pressure LC (UHPLC) peak widths and generate $MS^n$ trees for profiling and structure elucidation of metabolites using mzCloud, an $MS^n$ library for high- and low-resolution data [36].

A comprehensive method was developed using both $MS^n$ spectra and fragmentation trees for metabolite identification [37]. A very extensive dd-ITe was programed to perform 107 mass spectra per analyte with a maximum $MS^n$ depth of n = 5 by data-dependent fragmentation using the five most abundant product ions at the $MS^2$ and $MS^3$ levels, and the three most abundant product ions at the $MS^4$ level [37]. Such dd-ITe was used to investigate structures of intact polar lipids of microbes found in two regions in an artesian sulfur-rich spring source [38]. DD-ITe was used for structural analyses of seven vergauamide compounds in marine cyanobacteria, using a maximum $MS^3$ depth [39]. Non-genotoxic carcinogens in rats and lipid species were also structurally elucidated and confirmed using dd-ITe [40]. However, there are alternatives to ion-trap-based $MS^n$ studies. Using quadrupole time-of-flight (TOF)-MS, pseudo-$MS^3$ spectra can be acquired using high energy to produce in-source fragmentation and subsequent collision-induced fragmentation with accurate mass product-ion analysis. As the precursor ion and initial neutral-loss product ion are fragmented simultaneously, composite $MS^3$ spectrum are generated [41]. $MS^n$ data were also acquired using a triple-quadrupole mass spectrometer to identify and to characterize glycerophosphatidylethanolamine lipids structurally [42], specifically to characterize substituents on the fatty acyl chains.

A second alternative method combines LC with $MS^n$ and subsequent solid-phase extraction (SPE) with NMR spectroscopy [43]. This method was used for structural elucidation of tomato flavonoids present at sub-µg amounts in crude extracts (e.g., quercetin-3-O-glucoside) [44]. A combination of LC-FT ion cyclotron $MS^n$ and LC-TOF-MS-SPE-NMR proved to be successful in two studies: 138 urinary metabolites were annotated and 36 phenolic conjugates were structurally elucidated in a study on consumption of black or green teas [45]. In a related study, 177 phenolic compounds in tea products were annotated to be derivatives of flavan-3-ols and flavonols using spectral trees to profile conjugates and derivatives [46].

LC with high-resolution $MS^3$ has been combined with a mass spectral tree similarity-filter technique (MTSF) to identify 68 compounds in traditional Chinese medicine [47]. Some 14 reference compounds were used to generate mass spectral trees and build a user library in Mass Frontier version 7.0 software. MTSF workflow includes collecting $MS^n$ data on all detected compounds and using Mass

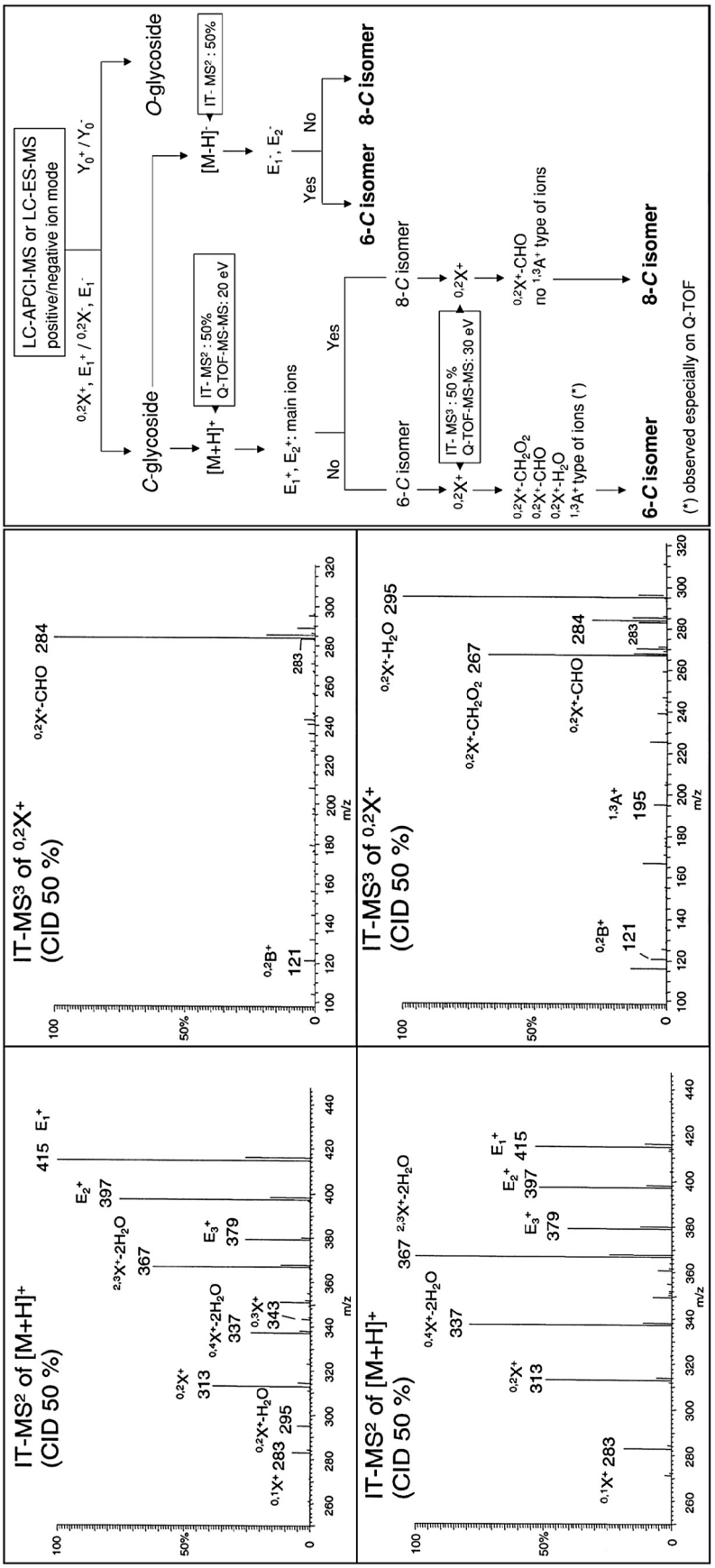


**Fig. 3.** $MS^2$ and $MS^3$ ion-trap spectra for the 8-C-glycosidic flavonoid vitexin (top left and top right) and the 6-C-glycosidic flavonoid isovitexin (bottom left and bottom right). The product ion in $MS^2$ *m/z* 313 produced many different fragment ions that can be used as diagnostic ions to differentiate the two C-glycoside isomers. Using the fragment ions, a decision tree (right) was made to differentiate 6-C and 8-Cglycosidic flavonoid isomers. {Reproduced with permission from [31]}.

