

## Uso de Redes Moleculares para Anotações de Compostos em Metabolômica

### *Use of Molecular Networking for Compound Annotation in Metabolomics*

Mariana de Souza Moura,<sup>a</sup>  Barbara Sayuri Bellele,<sup>b</sup>  Lucas Campos Curcino Vieira,<sup>c</sup>  Olívia Moreira Sampaio<sup>a,\*</sup> 

<sup>a</sup> Universidade Federal de Mato Grosso, Instituto de Ciências Exatas e da Terra, Departamento de Química, Campus Cuiabá, CEP 78068-600, Cuiabá-MT, Brasil.

<sup>b</sup> Universidade Federal de Lavras, Instituto de Ciências Naturais, Departamento de Química, CEP 37200-900, Lavras-MG, Brasil;

<sup>c</sup> Universidade Federal de Mato Grosso, Faculdade de Engenharia, CEP 78156-105, Várzea Grande-MT, Brasil.

\*E-mail: [olivia.sampaio@ufmt.br](mailto:olivia.sampaio@ufmt.br)

Recebido em: 01 de Março de 2021

Aceito em: 18 de Maio de 2021

Publicado online: 19 de Novembro de 2021

The metabolomics community and natural product researchers usually use hyphenated analytical techniques to investigate the metabolic profile in complex matrices. Modern analytical instruments used in metabolomic strategies usually generally provide a large amount of data, and these results processing can be optimized by computational tools. The use of classical metabolomics tools, such as consulting the reference databases represents one of the commonly used strategies, however, this approach can lead to a considerable number of unidentified metabolites, in addition to being a time-consuming task if performed manually. In 2012, an important computational strategy for metabolomic study was developed at The Dorrestein Laboratory by Pieter Dorrestein's research group, this approach called molecular networking is generated by the GNPS platform (Global Natural Products Social) based on the hypothesis that structurally related molecules produce similar fragmentation patterns, thus they are integrated in the same group. The analysis of molecular networks makes possible to annotate known molecules based on the identification of potential similarities between all mass spectra within the data set, and to extend annotations to unknown molecules, but with very similar fragmentation patterns. The use of advanced molecular network annotation tools such as the MolNetEnhancer, the network annotation propagation (NAP) and the MS2LDA-MOTIF, contribute to the identification of unannotated ions, being of great value for metabolomic studies. These tools allow the simultaneous analysis of numerous mass spectra obtained in several experiments, reducing the data processing time. This work aims to present classical and advanced computational tools available for research in metabolomics, as well as to discuss the analysis of necessary parameters for this molecular profile approach.

**Keywords:** Metabolomics; GNPS; molecular networking; MS2LDA-Motif; NAP; MolNetEnhancer.

### 1. Estudo Metabolômico por meio de Ferramentas Computacionais

A metabolômica é a mais recente das conhecidas ciências “ômicas”, cujos objetivos são o estudo do metaboloma, o qual representa o conjunto de todas as substâncias de baixo peso molecular presentes em organismos vivos, assim como avaliar e interpretar suas alterações.<sup>1,2</sup> O conhecimento dos compostos químicos presentes em amostras biológicas, torna a metabolômica uma importante estratégia aplicada em diversas áreas de pesquisa, como a química de produtos naturais,<sup>3,4</sup> medicina,<sup>5</sup> farmacologia,<sup>6</sup> entre outras.<sup>7-10</sup>

Os estudos metabolômicos podem ser divididos em duas categorias principais: alvo (do inglês, *targeted*) e não-alvo (do inglês, *untargeted*). As análises alvo destinam-se a um grupo específico de metabólitos de interesse, enquanto as análises não-alvo são aplicadas à detecção do maior número possível de metabólitos em uma determinada amostra.<sup>2,11</sup> Em particular, a investigação de um metaboloma por meio da análise não-alvo demanda um enorme esforço dos pesquisadores devido à grande quantidade e diversidade química dos metabólitos presentes nas amostras. Considerando a complexidade apresentada para as análises não-alvo, a comunidade científica, em especial os pesquisadores da área de metaboloma de produtos naturais, utilizam diversas plataformas e *softwares* para o processamento e análise dos dados experimentais, visando flexibilizar o refinamento das anotações metabólicas de acordo com os tipos e objetivos de cada estudo.<sup>2</sup>

A obtenção de dados experimentais para o estudo de metaboloma pode ser realizada pelas técnicas de espectrometria de massas (EM) e espectroscopia de ressonância magnética nuclear, sendo ambas comumente utilizadas pela comunidade científica. A espectrometria de massas, técnica abordada nessa revisão, apresenta vantagens e desvantagens específicas do seu uso na metabolômica.<sup>12</sup> A principal vantagem da EM é a sensibilidade, já que os espectrômetros

de massa de última geração podem detectar íons nas concentrações de femtomolar ( $10^{-15}$  mol/L) e attomolar ( $10^{-18}$  mol/L). Outra vantagem da EM, é a possibilidade do uso de técnicas hífenadas com cromatografia líquida (CL-EM) ou gasosa (CG-EM), permitindo a separação e identificação de centenas de moléculas individuais em uma única amostra. No entanto, a quantificação das substâncias presentes na amostra representa uma das desvantagens da EM na metabolômica, pois a intensidade dos íons presentes nos espectros de massas é dependente do tipo de preparo da amostra e da estrutura molecular das substâncias.<sup>13</sup>

Considerando a possibilidade do uso de diferentes técnicas para a aquisição de dados no estudo metabolômico, o que implica em uma grande diversidade de informações a serem analisadas, o Grupo de Trabalho de Análise Química (*Chemical Analysis Working Group*) tem elaborado uma padronização das condições experimentais na obtenção dos dados (preparo da amostra, análise experimental, validação do método, pré-processamento dos dados, entre outros) para uma melhor descrição das práticas de análise química relacionadas aos aspectos da investigação de metabolomas.<sup>14</sup>

Dentre as ferramentas utilizadas no processamento dos dados experimentais adquiridos empregando as técnicas CL-EM e CG-EM, destacam-se os algoritmos XCMS, MZmine, OpenMS, MetSign e MetAlign, os quais incluem quatro módulos básicos de funcionamento, como filtragem de ruído e correção de linha de base, detecção de pico e deconvolução espectral, alinhamento e normalização. Para a análise dos dados espectrais, as plataformas e *softwares* assim como os bancos de dados disponíveis mais utilizados para correlação espectral são: MetaboAnalyst2, METLIN, Metabolite detector, MassBank, NIST MS/MS Library, ReSpect, dentre outras.<sup>15-17</sup>

