

Molecular Networking: uma importante ferramenta computacional para o processo de identificação metabólica a partir de dados de Espectrometria de Massas de alta resolução

Alexandre S. Avincola¹, Ana Carolina A. T. Bialetzki¹, Tatyane Caruso Fernandes¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Paraná – IFPR

87703-536 – Paranavaí – PR – Brasil

alexandre.avincola@ifpr.edu.br,
ana.tomaz@ifpr.edu.br, tatyane.fernandes@ifpr.edu.br

Abstract. *Metabolite characterization through High Resolution Mass Spectrometry has been widely used nowadays. This technique uses the fragmentation profile of some ion to identify it, however, the identification process to mass fragmentation spectra needs to be compared to the literature and database specific for mass spectrometry, prolonging the time required to identify substances. Molecular Networking is a computational tool, which makes possible to compare many mass fragmentation spectra among them and to databases simultaneously, reducing the time for analysing of the mass fragmentation spectra and speeding up the metabolites identification process.*

Resumo. *A caracterização de metabólitos através da Espectrometria de Massas de alta resolução tem sido a principal técnica utilizada atualmente. Esta técnica utiliza-se do perfil de fragmentação de um íon para identificá-lo, contudo o processo de identificação dos espectros de fragmentação necessita de comparação com bancos de dados específicos para espectrometria de massas e literatura científica, prolongando o tempo de identificação da substância. O Molecular Networking é uma ferramenta computacional que possibilita a comparação de vários espectros de fragmentação entre si e com bancos de dados simultaneamente, reduzindo o tempo análise dos espectros de fragmentação, agilizando o processo de identificação de metabólitos.*

1. Introdução

1.1. A Espectrometria de Massas

A espectrometria de massas (EM), é uma técnica que se baseia na análise de íons em fase gasosa, mensurando seus valores de razão massa/carga (m/z). Ela apresenta vantagens como alta sensibilidade e praticidade, combinadas à possibilidade de se confirmar a identidade dos componentes presentes em amostras biológicas de baixas concentrações [Bouslimani, 2014].

Para a geração dos íons é necessária uma fonte de ionização no espectrômetro de massas sendo a responsável pela dessorção/ionização dos analitos. Uma das fontes de ionização mais conhecida e amplamente utilizada é a fonte do tipo Electrospray (ESI do inglês, *Electrospray Ionization*). A ionização por Electrospray é uma técnica branda de

geração de íon que ocorre em pressão atmosférica, na qual uma solução com os analitos passam através de um capilar carregado com um alto potencial elétrico, na ordem de kV (quilo-volts) [Bhardwaj e Hanley 2014]. A vantagem desta técnica de ionização é obtenção de espectros de fragmentação (EM/EM) que apresentam além da razão massa/carga do íon precursor os sinais relativos aos íons-fragmentos (íons gerados no processo de fragmentação) [Jarmusch e Cooks 2014].

Neste trabalho de revisão pretende-se expor de modo teórico o processo de obtenção do Molecular Networking, partindo do tratamento de dados iniciais provenientes das análises de EM/EM até a obtenção da rede de correlação e interpretação dos dados de fragmentação para a caracterização dos metabólitos analisados. Deste modo é possível evidenciar a contribuição que esta ferramenta computacional desempenha no processo de identificação de substâncias, por agilizar o tempo de tratamento dos dados.

1.2 A Identificação de metabólitos por Espectrometria de Massas

A confirmação da identidade química de metabólitos via Espectrometria de Massas pode ser obtida como auxílio de experimentos de fragmentação uma vez que o perfil de fragmentação de cada molécula é um fator intrínseco de sua estrutura química. Este tipo de análise é importante em estudos de caracterização de metabólitos, pois além de ser utilizado como uma forma de identificar a molécula, o perfil de fragmentação baseado nos íons fragmentos garante a caracterização mais precisa das substâncias de interesse [Ernst et al. 2014].

Alguns íons-fragmentos podem trazer informações importantes a respeito das classes a que a substância investigada pertence, bem como outras informações que auxiliam no processo de elucidação estrutural de metabólitos em estudos de produtos naturais [Capriotti et al. 2012].