Frontier 7.0 to calculate the similarity between the $MS^n$ tree of the reference compound and the detected compound and to assign a similarity score. After the candidate compound was determined, accurate *m/z* value and fragmentation rules were used to determine the identity further [47]. Similarly, Jia et al. [48] characterized and identified 38 compounds in *Saussurea involucrate* using the same method of combining LC-high-resolution $MS^n$ with MSTF. In human-urine samples, 10 compounds were annotated from 30 target unknown molecules using $MS^n$ trees and similarity matching to find and to verify sub-structures and validate that unknown metabolites belong to a specific compound class [49]. Similarly, 127 phenolics and glucosinolates were identified by $MS^n$ with MTSF matching from crude extracts of tomatoes and *Arabidopsis* leaves [50]. Wang et al. [51] also used the MSTF method to report detection and confirmation of illegal adulterants in health foods and herbal medicines.

## 5. Computational tools for $MS^n$ and fragmentation trees

When accounting for the huge chemical complexity of natural products in plants and microbes and biotransformations in microbial communities, such as the human gut, there are millions of small molecules in nature. It is impossible to acquire reference mass spectra for all these compounds in libraries [16,52]. Instead, *in-silico* prediction tools can be used to generate much larger virtual MS/MS and $MS^n$ spectral libraries [52,53]. Such prediction tools could be developed and validated by mass spectral libraries of authentic compounds. Especially the large LipidBlast library of over 200,000 MS/MS spectra of complex lipids is a good example how rule-based generation of virtual spectra may upend the dominance of small reference spectral libraries used in metabolomics [54]. Computational MS is a necessity for big data initiatives involving MS/MS and $MS^n$ analyses and fragmentation trees [37,52,55,56]. Use of such tools is evaluated in a new initiative, the Critical Assessment of Small Molecule Identification contest (CASMI) [57,58].

Fragmentation and ion trees aim to identify the molecular formulas of compounds, elemental compositions of fragment ions and neutral losses, to perform automatic annotations on MS/MS or $MS^n$ spectra, to aid in structure and sub-structure elucidation by *in-silico* fragmentation, to predict molecular fingerprints and to provide a *de-novo* identification strategy. Fragmentation trees have been primarily calculated from MS/MS data, GC electron-ionization-MS data, and $MS^n$ data [2,23,59–62]. Fragmentation trees were developed for MS/MS spectra annotation by representing each peak by a node with a molecular formula [23]. Similarly, fragmentation trees were computed to show the dependencies between fragment ions found in CID spectra [63]. More recently, the SIRIUS$^2$ software (Sum Formula Identification by Ranking Isotope Patterns Using Mass Spectrometry) was released to determine the molecular formula aided by fragmentation trees [64–66] but limited to $MS^2$ data. Calculation of fragmentation trees is more difficult from $MS^n$ data because the number of relationships between ions explodes with $MS^n$ data [61,62]. Molecular formulas are computationally generated and assigned to each fragments on a fragmentation tree. This calculation requires an input of $MS^n$ data. Fig. 4 shows the multiple relationships and fragment dependencies found between $MS^2$ and $MS^3$ spectra.

A published approach to calculate fragmentation trees used a tree-completion heuristic method [67] and a strategy was developed that computed fragmentation trees and used kernel-based machine-learning techniques to improve identification of metabolites [68]. Alignments of fragmentation trees are required to find similar fragmentation pathways across different compounds. Compounds that share similar fragmentation patterns correlate with strong chemical similarity. Tentative structural information can be gained by the alignment of unknown compounds to known compounds [69]. The fragmentation tree basic logic alignment search tool (FT-BLAST) compares fragmentation patterns and groups compounds based on those

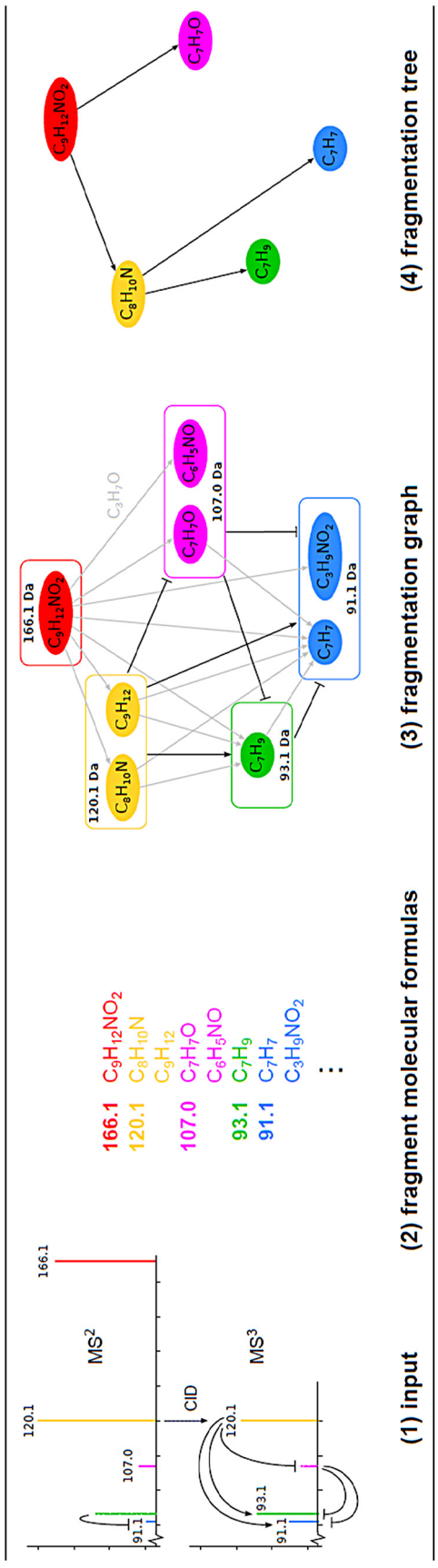


**Fig. 4.** Workflow for generating fragmentation trees from $MS^n$ data. $MS^n$ ions are assigned by molecular formulas, organized into a fragmentation graph and ultimately assembled into a tree that best explains likely fragmentation steps. {Reproduced with permission from [62]}.