Considerando que os experimentos de metaboloma são realizados em replicatas, torna-se necessário a realização de um pré-processamento dos dados por análises estatísticas uni e multivariadas, visando dimensionar informações relevantes sobre o metaboloma. Dentre os tratamentos estatísticos utilizados nos estudos metabolômicos, destacam-se métodos supervisionados e não supervisionados, como a análise por componentes principais (PCA) e análise discriminante por mínimos quadrados parciais (PLS-DA). Nas análises quimiométricas, a classificação e discriminação de metabólitos responsáveis por distinguir os grupos de amostras são realizadas através da avaliação do conjunto da matriz de dados extraídos na etapa de pré-processamento. A análise de PCA é adequada por fornecer uma visão geral do conjunto de dados e as relações entre grupos, enquanto a PLS-DA é particularmente adequada para dados lineares e colineares, contribuindo para a classificação das amostras.<sup>15,18</sup>

Após o pré-processamento dos dados experimentais, o uso de ferramentas clássicas de metabolômica como a comparação dos resultados com os bancos de dados de espectrometria de massas, NCBI, NIST ou Mass Bank, é uma das abordagens empregadas, contudo, essa

abordagem pode resultar em uma quantidade considerável de metabólitos não identificados.<sup>19,20</sup> Com o objetivo de superar essa adversidade, faz-se necessária a aplicação de ferramentas computacionais de análise, que combinadas com a versatilidade da espectrometria de massas, otimizam o processo de reconhecimento de moléculas em uma matriz complexa.<sup>20-23</sup> Nesse contexto, as redes moleculares (do inglês, *molecular networking*) têm sido propostas como importantes ferramentas para auxiliar e otimizar a investigação de metabolomas.<sup>23</sup> As redes moleculares geradas têm como base a hipótese de que moléculas estruturalmente relacionadas produzem padrões de fragmentação semelhantes, sendo assim integradas em uma mesma rede molecular.<sup>24,25</sup>

O uso de redes moleculares para a análise de espectros de massa de moléculas de baixo peso molecular foi introduzido em 2012 pelo grupo de pesquisa coordenado por Pieter Dorrestein no Laboratório Dorrestein (*The Dorrestein Laboratory*). Dorrestein e colaboradores utilizaram um algoritmo para analisar a produção de metabólitos por um conjunto diversificado de colônias microbianas, o que permitiu o mapeamento da variedade química estrutural observada por meio de um experimento de EM de análise não-alvo.<sup>26</sup> Com a finalidade de tornar o algoritmo usado na produção das redes moleculares acessível à comunidade científica, o *script* deste algoritmo foi convertido em uma plataforma da web, possibilitando assim, a criação de uma infraestrutura comunitária que apoiava o compartilhamento de um banco de dados de espectros de massas. Como resultado, foi criada a plataforma *Global Natural Products Social* (GNPS), a qual teve início em 2014 e foi publicada em 2016 (<http://gnps.ucsd.edu>).<sup>25,27</sup>

A plataforma GNPS permite a construção de redes moleculares que podem ser aplicadas para reunir em famílias moleculares, grupos de compostos com semelhanças encontradas a partir das análises entre os espectros MS/MS.<sup>25</sup> A GNPS possui uma base de dados com acesso aberto para toda comunidade científica compartilhar dados de espectrometria de massa brutos, processados ou identificados, sendo amplamente aplicada em pesquisas de metabolômica,<sup>6,25</sup> o que torna possível a anotação de substâncias conhecidas, pois permite analisar, combinar e comparar as informações espectrais dos usuários ao bancos de dados disponíveis gratuitamente.<sup>25</sup> Dentre os bancos de dados presentes na biblioteca espectral da GNPS estão o Mass Bank, MONA, ReSpect, NIST, METLIN, Atlas, dentre outras.<sup>6,20,25</sup> Por esses motivos, estudos realizados na plataforma GNPS têm levado a anotação de metabólitos em diversas áreas do conhecimento como biomedicina, ciência ambiental, ecologia, ciência forense, microbiologia, química e outros.<sup>5,19,26,28-31</sup>

As redes moleculares também têm sido empregadas no processo de desreplicação, uma abordagem analítica composta por um conjunto de metodologias rápidas de reconhecimento de substâncias presentes em uma matriz complexa de substâncias, com o objetivo de evitar o

reisolamento de compostos já descritos na literatura.<sup>32</sup> A técnica de desreplicação, quando comparada com a abordagem clássica de purificação, isolamento e elucidação estrutural, apresenta inúmeras vantagens devido à sua versatilidade, rapidez e a economia de amostra, materiais e solventes.<sup>19,32,33</sup> The dereplication of crude extracts performed prior to isolation work is of crucial importance for avoiding the tedious isolation of known constituents. In this respect, chemical screening strategies based on hyphenated techniques such as liquid chromatography - ultraviolet photodiode array detection, liquid chromatography-mass spectrometry, liquid chromatography tandem mass spectrometry and liquid chromatography - nuclear magnetic resonance (LC-NMR) Neste contexto, diversos trabalhos científicos têm sido reportados mostrando o emprego de redes moleculares para a anotação de substâncias bioativas em amostras de plantas e microrganismos, baseado na identificação de potenciais semelhanças entre os espectros de massas obtidos das amostras biológicas com o banco de dados espectrais disponíveis.<sup>5,23,28,34-40</sup>

A plataforma GNPS também disponibiliza em seu domínio ferramentas como a *Feature-Based Molecular Networking* (FBMN), um método avançado e integrado a rede molecular que analisa a abundância relativa dos íons (área do pico) para análises estatísticas, fornecendo suporte para a resolução de isômeros dentro da rede molecular e também colabora nos ensaios com analisadores do tipo *ion mobility* (mobilidade de íons).<sup>41</sup> Para a investigação de espectros sem anotação, análises *in silico* foram desenvolvidas para correlacionar os espectros de MS/MS desconhecidos com estruturas químicas putativas. Dentre os métodos mais utilizados neste estudo, destacam-se a propagação de anotação de rede (NAP - *Network Annotation Propagation*), ferramenta que utiliza as redes moleculares de interação para propagar informações de equivalência de bibliotecas espectrais.<sup>42,43</sup>

A análise das redes moleculares combinada aos métodos *in silico* como o NAP e a ferramenta MS2LDA a qual permite a descoberta de subestruturas a partir dos picos de fragmentos comuns e também de perdas neutras, possibilita uma maior eficiência na análise MolNetEnhancer, uma refinada abordagem usada para a anotação, visualização e descoberta da diversidade estrutural química dentro das famílias moleculares.<sup>44</sup> O MolNetEnhancer tem mostrado ser uma ferramenta útil no estudo de metaboloma por permitir a anotação de famílias, classes e subclasses moleculares, assim como sutis diferenças estruturais entre membros de uma mesma família molecular (FM).<sup>45-48</sup> especially prenatal and perinatal events. Newborn screening dried blood spot (DBS

## 2. Redes Moleculares

As redes moleculares representam uma estratégia computacional empregada na interpretação de misturas

complexas de dados adquiridos a partir da EM, auxiliando na visualização da diversidade química presente em uma determinada amostra.<sup>25</sup> Esta técnica utiliza algoritmos matemáticos para organizar e interpretar os dados de EM, agrupando-os com base no perfil de fragmentação referente a espectrometria de massas sequencial, associando as substâncias estruturalmente semelhantes como sendo de uma mesma FM.

