

REDES MOLECULARES: UMA ANÁLISE SOBRE ANOTAÇÕES E DESCOBERTA DE NOVOS ATIVOS

Alan C. Pilon^{a,b}, Natália C. Vieira^a, Juliano G. Amaral^c, Afif F. Monteiro^d, Ricardo R. da Silva^b, Laila S. Spíndola^e, Ian Castro-Gamboa^a e Norberto P. Lopes^{b,*} 

^aDepartamento de Química Orgânica, Instituto de Química, Universidade Estadual Paulista, 14800-060 Araraquara – SP, Brasil

^bDepartamento de Ciências Biomoleculares, Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, 14040-903 Ribeirão Preto – SP, Brasil

^cInstituto Multidisciplinar em Saúde, Campus Anísio Teixeira, Universidade Federal da Bahia, 45029-094 Vitória da Conquista – BA, Brasil

^dInstituto de Química de São Carlos, Universidade de São Paulo, 13560-970 São Carlos – SP, Brasil

^eUniversidade de Brasília, Campus Universitário Darcy Ribeiro, 70910-900 Brasília – DF, Brasil

Recebido em 25/03/2021; aceito em 22/04/2021; publicado na web em 15/06/2021

MOLECULAR NETWORKS: AN ANALYSIS ON ANNOTATIONS AND DISCOVERY OF NEW ASSETS. To speed up the discovery of bioactive natural products (NP), chemists have sought advanced approaches in analytical and computational chemistry in attempt to organize and extract information from large data sets. In this sense, the molecular networks (MN) successfully organized enormous sets of mass spectrometry (MS) data together with samples metadata information in an intuitive visualization in the spectral similarity networks format. GNPS (Global Natural Products Social Molecular Networking), a free online platform for storing and processing MSⁿ data, is a leading application of spectral matching with public databases aimed at the dereplication and discovery of new bioactive products through molecular networks. In this review, we address the concept of GNPS spectral similarity networks, as well as their complementary computational tools, benefits and limitations applied in NP studies associated with dereplication, chemical ecology, functional genetics and determination of biosynthetic pathways.

Keywords: natural products; tandem mass spectrometry; molecular networking; dereplication; chemical ecology.

INTRODUÇÃO

A busca por procedimentos ágeis e eficientes para descoberta de novos produtos naturais bioativos impulsionou, no final dos anos de 90, o avanço das tecnologias analíticas em separação e aquisição espectral.¹ O desenvolvimento de ferramentas computacionais, como as técnicas de elucidação de estrutura assistida por computador (CASE – do inglês *Computer Assisted Structure Elucidation*), permitiu também acelerar a identificação de compostos por meio de comparações individuais entre espectros adquiridos e repositórios espectrais de compostos.²

Embora essa comparação individual entre espectros tenha contribuído para anotação e desreplicação de muitas substâncias, o procedimento continua sendo laborioso, demandando longos períodos de curagem e verificação dos dados.^{3,4} Se consideramos as abordagens modernas aplicadas aos produtos naturais, como metabolômica^{5,6} ou a biologia de sistemas, a situação fica ainda mais complexa e impraticável. Plantas, microrganismos (terrestres e marinhos) e bactérias podem expressar, dependendo das condições ambientais e genéticas, centenas ou até milhares de compostos, tornando a tarefa de identificar/anotar compostos por meio das estratégias tradicionais em PN um desafio humanamente inviável.^{1,7}

A química de produtos naturais voltada as ciências ômicas, ou ao entendimento de eventos multifatoriais, tem buscado na análise estatística, bioinformática e na quimiometria formas alternativas de organizar, armazenar, mapear e analisar conjuntos complexos de dados, com rapidez e confiabilidade.⁸ Análise por redes moleculares (do inglês *Molecular Networking*) surge neste contexto como uma ferramenta de organização e visualização de grandes conjuntos de dados.⁹⁻¹² A partir de dados espectrais, a rede molecular facilita a identificação de padrões

(similaridades e diferenças espectrais) entre conjuntos amostrais. Isso permite reconhecer classes metabólicas, substituintes e funções orgânicas em estruturas complexas de produtos naturais.¹³⁻²⁸

A incorporação de bibliotecas de compostos (espectrais) nas redes moleculares têm auxiliado a anotação de compostos conhecidos por comparação direta (dado adquirido e as bibliotecas) mas também tem fornecido subsídios para anotação de moléculas desconhecidas.^{3,9} O aumento em número e quantidade de dados presentes em repositórios espectrais, como por exemplo o PubChem,²⁹ METLIN,³⁰ NIST,³¹ GOLM,³² BMRB,³³ Super Natural II³² e a NuBBE_{DB}³⁴ também tem ajudado no desenvolvimento de abordagens computacionais complementares as redes moleculares.

A *Global Natural Products Social Networking*³ (GNPS) é uma plataforma *online* e de livre acesso para o cálculo de redes moleculares a partir de dados de espectrometria de massas.³ Essa plataforma, inicialmente desenvolvida para estudos proteômicos,³⁵ foi, a princípio, adaptada para análise metabólica de organismos marinhos e atualmente é utilizada por toda comunidade de PN.³⁶ Sua eficiência em organizar e resumir grandes conjuntos de dados de espectrometria de massas (EM) levaram a expansão de seu protótipo de aplicação em linha de comando para uma abordagem de aplicação *web*, com uma interface simples e amigável a não programadores. A infraestrutura computacional do *Mass spectrometry Virtual Interactive Environment* (MassIVE), um repositório de dados comunitário, parte do consórcio global *ProteomeXchange*, foi utilizado como suporte para criação do GNPS, o que transformou esta plataforma em uma rede internacional de tratamento e visualização de grandes conjuntos de dados.^{13,37,38} É importante destacar que plataformas como o GNPS podem ser de grande utilidade para a anotação espectral, o que não necessariamente significa identificação estrutural.³⁹

O objetivo deste trabalho é descrever como as redes moleculares, a partir da plataforma GNPS, tem impactado no desenvolvimento

*e-mail: npelopes@gmail.com

da química de produtos naturais do século XXI. Esta revisão traz informações sobre a criação das redes moleculares e uso das diferentes ferramentas associadas, como também exemplos interdisciplinares de aplicação em produtos naturais.

CONCEITOS QUE ENVOLVEM AS REDES MOLECULARES

O conceito de redes pode ser aplicado a qualquer campo acadêmico interdisciplinar no estudo de dados complexos. Seu primeiro relato é descrito por Leonhard Euler em 1736 no caso matemático das *Sete Pontes de Königsberg*.⁴⁰ As redes moleculares por sua vez visam, através de uma estratégia computacional, ajudar a visualização e interpretação de dados espectrais complexos. Utilizando de fatores correlacionais, como por exemplo o grau de similaridade entre espectros, é possível gerar redes moleculares que, em última análise, podem ajudar na determinação de classes metabólicas, funções orgânicas, propriedades biológicas e quimiotaxonômicas.^{8,11}

A plataforma GNPS³ é voltada à análise de grandes conjuntos de dados (a partir de íons fragmento) em experimentos de espectrometria de massas. Esses dados podem ser gerados a partir de aquisições dependentes de dados (DDA), onde os experimentos de EMⁿ são realizados pelo isolamento prévio de íon(s) ou, pela aquisição independente de dados (DIA), em que os espectros de EMⁿ são adquiridos sem isolamento prévios dos íons precursores. Diversas fontes de ionização podem ser utilizadas para a criação das redes moleculares, como ESI, APCI, MALDI e EI, bem como diversas fontes de fragmentação, como *collision induced dissociation* (CID), *Higher-energy collisional dissociation* (HCD) e *Electron-transfer dissociation* (ETD). Contudo, as informações da EM de uma substância podem variar significativamente dependendo da técnica de ionização⁴¹ ou da forma de ativação colisional,⁴² o que impossibilita o uso de diferentes protocolos/instrumentação para análise de um conjunto de amostras.

O conjunto de algoritmos do GNPS organiza cada espectro adquirido (por exemplo em CLAE-EM e EMⁿ) e os classifica a partir de um valor similaridade, que nesse caso é calculado pelo valor de cosseno entre espectros presentes na(s) amostra(s) ou entre a(s) amostra(s) e o(s) espectro(s) de referência presentes nos bancos de dados. Dessa forma, são gerados agrupamentos espectrais com base em um valor pré-estabelecido de cosseno (valor que varia de 0 a 1 – sendo que 1 indica 100% similaridade ou identidade e 0, totalmente diferentes). Se o pareamento entre dois espectros possuir uma similaridade de cosseno maior ou igual ao valor limite pré-estabelecido pelo usuário, uma conexão ou um arco entre os dois espectros é estabelecido.^{3,10}

Como cada substância pode ser constituída por um conjunto de espectros (em experimentos CLAE-EM e EMⁿ podem ocorrer várias fragmentações de uma mesma substância devido a largura do sinal cromatográfico) é necessária uma etapa de agrupamento de íons de uma mesma substância, também chamado de espectro consenso. Esse espectro é representado por um nodo (todos os espectros de uma mesma substância ficam armazenados neste nodo) enquanto os arcos que os conectam representam os valores de similaridade entre os nodos.¹²

Finalmente, os nodos da rede são anotados utilizando diferentes bibliotecas de compostos, com os mesmos critérios de pareamento espectral, número mínimo de picos compartilhados (no espectro de massas) e a maximização da similaridade do cosseno, considerando a diferença das massas entre os íons precursores.^{3,10,12}

A Figura 1 representa o processo de construção das redes moleculares a partir de dados de CLAE-EM e EMⁿ. São extraídos dos sinais (cromatográficos) os espectros de EMⁿ comparados, formando os espectros consenso e a rede molecular a partir de um valor pré-estabelecido de similaridade de cosseno.

A plataforma também oferece armazenamento e análise de dados brutos (cromatogramas completos em formatos públicos *mzML* e *mzXML*) através do repositório de dados MassIVE, possibilitando o compartilhamento com toda a comunidade.³ Após a geração dos nodos e arcos, estes são representados como uma rede, Figura 1, permitindo que a informação química seja interpretada de forma intuitiva, associando compostos análogos, e suas respectivas particularidades químicas (diferenças de massas associadas aos arcos) sejam agrupadas em uma representação condensada.

A Figura 2 representa a estrutura da plataforma de redes moleculares (GNPS), considerando desde o carregamento dos dados, armazenamento, rede social e comunitária da plataforma como as ferramentas de análise. Como a plataforma está constantemente disponibilizando novas ferramentas, os usuários são aconselhados a verificar o fórum oficial da comunidade para eventuais dúvidas (https://groups.google.com/forum/#!forum/molecular_networking_bug_reports).