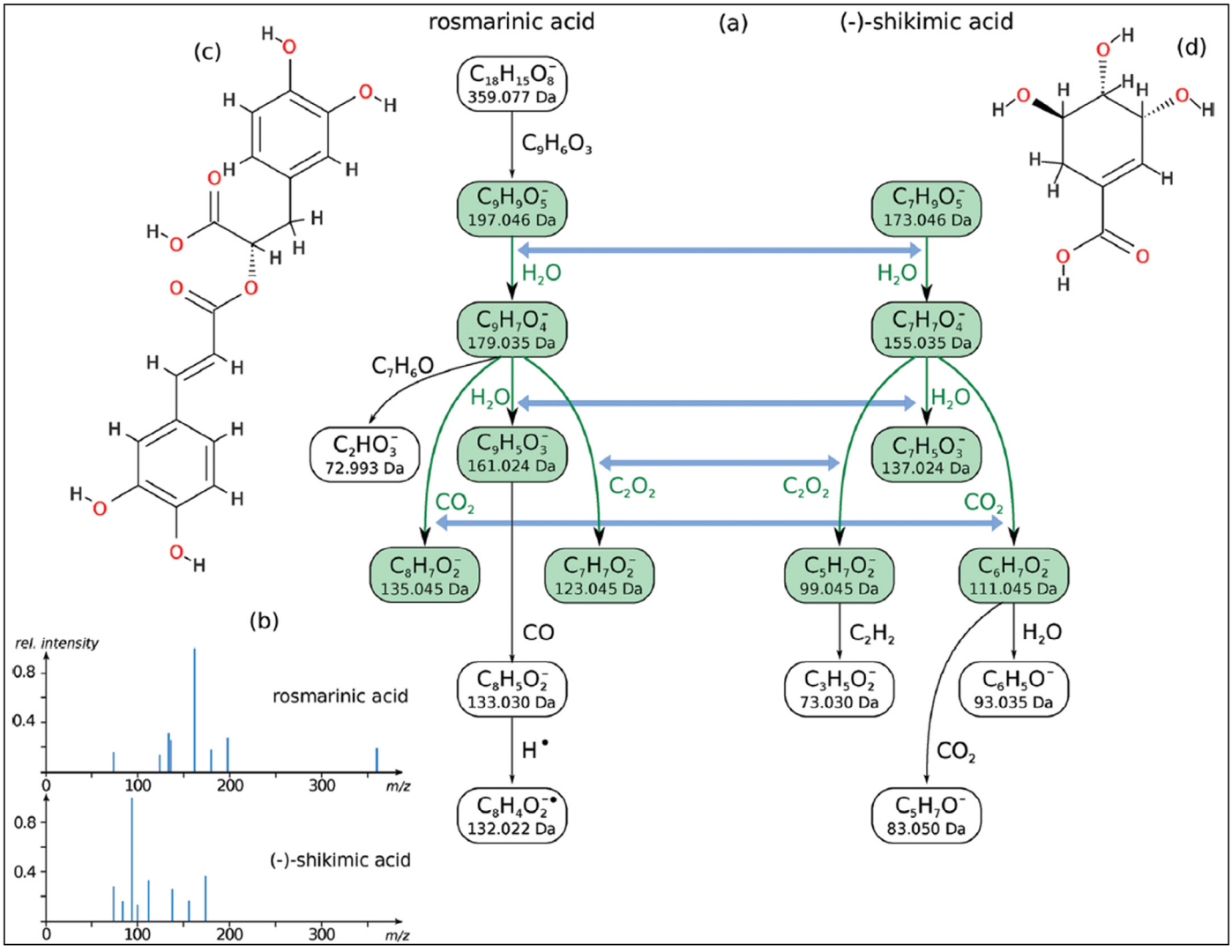


**Fig. 5.** An example how FT-BLAST generating fragmentation trees identifies sub-structures from complex mass spectra data. (a) Alignment of CID mass spectra shown in panel (b) of rosmarinic acid (c) and (-)-shikimic acid (d). While ions do not directly align in the mass spectra, the fragmentation-tree alignment shows structural similarity and similar key losses of: $H_2O$, $CO_2$, and $C_2O_2$. {Reproduced with permission from [70], ©2012 American Chemical Society}.

similarities (Fig. 5). Eight compounds from 89 *m/z* features were annotated from $MS^n$ data of extracts of Icelandic poppies (*P. nudicaule*) [70]. As tree alignments present multiple solutions, comparison results need to be scored. New methods may prove useful for both metabolite identification and searching databases [71].

Structure and sub-structure annotations based on *in-silico* fragmentations are based on known chemical rules or use combinatorial approaches. The state-of-the-art commercial software Mass Frontier generates fragments based on rule-based predictions, produces mass spectral trees, calculates fragmentation pathways, searches for sub-structures by FISh (Fragment Ion Search), calculates molecular formulas, and develops fragmentation rules. For fragment-structure predictions, Mass Frontier employs common fragmentation and rearrangement rules and literature-based assignments. Mass spectral trees are linked to predicted fragments and mechanisms. Fragmentation rules obtained from such $MS^n$ trees are best applied to sets of specific compound classes. A novel mass spectral database, such as mzCloud, supports the precursor-ion-fingerprinting (PIF) algorithm to interpret mass spectra by performing library searches for the precursor ion, generating spectral trees, and generating $MS^n$ tree libraries [26]. As an alternative to Mass Frontier, academic software was released to analyze $MS^n$ spectra and generate spectral trees [37]. This software was specifically designed to remove or filter artefacts present from the LTQ-Orbitrap XL. Such artefacts have been previously described as a result of electronic interference found in Fourier-transformation instruments, such as ion-cyclotron resonance and Orbitrap mass spectrometers [72]. Recently, an *iontree* R-package was developed for handling MS/MS and $MS^3$ spectra, comparing $MS^n$ spectra and building ion-tree libraries [73]. The *iontree* package is platform independent, which improves the capability of managing $MS^n$ data from different instruments and laboratories [73].

