

**(DES)CONSTRUINDO A METABOLÔMICA EM PRODUTOS NATURAIS: UM CONVITE A DISCUSSÃO****Ricardo Moreira Borges<sup>a,\*</sup> e João Victor Mendes Resende<sup>a</sup>**<sup>a</sup>Instituto de Pesquisas de Produtos Naturais Walter Mors, Universidade Federal do Rio de Janeiro, 21941-903 Rio de Janeiro – RJ, Brasil

Recebido em 18/02/2021; aceito em 06/04/2021; publicado na web em 16/06/2021

DECONSTRUCTING METABOLOMICS WITHIN NATURAL PRODUCTS: AN INVITATION FOR A DISCUSSION. The metabolomics approach is widespread and accepted in several lines within life sciences. However, much of what has been called metabolomics could be better defined as a comparison study, regardless of its purpose. Metabolomics finds its definition in the understanding of biological properties and effects on a given organism. In any case, the tools used in metabolomics are undoubtedly useful and should be used in natural products studies, both those already established and those that are under development. Natural products research is in a continuous process of evolution to become a discipline driven by multivariate data with a solid influence of “omic” technologies.

Keywords: metabolomics; natural products; statistics; compound identification.

**INTRODUÇÃO**

Talvez a melhor forma de iniciar esse documento seja deixando claro que nem toda comparação se trata de metabolômica. Durante anos, estudos envolvendo Produtos Naturais (PN) vem sendo descritos como metabolômica apenas por utilizarem uma abordagem não reducionista, que deriva da tradicional abordagem de fracionamento bioguiado, para comparar dois ou mais grupos de amostras.<sup>1</sup> É válido deixar claro que a própria definição de metabolômica como abordagem de pesquisa evoluiu nessa última década. Hoje a metabolômica é mais bem definida como sendo o estudo dos processos químicos envolvendo o metaboloma frente a uma determinada situação, ou seja, a metabolômica está mais ligada ao estudo da importância dos metabólitos à vida e às funções bioquímicas em um determinado organismo.<sup>2</sup> Metaboloma é o nome que se dá para o conjunto de metabólitos (ou pequenas moléculas) produzidos por um indivíduo.<sup>3,4</sup> Define-se como pequenas moléculas aquelas com massa molecular abaixo de 2000 Da ou, fazendo uma referência à técnica de RMN, aquelas moléculas cujo constante de tempo de relaxamento transversal (T2) está na mesma ordem de magnitude que sua constante de tempo de relaxamento longitudinal (T1) em uma solução líquida.<sup>5</sup>

Os conceitos de estatística, principalmente de estatística multivariada, estão intimamente relacionados à metabolômica logo que sempre se deseja inferir alguma informação a uma população de onde as amostras foram obtidas. Portanto, só faz sentido um estudo ser definido como metabolômica se ele permitir uma conclusão estatisticamente significativa acerca do comportamento ou modulação metabólica de determinado organismo em determinada situação. A etapa mais importante nesses estudos, devido ao peso que a estatística tem, é a do planejamento do estudo. Ele deve estar muito bem conectado à hipótese biológica para que os resultados sejam, de fato, conclusivos e reproduzíveis. Note que a definição estabelece que o resultado deve relatar alguma inferência estatística e ela deve ser significativa. Por esse motivo, o rigor estatístico deve ser considerado na etapa inicial de planejamento do estudo o que leva a necessidade de estabelecer o número de réplicas a serem utilizadas. O uso de amostras de controle de qualidade (CQ) é um tópico a parte que merece um documento próprio para discussão.<sup>6-8</sup>