Além do perfil de fragmentação, o valor de massa exata dos íons é útil no processo de desrepliação de uma amostra complexa, e está ligado à massa monoisotópica dos elementos presentes no íon, sendo um parâmetro usado na caracterização de metabólitos. Algumas vezes os processos de medições podem apresentar erros, em espectrometria de massas o erro de massa pode ser calculado pela equação (1) a seguir:

$$\text{Erro (ppm)} = \frac{(\text{Massa Monoisot. teor}) - (\text{Massa Monoisot. obs})}{\text{Massa Monoisot. teor}} \times 1000\,000 \quad (1)$$

onde o valor de erro (ppm) é dado pela razão da diferença das massas monoisotópicas, teórica e observada nos espectros de EM/EM, dividido pelo valor da massa monoisotópica teórica e este resultado multiplicado por 1.000.000.

Por esta razão dados obtidos em alta resolução são amplamente utilizados nos estudos de prospecção pois restringem um determinado valor de razão massa/carga a uma pequena quantidade de possíveis substâncias, o que facilita e agiliza no processo de caracterização, porém muitas vezes a identidade só pode ser confirmada com o auxílio de análises de EM/EM [Lei, Huhman e Sumner 2011], [Saito e Matsuda 2010].

O processo natural de interpretação de dados de espectrometria de massas basicamente é feito de forma humanizada, pois a comparação dos dados experimentais com a literatura e/ou bancos de dados é realizada manualmente espectro a espectro e muitas vezes torna-se exaustiva e pouco rentável. Além disso, o processo de comparação dos dados de EM/EM pelo processo humano está sujeito a erros não controlados e imprevistos. O processo humano de análise e comparação dos dados muitas vezes não é minucioso e preciso, dificultando na descoberta de novas moléculas ou identificação do perfil químico de amostra [Yang 2013].

Nesse contexto o Molecular Networking (MN) é uma potente ferramenta computacional desenvolvida pelo professor Dr. Peter Dorrestein da UCSD (Universidade da Califórnia em San Diego) e colaboradores para aplicação em dados de Espectrometria de Massas, e tem como objetivo comparar o perfil de fragmentação dos espectros de EM/EM para os íons precursores e agrupá-los de acordo com sua semelhança de fragmentação, agilizando o processo de identificação e caracterização metabólica [Yang 2013].

2. Metodologia

2.1. Conversão dos dados de EM e obtenção do Molecular Networking

Os dados de EM de alta resolução, para serem visualizados no MN devem seguir o esquema de conversão representado na figura 1.