$MS^n$ data are also analyzed by the Multistage Elemental Formula (MEF) software, which determines elemental compositions for precursor ion, fragment ion and neutral losses [49]. This software annotated several candidate structures from human urine by matching spectral sub-trees of known reference compounds to sub-trees of unknown metabolites. Such sub-trees are portions of the overall $MS^n$ tree that have sub-structures in common.

Recently, MAGMa (MS Automation based on in silico Generated Metabolites) was introduced as a tool for LC-$MS^n$ spectra annotation [43,74] and to provide structure elucidation and was reported to be the best automated tool in CASMI 2013 [58]. Sub-structures and spectral tree annotation using MAGMa are found by using systematic bond dissociation to fragment a candidate

structure. In itself, systematic bond dissociation has been a popular method for enumerating all possible fragment-candidate structures.

MetFrag was developed to generate all *in-silico* fragments using a bond-disconnection method [55] and was reviewed [27,75,76]. MetFrag generates fragmentation trees to reduce the number of calculated fragments. The problem with bond-disconnection-method approaches is that *all* bonds are fragmented, without initially accounting for bond strengths or bond-dissociation energies. They are taken into account when the scores of all *in-silico* candidate fragments are computed. MetFrag is limited to accurate mass data and performs best when generating a fragmentation tree with a maximum depth of 2 [55]. Subsequently, MetFusion was developed to improve compound identification by combining both MetFrag and mass spectral searching in MassBank [77].

The alternative fragmenter software ISIS (In Silico Identification Software) was developed for fragmentation-pattern analysis using artificial neural network machine learning and kinetic Monte Carlo algorithms to learn bond cleavages from ion-trap spectra in order to predict *in-silico* MS/MS spectra [78]. Systematic bond dissociation is also implemented in MIDAS (Metabolite Identification via Database Searching) software [79].

Originally, EPIC (Elucidation of Product Ion Connectivity) was created as a program to assign automatically sub-structures generated by systematic bond dissociation to only MS/MS data [80]. A further algorithm finds parent/fragment ion pairs and also helps to distinguish peaks from unrelated compounds or contaminants [81].

Recently, CASS (Chemically Aware Sub-structure Search) was developed to provide a tool that automatically detects functional groups in compound libraries [82]. CASS is also designed to create a functional group-resolved metabolite database. CASS is not hard-coded and is flexible to customize with additional functional groups.

Sub-structure generation via exhaustive combinatorial tools, such as MOLGEN (MOLecular structure GENeration) [83], quickly leads to computational performance problems and may not explain all peaks found in experimental or *in-silico* fragmentation spectra due to the vast number of similar candidate structures and the lack of rich sub-structure information from spectra [84]. Sub-structures and characteristic product ions are searchable using MS2Analyzer [85].

Overall, fragmentation trees and $MS^n$ trees combined may be best suited to search for sub-structures to provide annotations of unknown metabolites. Common sub-structures can provide evidence for annotation of compound classes and presence of specific functional groups.

Web-based analysis of $MS^n$ data has been made available by the MetiTree (Metabolite Identification Tree) application. Spectral data, fragmentation trees, and fragmentation reactions can be simultaneously explored and deciphered for both structure and sub-structure identification [86,87]. MetiTree was used to investigate secondary metabolites in filamentous fungus *Penicillium chrysogenum* [88].

A different solution to find characteristic sub-structures was presented by the MoleculePuzzle software [89] to predict *in-silico* fragmentations, sub-structures and structural isomers via rule-based logic. Linking $MS^n$ and sub-structure trees [90] may indeed provide a systematic and efficient method for structure elucidation, since the sub-structure tree is linked to the hierarchical order of $MS^n$ data.

Additionally, LC retention-time information may be added to workflows. The main idea is that successfully identified metabolites in LC-MS data sets may be used as bait to fish for derivatives of these compounds. By searching for characteristic *m/z* differences (e.g., for hydroxylations) and predicting that hydroxylated derivatives of identified metabolites will elute earlier in reversed-phase LC, the CSPP algorithm suggests pairs of substrate/product candidate pairs that may be annotated by $MS^n$ spectra [91].

**Table 1**
Summary of software and computational tools for mass spectral and fragmentation trees.

| Software | Main function | Ref. |
|---|---|---|
| FiD | Substructure Prediction | [92] |
| PIF | Generates spectral trees | [26] |
| Mass Frontier | Generates spectral trees | [93] |
| MetFrag | Fragmenter | [55] |
| MEF | Calculates molecular formulas | [86] |
| MoleculePuzzle | Predicts fragmentation of a compound | [89] |
| MetiTree | Generates spectral trees | [87] |
| ISIS | Fragmenter | [78] |
| MetFusion | Combines MetFrag with spectral library search | [52] |
| R-package "iontree" | Generates spectral trees | [73] |
| MAGMa | Fragmenter and annotates spectra | [74] |
| CSPP | Spectra annotation tool | [91] |
| SIRIUS$^2$ | Calculates molecular formulas and fragmentation trees | [64] |

Overall, recent years saw very active developments of a range of algorithms, approaches and software tools to use directly the tree-based approaches that are summarized in Table 1. CASMI may be a suitable test bed for comparing this array of novel opportunities in identification of unknowns.

Apart from the direct application of trees in computational MS, data and information may be generated that can be used indirectly to identify unknown metabolites. The MZmine 2 framework [94] combines heuristic rules, fragmentation-pattern analysis, and isotopic pattern matching to predict molecular formula from HRMS data. FingerID uses MS/MS data to predict molecular fingerprints or properties of a metabolite that are subsequently matched against PubChem to provide a metabolite identity [92,95]. Most recently, CFM (Competitive Fragmentation Modelling) and CFM-ID (Competitive Fragmentation Modelling Identification) were specifically developed to predict MS/MS spectra based on machine learning and probabilistic generative models [22,96]. Results of such predictions appear to outperform MetFrag and FingerID.