A plataforma GNPS foi desenvolvida para processar dados experimentais MS/MS, e muitas das ferramentas de análise foram destinadas a conjuntos de dados com pelo menos centenas de espectros. Embora seja possível analisar uma pequena quantidade de dados, determinados métodos de análise, como as redes moleculares, não fornecerão bons resultados em escalas tão pequenas. Para conjuntos de dados pequenos, os desenvolvedores da plataforma recomendam o uso da ferramenta de análise MASST.<sup>27</sup> Para dados obtidos em um experimento completo de CL-EM, é possível realizar todas as análises disponíveis na plataforma GNPS. Inicialmente, os arquivos de saída adquiridos por meio da EM (.raw, .RAW, .wiff, .d e .lcd) devem ser convertidos nas extensões .mzXML, .mzML e .mgf, as quais são suportadas pela plataforma.<sup>27</sup> O MSConvert é um programa gratuito que pode ser utilizado na conversão de dados obtidos em experimentos utilizando equipamentos das marcas AB Sciex, Agilent, ThermoFisher, Waters e Shimadzu. Para dados obtidos em equipamentos da Bruker, a conversão pode ser realizada no CompassXport, uma ferramenta disponível no programa DataAnalysis.<sup>49</sup> pathway-targeted tandem MS (MS/MS Então, os dados convertidos são transferidos de modo eficiente para a plataforma GNPS por meio do protocolo de transferência de arquivos – FTP, utilizando dentre os servidores mais empregados, WinSCP, CoreFTP e CoffeeCup Free.

Para a realização do fluxo de análises na plataforma GNPS, é necessário adicionar informações como o tipo de espectrômetro de massas utilizado na aquisição dos dados (alta ou baixa resolução) e a tolerância de massa do íon precursor para o espectro de consenso, sendo recomendado  $\pm 0,02$  Da e  $\pm 2,0$  Da para equipamentos de alta e baixa resolução, respectivamente. A tolerância de massa de íons de fragmentos também faz parte das opções básicas de parâmetros, sendo os valores recomendados para equipamento com alta resolução  $\pm 0,02$  Da e baixa resolução  $\pm 0,5$  Da.<sup>5</sup> Em relação aos parâmetros avançados, são definidos valores mínimos de íons produtos emparelhados, que consiste na pontuação mínima de íons fragmentos comuns compartilhados entre dois espectros de consenso (o valor padrão é 6) e o número mínimo de espectros de consenso necessários para ser considerado na rede molecular (dependente do tamanho do conjunto de dados).<sup>25,50</sup>

As estruturas químicas de um perfil metabólico determinam como as mesmas serão agrupadas, pois moléculas estruturalmente semelhantes produzem espectros de fragmentação com características comuns.<sup>29</sup> As redes

moleculares organizam esses dados e relacionam os compostos, mapeando as correlações detectadas na análise dos dados experimentais de MS/MS.<sup>24</sup>requiring to be sorted, curated, and statistically analyzed by dedicated software. Data-dependent acquisition mode including inclusion and exclusion rules for tandem mass spectrometry is routinely used to perform such analyses. While acquisition parameters are well described for proteomics, no general rule is currently available to generate reliable metabolomic data for molecular networking analysis on the Global Natural Product Social Molecular Networking platform (GNPS). Assim, as substâncias estruturalmente relacionadas em uma rede molecular são descritas como FMs, diferindo por variações estruturais decorridas de transformações químicas, como por exemplo, processos de glicosilação, alquilação, oxidação e redução, dentre outros.<sup>25,29</sup>

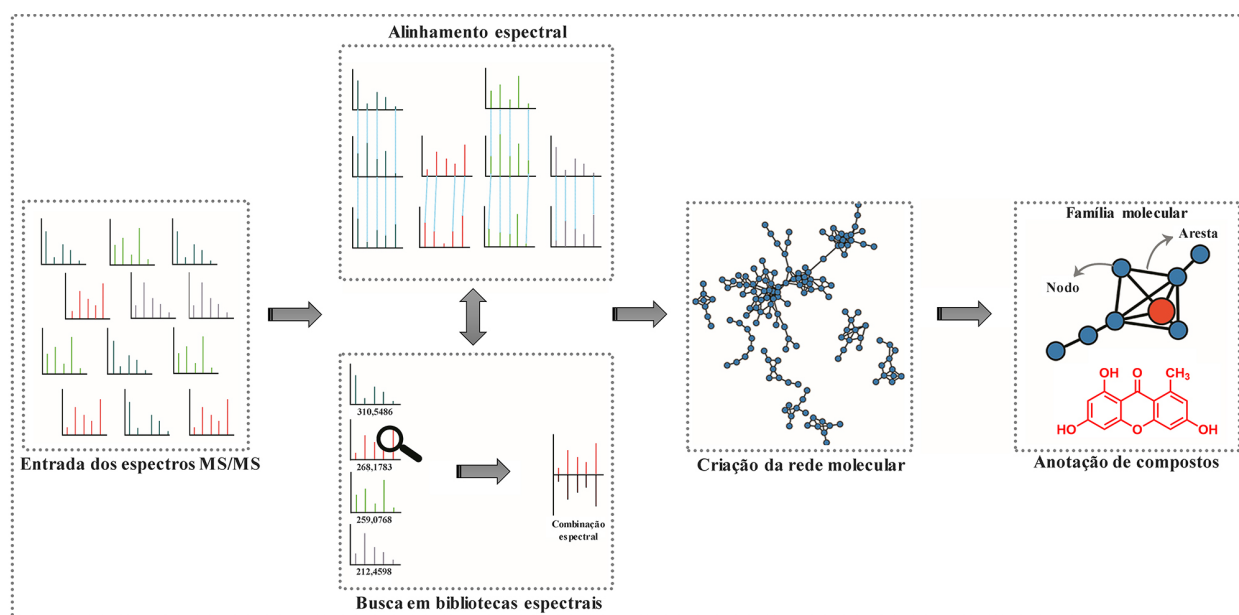
A Figura 1 ilustra o processo de construção das redes moleculares na plataforma GNPS, onde são realizadas comparações entre todos os espectros de consenso pertencentes a matriz de análise, no intuito de encontrar semelhanças entre os dados, agrupando os espectros de acordo com seus padrões de fragmentação. Conjuntamente a esse processo, os espectros são também correlacionados às bibliotecas espectrais para encontrar semelhanças entre os espectros em análise com substâncias de referência. Os resultados dessa análise podem ser visualizados em gráficos de similaridade espectral ou redes moleculares.<sup>27</sup>