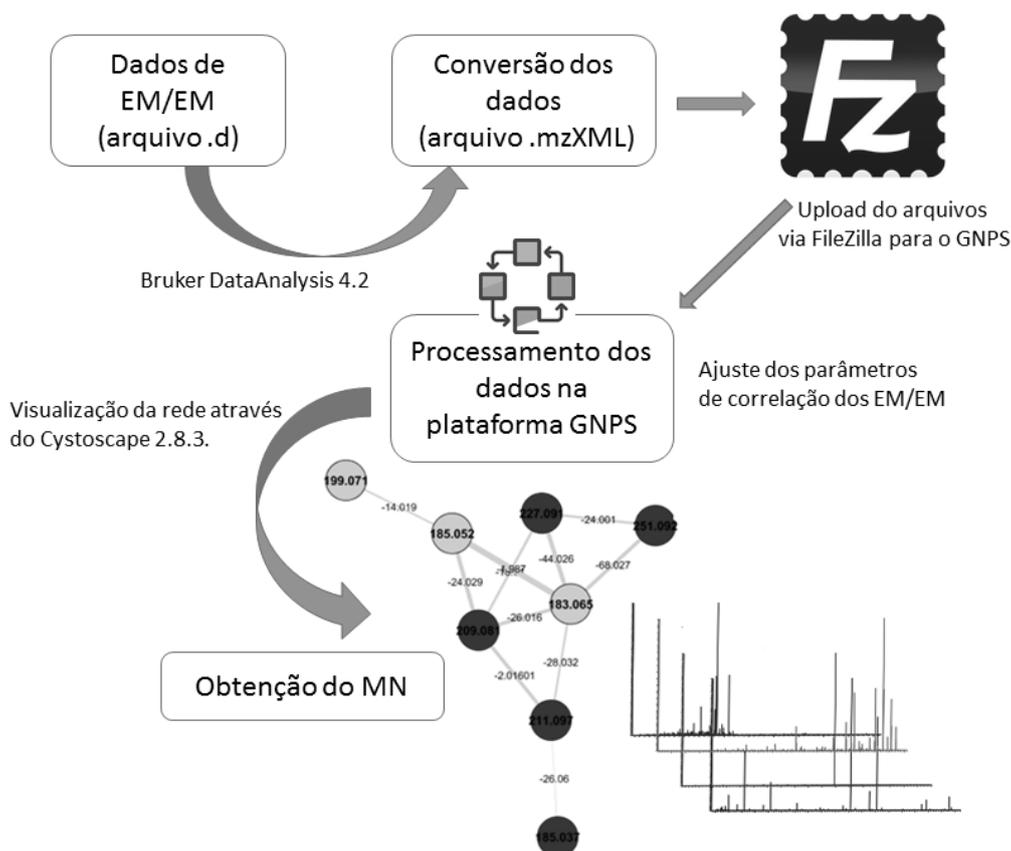


Figura 1: Organograma de conversão dos dados de EM/EM de alta resolução e obtenção do Molecular Networking.

Inicialmente os dados de EM/EM são gerados num formato de extensão D quando obtidos em equipamento Bruker. Para a visualização dos dados e espectros de massas é utilizado o software DataAnalysis 4.2 onde é possível aplicar ajustes à visualização gráfica dos dados de EM ou EM/EM bem como manipular os espectros. Os dados de EM para serem visualizados no Molecular Networking devem apresentar um outro formato de extensão, diferente do formato D. Neste caso os dados devem ser convertidos para o formato mzXML no próprio software DataAnalysis.

O MN para ser gerado precisa ser processado numa plataforma (<https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp>) de livre acesso denominada GNPS (do inglês *The Global Natural Product Social Molecular Networking*) onde um algoritmo específico irá agrupar os dados, compará-los entre si de acordo com suas similaridades e ainda compará-los com dados de EM/EM de compostos depositados em outros bancos de dados também específicos para espectrometria de massas.

Para que os dados, após convertidos, no formato mzXML sejam processados na plataforma é necessário realizar o envio destes até o servidor do GNPS. Assim, utiliza-se o FileZilla, um aplicativo de código aberto para o envio dos arquivos até o servidor para que os dados de EM/EM sejam processados utilizando o algoritmo da plataforma do GNPS.

Uma vez presente na plataforma do GNPS, alguns ajustes devem ser realizados para que os dados possam ser processados, como por exemplo o grau de similaridade dos agrupamentos, a intensidade dos sinais comparados e a quantidade de sinais comuns entre os espectros. Neste caso um dos principais fatores a serem considerados é o valor do cosseno de correlação que varia de 0 a 1, onde espectros totalmente similares apresentam um cosseno igual a 1 e espectros totalmente diferentes um cosseno igual a 0 [Yang 2013].

Após os arquivos serem processados, uma rede de dados correlacionados entre si por linhas é observada e os espectros de massas, cada um, convertidos em nodes que estão ligados por tais linhas. Esta rede só pode ser visualizada no software Cytoscape, sendo a versão 2.8.3 uma das mais usadas. No Cytoscape é possível separar as amostras por colorações atribuindo a uma determinada amostra analisada uma coloração, com isso facilitando a identificação da origem de cada metabólito observado na rede de acordo com a amostra de origem. É possível ainda organizar o layout de visualização do MN selecionando regiões de maior interesse em caixas ou destacando-as com o uso de colorações ou linhas para diferenciá-las ao longo de toda a rede [Watrous et al. 2012].