Dessa forma, quando se compara, por exemplo, grupos taxonômicos diferentes trata-se de um estudo de comparação e, talvez, ele até tenha um objetivo de se encontrar substâncias que caracterizam o agrupamento diferencial entre os grupos diferentes; o marcador químico ou biológico. Ou quando se analisa a correlação entre resultados de perfil metabólico de uma série de amostras com resultados de atividade biológica, trata-se de um estudo de correlação entre variáveis de uma série de experimentos. Se um estudo tem como objetivo a caracterização, ainda que qualitativa e quantitativa, de todos os metabólitos de um indivíduo (mesmo que isso seja impossível), trata-se de um estudo de caracterização metabólica e não exatamente de metabolômica. Nesses casos, não há inferência alguma sobre uma hipótese fisiológica ou bioquímica que tenha levado aquele(s) organismo(s) utilizado(s) a produzir substâncias ativas frente aquele ensaio biológico realizado. Siuzdak e colaboradores sugerem o nome activity metabolomics para se referir a estudos realizados a fim de se compreender a importância de pequenas moléculas e seu papel como agentes bioativos,<sup>3</sup> que é muito comum em PN. Sem dúvida, o uso de ferramentas muito utilizadas em metabolômica na busca por substâncias bioativas ou para caracterização taxonômica é benéfico e pode elevar o número de descobertas em PN. De fato, a química de PN está em um processo ativo de evolução para se tornar uma disciplina dirigida por dados multivariados com muita influência das tecnologias “ômicas”. Tomando como referência o Brasil que possui um histórico de longa data com a química de produtos naturais,<sup>9</sup> o número de revisões no assunto publicado em língua portuguesa no próprio periódico Química Nova<sup>10-12</sup> evidencia o interesse por diversos grupos de pesquisa nacionais o que justifica esta reflexão.

**METABOLÔMICA OU APLICAÇÃO COERENTE DA ESTATÍSTICA?**

A estatística é definida como sendo a disciplina que envolve o planejamento, a coleta, a organização, a análise, a interpretação e a apresentação de dados com objetivo de produzir conhecimento a partir desses dados.<sup>13</sup> Logo, a estatística não pode ser considerada uma simples ramificação da matemática, muito embora use a matemática para produzir resultados de maior robustez e capazes de serem reproduzidos uniformemente através de diferentes linguagens e conceitos. A estatística, então, permeia todos os ramos da pesquisa

\*e-mail: ricardo\_mborges@ufrj.br

científica cujo objetivo é de incrementar progressivamente o conhecimento de modo geral. Se considerarmos a matemática como uma linguagem, a estatística pode ser entendida como um meio de tornar padronizada o estudo científico dentre diferentes grupos de pesquisa. A reprodutibilidade é o conceito chave pelo qual a ciência é estabelecida como conhecimento. De forma pragmática, um estudo que fora encaminhado sem o devido rigor analítico e que acaba por não ser passível de reprodução, mesmo quando ele resulta em uma publicação, não contribui para o corpo do conhecimento; é possível até que contribua negativamente.

Note que a definição de Estatística como disciplina se confunde com a própria definição de pesquisa científica e, em parte, da metabolômica. Como na vida, em ciência, termos novos chamam atenção e podem ser usados sem problemas. O objetivo deste documento não é criticar o uso da nomenclatura, mas convidar a comunidade a uma discussão saudável sobre o tema a fim de explicitar pontos que precisam ser dominados para que a pesquisa científica brasileira tenha maior destaque mundial.

Em geral, um estudo envolvendo a metabolômica deve ser planejado detalhadamente à luz da hipótese criada e ela deverá ser testada frente a uma hipótese nula com o devido rigor estatístico. Uma questão importante é a determinação do número de réplicas que deverá ser utilizado e a maneira mais direta de se obter essa determinação é através de um estudo-piloto. Esse estudo-piloto, feito com um número de réplicas reduzido, produzirá as estatísticas necessárias para que o Poder Estatístico seja definido e o tamanho da amostra necessário para se obter o poder estatístico desejável. Por definição, o poder estatístico é a probabilidade de se detectar um efeito quando este efeito está presente.<sup>14</sup> Logo, é de se imaginar que um estudo com um poder estatístico baixo pode não ser reprodutivo. A importância de se documentar o poder estatístico nos estudos é óbvia.

O poder estatístico depende da magnitude do efeito estudado ou observado, do critério de significância estatística escolhido e do tamanho da amostra usado no estudo. O *Metaboanalyst*<sup>15</sup> utiliza um método de cálculo do poder estatístico e tamanho de amostra de forma intuitiva para o usuário. O método requer o uso de um estudo-piloto para explorar a magnitude do efeito desejado sob análise e o critério de significância estatística usado no teste. Em geral, quando o efeito a ser observado é muito grande, a quantidade de réplicas necessárias para produzir um poder estatístico satisfatório é menor. Muitos estudos assumem 80% como sendo um poder estatístico aceitável. O problema está em estudos nos quais o efeito que se deseja observar é muito pequeno. Nesses casos, estudos com um número reduzido de amostras irão produzir, por exemplo, um resultado de PLS-DA (técnica multivariada muito utilizada em metabolômica) com parâmetros de validação desfavoráveis, refletindo um caso comum de superajuste (*overfitting*), quando o modelo perde a capacidade de generalização.