## 6. Tandem and $MS^n$ spectral libraries and databases

Accuracy of spectral predictions by any of the aforementioned algorithms or software programs can best be validated by authentic, curated mass spectral repositories, such as the NIST14 library that currently holds MS/MS spectra for 8171 distinct compounds acquired on ion traps and 7692 distinct compounds acquired on QTOF or triple-quadrupole mass spectrometers. Such libraries and databases require high-resolution and high mass accuracy data for annotation of metabolites. Software that performs structural annotation relies on structural databases that may be larger than spectral libraries but nonetheless still incomplete. However, most mass spectral libraries and databases do not store $MS^n$ data, except for HighChem's commercial Spectral Tree library with currently 2740 spectral trees, mzCloud [36] (a freely available Web interface) with currently 2625 spectral trees, and the open access MassBank database [77] that contains 2.2% $MS^3$ or $MS^4$ spectra. Both High Chem's Spectral Tree library and mzCloud support precursor-ion fingerprinting [26] and are fully integrated with Mass Frontier 7.0.

HAMMER (High-throughput AutoMation of Mass frontiER) is freely available software that was developed to compensate for the lack of chemical space that current $MS^n$ spectral libraries cover [56]. HAMMER allows users to control Mass Frontier 7.0 to build *in-silico* $MS^n$ mass spectral libraries.

The ground-breaking FragLib [97] library was developed in 2005 for the characterization of glycans and oligosaccharides as the first repository that was structured to encompass $MS^n$ data and to build $MS^n$ fragmentation trees from $MS^n$ data. This glycan $MS^n$ spectral library was built to differentiate isobars, characterize sub-structures

due to extensive fragmentation, and allow for complete structure elucidation from these sub-structures, leading to discovery of novel compounds. Another glycan and glycolipid library is GMDB, which holds mas spectra up to the $MS^4$ level [98].

## 7. Conclusions

While over 95% of all acquired LC/MS fragmentation studies currently remain at the MS/MS level, the lack of standardized mass spectral libraries and the huge number of unidentified metabolites limit overall progress in metabolomics. The bottleneck of compound identification in metabolomics cannot be overcome without better mass spectral prediction tools. Fragmentation trees and $MS^n$ mass spectral trees may here give the answer. In recent years, a large increase in efforts was noted for acquisition of $MS^n$ data and developments of tools for structure elucidation and spectra annotations. Advancements have come to a point where bottlenecks may be limited by the number of publicly-available data with respect to authentic and curated $MS^n$ spectra of natural products as well as high-quality $MS^n$ data sets from metabolomics studies. The increasing interest in metabolomics by researchers and funding agencies raises hope that larger data sets may soon be available to test, to validate and to compare the multitude of algorithms and software tools that promise to yield accurate compound-annotation results. Computational contests, such as CASMI, will aid in developing standards similar to developments in proteomics and prediction of protein-crystal structures in the past.

## Authors' contribution

Both authors contributed equally to the work.

## Acknowledgements

This study was supported by US National Institutes of Health (U24 DK097154), and US National Science Foundation (Grants MCB 1139644 and MCB 1153491).