REDES MOLECULARES - MODO CLÁSSICO GNPS (CMN)

A rede molecular clássica, conhecida como *Classic Molecular Networking* no GNPS, foi a primeira ferramenta desenvolvida na plataforma dispensando a necessidade de estágios de pré-processamento (avançado) como a quantificação dos *features*. Neste modo é possível carregar dados de CLAE ou CLUE-EM e EMⁿ e MALDI-EM em modo DDA.^{3,10,43}

A criação das redes moleculares começa pelo processo de remoção de ruídos dos espectros de fragmentação (EMⁿ), com a seleção dos

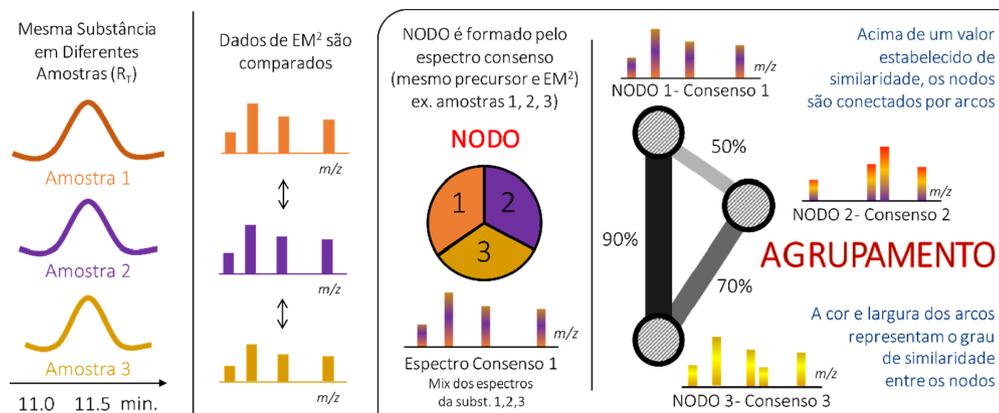


Figura 1. Formação de uma rede molecular a partir da inserção dos dados de EMⁿ, comparação dos espectros de massas para a geração do espectro de consenso e formação de agrupamento por similaridade (clusters)

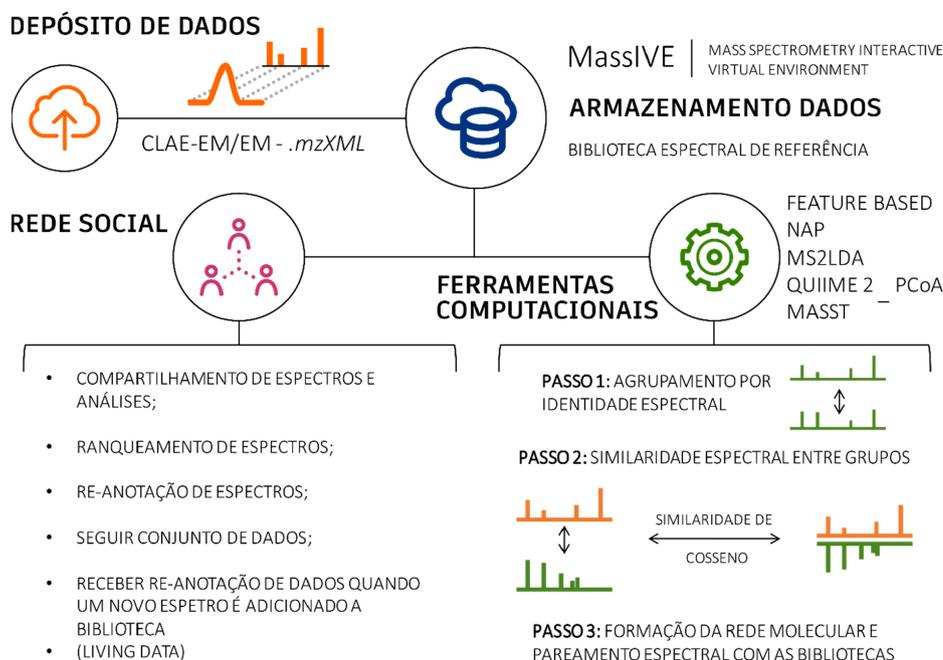


Figura 2. Representação esquemática da estrutura da plataforma GNPS. A plataforma conta com duas bases de dados, que armazenam dados brutos experimentais e dados de referência. Além das bases de dados existe um componente analítico, que está em constante expansão, com a adição de novos módulos e um componente social, onde os usuários podem trocar informações, depositar e analisar dados

N picos mais intensos (parâmetro definido pelo usuário) em janelas de ± 50 unidades de m/z ao longo do espectro. Em seguida, para a criação dos espectros consenso (espectros pertencentes a um único sinal cromatográfico), o GNPS oferece o algoritmo MSCluster,⁴⁴ o qual é possível ajustar a tolerância de massa para agrupamento de espectros referentes ao íon precursor e aos íons produto (valor padrão de 1.0 Da). Esses valores dependem da performance do instrumento de massas (alta ou baixa resolução). Assim, equipamentos calibrados com maior poder de resolução utilizam tolerâncias menores que 1.0 Da, como por exemplo instrumentos com analisadores do tipo *Time of Flight* (ToF) ou Orbitrap.

A vantagem desta abordagem é que ela requer um número menor de etapas de pré-processamento de dados, pode ser utilizada em pequenos e grandes conjuntos de amostras e não necessita de um protocolo de quantificação para as análises experimentais. A desvantagem é que o algoritmo MSCluster não leva em consideração o tempo de retenção dos sinais, o que pode levar ao agrupamento de isômeros e isôbaros que possuem algum nível de similaridade espectral. Ele também não é indicado para a criação de redes que buscam identificar flutuações na quantidade e na covariância dos compostos da rede.^{45,46}

Uma das primeiras publicações em produtos naturais utilizando de redes moleculares foi desenvolvido por Watrous e colaboradores,¹² trabalhando com entendimento de interações de colônias de bactérias. O potencial da técnica foi explorado para detecção, visualização e classificação química dos compostos.

Dentre as possíveis aplicações em produtos naturais do modo clássico destacam-se a anotação de compostos com eficiência e rapidez, estudos de quimiotaxonomia e descoberta de novas substâncias.^{18,37,47-49} Como as redes moleculares formam agrupamentos por similaridade espectral, muitas substâncias de uma mesma classe metabólica (perfil de fragmentação similar) podem ser arranjadas dentro de um mesmo agrupamento. Isso é bastante vantajoso quando se analisa grandes conjuntos amostrais (>15 amostras), pois a rede permite uma avaliação simultânea de todos os espectros da rede.

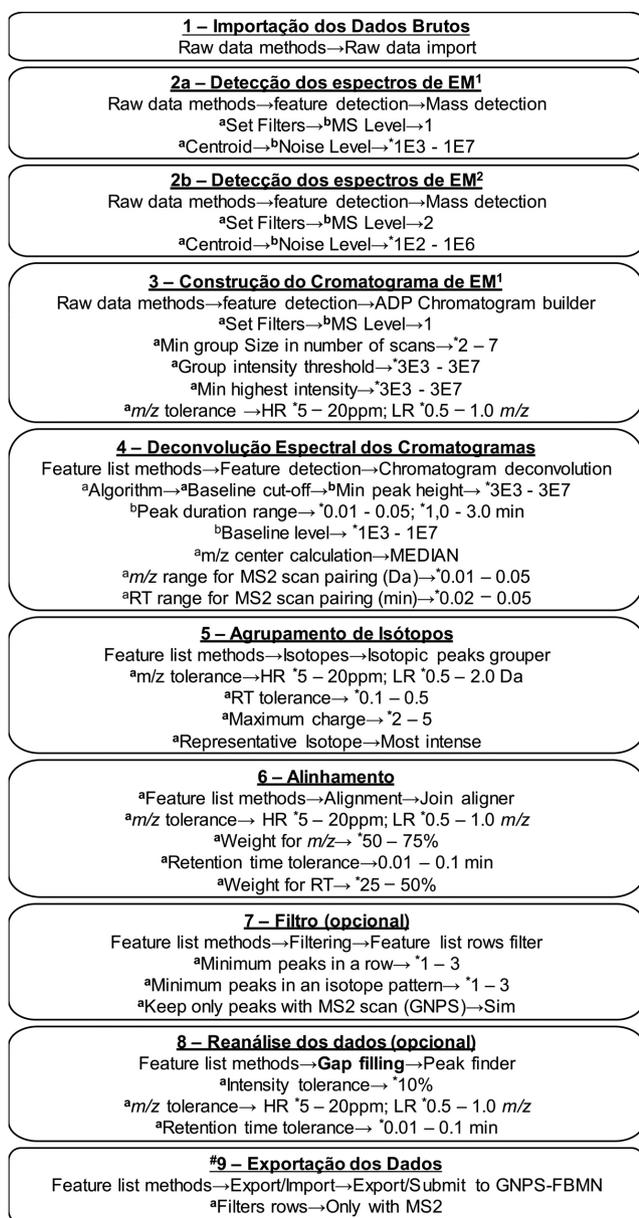
É possível ainda utilizar o banco de dados para aumentar anotação de várias classes metabólicas, priorizando compostos inéditos que

possuem similaridade espectral com compostos conhecidos. Esse procedimento é chamado de identificação de desconhecidos por compostos conhecidos, uma estratégia útil e rápida para a descoberta de novos produtos naturais. Se os compostos conhecidos de interesse possuem informações sobre alguma atividade biológica, essa priorização pode carregar novas informações biológicas a um dado organismo. O trabalho pioneiro de Yang e colaboradores¹¹ utilizando de redes moleculares para o estudo de microrganismos marinhos e terrestres, desreplicou mais 58 moléculas bioativas a partir de análogos dos bancos de dados, abrindo espaço para o crescimento dessa estratégia.

REDES MOLECULARES - MODO FEATURE BASED MOLECULAR NETWORKING GNPS (FBMN)

Para contornar os problemas causados pelo uso do algoritmo MSCluster e, também considerar a flutuação e quantificação metabólica, o GNPS disponibilizou um módulo de análise chamado de *Feature-Based Molecular Networking* (FBMN).⁵⁰ O FBMN requer experimentos quantitativos (quantidade relativa) e uma etapa de pré-processamento de dados espectrais de modo que espectros de massas sejam convertidos a um *feature*. Nessa abordagem é necessário que as análises experimentais sejam realizadas por técnicas acopladas, como por exemplo, CLAE-EM e EMⁿ, uma vez que o tempo de retenção será associado aos espectros (*feature* é um conjunto de características que alia diferentes fontes de dados, por exemplo, sinais cromatográficos e espectrometria de massas).

Existem diversas ferramentas que realizam os pré-processamentos de dados espectrais como MZmine,^{51,52} OpenMS,⁵³ MS-DIAL,⁵⁴ MetaboScape, XCMS,⁵⁵ Progenesis QI ou mzTab-M⁵⁶ entre outras.⁴⁶ Contudo, MZmine⁵⁷ é a mais divulgada e utilizada pela comunidade de espectrometria de massas. Em geral, o pré-processamento necessário para a geração de *features* passa por nove etapas antes de ser carregada no FBMN.⁵⁰ O Esquema 1 mostra, em detalhes, as etapas e os parâmetros (sugeridos) pelo fórum da comunidade de redes moleculares, para o pré-processamento de dados de EMⁿ no MZmine.