3. Discussão

O processo de identificação de metabólitos atualmente vem utilizando diversas técnicas, como por exemplo análises em ressonância magnética nuclear (RMN), infravermelho (IV), mas principalmente técnicas hífenadas como a cromatografia líquida acoplada à espectrometria de massas sequencial (CLAE-EM/EM) e outras técnicas combinada [El-Elmat 2013]. O denominador comum destas estratégias é a tentativa de se aproveitar o fato de que, estruturalmente, moléculas semelhantes compartilham características físicas semelhantes, tais como perfil de UV-vis, o tempo

de retenção cromatográfica, a massa molecular, o perfil químico de fragmentação, os deslocamentos químicos de RMN ou ainda propriedades biológicas [Nielsen 2011].

No entanto, mesmo quando várias características são experimentalmente verificadas, a elucidação estrutural completa de uma substância necessita de um longo período de tempo, muitas vezes sendo a principal dificuldade encontrada na caracterização de metabólitos e, em muitos casos, impossibilitando a caracterização de moléculas devido a complexidade observada ou à grande quantidade de dados a serem avaliados [Yang 2013].

Assim, estratégias para facilitar o processo de identificação precisam ser desenvolvidas e aplicadas nos estudos de investigação metabólica, para auxiliar na caracterização de substâncias presentes em amostras complexas de forma reprodutiva e rápida.

Neste contexto, o sistema de alinhamento e comparação de dados - Molecular Networking é uma poderosa ferramenta que auxilia no processo de identificação metabólica de amostras com alto grau de complexidade, como por exemplo extratos de produtos naturais, diminuindo o tempo gasto na interpretação e tratamento de dados [Nguyen 2013].

Além disso, ao inserir arquivos na plataforma do GNPS, os dados de EM/EM são comparados com dados de EM depositados em diversos bancos de dados como MassBank [Horai 2010], Metlin [Smith 2005] e HMDB [Wishart 2013]. Deste modo, além de correlacionar espectros de massas de acordo com suas semelhanças, facilitando no processo de identificação ou elucidação de metabólitos, o MN exclui o processo humano de comparação de dados experimentais com bancos de dados, pois o ele consegue simultaneamente agrupar e comparar dados de fragmentação da amostra com os bancos de dados.

Estes agrupamentos podem ser visualizados em uma rede (figura 2), de modo que os compostos que exibem espectros de fragmentação semelhante fiquem agrupados em clusters. Estes clusters são formados por um conjunto de nodes, que são os círculos observados no MN. Cada círculo é atribuído a um dado valor de m/z para um íon e a comunicação entre os nodes é em relação às suas semelhanças observadas nos espectros de fragmentação destes íons.

A correlação entre os nodes é baseada na similaridade no perfil de fragmentação entre os espectros. Isso se deve ao fato de que cada espectro de fragmentação é convertido em um vetor unitário no espaço dimensional. Os vetores, aos pares, são comparados através do produto vetorial entre eles, que por definição inclui o cosseno do ângulo dos vetores. Se o valor do cosseno for ajustado para um valor igual a 1 poucas correlações serão observadas pois os clusters formados serão aqueles entre compostos iguais que foram encontrados a partir da comparação entre amostra e os bancos de dados. Ao diminuir o valor do cosseno, um maior número de correlações será observado devido a permissividade de correlação estar em torno de uma menor similaridade, o que possibilita uma rede mais complexa. Assim o cosseno do produto vetorial é o valor que define quão seletiva será essa correlação. Muitas vezes um MN eficiente é o que mostra várias correlações entre nodes de compostos conhecidos e desconhecidos, pois além das

correlações estabelecidas, a diferença de massa exata entre os nodes auxilia no processo de identificação metabólica [Yang 2013].

A comunicação entre os nodes no MN é visualizada na forma de linhas, evidenciando quais íons estão correlacionados entre si. Além disso, com o auxílio das linhas é observado a diferença de massa entre os nodes, sugerindo assim os possíveis grupamentos (CH_2 , CH_3 , O_2 entre outros) que diferem estruturalmente os íons correlacionados. Se a diferença entre dois nodes ligados é por exemplo de 28,123 u, este valor é sugestível para uma diferença estrutural entre os íons atribuída à unidade $-\text{CH}_2\text{CH}_2-$ que está ligada à uma das estruturas causando o acréscimo de massa de 28,123 u. Como a única diferença estrutural entre os íons é devido a este acréscimo, logo o perfil de fragmentação será semelhante, justificando a comunicação entre os nodes [Nguyen 2013].