## References

[1] T. Kind, O. Fiehn, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry, BMC Bioinformatics 8 (2007) 20.
[2] S. Böcker, F. Rasche, T. Steijger, Annotating fragmentation patterns, in: S.L. Salzberg, T. Warnow (Editors), Algorithms in Bioinformatics, Proceedings, Springer-Verlag, Berlin, 2009, pp. 13–24.
[3] O. Fiehn, D. Robertson, J. Griffin, M. van der Werf, B. Nikolau, N. Morrison, et al., The metabolomics standards initiative (MSI), Metabolomics 3 (2007) 175–178.
[4] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, et al., Proposed minimum reporting standards for chemical analysis, Metabolomics 3 (2007) 211–221.
[5] D.J. Creek, W.B. Dunn, O. Fiehn, J.L. Griffin, R.D. Hall, Z.T. Lei, et al., Metabolite identification: are you sure? And how do your peers gauge your confidence?, Metabolomics 10 (2014) 350–353.
[6] T. Kind, O. Fiehn, Advances in structure elucidation of small molecules using mass spectrometry, Bioanal. Rev. 2 (2010) 23–60.
[7] T.W.-M. Fan, A.N. Lane, R.M. Higashi (Editors), The Handbook of Metabolomics, Springer, New York, 2012.
[8] M. Lämmerhofer, W. Weckwerth (Editors), Metabolomics in Practice: Successful Strategies to Generate and Analyze Metabolic Data, Wiley, New York, 2013.
[9] W.J. Griffiths (Editor), Metabolomics, Metabonomics and Metabolite Profiling, RSC Publishing, London, 2008.
[10] T. Kind, O. Fiehn, Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm, BMC Bioinformatics 7 (2006) 10.
[11] A.H. Lipkus, M.E. Munk, Automated classification of candidate structures for computer-assisted structure elucidation, J. Chem. Inf. Comput. Sci. 28 (1988) 9–18.
[12] M.E. Munk, C.A. Shelley, H.B. Woodruff, M.O. Trulson, Computer-assisted structure elucidation fresen, Z. Anal. Chem. 313 (1982) 473–479.
[13] M. Jaspars, Computer assisted structure elucidation of natural products using two-dimensional NMR spectroscopy, Nat. Prod. Rep. 16 (1999) 241–247.
[14] W.A. Warr, Computer-assisted structure elucidaton. 1. Library search and spectral data collections, Anal. Chem. 65 (1993) A1045–A1050.
[15] J.L. Little, C.D. Cleven, S.D. Brown, Identification of "known unknowns" utilizing accurate mass data and chemical abstracts service databases, J. Am. Soc. Mass Spectrom. 22 (2011) 348–359.
[16] S. Rochfort, Metabolomics reviewed: a new "Omics" platform technology for systems biology and implications for natural products research, J. Nat. Prod. 68 (2005) 1813–1820.
[17] F.E. Koehn, G.T. Carter, The evolving role of natural products in drug discovery, Nat. Rev. Drug Discov. 4 (2005) 206–220.
[18] J.H. Gross, Mass Spectrometry: A Textbook, Springer, Berlin, 2004.
[19] P. Rudewicz, K.M. Straub, Rapid structure elucidation of catecholamine conjugates with tandem mass-spectrometry, Anal. Chem. 58 (1986) 2928–2934.
[20] F. Cuyckens, R. Rozenberg, E. de Hoffmann, M. Claeys, Structure characterization of flavonoid O-diglycosides by positive and negative nano-electrospray ionization ion trap mass spectrometry, J. Mass Spectrom. 36 (2001) 1203–1210.
[21] E.H. Kerns, K.J. Volk, S.E. Hill, M.S. Lee, Profiling taxanes in taxus extracts using LC/MS and LC-MS/MS techniques, J. Nat. Prod. 57 (1994) 1391–1403.
[22] F. Allen, R. Greiner, D. Wishart, Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification, Metabolomics (2014) 1–13.
[23] S. Böcker, F. Rasche, Towards de novo identification of metabolites by analyzing tandem mass spectra, Bioinformatics 24 (2008) I49–I55.
[24] X.H. Guo, A.P. Bruins, T.R. Covey, Characterization of typical chemical background interferences in atmospheric pressure ionization liquid chromatography-mass spectrometry, Rapid Commun. Mass Spectrom. 20 (2006) 3145–3150.
[25] A. Schwarzenberg, F. Ichou, R.B. Cole, X. Machuron-Mandard, C. Junot, D. Lesage, et al., Identification tree based on fragmentation rules for structure elucidation of organophosphorus esters by electrospray mass spectrometry, J. Mass Spectrom. 48 (2013) 576–586.
[26] M.T. Sheldon, R. Mistrik, T.R. Croley, Determination of ion structures in structurally related compounds using precursor ion fingerprinting, J. Am. Soc. Mass Spectrom. 20 (2009) 370–376.
[27] F. Hufsky, K. Scheubert, S. Böcker, Computational mass spectrometry for small-molecule fragmentation, TrAC, Trends Anal. Chem. 53 (2014) 41–48.
[28] R.A. Yost, D.D. Fetterolf, Tandem mass spectrometry (MS/MS) instrumentation, Mass Spectrom. Rev. 2 (1983) 1–45.
[29] V.V. Tolstikov, O. Fiehn, Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry, Anal. Biochem. 301 (2002) 298–307.
[30] N. Fabre, I. Rustan, E. de Hoffmann, J. Quetin-Leclercq, Determination of flavone, flavonol, and flavanone aglycones by negative ion liquid chromatography electrospray ion trap mass spectrometry, J. Am. Soc. Mass Spectrom. 12 (2001) 707–715.
[31] P. Waridel, J.L. Wolfender, K. Ndjoko, K.R. Hobby, H.J. Major, K. Hostettmann, Evaluation of quadrupole time-of-flight tandem mass spectrometry and ion-trap multiple-stage mass spectrometry for the differentiation of C-glycosidic flavonoid isomers, J. Chromatogr. A 926 (2001) 29–41.
[32] P.Y. Shi, Q. He, Y. Song, H.B. Qu, Y.Y. Cheng, Characterization and identification of isomeric flavonoid O-diglycosides from genus Citrus in negative electrospray ionization by ion trap mass spectrometry and time-of-flight mass, spectrometry, Anal. Chim. Acta 598 (2007) 110–118.
[33] M. Beckmann, D. Parker, D.P. Enot, E. Duval, J. Draper, High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry, Nat. Protoc. 3 (2008) 486–504.
[34] Y.S. Ting, S.A. Shaffer, J.W. Jones, W.V. Ng, R.K. Ernst, D.R. Goodlett, Automated lipid A structure assignment from hierarchical tandem mass spectrometry data, J. Am. Soc. Mass Spectrom. 22 (2011) 856–866.
[35] D.J. Ashline, A.J. Lapadula, Y.-H. Liu, M. Lin, M. Grace, B. Pramanik, et al., Carbohydrate structural isomers analyzed by sequential mass spectrometry, Anal. Chem. 79 (2007) 3830–3842.
[36] J. Wang, D.A. Peake, R. Mistrik, Y. Huang, G.D. Araujo, A platform to identify endogenous metabolites using a novel high performance orbitrap MS and the mzCloud library, (2013).
[37] P.T. Kasper, M. Rojas-Chertó, R. Mistrik, T. Reijmers, T. Hankemeier, R.J. Vreeken, Fragmentation trees for the structural characterisation of metabolites, Rapid Commun. Mass Spectrom. 26 (2012) 2275–2286.
[38] S. Bühring, S. Sievert, H. Jonkers, T. Ertefai, M. Elshahed, L. Krumholz, et al., Insights into chemotaxonomic composition and carbon cycling of phototrophic communities in an artesian sulfur-rich spring (Zodletone, Oklahoma, USA), a possible analog for ancient microbial mat systems, Geobiology 9 (2011) 166–179.
[39] E. Mevers, W.-T. Liu, N. Engene, H. Mohimani, T. Byrum, P.A. Pevzner, et al., Cytotoxic veraguamides, alkynyl bromide-containing cyclic depsipeptides from the marine cyanobacterium cf. Oscillatoria margaritifera, J. Nat. Prod. 74 (2011) 928–936.
[40] Z. Ament, C.L. Waterman, J.A. West, C. Waterfield, R.A. Currie, J. Wright, et al., A metabolomics investigation of non-genotoxic carcinogenicity in the rat, J. Proteome Res. 12 (2013) 5775–5790.
[41] A.M. Palumbo, S.A. Smith, C.L. Kalcic, M. Dantus, P.M. Stemmer, G.E. Reid, Tandem mass spectrometry strategies for phosphoproteome analysis, Mass Spectrom. Rev. 30 (2011) 600–625.
[42] T.A. Lydic, J.V. Busik, W.J. Esselman, G.E. Reid, Complementary precursor ion and neutral loss scan mode tandem mass spectrometry for the analysis of glycerophosphatidylethanolamine lipids from whole rat retina, Anal. Bioanal. Chem. 394 (2009) 267–275.