Esquema 1. Pré-processamento de dados realizados no pacote Mzmine 2 para a geração de redes moleculares (FBMN). ^a representa a janela principal e ^b secundária. ^{*}os parâmetros aqui sugeridos não são absolutos, dependem do tipo de amostra, equipamentos e métodos; HR – alta resolução; LR – baixa resolução; # deve selecionar um arquivo no formato desejado

A primeira etapa consiste na importação dos dados brutos de CLAE-EM e EMⁿ (ou técnicas cromatográficas similares como CLUE-EM e EMⁿ) e sua conversão em formato *mzXML*; o segundo estágio corresponde a detecção e filtragem de íons presentes nos dados brutos. Nesse processamento define-se os limites de intensidade dos íons em EM¹ e EMⁿ visando excluir a maior quantidade possível de sinais de ruído.^{5,51} Na sequência, são reconstruídos os cromatogramas de íons extraídos a partir de dados adquiridos, isto é, os íons extraídos anteriormente são rearranjados sequencialmente dando origem a um novo cromatograma.⁵⁸ O quarto estágio compreende-se a deconvolução dos cromatogramas. Esse processamento é importante para reconhecer e separar os isóbaros que possuem tempos de retenção diferentes e também promover a associação dos íons precursores com seus respectivos íons produto.⁵¹ Agrupar os íons que participam de um mesmo perfil isotópico em um único *feature* é a etapa seguinte, isótopos podem ser interpretados como substâncias diferentes, por

isso é necessário reconhecê-los e agrupá-los sob um único *feature*. No sexto estágio ocorre o alinhamento dos *features* de cada amostra em uma única lista, nesta etapa é possível reconhecer as amostras que apresentam os mesmos *features*.⁵¹ O sétimo consiste em executar uma filtração na lista de dados que visa excluir os *features* provenientes de ruídos e interferentes. O oitavo estágio consiste em uma reanálise dos dados (*Gap Filling*), e tem o propósito de reestabelecer *features* de interesse que foram excluídos ou não selecionados por conta dos valores dos parâmetros utilizados. Entretanto, deve-se ter atenção com os parâmetros aplicados para não recuperar sinais que não são verdadeiros.^{50,51} Por fim os dados são exportados em dois arquivos, um arquivo no formato *csv*, contendo as intensidades dos espectros de EM¹ (com seus respectivos números de *scans*, possibilitando gerar maior interação e adições de novas informações aos *features*), e um no formato *mgf*, contendo todas as informações dos espectros de EMⁿ. Uma vez finalizado, esses arquivos podem ser submetidos ao fluxo de trabalho do FBMN.^{50,51,57,58}

É possível ainda incluir uma tabela de metadados⁵⁹ tanto no modo CMN quanto no FBMN.

REDES MOLECULARES – ANÁLISE POR CG-IE-EM

Redes moleculares a partir de dados de CG-IE-EM é uma opção recente na plataforma GNPS. Como os dados de massas oriundos de experimentos em IE não apresentam pré-seleção de íons precursores (formato de aquisição denominado de DIA) certas considerações são necessárias. Diversos conjuntos de dados podem ser combinados, entretanto é exigido um mesmo protocolo de análise, como o programa de temperatura, coluna, modo de injeção e energia de ionização.^{3,10,60}

A plataforma do GNPS disponibiliza duas sessões para iniciar o trabalho de análise de dados CG-IE-EM. Na primeira é oferecido ao usuário a possibilidade de realização de um pré-processamento utilizando o MS-Hub, que além da deconvolução dos cromatogramas, típico de experimentos realizados em modo DDA, promove uma deconvolução espectral por meio de uma transformação rápida de Fourier, multiplicação e transformação inversa de Fourier para cada íon, seguida por uma fatoração de matriz não supervisionada.⁶⁰ Todo processo é realizado de forma automática e dispensando qualquer expertise em estatística e/ou bioinformática. Entretanto para garantir a eficiência do processo, é crucial que no mínimo 10 amostras sejam utilizadas. A falta de informação pode levar a uma deconvolução instável, gerando espectros de baixa qualidade.⁶⁰ É possível utilizar outras opções de deconvolução como MZMine2/ADAP,^{18,51,61} MS-Dial⁵⁴ ou XCMS,⁵⁵ as quais não exigem um número mínimo de amostras.

Transcorrida a etapa de pré-processamento, é gerado um arquivo *mgf* contendo os espectros alinhados, deconvoluídos e com os tempos de retenção, bem como uma tabela contendo as áreas de pico detectados. Utilizando o MS-Hub, o arquivo é automaticamente introduzido no fluxo de trabalho do GNPS para comparação com a biblioteca espectral e, para construção da rede molecular, não exigindo etapas de *download* e *upload* no sistema. O usuário deve somente selecionar o link: *Search Spectral Library and Molecular Network*, disponível após a deconvolução dos espectros. O usuário também pode realizar seu *download* e utilizá-lo para explorar os resultados empregando diversas outras ferramentas ou bibliotecas,⁶⁰ por exemplo MetExpert,⁶² NIST 2017⁶³ e Wiley.

Tendo o usuário optado por utilizar o MS-Hub, o fluxo de trabalho para busca nas bibliotecas e criação da rede molecular será exibido automaticamente exigindo a definição de tolerância de massa para agrupamento dos espectros dos íons produtos, o número mínimo de íons produto e o valor mínimo de cosseno. Opções avançadas

estão disponíveis, como da inclusão dos valores de referências de hidrocarbonetos para cálculo do índice de retenção (Kovats).⁶⁰

COMO UTILIZAR A PLATAFORMA GNPS E A FERRAMENTA DE VISUALIZAÇÃO CYTOSCAPE

Para acessar a documentação e as ferramentas do GNPS é necessário inicialmente fazer um cadastro através do website <https://gnps.ucsd.edu>. Para a construção da rede molecular, as análises e processamentos devem ser realizados de acordo com as recomendações da plataforma. Os dados devem inicialmente ser convertidos em formatos do tipo *mzXML*, *mzML*, ou *mgf* – pacotes como *MSConvert*, do grupo de ferramentas *ProteoWizard* realiza tais conversões, além da funcionalidade de conversão *online* usando o recurso arrastar e soltar disponível na plataforma GNPS, Figura 3. Existem duas opções para o carregamento dos dados na plataforma (*upload*): se os dados não ultrapassarem 20 MB, o *upload* pode ser realizado diretamente pelo site utilizando a função *Drag and Drop*; para arquivos maiores que 20 MB, protocolos de transferência do tipo FTP (*File Transfer Protocol*) são recomendados. O pacote livre *WinSCP* é indicado pela plataforma, onde o usuário acessa o servidor *PreoteoSAFE* – hospeda o GNPS e o *MassIVE* – e transfere os dados para a plataforma (outras ferramentas podem ser utilizadas como *CoreFTP* e *CoffeeCup Free FTP*).³

Uma vez convertidos e carregados, segue-se para etapa de configuração de parâmetros para a criação da rede molecular (para mais informações segue o link da documentação: <https://ccms-ucsd.github.io/GNPSDocumentation/>). De acordo com a particularidade do estudo, como o instrumento, técnica e poder de resolução, serão necessárias alterações nos parâmetros. O GNPS oferece campos de instrução para cada um deles.^{3,10}

Não é recomendado o processamento simultâneo de dados obtidos em diferentes modos de ionização, porém o recurso *Merge Network Polarity* possibilita mesclar duas redes moleculares criadas em modos diferentes de ionização. Esse recurso pode ser utilizado tanto em redes criadas no CMN quanto no FBMN.

Embora a plataforma GNPS forneça a possibilidade de avaliação por ferramentas *web browser*, o uso de pacotes *standalone* faz-se

necessário em casos de avaliação de redes em modo *off-line* e para conjuntos amostrais que necessitem maior poder de processamento. O *Cytoscape*⁶⁴ é um destes programas (recomendado) de aplicação livre, feito pela comunidade científica para a visualização de grandes conjuntos de dados no formato de redes. Nele é possível realizar uma navegação dinâmica de gráficos, bem como utilizar de ferramentas de edição de rótulos, cores e formatos (os nodos são frequentemente associados a massa do íon precursor). Os arcos da rede frequentemente correspondem diferença de massa entre dois íons precursores ou o grau de similaridade entre eles.

A opção por *web browser* possibilita a rápida filtragem de tabelas do tipo *html* com ligações dinâmicas a bancos de dados espectrais, visualização de famílias moleculares (componentes conectados à rede), visualização dos espectros de fragmentação, comparação de espectros, geração de histogramas, além do gerenciamento e compartilhamento simplificado pelos *web links*. A possibilidade de gerenciar as tarefas, repetir as análises (clonar os parâmetros) e compartilhar com colaboradores tem motivado a ampla adoção da ferramenta pela comunidade científica.^{3,10}

BIBLIOTECAS ESPECTRAIS E AS REDES MOLECULARES

Um dos grandes desafios no processo anotação/identificação em química de produtos naturais é a carência de banco de dados públicos de espectros de massas sequencial (EMⁿ).⁶⁵ Na tentativa de contornar esse obstáculo, muitos laboratórios acabam criando suas próprias bibliotecas (*in house*), que em sua maioria não são compartilháveis e/ou são incompatíveis com outras bibliotecas. Nesse sentido, o GNPS tem se empenhado na construção de uma biblioteca de produtos naturais comunitária, em que cada pesquisador pode acessar os espectros de referência gratuitamente, mas também depositar os seus próprios em um formato padronizado.^{3,10,65} Além disso, o GNPS disponibiliza informações de importantes bibliotecas de compostos como LDB Lichen Database, MIADB Spectral Library,^{12,66,67} Sumner Spectral Library,⁶⁸ CASMI Spectral Library,^{66,69} MassBank (Japan (<http://massbank.jp>)),⁷⁰ EU Mass Bank (<https://massbank.eu/MassBank/>)⁷¹ e North America Mass Bank (<http://mona.fiehnlab.ucdavis.edu/>).

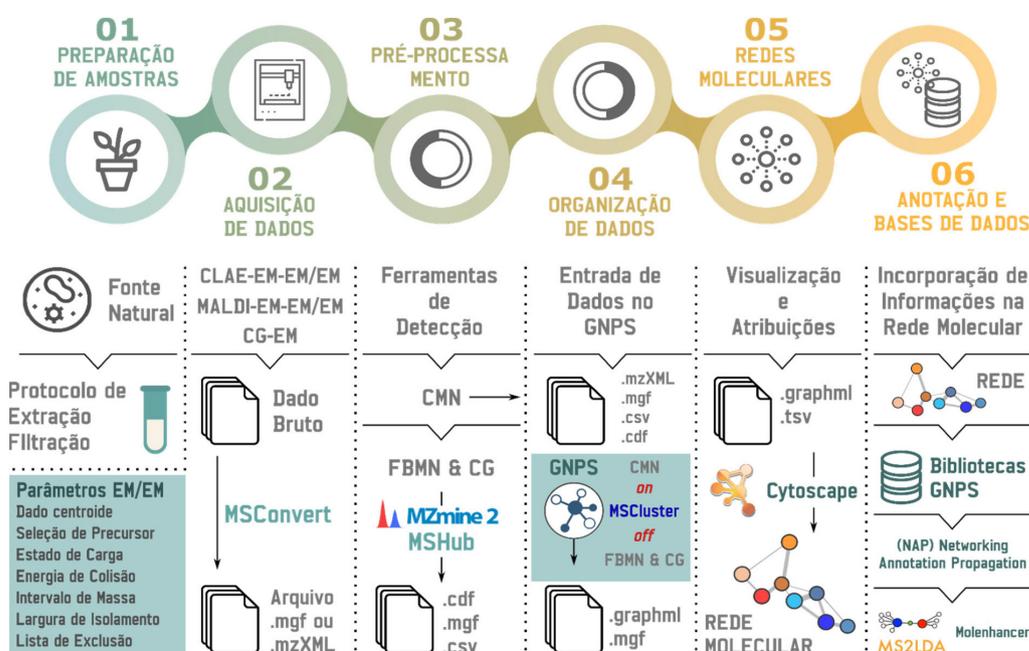


Figura 3. Ilustração do uso da ferramenta GNPS. São necessários 6 etapas entre a inserção e interpretação dos resultados. CMN – redes moleculares no modo clássico; FBMN – feature-based molecular networking; CG – redes moleculares para dados de CG

A forma mais comum de utilização da biblioteca é através da comparação automática dos espectros consenso das redes CMN,¹⁰ FBMN⁵⁰ ou CG-IE-EM,⁶⁰ contra os espectros da biblioteca espectral GNPS.³ Durante o fluxo de trabalho na plataforma o usuário pode selecionar as bibliotecas definindo o valor mínimo de cosseno (similaridade) e o número mínimo de íons fragmento comuns (entre o dado adquirido e o banco de espectros).^{10,50,60} Também é possível realizar a reanálise de um espectro específico empregando a ferramenta MASST, ela realiza uma nova busca na biblioteca espectral e permite alteração nos parâmetros de busca de forma individual.⁷²

Com *Library Search* é possível fazer buscas e comparações espectrais a partir de um único ou um conjunto espectros frente a biblioteca espectral. O usuário deve previamente converter seus dados para um formato compatível (*mzXML* ou *mzML*) antes de carregá-los na plataforma. Em seguida, seleciona-se os arquivos, define-se a tolerância de massa em relação ao íon precursor e aos íons produto, o número mínimo de íons produto e o valor mínimo de cosseno.³ A principal diferença desse recurso daqueles que geram uma rede molecular é que não há alinhamento de espectros de consenso, ou seja, os espectros adquiridos de uma mesma substância serão considerados únicos.