Outro objetivo importante a ser definido na etapa de planejamento do estudo é a minimização da variância biológica e técnica de um experimento. Para que a distinção entre dois grupos estudados seja relevante, é preciso mostrar que a variância entre amostras do mesmo grupo é menor do que a diferença entre os grupos. Muitas vezes a distinção entre os grupos é sutil e nestes casos, o número de réplicas deverá ser maior. É nesta etapa pré-experimental que as amostras deverão ser organizadas em uma ordem aleatória. O objetivo dessa ordem aleatória é garantir que todas as amostras tenham uma probabilidade igual de serem escolhidas e a consequência óbvia disso é que possíveis efeitos de bloco serão menos ocorrentes, assim como a variância técnica do experimento.

Muitos estudos em PN acabaram migrando para o uso de uma abordagem metabolômica da maneira mais direta possível. Em muitos casos, estudos já em andamento foram modificados para “se tornarem” metabolômica e esse tipo de conversão tende a ser incompleta. Nesses casos, a etapa de planejamento do estudo é quase sempre feita às

pressas e desconsideram questões importantes como: (1) objetividade da hipótese; (2) número de réplicas e poder estatístico; e (3) controle de qualidade. Nesta sessão, o tema controle de qualidade será abordado de forma geral, dado que não é objetivo deste documento criar um guia. Sugerimos aqui algumas referências.<sup>6,7,16</sup>

Brevemente, a aplicação de métodos de controle de qualidade e garantia de qualidade provê mecanismos para julgar a qualidade de um experimento e acessar a variância analítica dos dados produzidos.<sup>6</sup> O controle de qualidade não é, de forma alguma, um tema exclusivo da metabolômica, mas é um requerimento para qualquer estudo científico, principalmente aqueles que envolvem química analítica. De forma prática, faz parte do controle de qualidade a determinação das fontes de variâncias técnicas: da etapa de extração, do preparo das amostras, da instrumentação analítica e do processamento dos dados.

As amostras combinadas (*pooled samples*) de todos os indivíduos do estudo são produzidas para representarem todos os metabólitos presentes. Elas são produzidas retirando-se uma alíquota de cada uma das amostras do estudo e as combinando em um *vial* separado. Em seguida, essa amostra combinada, que terá um volume maior do que as amostras do estudo, será dividida e distribuída ao longo da sequência de análise para que as variâncias dos dados produzidos por elas sejam avaliados e indiquem a qualidade do procedimento experimental seguido. Essas amostras não são parte do estudo (não contribuem para o tamanho da amostra) e, por isso, não devem ser incluídas como réplicas. Quando isso é feito, diz-se que os dados estão inflados, por contribuírem a uma distribuição com uma variância falsamente reduzida, o que pode levar a conclusões equivocadas sobre determinado estudo.

Uma opção para a determinação da variância das etapas de extração e preparo de amostra é a inclusão de uma amostra-estoque (e homogênea) externa ao estudo e, se possível, padronizada; de preferência usando a matriz mais semelhante possível e de fácil acesso.

Amostras-branco são amostras preparadas de forma a não conter os metabólitos que se pretende analisar. Elas podem ser feitas usando frascos vazios juntos à sequência inicial de preparo de amostras ou até utilizando placas com meios de cultura sem o organismo que está sendo estudado. Aqueles sinais detectados nas amostras-branco podem ser excluídos (com cautela) como sendo artefatos do preparo das amostras ou substâncias do meio de cultura que não são de interesse. Ainda, o uso de amostras-branco na sequência analítica permitirá a avaliação de contaminantes de arraste (*carryover*), quando um método cromatográfico é utilizado.

## PROPOSTA PARA UM PLANO NACIONAL DE CATALOGAÇÃO EM PRODUTOS NATURAIS

Considerando que um dos objetivos mais comuns em estudos envolvendo PN é a contribuição ao conhecimento dos constituintes químicos de determinada espécie ou grupo de espécies, nós poderíamos desenvolver um trabalho mais eficiente em conjunto. Ora, se existem diversos projetos pelo país usando a mesma justificativa de determinar o perfil químico de uma (ou mais) espécie(s), seria mais eficiente se todos os grupos trabalhassem de forma concertada. Obviamente, isso criaria um problema de financiamento, mas há de se aceitar que já existe um problema sério de financiamento à pesquisa científica no Brasil neste momento. A proposta que pretendemos apresentar aqui promete acelerar de forma progressiva o fluxo de identificação de substâncias conhecidas em amostras da biodiversidade nacional.