[43] J.J.J. van der Hooft, R.C.H. de Vos, L. Ridder, J. Vervoort, R.J. Bino, Structural elucidation of low abundant metabolites in complex sample matrices, Metabolomics 9 (2013) 1009–1018.
[44] J.J.J. van der Hooft, V. Mihaleva, R.C.H. de Vos, R.J. Bino, J. Vervoort, A strategy for fast structural elucidation of metabolites in small volume plant extracts using automated MS-guided LC-MS-SPE-NMR, Magn. Reson. Chem. 49 (2011) S55–S60.
[45] J.J.J. van der Hooft, R.C.H. de Vos, V. Mihaleva, R.J. Bino, L. Ridder, N. de Roo, et al., Structural elucidation and quantification of phenolic conjugates present in human urine after tea intake, Anal. Chem. 84 (2012) 7263–7271.
[46] J.J.J. van der Hooft, M. Akermi, F.Y. Ünlü, V. Mihaleva, V.G. Roldan, R.J. Bino, et al., Structural annotation and elucidation of conjugated phenolic compounds in black, green, and white tea extracts, J. Agric. Food Chem. 60 (2012) 8841–8850.
[47] C.-H. Wang, C.-S. Wu, H.-L. Qin, J.-L. Zhang, Rapid discovery and identification of 68 compounds in the active fraction from Xiao–Xu–Ming decoction (XXMD) by HPLC–HRMS and MTSF technique, Chin. Chem. Lett. (2014).
[48] Z. Jia, C. Wu, H. Jin, J. Zhang, Identification of the chemical components of Saussurea involucrata by high-resolution mass spectrometry and the mass spectral trees similarity filter technique, Rapid Commun. Mass Spectrom. 28 (2014) 2237–2251.
[49] J.E. Peironcely, M. Rojas-Cherto, A. Tas, R. Vreeken, T. Reijmers, L. Coulier, et al., Automated pipeline for de novo metabolite identification using mass-spectrometry-based metabolomics, Anal. Chem. 85 (2013) 3576–3583.
[50] J.J.J. van der Hooft, J. Vervoort, R.J. Bino, R.C.H. de Vos, Spectral trees as a robust annotation tool in LC–MS based metabolomics, Metabolomics 8 (2012) 691–703.
[51] Z. Wang, C. Wu, G. Wang, Q. Zhang, J. Zhang, A novel strategy for the determination of illegal adulterants in health foods and herbal medicines using high-performance liquid chromatography with high-resolution mass spectrometry, J. Sep. Sci. (2015).
[52] M. Gerlich, S. Neumann, MetFusion: integration of compound identification strategies, J. Mass Spectrom. 48 (2013) 291–298.
[53] S. Neumann, S. Böcker, Computational mass spectrometry for metabolomics: identification of metabolites and small molecules, Anal. Bioanal. Chem. 398 (2010) 2779–2788.
[54] T. Kind, K.-H. Liu, D.Y. Lee, B. DeFelice, J.K. Meissen, O. Fiehn, LipidBlast in silico tandem mass spectrometry database for lipid identification, Nat. Methods 10 (2013) 755–758.
[55] S. Wolf, S. Schmidt, M. Muller-Hannemann, S. Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra, BMC Bioinformatics 11 (2010) 12.
[56] J.R. Zhou, R.J.M. Weber, J.W. Allwood, R. Mistrik, Z.X. Zhu, Z. Ji, et al., HAMMER: automated operation of mass frontier to construct in silico mass spectral fragmentation libraries, Bioinformatics 30 (2014) 581–583.
[57] E.L. Schymanski, S. Neumann, CASMI: and the winner is, Metabolites 3 (2013) 412–439.
[58] T. Nishioka, T. Kasama, T. Kinumi, H. Makabe, F. Matsuda, D. Miura, et al., Winners of CASMI2013: automated tools and challenge data, Mass Spectrom. 3 (2014) S0039.
[59] F. Hufsky, S. Böcker, Comparing fragmentation trees from electron impact mass spectra with annotated fragmentation pathways. 26 (2012).
[60] F. Hufsky, M. Rempt, F. Rasche, G. Pohnert, S. Böcker, De novo analysis of electron impact mass spectra using fragmentation trees, Anal. Chim. Acta 739 (2012) 67–76.
[61] K. Scheubert, F. Hufsky, F. Rasche, S. Böcker, Computing fragmentation trees from metabolite multiple mass spectrometry data, J. Comp. Biol. 18 (2011) 1383–1397.
[62] K. Scheubert, F. Hufsky, S. Böcker, Multiple mass spectrometry fragmentation trees revisited: boosting performance and quality, in: D. Brown, B. Morgenstern (Editors), Algorithms in Bioinformatics, Springer, New York, 2014, pp. 217–231.
[63] F. Rasche, A. Svatoš, R.K. Maddula, C. Böttcher, S. Böcker, Computing fragmentation trees from tandem mass spectrometry data, Anal. Chem. 83 (2010) 1243–1251.
[64] K. Dührkop, K. Scheubert, S. Böcker, Molecular formula identification with SIRIUS, Metabolites 3 (2013) 506–516.
[65] K. Dührkop, F. Hufsky, S. Böcker, Molecular formula identification using isotope pattern analysis and calculation of fragmentation trees, Mass Spectrom. 3 (2014) S0037.
[66] M.A. Stravs, E.L. Schymanski, H.P. Singer, J. Hollender, Automatic recalibration and processing of tandem mass spectra using formula annotation, J. Mass Spectrom. 48 (2013) 89–99.
[67] I. Rauf, F. Rasche, F. Nicolas, S. Böcker, Finding maximum colorful subtrees in practice, J. Comp. Biol. 20 (2013) 311–321.
[68] H. Shen, K. Dührkop, S. Böcker, J. Rousu, Metabolite identification through multiple kernel learning on fragmentation trees, Bioinformatics 30 (2014) i157–i164.
[69] F. Hufsky, K. Dührkop, F. Rasche, M. Chimani, S. Böcker, Fast alignment of fragmentation trees, Bioinformatics 28 (2012) i265–i273.