Outra opção é o *Molecular Explorer*, ferramenta que permite o contato entre pesquisadores que possuem dados em comum. Utilizando do conceito de “dados vivos”, o usuário pode buscar substâncias presentes em suas amostras, que foram relatadas para outros conjuntos de dados públicos.^{3,10} Por se tratar de um repositório público e colaborativo, também é possível acessar, explorar e realizar *download* de todos os espectros depositados na biblioteca espectral do GNPS.³

Além das buscas é possível contribuir com o banco comunitário de substância do GNPS. Todas as contribuições são revisadas e categorizadas baseadas na qualidade dos espectros e na confiabilidade das anotações.^{3,73} O GNPS classifica as contribuições utilizando três níveis de qualidade: (i) *Gold* (ouro) caracterizados por espectros de substâncias puras de origem sintética ou isolada de produtos naturais com caracterização estrutural completa, (ii) *Silver* (prata) os espectros são suportados por uma publicação e o (iii) *Bronze* (bronze), refere-se a caracterização incompleta ou não suportada por publicação. Os níveis ouro e prata, só podem ser atribuídos por membros autorizados da comunidade, mesmo se tratando de espectros de substâncias que preenchem os requisitos. A comunidade faz o depósito na categoria bronze e, após revisão, eles podem ser promovidos aos níveis superiores. Os formatos aceitos para depósito são *mzXML*, *mzML* e *mgf*.³

FERRAMENTAS INTEGRADAS AS REDES MOLECULARES

As redes moleculares criaram uma forma estruturada de extrair e organizar dados de EMⁿ contribuindo para a solução de muitos desafios em produtos naturais.⁷⁴ No entanto, a anotação em larga escala de matrizes complexas continua sendo um obstáculo para a descoberta de novas entidades químicas bioativas. A taxa de anotação média, em estudos de redes moleculares pelo GNPS era de 1-2% em 2012, data de sua primeira publicação, saltando para 5-6% em nossos dias.^{10,74}

Os metabólitos não anotados, também chamados de “matéria escura – alusão a matéria desconhecida no cosmo”, representa o maior desafio para a química de produtos naturais, bioinformática e das ferramentas computacionais.^{18,75} Na seção a seguir são mostradas ferramentas complementares as redes moleculares e ao processo de anotação: *Network Annotation Promotion* (NAP),⁷⁴ MS2LDA,^{76,77} MolNetEnhancer⁷⁸ e Qemistree.⁷⁹

Networking Annotation Propagation (NAP)

O NAP é baseado no algoritmo de fragmentação do MetFrag⁸⁰ e de ligação do MetFusion.⁸¹ Ele utiliza a rede molecular para melhorar a precisão das previsões *in silico* por meio da propagação de anotações estruturais ampliando as probabilidades de anotar corretamente as substâncias, mesmo sem uma correspondência com as bibliotecas espectrais. Isso é realizado através da criação de uma lista de candidatos que são reclassificados usando a topologia da rede molecular e similaridade estrutural, melhorando assim significativamente a classificação das anotações *in silico*.⁷⁴ Apresentando uma interface de fácil aplicação, além do fato de estar integrado à plataforma GNPS, o pacote NAP⁷⁴ pode ser utilizado logo após a construção das redes moleculares.^{10,50} O NAP pode ser utilizado acessando o link <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-theoretical.jsp>, inserindo o JobID (identificador do processamento de uma rede, *i.e.*, quando se cria uma rede no GNPS é gerado um ID) definindo alguns critérios para análise, como por exemplo, o valor mínimo de cosseno, modo de aquisição e a precisão de massa em ppm. Finalizado o processamento, os resultados podem ser explorados na própria plataforma ou exportado para o *software* Cytoscape.⁶⁴ NAP vem sendo empregado em diversos estudos na área de produto naturais, em que podemos destacar: avaliação da diversidade química de espécies do gênero *Alnus*,⁸² no estudo *multi-omics* para avaliar o potencial de cianobactérias marinhas filamentosas tropicais⁸³ e para anotação *in silico* dos marcadores de três espécies do gênero *Zanthoxylum*.⁸⁴

MS2LDA

Inspirado na modelagem de texto por *Latent Dirichlet Allocation* (LDA),⁸⁵ MS2LDA é uma ferramenta que decompõe dados de fragmentação molecular e identifica padrões de fragmentação de massas. Fragmentos co-ocorrentes e/ou perdas neutras que representam subestruturas moleculares relevantes (p. ex., hexose ou a perda de um grupo carboxila) potencialmente indicativos de famílias estruturais, denominados Mass2Motifs, são estruturalmente caracterizados e utilizados para anotar moléculas para as quais não existem espectros de referência.^{76,77}

A ferramenta MS2LDA está integrada à plataforma GNPS e pode ser selecionada através da opção *Analyze with MS2LDA* na página resultados, disponível no final do processamento da rede molecular. Uma vez que o processamento seja finalizado é possível explorar os resultados tanto na plataforma ou exportando para o *software* Cytoscape⁶⁴ ou na página do MS2LDA.^{76,77}

Assim como as bibliotecas comunitárias nas quais o usuário pode depositar/anotar uma substância com base no seu espectro, o MS2LDA oferece um serviço de anotação de *motifs*. Essa funcionalidade é uma das grandes inovações da aplicação, uma vez que facilita e acelera a anotação de substâncias a partir dos *motifs* identificados (fragmentos).

Recentemente um trabalho utilizando redes moleculares e MS2LDA foi capaz de indicar a presença de novos ésteres de diterpenos em espécies do gênero *Euphorbia* com atividade inibitória contra o vírus Chikungunya.⁴⁸

MolNetEnhancer

A fim de integrar algumas ferramentas complementares disponíveis no GNPS às redes moleculares foi desenvolvido a ferramenta MolNetEnhancer.⁷⁸ Esta aplicação proporciona a combinação de uma rede molecular com ferramentas *in silico* como o NAP⁷⁴ e o MS2LDA⁷⁶ e também automatiza a

classificação química da ferramenta ClassyFire.⁸⁶ Isso permite a visualização de classes e subclasses químicas presente nos *clusters*, incrementando as informações estruturais.⁷⁸ Na seção das ferramentas avançadas, subseção ferramentas experimentais, pode-se utilizar o MolNetEnhancer inserindo os JobsIDs das redes moleculares, NAP e MS2LDA. Ao término do processamento é possível exportá-los para o Cytoscape.⁶⁴

Uma das primeiras aplicações do MolNetEnhancer foi no estudo metabólico de *superfoods*. Os autores observaram que os *superfoods* tiveram uma maior taxa de anotação de compostos quando comparados aos não *superfoods*.⁸⁷

Qemistree

O Qemistree⁷⁹ é uma ferramenta que combina as redes moleculares do tipo FBMN⁵⁰ com as ferramentas *in silico*, SIRIUS⁸⁸ e CSI:FingerID,⁸⁹ e constrói uma árvore química com base nas impressões digitais moleculares. Isso permite avaliações ecológicas de dados químicos em relação as sequências de DNA geradas para árvores taxonômicas (disponíveis) de espécies, gêneros e famílias.⁷⁹ Para simplificar sua aplicação, o fluxo de trabalho do Qemistree pode ser acessado diretamente da página de resultados da rede molecular processada através da opção: *Visualize with Qemistree*. Automaticamente o fluxo de trabalho será carregado requisitando apenas um arquivo *mgf*, contendo as informações dos espectros gerados para a plataforma do SIRIUS através do pacote MZmine 2 (Esquema 1), seleção do tipo de analisador e modo de ionização dos experimentos. Uma vez que o processamento seja finalizado, as árvores químicas são visualizadas e exploradas através das ferramentas iTOL⁹⁰ e QIIME2.⁹¹

APLICAÇÃO DE REDES MOLECULARES A QUÍMICA DE PRODUTOS NATURAIS

Um levantamento realizado na *Web of Science* (ISI), empregando como palavras-chave, “*molecular networking*” e “*natural products*” apontou que 711 trabalhos foram publicados em periódicos internacionais especializados entre 2013 e 2021 (01 de janeiro de 2013 a 20 março de 2021), como mostrado na Figura 4. Uma inspeção detalhada dos resultados confirmou a aplicação de redes moleculares em estudos de desreplicação, priorização de extratos e frações, exploração de interações ecológicas, caracterização de genes e vias biossintéticas. A versatilidade das redes moleculares, frequentemente

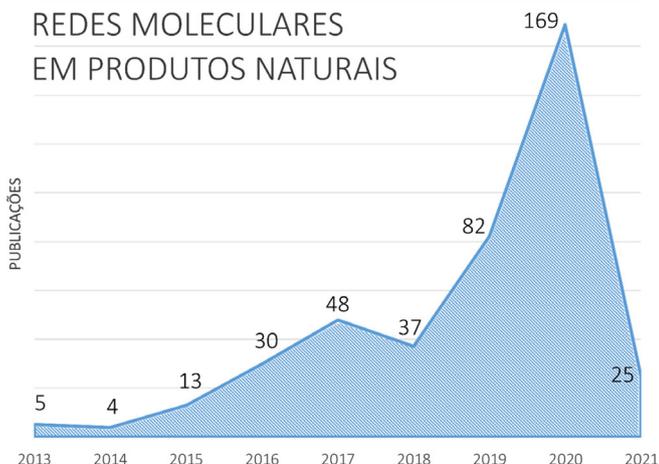


Figura 4. Levantamento de artigos científicos publicados entre 2013 e 2021 pela *Web of Science* utilizando como palavras-chave: *Molecular Networking* e *Natural Products*

associada a metabólica e aos produtos naturais, reflete o crescente número e qualidade das publicações.

Contudo, o surgimento de novas ferramentas exige um período de adaptação. Até o momento, a capacidade média de anotação do GNPS não ultrapassou 20% dos íons totais para um estudo. Por outro lado, a anotação com base em reações de fragmentação de um íon,⁴² presente em um nodo, pode levar a novas estruturas ou permitir o isolamento guiado para a completa caracterização de novos produtos naturais.⁴⁷

Estudos bioguiados e priorização de frações e extratos para anotação e isolamento de novos quimiotipos bioativos

Estudos de desreplicação e descoberta de novos agentes bioativos a partir de matrizes biológicas complexas (produtos naturais) são os tipos de estudos mais aplicados em redes moleculares. Naman e colaboradores⁹² analisaram 10 cianobactérias do gênero *Symploca* sp. na busca por substâncias citotóxicas. As frações analisadas pelas redes moleculares e a biblioteca do GNPS levou a identificação de um peptídeo inédito com ação citotóxica.