Nas redes moleculares, cada nodo representa um espectro de consenso e as arestas que ligam os nodos indicam o grau de similaridade entre eles.<sup>5,20,25</sup> Durante o emparelhamento dos espectros de EM, para cada espectro de consenso em um determinado conjunto de dados são atribuídos valores nos eixos x, y e z, e então ocorre a correlação com todos os outros espectros.<sup>51</sup> O valor que

representa a similaridade espectral MS/MS entre dois nodos em uma rede molecular é o cosseno, considerado o filtro mais básico de pontuação de alinhamento. Os valores de cosseno podem variar de 0 a 1, sendo que o valor numérico 1 representa espectros idênticos e o valor 0 indica que não há nenhuma similaridade entre os espectros. São considerados alinhamentos entre espectros semelhantes quando o valor de cosseno entre os vetores dos espectros é maior ou igual a 0,7.<sup>27</sup> Algoritmos de conformidade de alinhamento de todos os espectros de fragmentação devem utilizar filtros como o valor de consenso (cosseno espectral), K vizinho mais próximos (um algoritmo simples que armazena todos os casos disponíveis e classifica novos casos com base em uma medida de similaridade), teste de classificação e tamanho do componente conectado. Esses filtros reduzem o número de alinhamentos falsos e conexões falso positivas, produzindo redes moleculares confiáveis.<sup>20,25,27</sup>

Os dados de correlação espectral obtidos usando a plataforma GNPS podem ser visualizados utilizando o software Cytoscape, uma interessante ferramenta para a representação das redes moleculares, onde é possível o ajuste dos nodos e arestas a fim de facilitar a interpretação dos resultados, como por exemplo, melhorando a apresentação de atributos como espessura das arestas (proporcional à pontuação do cosseno), tamanho do nodo (correlacionado à intensidade do íon precursor), cor do nodo (usado para mapear vários metadados associados às amostras).<sup>5</sup>

A análise das redes moleculares baseada apenas na distribuição das FMs por agrupamento é classificada como análise básica. No entanto, é possível o emprego de ferramentas avançadas como o NAP, MS2LDA e MolNetEnhancer para uma melhor compreensão e anotação dos espectros relacionados nas redes moleculares,



**Figura 1.** Representação esquemática do processo de criação de uma rede molecular a partir de espectros de massas MS/MS, com alinhamento espectral e buscas nas bases espectrais para anotação de famílias moleculares e identificação de metabólitos em amostras complexas

possibilitando a identificação de um maior número de compostos em uma amostra complexa.

### 3. Propagação de Anotação de Rede - NAP

---

A investigação de um perfil metabólico analisando apenas as massas moleculares dos constituintes da amostra é uma abordagem insuficiente para uma confiável atribuição das estruturas químicas dos metabólitos. As informações estruturais podem ser mais precisas pelo uso de fragmentação, ou seja, explorando os espectros de MS/MS.<sup>30</sup> No cálculo de uma rede molecular, uma grande porcentagem dos espectros de consenso permanece sem anotação, devido ao limitado domínio de substâncias químicas de baixo peso molecular conhecidas presentes nas bibliotecas de fragmentação disponíveis na base de dados da GNPS.<sup>52</sup>

As taxas de anotação de espectros podem ser aumentadas com o uso de métodos *in silico*, os quais combinam os espectros de fragmentação não anotados com estruturas químicas putativas. Os cálculos *in silico* funcionam como um modelo preditor de fragmentação, expandindo a anotação primária para outras moléculas sem nenhum espectro de referência depositado nos bancos de dados disponíveis.<sup>53</sup> Usando o alinhamento espectral para criar redes moleculares, as moléculas podem ser agrupadas em FMs (análogos estruturais) e as anotações de compostos conhecidos podem ser propagadas para espectros conectados (vizinhos) na rede.<sup>43</sup>

Dentre as ferramentas *in silico* empregadas na anotação de espectros, destacam-se o NAP, MetFrag, MetFusion, SIRIUS, CSI: FingerID, MS-Finder e Dereplicator, as quais podem ser integradas às análises de redes moleculares na plataforma GNPS.<sup>27,43,52</sup> Esses métodos permitem ampliar as informações estruturais obtidas dos dados de espectrometria de massas, aumentando as taxas de anotação das FM nas redes moleculares.<sup>52</sup>

O método NAP utiliza as redes moleculares de interação para propagar informações de bibliotecas espectrais, com finalidade de melhorar a anotação dos compostos candidatos. Essa ferramenta de análise faz a propagação automatizada por meio do MetFrag, que é uma abordagem combinatória desenvolvida para realizar a fragmentação *in silico* de estruturas moleculares, gerando várias etapas de pesquisa. Então, é realizado o cálculo da pontuação que qualifica os compostos candidatos aos espectros de MS/MS analisados.<sup>54</sup>

Ademais, são realizadas buscas em bibliotecas estruturais de produtos naturais para comparação com os dados espectrais da amostra, como por exemplo, os bancos de dados Super Natural II e DNP (*dictionary of natural products*), classificando as melhores estruturas candidatas para cada nodo.<sup>52</sup> Concomitantemente, NAP faz uma nova classificação desses candidatos com base na propagação da similaridade estrutural dos nodos vizinhos, assim obtendo

uma melhor qualificação dos compostos candidatos em relação aos espectros analisados.<sup>43</sup>

### 4. Análise não Supervisionada de Subestruturas - MS2LDA

---

Em 2017, Van Der Hooft e colaboradores, apresentaram o MS2LDA, um método de análise não supervisionado para a extração de padrões comuns de fragmentos de massas e perdas neutras.<sup>55</sup> Este método busca encontrar padrões de fragmentação de massas relacionados às subestruturas moleculares encontrados nos dados obtidos em análises de espectrometria de massas.<sup>30</sup>

O MS2LDA decompõe cada molécula em um ou mais grupos de fragmentos de massas e perdas neutras chamados de Mass2Motifs, possibilitando a exploração e o agrupamento molecular mais eficiente dos dados espectrais. Os fragmentos e perdas neutras presentes em múltiplos espectros frequentemente correspondem às mesmas subestruturas químicas. Os espectros que compartilham de um mesmo Mass2Motif podem ser agrupados, mesmo que o Mass2Motif, corresponda a uma pequena parte dos espectros, e a correlação entre vários Mass2Motifs, possibilita que os dados espectrais podem pertencer a várias FMs.<sup>55</sup>

Então, as informações químicas estruturais obtidas dos espectros experimentais, como os padrões de fragmentação MS/MS, são comparadas diretamente com dados de referência das bibliotecas espectrais, correlacionando os mesmos com subestruturas de potenciais moléculas candidatas. Em comparação com a anotação estrutural de moléculas inteiras, o uso dos Mass2Motifs fornece informações mais diretas e menos complexas, pois estes representam subestruturas menores.<sup>30</sup>