[70] F. Rasche, K. Scheubert, F. Hufsky, T. Zichner, M. Kai, A. Svatoš, et al., Identifying the unknowns by aligning fragmentation trees, Anal. Chem. 84 (2012) 3417–3426.
[71] K. Dührkop, S. Böcker, Fragmentation trees reloaded, ARXIV (2014).
[72] R. Mathur, P.B. O'Connor, Artifacts in Fourier transform mass spectrometry, Rapid Commun. Mass Spectrom. 23 (2009) 523–529.
[73] M. Cao, K. Fraser, S. Rasmussen, Computational analyses of spectral trees from electrospray multi-stage mass spectrometry to aid metabolite identification, Metabolites 3 (2013) 1036–1050.
[74] L. Ridder, J.J. van der Hooft, S. Verhoeven, Automatic compound annotation from mass spectrometry data using MAGMa. 3 (2014) S0033.
[75] F. Hufsky, K. Scheubert, S. Böcker, New kids on the block: novel informatics methods for natural product discovery, Nat. Prod. Rep. 31 (2014) 807–817.
[76] K. Scheubert, F. Hufsky, S. Böcker, Computational mass spectrometry for small molecules, J. Cheminform. 5 (2013) 12.
[77] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, et al., MassBank: a public repository for sharing mass spectral data for life sciences, J. Mass Spectrom. 45 (2010) 703–714.
[78] L.J. Kangas, T.O. Metz, G. Isaac, B.T. Schrom, B. Ginovska-Pangovska, L. Wang, et al., In silico identification software (ISIS): a machine learning approach to tandem mass spectral identification of lipids, Bioinformatics 28 (2012) 1705–1713.
[79] Y. Wang, G. Kora, B.P. Bowen, C. Pan, MIDAS: a database-searching algorithm for metabolite identification in metabolomics, Anal. Chem. 86 (2014) 9496–9503.
[80] A.W. Hill, R.J. Mortishire-Smith, Automated assignment of high-resolution collisionally activated dissociation mass spectra using a systematic bond disconnection approach, Rapid Commun. Mass Spectrom. 19 (2005) 3111–3118.
[81] A. Ipsen, E.J. Want, J.C. Lindon, T.M. Ebbels, A statistically rigorous test for the identification of parent – fragment pairs in LC-MS datasets, Anal. Chem. 82 (2010) 1766–1778.
[82] J.M. Mitchell, T.W.-M. Fan, A.N. Lane, H.N. Moseley, Development and in silico evaluation of large-scale metabolite identification methods using functional group detection for metabolomics, FGENE 5 (2014).
[83] C. Benecke, T. Gruner, A. Kerber, R. Laue, T. Wieland, MOLecular structure GENeration with MOLGEN, new features and future developments, Fresen. J. Anal. Chem. 359 (1997) 23–32.
[84] M. Meringer, E.L. Schymanski, Small molecule identification with MOLGEN and mass spectrometry, Metabolites 3 (2013) 440–462.
[85] Y. Ma, T. Kind, D. Yang, C. Leon, O. Fiehn, MS2Analyzer: a software for small molecule sub-structure annotations from accurate tandem mass spectra, Anal. Chem. 86 (2014) 10724–10731.
[86] M. Rojas-Cherto, P.T. Kasper, E.L. Willighagen, R. Vreeken, T. Hankemeier, T.H. Reijmers, Elemental composition determination based on MSn, Bioinformatics 27 (2011) 2376–2383.
[87] M. Rojas-Chertó, M. van Vliet, J.E. Peironcely, R. van Doorn, M. Kooyman, T. te Beek, et al., MetiTree: a web application to organize and process high-resolution multi-stage mass spectrometry metabolomics data, Bioinformatics 28 (2012) 2707–2709.
[88] H. Ali, M.I. Ries, J.G. Nijland, P.P. Lankhorst, T. Hankemeier, R.A. Bovenberg, et al., A branched biosynthetic pathway is involved in production of roquefortine and related compounds in Penicillium chrysogenum, PLoS ONE 8 (2013) e65328.
[89] M. Ludwig, F. Hufsky, S. Elshamy, S. Böcker, Finding characteristic sub-structures for metabolite classes. 26 (2012).
[90] L. Ridder, J.J.J. van der Hooft, S. Verhoeven, R.C.H. de Vos, R. van Schaik, J. Vervoort, Sub-structure-based annotation of high-resolution multistage MSn spectral trees, Rapid Commun. Mass Spectrom. 26 (2012) 2461–2471.
[91] K. Morreel, Y. Saeys, O. Dima, F. Lu, Y. Van de Peer, R. Vanholme, et al., Systematic structural characterization of metabolites in arabidopsis via candidate substrate-product pair networks, Plant Cell 26 (2014) 929–945.
[92] M. Heinonen, H.B. Shen, N. Zamboni, J. Rousu, Metabolite identification and molecular fingerprint prediction through machine learning, Bioinformatics 28 (2012) 2333–2341.
[93] R. Mistrik, HighChem mass frontier, (1998).
[94] T. Pluskal, T. Uehara, M. Yanagida, Highly accurate chemical formula prediction tool utilizing high-resolution mass spectra, MS/MS fragmentation, heuristic rules, and isotope pattern matching, Anal. Chem. 84 (2012) 4396–4403.
[95] H. Shen, N. Zamboni, M. Heinonen, J. Rousu, Metabolite identification through machine learning – tackling CASMI challenge using FingerID, Metabolites 3 (2013) 484–505.
[96] F. Allen, A. Pon, M. Wilson, R. Greiner, D. Wishart, CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra, Nucleic Acids Res. (2014) W94–W99.
[97] H. Zhang, S. Singh, V.N. Reinhold, Congruent strategies for carbohydrate sequencing. 2. FragLib: an MS n spectral library, Anal. Chem. 77 (2005) 6263–6270.
[98] H. Ito, Y. Chiba, A. Kameyama, T. Sato, H. Narimatsu, Chapter five – in vitro and in vivo enzymatic syntheses and mass spectrometric database for N-glycans and O-glycans, Methods Enzymol. 478 (2010) 127–149.