Palicourea sessilis (Família Rubiaceae) foi analisada por redes moleculares, com a utilização do *Dictionary of Natural Products* combinado a ferramenta de fragmentação *in silico* DNP-ISDB. Os autores identificaram e, posteriormente, confirmaram por isolamento, oito alcaloides monoterpênicos, sendo três inéditos com moderada atividade anticolinesterásica.⁹³

Olivon e colaboradores⁹⁴ utilizaram as redes moleculares em combinação com dados de bioatividade contra o vírus Chikungunya e sua relação de sinalização Wnt (regulador de eventos embrionários) sob um conjunto taxonômico de 107 espécies de Euphorbiaceae (Nova Caledonian). 292 extratos (ativos e inativos) foram gerados, entre extratos de folhas, cascas, galhos, plantas e frutos inteiros. O estudo levou a priorização e posteriormente ao isolamento e elucidação de quatro ésteres de 12-desoxiforbol (dois inéditos) com atividade contra o vírus Chikungunya.

O estudo bioguiado com *Angelica keiskei* (Apiaceae) e redes moleculares levou à descoberta de chalconas com propriedades antimicrobianas contra *Staphylococcus aureus*. Os resultados da rede foram analisados por ferramentas quimiométricas, indicando o agrupamento das chalconas como responsável pela ação antimicrobiana dos extratos e frações. Foram isolados análogos como o xanthoangelol K, cuja atividade antimicrobiana foi pela primeira vez reportada.⁹⁵

Hou e colaboradores,⁹⁶ utilizando dados de CLAE-EM e EMⁿ, redes moleculares e dados de RMN de ¹H, identificaram ciclohexadepsipeptídeos bioativos (crysogeaminas A –G) de *Penicillium chrysogenum*.

Redes moleculares combinadas a bancos de dados *in house* foram utilizadas para estudos de cascas de *Geissospermum laeve* (Família Apocynaceae). Três novos alcaloides indólicos monoterpênicos foram encontrados. O estudo ainda realizou ensaios antiparasitários e citotóxicos para as substâncias isoladas.³⁶

A estratégia metabólica baseada em espectrometria de massas e redes moleculares também foi recentemente utilizada para caracterizar compostos com atividade larvicida contra *Aedes aegypti*. Utilizando o MetaboAnalyst, redes moleculares e ensaios biológicos para os extratos da espécie *Annona crassiflora* foram isolados e caracterizados diversas acetogeninas ativas contra as larvas do mosquito *Aedes aegypti* (principal responsável pela transmissão de dengue e outras arboviroses no Brasil).³⁸

Redes moleculares têm sido aplicadas também para a caracterização de compostos anticancer⁹⁷ e antibiofilme⁹⁸ de organismos marinhos. Contudo, o grande potencial dessa estratégia pode ser verificado pelo trabalho de Floros e colaboradores⁹⁹ analisando cerca

de 1000 microrganismos marinhos, dos quais mais de 70% não haviam sido taxonomicamente caracterizados. Os autores avaliaram o espaço químico de produtos naturais marinhos priorizando extratos na ausência de informações genômicas ou de atividade biológica. Agrupamentos específicos para algumas cepas levaram ao isolamento e elucidação de duas substâncias novas, os ácidos marídricos A e B.⁹⁹

Uma derivação mais recente das redes moleculares, denominada *bioactive molecular network*, foi desenvolvida por Nothias e colaboradores para acelerar a priorização e isolamento de substâncias em estudos bioguiados.¹⁰⁰ O valor de bioatividade dos extratos (independente do teste de bioatividade) é associado as quantificações dos íons detectados nos experimentos de EM (EM¹) nas redes moleculares. Após identificar as correlações significativas, os dados são visualizados através da rede, indicando os nodos bioativos, facilitando assim seu processo de desrepliação. Contudo, é importante ressaltar que a intensidade dos íons detectados pode não refletir a realidade da concentração, uma vez que essa intensidade está diretamente relacionada a afinidade prótonica ou a capacidade de coordenação.¹⁰¹

Interações químicas, ecológicas e variação metabólica

Existem diversos estudos em redes moleculares relacionados à ecologia química, incluindo relações de comensalismo, mutualismo e simbiose, para as interações positivas, como também relações de predação, parasitismo e antibiose.^{8,102} Entre as diferentes ferramentas ômicas utilizadas para compreensão destes fenômenos,¹⁰³ as redes moleculares ganharam destaque para eventos de sinalização ou defesa química.

Caraballo-Rodríguez,¹⁰⁴ observou a expressão diferenciada de metabólitos produzidos por endófitos (actinobactérias e fungos) de *Lychnophora ericoides* (Asteraceae) sob diferentes formas de tratamento. Culturas isoladas e co-cultivos mostraram que actinobactérias afetaram negativamente o crescimento e desenvolvimento dos fungos. A rede molecular indicou algumas classes e análogos responsáveis pelo processo de inibição dos fungos.

Vallet *et al.*¹⁰⁵ notaram biotransformações de surfactinas em experimentos de co-cultivo entre *Bacillus subtilis* e o fungo *Paraconiothyrium variabile* (endófito de *Cephalotaxus harringtonia* - Taxaceae). Análogos presentes nos agrupamentos das surfactinas apontaram uma diferença de massa de 18 Da, indicando uma perda de água entre os nodos do *cluster*. Esse estudo sugere uma interação interespecífica, onde a perda de água das surfactinas produzidas pela bactéria aumenta toxicidade e resistência contra os fungos.

Em um estudo brasileiro com espécies de *Palitua*, que habitam o mesmo ambiente dos corais, foi observado um efeito ambiental na química das espécies estudadas.¹⁰⁶

Redes moleculares foram usadas para avaliação química de recifes de corais distribuídos em diferentes locais do planeta. Os autores apontaram diferenças de massas para grupos putativos como H₂, CH₂, COCH₂, dependendo da origem do táxon.¹⁰⁷ Dessa forma, a rede molecular foi capaz de ajudar a anotar de modo simplificado a variação metabólica que ocorre nesses holobiontes, mostrando como o ambiente afeta diretamente a produção metabólica.

Um estudo metabolômico utilizando redes moleculares, possibilitou a identificação dos metabólitos responsáveis pela resistência em sojas contaminadas pelo fungo causador da ferrugem asiática. O estudo mostra o potencial da estratégia para o controle de pragas.¹⁰⁸ Finalmente, um estudo utilizando de redes moleculares para caracterização de bileverdina, serviu de partida para a identificação de um complexo metabólito-proteína de cor azul, que sob a pele amarela das rãs, confere um tom esverdeado utilizado em processos de camuflagem em folhas.¹⁰⁹

Genômica funcional, estudos de biossíntese via redes moleculares

A associação da rede molecular com estudos genéticos e biossíntese também tem ganhado destaque na área de produtos naturais. A análise de três espécies de cianobactérias marinhas *Moorea producens* 3L, *Moorea producens* JHB, e *Moorea bouillonii* PNG levou à descoberta de três novas substâncias pertencentes a uma nova classe, chamadas de columbamidas A, B e C. As cianobactérias foram submetidas às análises genômicas comparando as informações obtidas com as análises metabolômica através das redes moleculares.¹¹⁰

Uma abordagem similar foi descrita por Maansson *et al.*,¹¹¹ a partir de estudos genômicos combinado às redes moleculares para análise de metabólitos bioativos de treze bactérias marinhas Gram-negativas. As análises de *pan-plots* e *core-plots* apontaram que apenas 2% das características foram compartilhadas entre todas as linhagens, enquanto 30% foram específicas a uma única linhagem. A rede molecular apontou um agrupamento característico das substâncias conhecidas e guiou a descoberta de dois novos análogos de tiomarínol.

Parkinson e colaboradores¹¹² também utilizaram a metabolômica, redes moleculares e genômica para a descoberta de novos produtos naturais. Dois peptídeos não ribossomais produzidos por espécies da actinobactéria do gênero *Streptomyces* e possíveis análogos estavam agrupados em um mesmo *cluster*, os quais foram posteriormente caracterizados pelas vias biossintéticas.

O fungo *Aspergillus nidulans* foi estudado utilizando aminoácidos marcados com isótopos estáveis (SILAA) e redes moleculares para caracterização de suas vias biossintéticas. O cultivo possibilitou a determinação da via da nudalina A e seus análogos, bem como a anotação de fungisporina, ainda não reportada para a espécie.¹¹³

Outro exemplo importante, nesse contexto, ocorreu pela aplicação de proteômica *top-down*, redes moleculares e marcação isotópica para a caracterização da via biossintética de um peptídeo não-ribossomal raro.¹¹⁴

A Tabela 1 mostra outros estudos não relatados até aqui empregando redes moleculares em conjunto com as ferramentas complementares em produtos naturais.

CONCLUSÕES

A abordagem por redes moleculares está entre os grandes avanços das ferramentas computacionais aplicada a química de produtos naturais do século XXI. A visualização de forma estruturada e intuitiva de grandes conjuntos amostrais facilita a compreensão de eventos multifatoriais encontrados em ecologia química, biossíntese, quimiotaxonomia, e principalmente na desrepliação e na descoberta de entidades químicas bioativas. Contudo, como em qualquer abordagem, seu uso indiscriminado e não supervisionado pode levar a erros de anotação. Portanto, recomenda-se uma curagem dos compostos de interesse usando uma abordagem de *double-check*.

O aspecto comunitário das redes moleculares, como a troca de informações entre autores e dados experimentais, é outra importante conquista para o campo dos produtos naturais. A expansão das bases de dados experimentais a partir de esforços individuais – geração de coleções espectrais de alta qualidade – podem ser agrupadas e disponibilizadas, aumentando o grau de confiabilidade das anotações em matrizes biológicas complexas. Os avanços nos protocolos e métodos de detecção ortogonais (tais como RMN, DAD, ELSD e CAD) são fatores necessários para uma completa caracterização dos compostos presentes nas redes. Nesse sentido, é importante ressaltar que a principal funcionalidade das redes moleculares é organização e visualização intuitiva de grandes conjuntos de dados,

Tabela 1. Trabalhos utilizando redes moleculares para estudos em produtos naturais

Fonte Natural	Tipo de Estudo	Relevância	Ferramentas Computacionais	Ref.
Abelhas e colônias microbianas	Ecologia Química	A importância dos flavonoides para sobrevivência de abelhas	CMN/PCA/OPLS-DA	Silva-Junior ¹¹⁵
Actinobactérias	Ecologia Química e Biotransformação	Halogenação natural de flavonoides por actinobactérias	CMN	Lee ¹¹⁶
Bactérias Marinhas	Bioatividade	Estudo da variação dos parâmetros de crescimento para a produção de metabólitos bioativos	FBMN	Conde-Martínez ¹¹⁷
Barata do Mar	Bioatividade	Analgésico e anti-inflamatório de derivados do ácido Z-9-hexadecenoico	CG-EM	Yue ¹¹⁸
Corais	Ecologia e Ambiental	Diferenciação do metabolismo de dois zoantários coletados ao longo da costa brasileira	CMN	Costa-Lotuf ¹⁰⁶
Esponja marinha	Caracterização	Quinoides azuis zwitteriônicos	CMN	Bonneau ¹¹⁹
Fungo fitopatogênico	Biossíntese	Elucidação da rota biosintética de dibenzoepirocetais	CMN	Zhu ¹²⁰
Fungos	Biotransformação	Oxidações do anéis B e C de galatos de epigalocatequinas	FBMN	Xie ⁴⁵
Fungos de formigas	Ecologia	Produção metabólica da interação de fungos e bactérias	CMN	Boya ¹²¹
Plantas	Bioatividade	alcaloides 3-aminospirostanos anotados por redes moleculares	CMN	Gazolla ⁴⁷
Plantas	Desreplicação e bioatividade	Elucidação e priorização de alcaloides quinolínicos com atividade tripanocida	CMN	Gaudry ¹²²
Plantas	Anotação	Estudo de fragmentação de flavonoides glicosilados	CMN	Pilon ⁹
Plantas	Biossíntese	Elucidação da rota metabólica de alcaloides eritrínicos	FBMN	Chacon ¹²³
Sapos	Ecologia Química	Primer relato de fluorescência em Sapos	CMN	Taboada ⁴⁹

e que o processo de anotação e principalmente de identificação de compostos é dependente de estratégias complementares, que podem ser através metodologias de desreplicação ou por meio de isolamento e caracterização utilizando de técnicas clássicas em PN.