O emprego destes algoritmos de agrupamento baseados em similaridade espectral MS/MS são fundamentais para o estudo de metabolômica não-alvo e estratégias de desreplacação. Dentre as vantagens do MS2LDA, é possível destacar que esta ferramenta de análise contribui para a anotação de moléculas as quais não existem espectros de referência nas bases de dados e ainda corrobora no estudo das relações bioquímicas existentes entre elas, fornecendo uma visão mais ampla da diversidade molecular em amostras complexas.<sup>30,31</sup> O método MS2LDA também pode ser utilizado como uma entrada opcional ao MolNetEnhancer, uma ferramenta complementar de análise das redes moleculares, que classifica as FMs existentes e anota as subclasses de estruturas para a formação das superclasses estruturais.<sup>44</sup>

### 5. MolNetEnhancer - Análise Avançada para Anotação de Metaboloma

---

O MolNetEnhancer representa um fluxo de análises o qual permite associar os dados das redes moleculares

com resultados obtidos em outras ferramentas de análise como o NAP e o MS2LDA, dentre outras. Esta abordagem integrada, associada a classificação química automatizada empregando o ClassyFire, possibilita uma melhor compreensão das informações químicas estruturais presentes nas redes moleculares, otimizando a identificação de metabólitos em misturas complexas, o que torna o MolNetEnhancer uma ferramenta importante na química de produtos naturais.<sup>31,46</sup>

Esse método de análise baseia-se nas informações de análogos moleculares, os quais são estruturalmente semelhantes e geram espectros de fragmentação que podem ser agrupados de acordo com as classes químicas presentes na amostra biológica. O agrupamento dos metabólitos por classes químicas existentes em um extrato de origem biológica facilita a identificação de íons não anotados, melhorando a anotação das substâncias e a capacidade de fornecer explicações para o comportamento biológico da matriz em relação a estrutura e atividade biológica.<sup>44,56</sup>

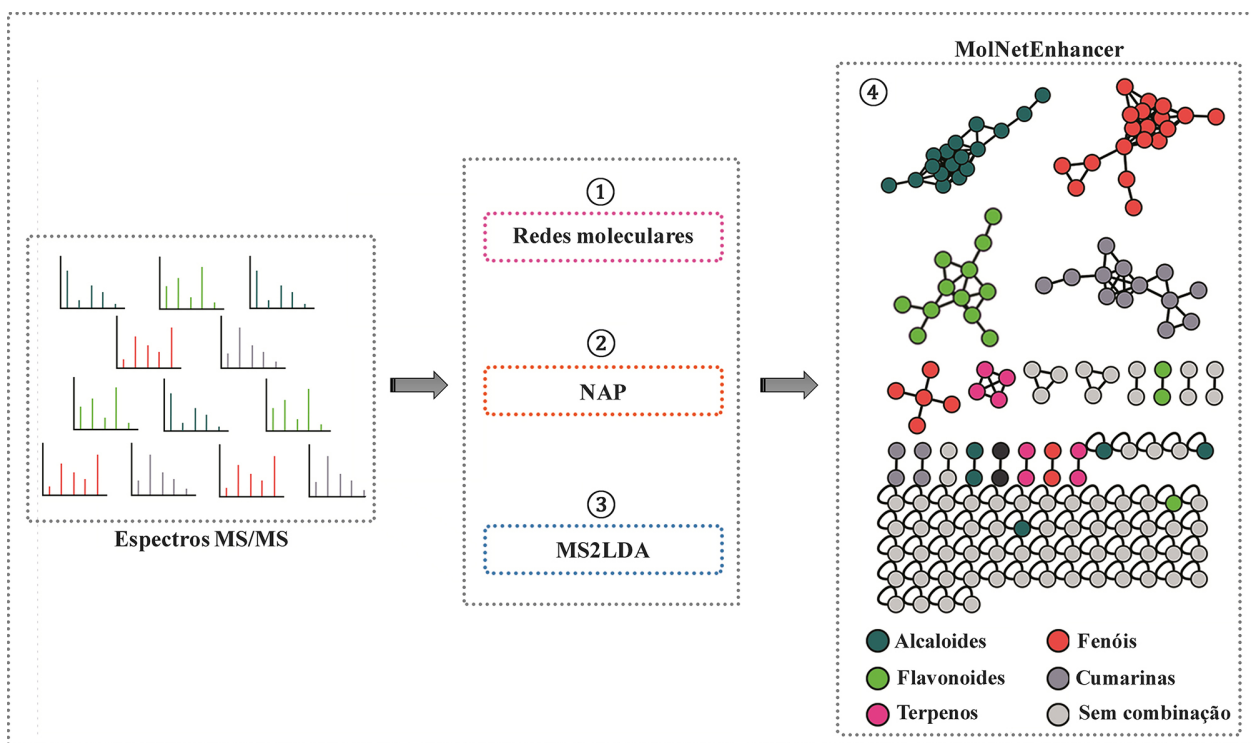
O modo de análise empregando MolNetEnhancer segue 4 etapas, iniciando pela análise dos espectros MS/MS obtidos da amostra em estudo para a elaboração das redes moleculares na plataforma GNPS (Figura 2). Esta primeira etapa é obrigatória para que o fluxo de análise MolNetEnhancer seja realizado. As etapas 2 e 3 estão relacionadas às anotações *in silico* NAP e não supervisionada de subestruturas MS2LDA, e são de uso facultativo, não sendo necessário às suas aplicações para que o MolNetEnhancer seja realizado. Então, o MolNetEnhancer é usado na etapa 4, mapeando os dados obtidos com as

ferramentas de análise empregadas (etapas 1-3), assim combinando estas informações em função de classes químicas em toda a rede molecular, com detalhes para as subestruturas dentro das famílias moleculares.<sup>44</sup>

Portanto, considerando a capacidade organizacional dos dados espectrais pelas redes moleculares e MolNetEnhancer, é possível o conhecimento de uma ampla variedade de classes químicas e as diferenças subestruturais presentes em uma amostra complexa de metabólitos. Estes métodos permitem ainda a identificação adequada de FMs que podem conter metabólitos secundários bioativos, o que torna as ferramentas de análise presentes na plataforma GNPS úteis para auxiliar em trabalhos com descrição de perfil químico.

## 6. Considerações Finais

A expansão dos estudos metabolômicos em diversas áreas de pesquisa tem levado ao desenvolvimento de plataformas e *softwares* para auxiliar o pré-processamento e análise dos dados espectrais obtidos de amostras complexas de metabólitos. As ferramentas de bioinformática abordadas neste trabalho, são empregadas na anotação de moléculas conhecidas e cadastradas em banco de dados, e também podem ser utilizadas para relacionar moléculas semelhantes e identificar padrões de fragmentação, auxiliando na identificação de moléculas ainda sem anotação. Outra importante vantagem no uso destas ferramentas de análise computacional é a possibilidade de investigar diversos espectros simultaneamente, o que reduz o tempo de



**Figura 2.** Representação esquemática geral do fluxo de análises para o cálculo do MolNetEnhancer (4) a partir dos dados espectrais MS/MS, elaboração das redes moleculares (1), propagação de anotação de rede - NAP (2) e análise não supervisionada MS2LDA (3)

processamento dos dados, economiza recursos e leva a um aumento na qualidade dos resultados.