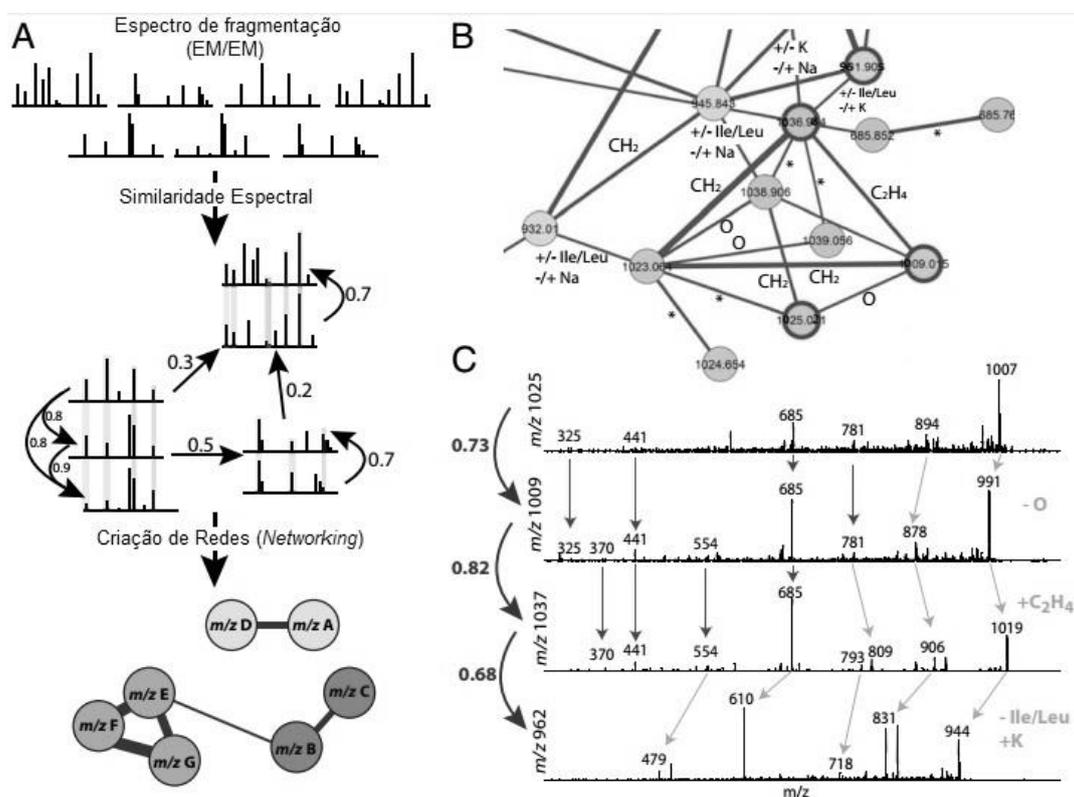


Figura 2: Alinhamento e correlação dos dados de EM/EM (A) para diferentes valores de cosseno e representação do MN e seus clusters (B) e comparação dos espectros de fragmentação (C) [adaptado de Watrous et al. 2012].

Com isso, metabólitos estruturalmente semelhantes, com padrões de fragmentação semelhantes, são agrupados e considerados famílias moleculares, as quais tendem a se agrupar dentro de um cluster na rede molecular [Nguyen 2013]. Esses modelos permitem a exploração visual simultânea de moléculas idênticas, análogos, e/ou famílias de moléculas, dentro de um conjunto de dados único ou múltiplo e de composto por diversas fontes biológicas [Yang 2013].