Como perspectiva, a abordagem pode incorporar outras funcionalidades como dados quimiométricos, genômicos e/ou farmacológicos de modo a gerar uma plataforma multifuncional para a pesquisa em produtos naturais.

AGRADECIMENTOS

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP - ACP [#2019/15889-0 e 2016/13286-8] e à RRS [#17/18922-2 e #19/05026-4]), à FINATEC pela bolsa concedida a ACP - #807/2020), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela bolsa concedida a AFM [153221/2018-6] e ao CNPq pelo projeto concedido a JGA (#436333/2018-0). Agradecemos também ao especial apoio da FAPESP ao projeto temático sob coordenação de NPL.

REFERÊNCIAS

- Gaudêncio, S. P.; Pereira, F.; *Nat. Prod. Rep.* **2015**, *32*, 779.
- Valli, M.; Russo, H. M.; Pilon, A. C.; Pinto, M. E. F.; Dias, N. B.; Freire, R. T.; Castro-Gamboa, I.; Bolzani, V. D. S.; *Phys. Sci. Rev.* **2019**, *4*.
- Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kaponov, 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, C. A. P.; 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.; Pogliano, K.; Lington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N.; *Nat. Biotechnol.* **2016**, *34*, 828.
- Allard, P. M.; Genta-Jouve, G.; Wolfender, J. L.; *Curr. Opin. Chem. Biol.* **2017**, *36*, 40.
- Pilon, A. C.; Selegato, D. M.; Fernandes, R. P.; Bueno, P. C. P.; Pinho, D. R.; Neto, F. C.; Freire, R. T.; Castro-Gamboa, I.; Bolzani, V. S.; Lopes, N. P.; *Quim. Nova* **2020**, *43*, 329.
- Bueno, P. C. P.; Lopes, N. P.; *ACS Omega* **2020**, *5*, 1752.
- Ernst, M.; Silva, D. B.; Silva, R. R.; Vêncio, R. Z. N.; Lopes, N. P.; *Nat. Prod. Rep.* **2014**, *31*, 784.