A plataforma GNPS está sob constante atualização, nas ferramentas de análise realizada pelos seus criadores, assim como o enriquecimento do banco de dados espectrais realizado pelos usuários. Contudo, como a rede molecular feita na GNPS usa dados MS/MS, está sujeita aos mesmos desafios enfrentados em experimentos de espectrometria de massas, como a baixa relação sinal-ruído, separação insuficiente de analitos e formato inadequados de picos.<sup>27,57</sup> Além disso, apenas informações qualitativas são obtidas a partir das redes moleculares clássicas, pois as análises são realizadas considerando somente dados MS/MS. Portanto, o estudo computacional de metabolomas usando a GNPS ainda enfrenta grandes desafios para a anotação completa de um metaboloma.

## Referências Bibliográficas

- Carriot, N.; Paix, B.; Greff, S.; Viguier, B.; Briand, J. F.; Culioli, G.; Integration of LC/MS-based molecular networking and classical phytochemical approach allows in-depth annotation of the metabolome of non-model organisms - The case study of the brown seaweed *Taonia atomaria*. *Talanta* **2021**, *225*, 121925. [[CrossRef](#)]
- Pilon, A.; Selegato, D.; Fernandes, R.; Bueno, P.; Pinho, D.; Carnevale Neto, F.; Freire, R.; Castro-Gamboa, I.; Bolzani, V.; Lopes, N.; Metabolômica de plantas: Métodos e desafios. *Química Nova* **2020**, *43*, 329. [[CrossRef](#)]
- de Souza, L. P.; Naake, T.; Tohge, T.; Fernie, A. R. From chromatogram to analyte to metabolite.; How to pick horses for courses from the massive web resources for mass spectral plant metabolomics. *GigaScience* **2017**, *6*, 1. [[CrossRef](#)] [[PubMed](#)]
- Urban, L.; Aarouf, J.; Bidel, L. P. R.; Assessing the Effects of Water Deficit on Photosynthesis Using Parameters Derived from Measurements of Leaf Gas Exchange and of Chlorophyll a Fluorescence. *Frontiers in Plant Science* **2017**, *8*, 2068. [[CrossRef](#)]
- Quinn, R. A.; Nothias, L. F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C.; Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends in Pharmacological Sciences* **2017**, *38*, 143. [[CrossRef](#)] [[PubMed](#)]
- Nothias, L. F.; Nothias-Esposito, M.; Da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C.; Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation. *Journal of Natural Products* **2018**, *81*, 758. [[CrossRef](#)] [[PubMed](#)]
- McCall, L. I.; Callewaert, C.; Zhu, Q.; Song, S. J.; Bouslimani, A.; Minich, J. J.; Ernst, M.; Ruiz-Calderon, J. F.; Cavallin, H.; Pereira, H. S.; Novoselac, A.; Hernandez, J.; Rios, R.; Branch, O. L. H.; Blaser, M. J.; Paulino, L. C.; Dorrestein, P. C.; Knight, R.; Dominguez-Bello, M. G.; Home chemical and microbial transitions across urbanization. *Nature Microbiology* **2020**, *5*, 108. [[CrossRef](#)] [[PubMed](#)]
- Sedio, B. E.; Recent breakthroughs in metabolomics promise to reveal the cryptic chemical traits that mediate plant community composition, character evolution and lineage diversification. *New Phytologist* **2017**, *214*, 952. [[CrossRef](#)] [[PubMed](#)]
- Wang, J. H.; Byun, J.; Pennathur, S.; Analytical approaches to metabolomics and applications to systems biology. *Seminars in Nephrology* **2010**, *30*, 500. [[CrossRef](#)] [[PubMed](#)]
- Olivier-Jimenez, D.; Chollet-Krugler, M.; Rondeau, D.; Beniddir, M. A.; Ferron, S.; Delhay, T.; Allard, P. M.; Wolfender, J. L.; Sipman, H. J. M.; Lücking, R.; Boustie, J.; Le Pogam, P.; A database of high-resolution MS/MS spectra for lichen metabolites. *Scientific Data* **2019**, *6*, 1. [[CrossRef](#)] [[PubMed](#)]
- Gauglitz, J. M.; Aceves, C. M.; Aksenov, A. A.; Aleti, G.; Almaliti, J.; Bouslimani, A.; Brown, E. A.; Campeau, A.; Caraballo-Rodríguez, A. M.; Chaar, R.; da Silva, R. R.; Demko, A. M.; Di Ottavio, F.; Elijah, E.; Ernst, M.; Ferguson, L. P.; Holmes, X.; Jarmusch, A. K.; Jiang, L.; Kang, K. Bin; Koester, I.; Kwan, B.; Li, J.; Li, Y.; Melnik, A. V.; Molina-Santiago, C.; Ni, B.; Oom, A. L.; Panitchpakdi, M. W.; Petras, D.; Quinn, R.; Sikora, N.; Spengler, K.; Teke, B.; Tripathi, A.; Ul-Hasan, S.; van der Hooff, J. J. J.; Vargas, F.; Vrbanac, A.; Vu, A. Q.; Wang, S. C.; Weldon, K.; Wilson, K.; Wozniak, J. M.; Yoon, M.; Bandeira, N.; Dorrestein, P. C.; Untargeted mass spectrometry-based metabolomics approach unveils molecular changes in raw and processed foods and beverages. *Food Chemistry* **2020**, *302*, 125290. [[CrossRef](#)]
- Patti, G. J.; Yanes, O.; Siuzdak, G.; Innovation: Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* **2012**, *13*, 263. [[CrossRef](#)] [[PubMed](#)]
- Veenstra, T. D.; Metabolomics: the final frontier? *Genome Medicine* **2012**, *4*, 40. [[CrossRef](#)]
- Fiehn, O.; Robertson, D.; Griffin, J.; vab der Werf, M.; Nikolau, B.; Morrison, N.; Sumner, L. W.; Goodacre, R.; Hardy, N. W.; Taylor, C.; Fostel, J.; Kristal, B.; Kaddurah-Daouk, R.; Mendes, P.; van Ommen, B.; Lindon, J. C.; Sansone, S. A.; The metabolomics standards initiative (MSI). *Metabolomics* **2007**, *3*, 175. [[CrossRef](#)]
- Yi, L.; Dong, N.; Yun, Y.; Deng, B.; Ren, D.; Liu, S.; Liang, Y.; Chemometric methods in data processing of mass spectrometry-based metabolomics: A review. *Analytica Chimica Acta* **2016**, *914*, 17. [[CrossRef](#)] [[PubMed](#)]
- Coble, J. B.; Fraga, C. G.; Comparative evaluation of preprocessing freeware on chromatography/mass spectrometry data for signature discovery. *Journal of Chromatography A* **2014**, *1358*, 155. [[CrossRef](#)] [[PubMed](#)]
- Alonso, A.; Marsal, S.; Julià, A.; Analytical methods in untargeted metabolomics: State of the art in 2015. *Frontiers in Bioengineering and Biotechnology* **2015**, *3*, 23. [[CrossRef](#)] [[PubMed](#)]
- Belinato, J.; Bazioli, J.; Sussulini, A.; Augusto, F.; Fill, T.; Metabolômica microbiana: Inovações e aplicações. *Química Nova* **2019**, *42*, 546. [[CrossRef](#)]
- Gaudêncio, S. P.; Pereira, F.; Dereplication: Racing to speed up the natural products discovery process. *Natural Product Reports* **2015**, *32*, 779. [[CrossRef](#)]

20. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C.; Molecular networking as a dereplication strategy. *Journal of Natural Products* **2013**, *76*, 1686. [[CrossRef](#)] [[PubMed](#)]
21. Cech, N. B.; Enke, C. G.; Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrometry Reviews* **2001**, *20*, 362. [[CrossRef](#)]
22. Aksenov, A. A.; Da Silva, R.; Knight, R.; Lopes, N. P.; Dorrestein, P. C.; Global chemical analysis of biology by mass spectrometry. *Nature Reviews Chemistry* **2017**, *1*, 1. [[CrossRef](#)]
23. Bouslimani, A.; Porto, C.; Rath, C. M.; Wang, M.; Guo, Y.; Gonzalez, A.; Berg-Lyon, D.; Ackermann, G.; Christensen, G. J. M.; Nakatsuji, T.; Zhang, L.; Borkowski, A. W.; Meehan, M. J.; Dorrestein, K.; Gallo, R. L.; Bandeira, N.; Knight, R.; Alexandrov, T.; Dorrestein, P. C.; Molecular cartography of the human skin surface in 3D. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, E2120. [[CrossRef](#)]
24. Olivon, F.; Roussi, F.; Litaudon, M.; Touboul, D.; Optimized experimental workflow for tandem mass spectrometry molecular networking in metabolomics. *Analytical and Bioanalytical Chemistry* **2017**, *409*, 5767. [[CrossRef](#)] [[PubMed](#)]
25. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya P, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N.; Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology* **2016**, *34*, 828. [[CrossRef](#)]
26. Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C.; Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences* **2012**, *109*, E1743. [[CrossRef](#)] [[PubMed](#)]
27. Aron, A. T.; Gentry, E. C.; McPhail, K. L.; Nothias, L. F.; Nothias-Esposito, M.; Bouslimani, A.; Petras, D.; Gauglitz, J. M.; Sikora, N.; Vargas, F.; van der Hooft, J. J. J.; Ernst, M.; Kang, K. Bin; Aceves, C. M.; Caraballo-Rodríguez, A. M.; Koester, I.; Weldon, K. C.; Bertrand, S.; Roullier, C.; Sun, K.; Tehan, R. M.; Boya P, C. A.; Christian, M. H.; Gutiérrez, M.; Ulloa, A. M.; Tejada Mora, J. A.; Mojica-Flores, R.; Lakey-Beitia, J.; Vásquez-Chaves, V.; Zhang, Y.; Calderón, A. I.; Tayler, N.; Keyzers, R. A.; Tugizimana, F.; Ndlovu, N.; Aksenov, A. A.; Jarmusch, A. K.; Schmid, R.; Truman, A. W.; Bandeira, N.; Wang, M.; Dorrestein, P. C.; Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols* **2020**, *15*, 1954. [[CrossRef](#)] [[PubMed](#)]
28. Raheem, D. J.; Tawfike, A. F.; Abdelmohsen, U. R.; Edrada-Ebel, R. A.; Fitzsimmons-Thoss, V.; Application of metabolomics and molecular networking in investigating the chemical profile and antitrypanosomal activity of British bluebells (*Hyacinthoides non-scripta*). *Scientific Reports* **2019**, *9*, 1. [[CrossRef](#)] [[PubMed](#)]
29. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodríguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kamenik, Z.; Kang, K. Bin; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin H., C.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyne, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; van der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C.; Feature-based molecular networking in the GNPS analysis environment. *Nature Methods* **2020**, *17*, 905. [[CrossRef](#)]
30. Rogers, S.; Ong, C. W.; Wandy, J.; Ernst, M.; Ridder, L.; van der Hooft, J. J. J.; Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. *Faraday Discussions* **2019**, *218*, 284. [[CrossRef](#)]
31. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S.; ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics* **2016**, *8*, 1. [[CrossRef](#)]
32. Wolfender, J. L.; Queiroz, E. F.; Hostettmann, K.; The importance of hyphenated techniques in the discovery of new lead compounds from nature. *Expert Opinion on Drug Discovery* **2006**, *1*, 237. [[CrossRef](#)]
33. Li, P.; Su, W.; Xie, C.; Zeng, X.; Peng, W.; Liu, M.; Rapid Identification and Simultaneous Quantification of Multiple