Esta nova abordagem comparativa dos dados de EM/EM pode agilizar o processo de prospecção metabólica presente em uma amostra complexa. Extratos vegetais, fúngicos ou amostras com alto grau de complexidade apresentam diversos metabólitos e a avaliação de todos os espectros de EM/EM dos metabólitos presentes torna-se um processo longo e inviável. Com isso, utilizando o MN é possível avaliar de uma forma abrangente a diversidade de substâncias presentes nestas amostras, e ainda compará-la com outras amostras ao mesmo tempo. Quanto maior o número de amostras presente na rede, maior são as possibilidades de correlações observáveis no MN, além disso substâncias com o perfil de fragmentação conhecido podem ser inseridos nesta rede auxiliando o processo de desreplicação metabólica.

Para os extratos obtidos a partir de fontes onde meios de cultivo são utilizados, é possível selecionar na rede quais são os íons atribuídos ao meio de cultura e discriminá-los dos demais íons, utilizando um sistema de coloração dos nodes. Além disso, para MN criado com diferentes amostras, é possível utilizar o mesmo sistema de coloração para identificar a origem de cada metabólito visualizado na rede de acordo com sua amostra, facilitando a identificação e análise dos íons observados para cada amostra.

4. Conclusão

Comparado aos métodos tradicionais de trabalho, os quais dependem da análise comparativa e individual dos dados de EM/EM frente a literatura e bancos de dados específicos para a Espectrometria de Massas, o Molecular Networking é uma importante ferramenta no processo de identificação e caracterização metabólica. Desta forma o processo de caracterização/elucidação de substâncias em sistemas complexos (extratos de plantas, fungos, bactérias) ocorre de forma mais rápida e abrangente, sendo atualmente uma das mais eficientes ferramentas computacionais dos estudos metabolômicos.

5. Referências

- Bhardwaj, C.; Hanley, L. (2014) Ion sources for mass spectrometric identification and imaging of molecular species, *Natural Product Reports*, vol. 31, num. 6, p. 756-767.
- Bousslimani, A. et al. (2014) Mass spectrometry of natural products: current, emerging and future technologies, *Natural Product Reports*, vol. 31, num. 6, p. 718-729.
- Capriotti, A. L. et al. (2012) Multiclass mycotoxin analysis in food, environmental and biological matrices with chromatography/mass spectrometry. *Mass spectrometry reviews*, vol. 31, num. 4, p. 466-503.
- El-Elimat, T. et al. (2013) High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products. *Journal of Natural Products*, vol. 76, num. 9, p. 1709–1716.
- Ernst, M. et al. (2014) Mass spectrometry in plant metabolomics strategies: from analytical platforms to data acquisition and processing. *Natural Product Reports*, vol. 31, num. 6, p. 784-806.
- Horai, H. et al. (2010) MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, vol. 45, num. 7, p. 703–714.

- Jarmusch, A. K.; Cooks, R. G. (2014) Emerging capabilities of mass spectrometry for natural products. *Natural Product Reports*, vol. 31, num. 6, p. 730-738.
- Lei, Z.; Huhman, D. V.; Sumner, L. W. (2011) Mass spectrometry strategies in metabolomics. *Journal of Biological Chemistry*, vol. 286, num. 29, p. 25435-25442.
- Nguyen, D. D. et al. (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proceedings of the National Academy of Sciences*, vol. 110, num. 28, p. E2611-E2620.
- Nielsen, K. F. et al. (2011) Dereplication of microbial natural products by LC-DAD-TOFMS. *Journal of Natural Products*, vol. 74, num. 11, p. 2338–2348.
- Saito, K.; Matsuda, F. (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annual review of plant biology*, vol. 61, p. 463-489.
- Smith, C. A. et al. (2005) METLIN: a metabolite mass spectral database. *Therapeutic Drug Monitoring*, vol. 27, num. 6, p. 747–751.
- Watrous, J. et al. (2012) From the Cover: PNAS Plus: Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences*, vol. 109, num. 26, p. E1743–E1752.
- Wishart, D. S. et al. (2013) HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Research*, vol. 41, num. D1, p. D801–D807.
- Yang, J. Y. et al. (2013) Molecular networking as a dereplication strategy. *Journal of Natural Products*, vol. 76, num. 9, p. 1686–1699.