8. Fox Ramos, A. E.; Evanno, L.; Poupon, E.; Champy, P.; Beniddir, M. A.; *Nat. Prod. Rep.* **2019**, *36*, 960.
9. Pilon, A. C.; Gu, H.; Raftery, D.; Bolzani, V. D. S.; Lopes, N. P.; Castro-Gamboa, I.; Carnevale Neto, F.; *Anal. Chem.* **2019**, *91*.
10. 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.; *Nat. Protoc.* **2020**, *15*, 1954.
11. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; De Felício, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C.; *J. Nat. Prod.* **2013**, *76*, 1686.
12. 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.; *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1743.
13. Cortelo, P. C.; Demarque, D. P.; Dusi, R. G.; Albernaz, L. C.; Braz-Filho, R.; Goncharova, E. I.; Bokesch, H. R.; Gustafson, K. R.; Beutler, J. A.; Espindola, L. S.; *Cells* **2021**, *10*, 691.
14. Mannochio-Russo, H.; Bueno, P. C. P.; Bauermeister, A.; de Almeida, R. F.; Dorrestein, P. C.; Cavalheiro, A. J.; Bolzani, V. S.; *J. Nat. Prod.* **2020**, *83*, 3239.
15. Silva, E.; Perez da Graça, J.; Porto, C.; Martin do Prado, R.; Nunes, E.; Corrêa Marcelino-Guimarães, F.; Conrado Meyer, M.; Jorge Pilau, E.; *Metabolites* **2021**, *11*, 179.
16. Tangerina, M. M. P.; Furtado, L. C.; Leite, V. M. B.; Bauermeister, A.; Velasco-Alzate, K.; Jimenez, P. C.; Garrido, L. M.; Padilla, G.; Lopes, N. P.; Costa-Lotufo, L. V.; Pena Ferreira, M. J.; *PLoS One* **2020**, *15*, e0244385.
17. Santos, C. L. G.; Angolini, C. F. F.; Neves, K. O. G.; Costa, E. V.; Souza, A. D. L.; Pinheiro, M. L. B.; Koolen, H. H. F.; Silva, F. M. A.; *Rapid Commun. Mass Spectrom.* **2020**, *34*.
18. Soares, V.; Taujale, R.; Garrett, R.; da Silva, A. J. R.; Borges, R. M.; *Phytochem. Anal.* **2019**, *30*, 132.
19. Calheiros de Carvalho, A.; De Camillis Rodrigues, L.; Ribeiro, A. I.; Fernandes da Silva, M. F. das G.; Soman de Medeiros, L.; Moura Veiga, T. A.; *Molecules* **2019**, *25*, 153.
20. Paulo, B. S.; Sigrist, R.; Angolini, C. F. F.; De Oliveira, L. G.; *ChemistrySelect* **2019**, *4*, 7785.
21. Paz, W. H. P.; de Oliveira, R. N.; Heerd, G.; Angolini, C. F. F.; S. de Medeiros, L.; Silva, V. R.; Santos, L. S.; Soares, M. B. P.; Bezerra, D. P.; Morgon, N. H.; Almeida, J. R. G. S.; da Silva, F. M. A.; Costa, E. V.; Koolen, H. H. F.; *J. Nat. Prod.* **2019**, *82*, 2220.
22. de Almeida, R. T. R.; do Prado, R. M.; Porto, C.; dos Santos, G. T.; Huws, S. A.; Pilau, E. J.; *Sci. Rep.* **2018**, *8*, 17971.
23. Bauermeister, A.; Velasco-Alzate, K.; Dias, T.; Macedo, H.; Ferreira, E. G.; Jimenez, P. C.; Lotufo, T. M. C.; Lopes, N. P.; Gaudêncio, S. P.; Costa-Lotufo, L. V.; *Front. Microbiol.* **2018**, *9*.
24. Philippus, A. C.; Zatelli, G. A.; Wanke, T.; Gabriela de A. Barros, M.; Kami, S. A.; Lhullier, C.; Armstrong, L.; Sandjo, L. P.; Falkenberg, M.; *RSC Adv.* **2018**, *8*, 29654.
25. 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.; *PLOS Comput. Biol.* **2018**, *14*, e1006089.
26. Trivella, D. B. B.; de Felício, R.; *mSystems* **2018**, *3*.
27. de Oliveira, G.; Carnevale Neto, F.; Demarque, D.; de Sousa Pereira-Junior, J.; Sampaio Peixoto Filho, R.; de Melo, S.; da Silva Almeida, J.; Lopes, J.; Lopes, N.; *Planta Med.* **2016**, *83*, 636.
28. Caraballo-Rodríguez, A. M.; Dorrestein, P. C.; Pupo, M. T.; *Sci. Rep.* **2017**, *7*, 5373.
29. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E.; *Nucleic Acids Res.* **2021**, *49*, D1388.
30. Smith, C. A.; Maille, G. O.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G.; *Ther. Drug Monit.* **2005**, *27*, 747.
31. Shen, V. K.; Siderius, D. W.; Kreckelberg, W. P.; Hatch, H. W.; *NIST Stand. Ref. Database* **2017**, 2014.
32. Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmüller, E.; Dörmann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.; Fernie, A. R.; Steinhauser, D.; *Bioinformatics* **2005**, *21*, 1635.
33. Schulte, C. F.; Tolmie, D. E.; Maziuk, D.; Nakatani, E.; Akutsu, H.; Yao, H.; others; *Nucleic Acids Res.* **2007**, *36*, D402.
34. Pilon, A. C.; Valli, M.; Dametto, A. C.; Pinto, M. E. F.; Freire, R. T.; Castro-Gamboa, I.; Andricopulo, A. D.; Bolzani, V. S.; *Sci. Rep.* **2017**, *7*.
35. Bandeira, N.; Tsur, D.; Frank, A.; Pevzner, P. A.; *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 6140.
36. Fox Ramos, A. E.; Alcover, C.; Evanno, L.; Maciuk, A.; Litaudon, M.; Duplais, C.; Bernadat, G.; Gallard, J. F.; Jullian, J. C.; Mouray, E.; Grellier, P.; Loiseau, P. M.; Pomel, S.; Poupon, E.; Champy, P.; Beniddir, M. A.; *J. Nat. Prod.* **2017**, *80*, 1007.
37. Mendes Resende, J. V.; de Sá, N. M. D.; de Oliveira, M. T. L.; Lopes, R. C.; Garrett, R.; Moreira Borges, R.; *Phytochem. Lett.* **2020**, *36*, 99.
38. Demarque, D. P.; Dusi, R. G.; de Sousa, F. D. M.; Grossi, S. M.; Silvério, M. R. S.; Lopes, N. P.; Espindola, L. S.; *Sci. Rep.* **2020**, *10*, 1051.
39. Aksenov, A. A.; Da Silva, R.; Knight, R.; Lopes, N. P.; Dorrestein, P. C.; *Nat. Rev. Chem.* **2017**, *1*, 1.
40. Gribovskaja, I.; Halskau, Ø.; Laporte, G.; *Networks* **2007**, *49*, 199.
41. Guaratini, T.; Vessecchi, R.; Pinto, E.; Colepicolo, P.; Lopes, N. P.; *J. Mass Spectrom.* **2005**, *40*, 963.
42. Demarque, D. P.; Crotti, A. E. M.; Vessecchi, R.; Lopes, J. L. C.; Lopes, N. P.; *Nat. Prod. Rep.* **2016**, *33*, 432.
43. Alcoriza-Balaguer, M. I.; García-Cañaveras, J. C.; López, A.; Conde, I.; Juan, O.; Carretero, J.; Lahoz, A.; *Anal. Chem.* **2019**, *91*, 836.
44. Frank, A. M.; Bandeira, N.; Shen, Z.; Tanner, S.; Briggs, S. P.; Smith, R. D.; Pevzner, P. A.; *J. Proteome Res.* **2008**, *7*, 113.
45. Xie, H.-F.; Kong, Y.-S.; Li, R.-Z.; Nothias, L.-F.; Melnik, A. V.; Zhang, H.; Liu, L.-L.; An, T.-T.; Liu, R.; Yang, Z.; Ke, J.-P.; Zhang, P.; Bao, G.-H.; Xie, Z.-W.; Li, D.-X.; Wan, X.-C.; Dai, Q.-Y.; Zhang, L.; Zhao, M.; An, M.-Q.; Long, Y.-H.; Ling, T.-J.; *J. Agric. Food Chem.* **2020**, *68*, 7995.
46. Nothias, L. F.; Nothias-Esposito, M.; da Silva, R.; Mingxun, W.; Protsyuk, I.; Zhang, Z.; Petras, D.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; Paolini, J.; Alexandrov, T.; Litaudon, M.; Dorrestein, P. C.; *J. Nat. Prod.* **2018**, *81*, 75.
47. Gazolla, M. C.; Marques, L. M. M.; Silva, M. G.; Araújo, M. T. M. F.; Mendes, R. L.; Silva Almeida, J. R. G.; Vessecchi, R.; Lopes, N. P.; *Rapid Commun. Mass Spectrom.* **2020**, *34*.
48. Nothias-Esposito, M.; Nothias, L. F.; Da Silva, R. R.; Retailleau, P.; Zhang, Z.; Leyssen, P.; Roussi, F.; Touboul, D.; Paolini, J.; Dorrestein, P. C.; Litaudon, M.; *J. Nat. Prod.* **2019**, *82*, 1459.
49. Taboada, C.; Brunetti, A. E.; Pedron, F. N.; Carnevale Neto, F.; Estrin, D. A.; Bari, S. E.; Chemes, L. B.; Peoporine Lopes, N.; Lagorio, M. G.; Faivovich, J.; *Proc. Natl. Acad. Sci.* **2017**, *114*, 3672.
50. 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.; Kamenfk, 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.; Neuweiger, 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.; *Nat. Methods* **2020**, *17*, 905.
51. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M.; *BMC Bioinformatics* **2010**, *11*, 395.
52. Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D.; *Anal. Chem.* **2017**, *89*, 7836.
53. Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, W.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O.; *Nat. Methods* **2016**, *13*, 741.
54. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; Vanderghelyst, J.; Fiehn, O.; Arita, M.; *Nat. Methods* **2015**, *12*, 523.
55. Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G.; *Anal. Chem.* **2012**, *84*, 5035.
56. Hoffmann, N.; Rein, J.; Sachsenberg, T.; Hartler, J.; Haug, K.; Mayer, G.; Alka, O.; Dayalan, S.; Pearce, J. T. M.; Rocca-Serra, P.; Qi, D.; Eisenacher, M.; Perez-Riverol, Y.; Vizcaíno, J. A.; Salek, R. M.; Neumann, S.; Jones, A. R.; *Anal. Chem.* **2019**, *91*, 3302.
57. Katajamaa, M.; Miettinen, J.; Orešič, M.; *Bioinformatics* **2006**, *22*, 634.
58. Myers, O. D.; Sumner, S. J.; Li, S.; Barnes, S.; Du, X.; *Anal. Chem.* **2017**, *89*, 8696.
59. Jarmusch, A. K.; Wang, M.; Aceves, C. M.; Advani, R. S.; Aguirre, S.; Aksenov, A. A.; Aleti, G.; Aron, A. T.; Bauermeister, A.; Bolleddu, S.; Bouslimani, A.; Caraballo Rodríguez, A. M.; Chaar, R.; Coras, R.; Elijah, E. O.; Ernst, M.; Gauglitz, J. M.; Gentry, E. C.; Husband, M.; Jarmusch, S. A.; Jones, K. L.; Kamenik, Z.; Le Gouellec, A.; Lu, A.; McCall, L. I.; McPhail, K. L.; Meehan, M. J.; Melnik, A. V.; Menezes, R. C.; Montoya Giraldo, Y. A.; Nguyen, N. H.; Nothias, L. F.; Nothias-Esposito, M.; Panitchpakdi, M.; Petras, D.; Quinn, R. A.; Sikora, N.; van der Hooft, J. J. J.; Vargas, F.; Vrbanc, A.; Weldon, K. C.; Knight, R.; Bandeira, N.; Dorrestein, P. C.; *Nat. Methods* **2020**, *17*, 901.
60. Aksenov, A.; Laponogov, I.; Zhang, Z.; Doran, S. L.; Belluomo, I.; Veselkov, D.; Bittremieux, W.; Nothias, L. F.; Nothias-Esposito, M.; Maloney, K.; Misra, B.; Melnik, A.; Jones, K.; Dorrestein, K.; Panitchpakdi, M.; Ernst, M.; van der Hooft, J.; Gonzalez, M.; Carrazzone, C.; Amézquita, A.; Callewaert, C.; Morton, J.; Quinn, R.; Bouslimani, A.; Albarracín Orío, A.; Petras, D.; Smania, A.; Couvillion, S.; Burnet, M.; Nicora, C.; Zink, E.; Metz, T.; Artaev, V.; Humston-Fulmer, E.; Gregor, R.; Meijler, M.; Mizrahi, I.; Eyal, S.; Anderson, B.; Dutton, R.; Lugan, R.; Boulch, P. Le; Guittou, Y.; Prevost, S.; Poirier, A.; Dervilly, G.; Bizec, B. Le; Fait, A.; Persi, N. S.; Song, C.; Gashu, K.; Coras, R.; Guma, M.; Manasson, J.; Scher, J.; Barupal, D.; Alseikh, S.; Fernie, A.; Mirnezami, R.; Vasiliou, V.; Schmid, R.; Borisov, R.; Kulikova, L.; Knight, R.; Wang, M.; Hanna, G.; Dorrestein, P.; Veselkov, K.; **2020**.
61. Smirnov, A.; Qiu, Y.; Jia, W.; Walker, D. I.; Jones, D. P.; Du, X.; *Anal. Chem.* **2019**, *91*, 9069.
62. Qiu, F.; Lei, Z.; Sumner, L. W.; *Anal. Chim. Acta* **2018**, *1037*, 316.
63. Remorosa, C. A.; Mak, T. D.; De Leoz, M. L. A.; Mirokhin, Y. A.; Stein, S. E.; *Anal. Chem.* **2018**, *90*, 8977.
64. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T.; *Genome Res.* **2003**, *13*, 2498.
65. Vargas, F.; Weldon, K. C.; Sikora, N.; Wang, M.; Zhang, Z.; Gentry, E. C.; Panitchpakdi, M. W.; Caraballo-Rodríguez, A. M.; Dorrestein, P. C.; Jarmusch, A. K.; *Rapid Commun. Mass Spectrom.* **2020**, *34*.
66. Schymanski, E. L.; Ruttkies, C.; Krauss, M.; Brouard, C.; Kind, T.; Dührkop, K.; Allen, F.; Vaniya, A.; Verdegem, D.; Böcker, S.; Rousu, J.; Shen, H.; *J. Cheminform.* **2017**, *1*.
67. Ramos, A. E. F.; Pogam, P. Le; Alcover, C. F.; Otogo, E.; Nang, N.; Cauchie, G.; Hazni, H.; Awang, K.; Bréard, D.; Echavarrén, A. M.; Frédérick, M.; Gaslonde, T.; Girardot, M.; Grougnet, R.; Kirillova, M. S.; Kritsanida, M.; Lémus, C.; Ray, A. Le; Lewin, G.; Litaudon, M.; Mambu, L.; Michel, S.; Miloserdov, F. M.; Muratore, M. E.; Richomme-peniguel, P.; Roussi, F.; *Sci. Data* **2019**, *6*, 1.
68. Lei, Z.; Jing, L.; Qiu, F.; Zhang, H.; Huhman, D.; Zhou, Z.; Sumner, L. W.; **2015**.
69. Nikolic, D.; Jones, M.; Dunn, L. S. and W.; *Curr. Metabolomics* **2017**, *5*, 5.
70. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Nishioka, T.; *J. Mass Spec.* **2010**, *45*, 703.
71. Stravs, M. A.; Schymanski, E. L.; Singer, H. P.; Hollender, J.; *J. Mass Spec.* **2013**, *48*, 89.
72. Hodgson, J.; *Nat. Biotechnol.* **2020**, *38*, 19.
73. Sumner, L. W.; Amberg, A. R.; Barrett, D.; Beale, Beger, R.; Daykin, C. A.; Fan, T. W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reilly, M. D.; Thaden, J. J.; Viant, M. R.; *Metabolomics* **2007**, *3*, 211.
74. 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.; *PLoS Comput. Biol.* **2018**, *14*, 1.
75. Da Silva, R. R.; Dorrestein, P. C.; Quinn, R. A.; *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12549.
76. Van Der Hooft, J. J. J.; Wandy, J.; Barrett, M. P.; Burgess, K. E. V.; Rogers, S.; *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 13738.
77. Wandy, J.; Zhu, Y.; Van Der Hooft, J. J. J.; Daly, R.; Barrett, M. P.; Rogers, S.; *Bioinformatics* **2018**, *34*, 317.
78. Ernst, M.; Kang, K. Bin; Caraballo-Rodríguez, A. M.; Nothias, L. F.; Wandy, J.; Wang, M.; Rogers, S.; Medema, M. H.; Dorrestein, P. C.; van der Hooft, J. J. J.; *bioRxiv* (2019), doi: 10.1101/654459.
79. Tripathi, A.; Vázquez-Baeza, Y.; Gauglitz, J. M.; Wang, M.; Dührkop, K.; Nothias-Esposito, M.; Acharya, D. D.; Ernst, M.; van der Hooft, J. J. J.; Zhu, Q.; McDonald, D.; Brejnrod, A. D.; Gonzalez, A.; Handelsman, J.; Fleischauer, M.; Ludwig, M.; Böcker, S.; Nothias, L. F.; Knight, R.; Dorrestein, P. C.; *Nat. Chem. Biol.* **2021**, *17*, 146.
80. Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S.; *J. Cheminform.* **2016**, *8*, 1.
81. Gerlich, M.; Neumann, S.; *J. Mass Spectrom.* **2013**, *48*, 291.
82. 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.; *Phytochemistry* **2020**, *173*.
83. Leão, T.; Wang, M.; Moss, N.; da Silva, R.; Sanders, J.; Nurk, S.; Gurevich, A.; Humphrey, G.; Reher, R.; Zhu, Q.; Belda-Ferre, P.; Glukhov, E.; Whitner, S.; Alexander, K. L.; Rex, R.; Pevzner, P.; Dorrestein, P. C.; Knight, R.; Bandeira, N.; Gerwick, W. H.; Gerwick, L.; *Mar. Drugs* **2021**, *19*.
84. Lee, J.; da Silva, R. R.; Jang, H. S.; Kim, H. W.; Kwon, Y. S.; Kim, J. H.; Yang, H.; *Food Chem.* **2019**, *295*, 368.
85. Blei, D. M.; Ng, A. Y.; Jordan, M. I.; *J. Mach. Learn. Res.* **2003**, *3*, 993.
86. 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.; *J. Cheminform.* **2016**, *8*, 1.