Essa proposta envolve o uso de procedimentos validados de análise em LC-HRMS/MS de modo que os tempos de retenção possam ser reproduzidos em diferentes laboratórios acreditados presentes em cada região do país. Esses cinco laboratórios teriam acesso a amostras de controle de qualidade específicas para garantir uma precisão intermediária. Além dessas amostras, um grupo de amostras de

marcação de tempos de retenção de diferentes substâncias com suas estruturas inequivocamente elucidadas poderia ser disponibilizado a cada grupo. Esses, por sua vez, seriam obtidos por doação de cada grupo usuário desse sistema que aceitem cumprir seu objetivo de contribuir com a ciência nacional. A marcação de tempos de retenção permitirá o desenvolvimento de um método de predição de tempos de retenção progressivamente mais acurado e preciso conforme o acúmulo de novas substâncias identificadas no grupo de treinamento de um aprendizado de máquina.<sup>17,18</sup> Então, em conjunto com os dados de MS/MS, que seriam sempre submetidos a bancos de dados predeterminados, permitirá maior confiança na desreplicação de substâncias conhecidas e facilitar a anotação de substâncias desconhecidas. Esses métodos de predição de tempos de retenção, assim como outras propriedades físico-químicas, utilizam uma série de substâncias conhecidas e com propriedades conhecidas como grupo de treinamento para derivar uma equação que permita prever aquelas propriedades para substâncias desconhecidas. Tal como um estudo de relação quantitativa estrutura-atividade (QSAR). Essa equação é produzida via mecanismos matemáticos (p.e. métodos de regressão) usando os descritores moleculares<sup>19</sup> das substâncias conhecidas e é aplicada nos descritores das moléculas das quais as aquelas propriedades se deseja descobrir.

Ainda nesse sentido, esse plano pode, e deve, ser expandido com o uso de dados de RMN para uma confiança ainda maior nessa identificação. Não diferente do crescente uso de dados catalogados de MS para desreplicação, em RMN também é muito importante que as informações obtidas, assim como os espectros, sejam sistematicamente catalogadas. Os métodos de predição (não aqueles ab initio) de dados espectrais e físico-químicos de substâncias orgânicas tem crescido a um passo interessante<sup>20</sup> (<https://pubs.acs.org/doi/abs/10.1021/acs.chemrev.0c00901>) e, com eles, métodos que usam essas ferramentas diretamente na identificação de substâncias em misturas.<sup>21,22</sup> Devido ao fato desses métodos serem baseados em bancos de dados experimentais para a predição de, por exemplo, os deslocamentos químicos de núcleos de uma determinada substância, a presença de outras substâncias similares nesses bancos de dados levarão a um resultado de predição mais acurado. Daí a vantagem de catalogar os dados de qualquer descoberta de forma passível de ser lida por programas computacionais. O mantimento de uma biblioteca de artigos, teses ou outros documentos impressos no escritório para que sejam consultados em busca de dados de deslocamento químico e variações é válido, mas consome um tempo atualmente desnecessário e torna-se pouco reprodutivo ou sistemático.

A dificuldade técnica para essas questões é apenas de logística de distribuição das amostras de controle de qualidade para garantir a precisão entre os diferentes instrumentos a serem utilizados e a aquisição de colunas cromatográficas e outros insumos específicos em comum. Quanto ao costume de catalogar os dados em bancos de dados dedicados, nós acreditamos que ele deva ser incentivado o quanto antes e os periódicos terão papel importante nesse esforço. Acredito que uma nova diretriz, principalmente referente ao plano de desreplicação, possa ser acertado junto ao CNPq e à própria Divisão de Produtos Naturais da Sociedade Brasileira (DPN-SBQ). É de se esperar que o avanço científico previsto por um projeto dessa magnitude é sem precedentes e resolve possíveis limitações impostas pela falta de infraestrutura de alguns grupos. Novamente, permitir um processo contínuo de identificação e catalogação de PN em um ciclo virtuoso poderá levar a resultados de maior impacto científico e ainda acelerar a exploração biotecnológica do arsenal químico da biodiversidade brasileira.

## CONCLUSÕES

Nesta discussão, a definição do termo metabólica está formalmente apresentada e questões importantes acerca do

planejamento do estudo foram levantadas na esperança de incentivar uma discussão para o desenvolvimento de pesquisa de maior impacto científico. Novamente, o objetivo deste documento é a maximização do conhecimento adquirido por estudos envolvendo a biodiversidade, seja eles taxados de metabólica ou não.