- Constituents in Nao-Shuan-Tong Capsule by Ultra-Fast Liquid Chromatography/Diode-Array Detector/Quadrupole Time-of-Flight Tandem Mass Spectrometry. *Journal of Chromatographic Science* **2015**, *53*, 886. [[CrossRef](#)]
34. Fox Ramos, A. E.; Evanno, L.; Poupon, E.; Champy, P.; Beniddir, M. A.; Natural products targeting strategies involving molecular networking: Different manners, one goal. *Natural Product Reports* **2019**, *36*, 960. [[CrossRef](#)]
  35. Kalinski, J.-C.; Waterworth, S.; Siwe Noundou, X.; Jiwaji, M.; Parker-Nance, S.; Krause, R.; McPhail, K.; Dorrington, R.; Molecular Networking Reveals Two Distinct Chemotypes in Pyrroloiminoquinone-Producing *Tsitsikamma favus* Sponges. *Marine Drugs* **2019**, *17*, 60. [[CrossRef](#)]
  36. Teta, R.; Della Sala, G.; Esposito, G.; Via, C. W.; Mazzocchi, C.; Piccoli, C.; Bertin, M. J.; Costantino, V.; Mangoni, A.; A joint molecular networking study of a *Smenospongia* sponge and a cyanobacterial bloom revealed new antiproliferative chlorinated polyketides. *Organic Chemistry Frontiers* **2019**, *6*, 1762. [[CrossRef](#)]
  37. Vizcaino, M. I.; Engel, P.; Trautman, E.; Crawford, J. M.; Comparative metabolomics and structural characterizations illuminate colibactin pathway-dependent small molecules. *Journal of the American Chemical Society* **2014**, *136*, 9244. [[CrossRef](#)] [[PubMed](#)]
  38. Nguyen, D. D.; Melnik, A. V.; Koyama, N.; Lu, X.; Schorn, M.; Fang, J.; Aguinaldo, K.; Lincecum, T. L.; Ghequire, M. G. K.; Carrion, V. J.; Cheng, T. L.; Duggan, B. M.; Malone, J. G.; Mauchline, T. H.; Sanchez, L. M.; Kilpatrick, A. M.; Raaijmakers, J. M.; De Mot, R.; Moore, B. S.; Medema, M. H.; Dorrestein, P. C.; Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poeamide B and the bananamides. *Nature Microbiology* **2016**, *2*, 1. [[CrossRef](#)]
  39. Melnik, A. V.; Da Silva, R. R.; Hyde, E. R.; Aksenov, A. A.; Vargas, F.; Bouslimani, A.; Protsyuk, I.; Jarmusch, A. K.; Tripathi, A.; Alexandrov, T.; Knight, R.; Dorrestein, P. C.; Coupling Targeted and Untargeted Mass Spectrometry for Metabolome-Microbiome-Wide Association Studies of Human Fecal Samples. *Analytical Chemistry* **2017**, *89*, 7549. [[CrossRef](#)]
  40. Floros, D. J.; Jensen, P. R.; Dorrestein, P. C.; Koyama, N.; A metabolomics guided exploration of marine natural product chemical space. *Metabolomics* **2016**, *12*, 1. [[CrossRef](#)]
  41. Phelan, V. V.; Em *Computational Methods and Data Analysis for Metabolomics. Methods in Molecular Biology*; Li, S., org.; Methods in Humana Press Inc.: New York, 2020, cap. 13.
  42. Mohimani, H.; Gurevich, A.; Mikheenko, A.; Garg, N.; Nothias, L. F.; Ninomiya, A.; Takada, K.; Dorrestein, P. C.; Pevzner, P. A.; Dereplication of peptidic natural products through database search of mass spectra. *Nature Chemical Biology* **2017**, *13*, 30. [[CrossRef](#)] [[PubMed](#)]
  43. da Silva, R. R.; Wang, M.; Nothias, L.-F.; van der Hooft, J. J. J.; Caraballo-Rodríguez, A. M.; Fox, E.; Balunas, M. J.; Klassen, J. L.; Lopes, N. P.; Dorrestein, P. C.; Propagating annotations of molecular networks using in silico fragmentation. *PLOS Computational Biology* **2018**, *14*, e1006089. [[CrossRef](#)]
  44. Ernst, M.; Kang, K. Bin; Caraballo-Rodríguez, A. M.; Nothias, L.-F.; Wandy, J.; Chen, C.; Wang, M.; Rogers, S.; Medema, M. H.; Dorrestein, P. C.; van der Hooft, J. J. J.; MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites* **2019**, *9*, 144. [[CrossRef](#)]
  45. Courraud, J.; Ernst, M.; Svane Laursen, S.; Hougaard, D. M.; Cohen, A. S.; Studying Autism Using Untargeted Metabolomics in Newborn Screening Samples. *Journal of molecular neuroscience : MN* **2021**. [[CrossRef](#)] [[PubMed](#)]
  46. Cheng, X.; Zhao, W.; Dong, W.; Le, G.; Chemical Space Charting of Different Parts of *Inula nervosa* Wall.: Upregulation of Expression of Nrf2 and Correlated Antioxidants Enzymes. *Molecules* **2020**, *25*, 4789. [[CrossRef](#)]
  47. Wang, Y.-K.; Xiao, X.-R.; Zhou, Z.-M.; Xiao, Y.; Zhu, W.-F.; Liu, H.-N.; Li, F.; A strategy combining solid-phase extraction, multiple mass defect filtering and molecular networking for rapid structural classification and annotation of natural products: characterization of chemical diversity in *Citrus aurantium* as a case study. *Analytical and Bioanalytical Chemistry* **2021**, *413*, 2879. [[CrossRef](#)]
  48. Brites-Neto, J.; Maimone, N. M.; Piedade, S. M. D. S.; Andrino, F. G.; Andrade, P. A. M. de; Baroni, F. de A.; Gomes, L. H.; Lira, S. P.; Scorpionicidal activity of secondary metabolites from *Paecilomyces* sp. CMAA1686 against *Tityus serrulatus*. *Journal of Invertebrate Pathology* **2021**, *179*, 107541. [[CrossRef](#)]
  49. Vizcaino, M. I.; Crawford, J. M.; Em *Nonribosomal Peptide and Polyketide Biosynthesis. Methods in Molecular Biology*; Evans, B. S., org.; Humana Press Inc.: New York, 2016, cap. 12.
  50. Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L. F.; Dührkop, K.; Bandeira, N.; Dorrestein, P. C.; Böcker, S.; Significance estimation for large scale metabolomics annotations by spectral matching. *Nature Communications* **2017**, *8*, 1. [[CrossRef](#)] [[PubMed](#)]
  51. Frank, A. M.; Bandeira, N.; Shen, Z.; Tanner, S.; Briggs, S. P.; Smith, R. D.; Pevzner, P. A.; Clustering Millions of Tandem Mass Spectra. *Journal of Proteome Research* **2008**, *7*, 113. [[CrossRef](#)]
  52. Kang, K. Bin; Park, E. J.; Da Silva, R. R.; Kim, H. W.; Dorrestein, P. C.; Sung, S. H.; Targeted Isolation of Neuroprotective Dicomaroyl Neolignans and Lignans from *Sageretia theezans* Using in Silico Molecular Network Annotation Propagation-Based Dereplication. *Journal of Natural Products* **2018**, *81*, 1819. [[CrossRef](#)] [[PubMed](#)]
  53. Kang, K. Bin; Woo, S.; Ernst, M.; van der Hooft, J. J. J.; Nothias, L. F.; da Silva, R. R.; Dorrestein, P. C.; Sung, S. H.; Lee, M.; Assessing specialized metabolite diversity of *Alnus* species by a digitized LC-MS/MS data analysis workflow. *Phytochemistry* **2020**, *173*, 112292. [[CrossRef](#)] [[PubMed](#)]
  54. Ruttkies, C.; Neumann, S.; Posch, S.; Improving MetFrag with statistical learning of fragment annotations. *BMC Bioinformatics* **2019**, *20*, 1. [[CrossRef](#)] [[PubMed](#)]
  55. Van Der Hooft, J. J. J.; Wandy, J.; Young, F.; Padmanabhan, S.; Gerasimidis, K.; Burgess, K. E. V.; Barrett, M. P.; Rogers, S. Unsupervised Discovery and Comparison of Structural Families Across Multiple Samples in Untargeted Metabolomics. *Analytical Chemistry* **2017**, *89*, 7569. [[CrossRef](#)] [[PubMed](#)]

56. van der Hooft, J. J. J.; Wandy, J.; Barrett, M. P.; Burgess, K. E. V.; Rogers, S.; Topic modeling for untargeted substructure exploration in metabolomics. *Proceedings of the National Academy of Sciences* **2016**, *113*, 13738. [[CrossRef](#)]
57. Taylor, P. J.; Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. *Clinical Biochemistry* **2005**, *38*, 328. [[CrossRef](#)] [[PubMed](#)]