87. Di Ottavio, F.; Gauglitz, J. M.; Ernst, M.; Panitchpakdi, M. W.; Fanti, F.; Compagnone, D.; Dorrestein, P. C.; Sergi, M.; *Food Chem.* **2020**, *313*.
88. Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S.; *Nat. Methods* **2019**, *16*, 299.
89. Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S.; *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12580.
90. Letunic, I.; Bork, P.; *Nucleic Acids Res.* **2019**, *47*, 256.
91. Bolyen, E.; Rideout, J. R.; Dillon, M. R.; Bokulich, N. A.; Abnet, C. C.; Al-Ghalith, G. A.; Alexander, H.; Alm, E. J.; Arumugam, M.; Asnicar, F.; Bai, Y.; Bisanz, J. E.; Bittinger, K.; Brejnrod, A.; Brislawn, C. J.; Brown, C. T.; Callahan, B. J.; Caraballo-Rodríguez, A. M.; Chase, J.; Cope, E. K.; Da Silva, R.; Diener, C.; Dorrestein, P. C.; Douglas, G. M.; Durall, D. M.; Duvallet, C.; Edwardson, C. F.; Ernst, M.; Estaki, M.; Fouquier, J.; Gauglitz, J. M.; Gibbons, S. M.; Gibson, D. L.; Gonzalez, A.; Gorlick, K.; Guo, J.; Hillmann, B.; Holmes, S.; Holste, H.; Huttenhower, C.; Huttley, G. A.; Janssen, S.; Jarmusch, A. K.; Jiang, L.; Kaehler, B. D.; Kang, K. Bin; Keefe, C. R.; Keim, P.; Kelley, S. T.; Knights, D.; Koester, I.; Kosciolk, T.; Kreps, J.; Langille, M. G. I.; Lee, J.; Ley, R.; Liu, Y. X.; Loftfield, E.; Lozupone, C.; Maher, M.; Marotz, C.; Martin, B. D.; McDonald, D.; McIver, L. J.; Melnik, A. V.; Metcalf, J. L.; Morgan, S. C.; Morton, J. T.; Naimey, A. T.; Navas-Molina, J. A.; Nothias, L. F.; Orphanian, S. B.; Pearson, T.; Peoples, S. L.; Petras, D.; Preuss, M. L.; Pruesse, E.; Rasmussen, L. B.; Rivers, A.; Robeson, M. S.; Rosenthal, P.; Segata, N.; Shaffer, M.; Shiffer, A.; Sinha, R.; Song, S. J.; Spear, J. R.; Swafford, A. D.; Thompson, L. R.; Torres, P. J.; Trinh, P.; Tripathi, A.; Turnbaugh, P. J.; Ul-Hasan, S.; van der Hooft, J. J. J.; Vargas, F.; Vázquez-Baeza, Y.; Vogtmann, E.; von Hippel, M.; Walters, W.; Wan, Y.; Wang, M.; Warren, J.; Weber, K. C.; Williamson, C. H. D.; Willis, A. D.; Xu, Z. Z.; Zaneveld, J. R.; Zhang, Y.; Zhu, Q.; Knight, R.; Caporaso, J. G.; *Nat. Biotechnol.* **2019**, *37*, 852.
92. Naman, C. B.; Rattan, R.; Nikoulina, S. E.; Lee, J.; Miller, B. W.; Moss, N. A.; Armstrong, L.; Boudreau, P. D.; Debonsi, H. M.; Valeriote, F. A.; Dorrestein, P. C.; Gerwick, W. H.; *J. Nat. Prod.* **2017**, *80*, 625.
93. Klein-Júnior, L. C.; Cretton, S.; Allard, P. M.; Genta-Jouve, G.; Passos, C. S.; Salton, J.; Bertelli, P.; Pupier, M.; Jeannerat, D.; Heyden, Y. Vander; Gasper, A. L.; Wolfender, J. L.; Christen, P.; Henriques, A. T.; *J. Nat. Prod.* **2017**, *80*, 3032.
94. Olivon, F.; Allard, P. M.; Koval, A.; Righi, D.; Genta-Jouve, G.; Neyts, J.; Apel, C.; Pannecouque, C.; Nothias, L. F.; Cachet, X.; Marcourt, L.; Roussi, F.; Katanaev, V. L.; Touboul, D.; Wolfender, J. L.; Litaudon, M.; *ACS Chem. Biol.* **2017**, *12*, 2644.
95. Caesar, L. K.; Kellogg, J. J.; Kvalheim, O. M.; Cech, R. A.; Cech, N. B.; *Planta Med.* **2018**, *84*, 721.
96. Hou, X. M.; Li, Y. Y.; Shi, Y. W.; Fang, Y. W.; Chao, R.; Gu, Y. C.; Wang, C. Y.; Shao, C. L.; *J. Org. Chem.* **2019**, *84*, 1228.
97. Velasco-Alzate, K. Y.; Bauermeister, A.; Tangerina, M. M. P.; Lotufo, T. M. C.; Ferreira, M. J. P.; Jimenez, P. C.; Padilla, G.; Lopes, N. P.; Costa-Lotufo, L. V.; *Mar. Drugs* **2019**, *17*, 671.
98. Bauermeister, A.; Pereira, F.; Grilo, I. R.; Godinho, C. C.; Paulino, M.; Almeida, V.; Gobbo-Neto, L.; Prieto-Davó, A.; Sobral, R. G.; Lopes, N. P.; Gaudêncio, S. P.; *Environ. Microbiol.* **2019**, *21*, 1099.
99. Floros, D. J.; Jensen, P. R.; Dorrestein, P. C.; Koyama, N.; Diego, S.; Jolla, L.; Diego, S.; Jolla, L.; Spectrometry, C. M.; Diego, S.; Jolla, L.; *Metabolomics* **2017**, *12*, 145.
100. 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.; *J. Nat. Prod.* **2018**, *81*, 758.
101. Crotti, A. E. M.; Vessecchi, R.; Lopes, J. L. C.; Lopes, N. P.; *Quim. Nova* **2006**, *29*, 287.
102. Braga, R. M.; Dourado, M. N.; Araújo, W. L.; *Braz. J. Microbiol.* **2016**, *47*, 86.
103. Brunetti, A. E.; Carnevale Neto, F.; Vera, M. C.; Taboada, C.; Pavarini, D. P.; Bauermeister, A.; Lopes, N. P.; *Chem. Soc. Rev.* **2018**, *47*, 1574.
104. Caraballo-Rodríguez, A. M.; Dorrestein, P. C.; Pupo, M. T.; *Sci. Rep.* **2017**.
105. Vallet, M.; Vanbellingen, Q. P.; Fu, T.; Le Caer, J. P.; Della-Negra, S.; Touboul, D.; Duncan, K. R.; Nay, B.; Brunelle, A.; Prado, S.; *J. Nat. Prod.* **2017**, *80*, 2863.
106. Costa-Lotufo, L. V.; Carnevale-Neto, F.; Trindade-Silva, A. E.; Silva, R. R.; Silva, G. G. Z.; Wilke, D. V.; Pinto, F. C. L.; Sahn, B. D. B.; Jimenez, P. C.; Mendonça, J. N.; Lotufo, T. M. C.; Pessoa, O. D. L.; Lopes, N. P.; *Chem. Commun.* **2018**, *54*, 1952.
107. Hartmann, A. C.; Petras, D.; Quinn, R. A.; Protsyuk, I.; Archer, F. I.; Ransome, E.; Williams, G. J.; Bailey, B. A.; Vermeij, M. J. A.; Alexandrov, T.; Dorrestein, P. C.; Rohwer, F. L.; *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 11685.
108. Silva, E.; da Graça, J. P.; Porto, C.; Martin do Prado, R.; Hoffmann-Campo, C. B.; Meyer, M. C.; de Oliveira Nunes, E.; Pilau, E. J.; *Sci. Rep.* **2020**, *10*, 1.
109. Taboada, C.; Brunetti, A. E.; Lyra, M. L.; Fitak, R. R.; Faigón Soverna, A.; Ron, S. R.; Lagorio, M. G.; Haddad, C. F. B.; Lopes, N. P.; Johnsen, S.; Faivovich, J.; Chemes, L. B.; Bari, S. E.; *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 18574.
110. Kleigrewe, K.; Almaliti, J.; Tian, I. Y.; Kinnel, R. B.; Korobeynikov, A.; Monroe, E. A.; Duggan, B. M.; Marzo, V. Di; Sherman, D. H.; Dorrestein, P. C.; Gerwick, L.; Gerwick, W. H.; *J. Nat. Prod.* **2015**, *78*, 1671.
111. Maansson, M.; Vynne, N. G.; Klitgaard, A.; Nybo, J. L.; Melchiorson, J.; Nguyen, D. D.; Sanchez, L. M.; Ziemert, N.; Dorrestein, P. C.; Andersen, M. R.; Grama, L.; *mSystems* **2016**, *1*, 1.
112. Parkinson, E. I.; Tryon, J. H.; Goering, A. W.; Ju, K. S.; McClure, R. A.; Kemball, J. D.; Zhukovsky, S.; Labeda, D. P.; Thomson, R. J.; Kelleher, N. L.; Metcalf, W. W.; *ACS Chem. Biol.* **2018**, *13*, 1029.
113. Klitgaard, A.; Nielsen, J. B.; Frandsen, R. J. N.; Andersen, M. R.; Nielsen, K. F.; *Anal. Chem.* **2015**, *87*, 6520.
114. Perez, C. E.; Crawford, J. M.; *Biochemistry* **2019**, *58*, 1131.
115. Silva-Junior, E. A.; Paludo, C. R.; Amaral, J. G.; Gallon, M. E.; Gobbo-Neto, L.; Nascimento, F. S.; Lopes, N. P.; *ACS Omega* **2019**, *4*, 15208.
116. Rak Lee, S.; Schalk, F.; Schwitalla, J. W.; Benndorf, R.; Vollmers, J.; Kaster, A.-K.; de Beer, Z. W.; Park, M.; Ahn, M.-J.; Jung, W. H.; Beemelmans, C.; Kim, K. H.; *J. Nat. Prod.* **2020**, *83*, 3102.
117. Conde-Martínez, N.; Bauermeister, A.; Pilon, A. C.; Lopes, N. P.; Tello, E.; *Mar. Drugs* **2019**, *17*.
118. Yue, Y.; Zhang, Q.; Wang, J.; *Mar. Drugs* **2019**, *17*, 395.
119. Bonneau, N.; Chen, G.; Lachkar, D.; Boufridi, A.; Gallard, J.-F.; Retailleau, P.; Petek, S.; Debitus, C.; Evanno, L.; Beniddir, M. A.; Poupon, E.; *Chem. - A Eur. J.* **2017**, *23*, 14454.
120. Zhu, G.; Hou, C.; Yuan, W.; Wang, Z.; Zhang, J.; Jiang, L.; Karthik, L.; Li, B.; Ren, B.; Lv, K.; Lu, W.; Cong, Z.; Dai, H.; Hsiang, T.; Zhang, L.; Liu, X.; *Chem. Commun.* **2020**, *56*, 10171.
121. A. Boya P., C.; Christian, M. H.; Fernández-Marín, H.; Gutiérrez, M.; *Nat. Prod. Commun.* **2019**, *14*, 1934578X1901400.
122. Gaudry, A.; Quirós, L.; Rutz, A.; Dounoue, M.; Kaiser, M.; David, B.; Marcourt, L.; Queiroz, E.; Wolfender, J.-L.; Allard, P.-M.; *Resumo da 67th International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product research (GA) in Cooperation with the French Society of Pharmacognosy AFERP*, Innsbruck, Austria, 2019.
123. Chacon, D. S.; Torres, T. M.; da Silva, I. B.; de Araújo, T. F.; Roque, A. de A.; Pinheiro, F. A. S. D.; Selegato, D.; Pilon, A.; Reginaldo, F. P. S.; da Costa, C. T.; Vilasboa, J.; Freire, R. T.; Voigt, E. L.; Zuanazzi, J. A. S.; Libonati, R.; Rodrigues, J. A.; Santos, F. L. M.; Scortecci, K. C.; Lopes, N. P.; Ferreira, L. D. S.; dos Santos, L. V.; Cavalheiro, A. J.; Fett-Neto, A. G.; Giordani, R. B.; *J. Adv. Res.* **2021**.