A necessidade de um conhecimento fundamental de estatística fica evidente no decorrer da leitura e ele é suficiente em grande parte dos estudos em PN, desde que sejam cuidadosamente considerados. Em casos mais complexos, é indicado o convite a um profissional de estatística para discussão e planejamento do estudo antes que este tenha sido iniciado.

Finalmente, o Brasil tem capacidade científico-intelectual no estado-da-arte e um poderio instrumental suficiente para explorar de forma mais eficiente a sua invejável biodiversidade. Sendo otimista, por mais difícil que pareça atualmente, a ciência de PN pode orquestrar um resgate tecnológico e industrial no país, pelo menos por parte dos cientistas.

## REFERÊNCIAS

1. Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S.; *Acc. Chem. Res.* **2012**, *45*, 288.
2. Leggett, A.; Wang, C.; Li, D. W.; Somogyi, A.; Brusweiler-Li, L.; Brusweiler, R.; *Methods Enzymol.* **2019**, *615*, 407.
3. Rinschen, M. M.; Ivanisevic, J.; Giera, M.; Siuzdak, G.; *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 353.
4. Nicholson, J. K.; Lindon, J. C.; Holmes, E.; *Xenobiotica* **1999**, *29*, 1181.
5. Reher, R.; Kim, H. W.; Zhang, C.; Mao, H. H.; Wang, M.; Nothias, L. F.; Caraballo-Rodriguez, A. M.; Glukhov, E.; Teke, B.; Leao, T.; Alexander, K. L.; Duggan, B. M.; Van Everbroeck, E. L.; Dorrestein, P. C.; Cottrell, G. W.; Gerwick, W. H.; *J. Am. Chem. Soc.* **2020**, *142*, 4114.
6. Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B.; *Metabolomics* **2018**, *14*, 72.
7. Brunius, C.; Shi, L.; Landberg, R.; *Metabolomics* **2016**, *12*, 173.
8. Naz, S.; Vallejo, M.; Garcia, A.; Barbas, C.; *J. Chromatogr. A* **2014**, *1353*, 99.
9. Pinto, A. C.; Rezende, C. M.; Garcez, F. R.; Epifanio, R. A.; *Quim. Nova* **2003**, *26*, 966.
10. Canuto, G.; Costa, J. L.; Cruz, P.; Souza, A.; Faccio, A.; Klassen, A.; Rodrigues, K.; Tavares, M.; *Quim. Nova* **2018**, *41*, 75.
11. Pilon, A.; Selegato, D.; Fernandes, R.; Bueno, P.; Pinho, D.; Carnevale Neto, F.; Freire, R.; Castro-Gamboa, I.; Bolzani, V.; Lopes, N.; *Quim. Nova* **2020**, *43*, 329.
12. Belinato, J.; Bazioli, J.; Sussulini, A.; Augusto, F.; Fill, T.; *Quim. Nova* **2019**, *42*, 546.
13. Dobson, A.; *Statistics in Medicine* **2004**, *23*, 1824.
14. Krzywinski, M.; Altman, N.; *Nat. Methods* **2013**, *10*, 1139.
15. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J.; *Nucleic Acids Res.* **2018**, *46*, W486.
16. Garcia, C. J.; Yang, X.; Huang, D.; Tomas-Barberan, F. A.; *Metabolomics* **2020**, *16*, 85.
17. Bonini, P.; Kind, T.; Tsugawa, H.; Barupal, D. K.; Fiehn, O.; *Anal. Chem.* **2020**, *92*, 7515.
18. Naylor, B. C.; Catrow, J. L.; Maschek, J. A.; Cox, J. E.; *Metabolites* **2020**, *10*, 237.
19. Chandrasekaran, B.; Abed, S. N.; Al-Attraqchi, O.; Kuche, K.; Tekade, R. K. Em *Dosage form design parameters*, 1<sup>st</sup> ed., Academic Press: Cambridge, 2018.
20. Jonas, E.; Kuhn, S.; *J. Cheminf.* **2019**, *11*, 50.
21. Kuhn, S.; Colreavy-Donnelly, S.; Quaresma, L. E. de A. S.; Quaresma, E. de A. S.; Borges, R. M.; *Metabolomics* **2020**, *16*, 123.
22. Kuhn, S.; Colreavy-Donnelly, S.; de Souza, J. S.; Borges, R. M.; *Faraday Discuss.* **2019**, *218*, 339.

