UNIVERSIDADE DE SÃO PAULO

FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE RIBEIRÃO PRETO

**Aplicação de ferramentas computacionais e analíticas na
construção de bibliotecas de produtos naturais da família**

Asteraceae

Tese de Doutorado apresentada ao
Programa de Pós-Graduação em Ciências
Farmacêuticas para obtenção do Título de
Doutora em Ciências.

Área de concentração: Produtos Naturais
e Sintéticos.

Orientada: Annylory Lima Rosa

Orientador: Prof. Dr. Fernando Batista da
Costa

Versão corrigida da Tese de Doutorado Direto apresentada ao Programa de Pós-Graduação em Ciências Farmacêuticas no dia 02/04/2018. A versão original encontra-se disponível na Faculdade De Ciências Farmacêuticas de Ribeirão Preto/USP.

Ribeirão Preto

2018

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Rosa, Annylory Lima

Aplicação de ferramentas computacionais e analíticas na construção de bibliotecas de produtos naturais da família Asteraceae

71 p.; 30cm.

Tese de Doutorado, apresentada à Faculdade de Ciências Farmacêuticas de Ribeirão Preto/USP – Área de concentração: Produtos Naturais e Sintéticos.

Orientador: Da Costa, Fernando Batista

1. Bibliotecas de produtos naturais. 2. Quimiometria. 3. Metabolômica.
4. Asteraceae. 5. CLAE-EM.

FOLHA DE APROVAÇÃO

Annylory Lima Rosa

Aplicação de ferramentas computacionais e analíticas na construção de bibliotecas de produtos naturais da família Asteraceae

Tese de Doutorado apresentada ao
Programa de Pós-Graduação em Ciências
Farmacêuticas para obtenção do Título de
Doutora em Ciências

Área de Concentração: Produtos Naturais
e Sintéticos.

Orientador(a): Prof. Dr. Fernando Batista
da Costa

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Agradecimento

Ao prof. Dr. Fernando Batista da Costa pela oportunidade, pelos ensinamentos e orientação. Agradeço muito pelas horas de discussões, por dar asas à criatividade e busca pelo conhecimento. Ao professor, minha enorme gratidão!

Aos meus familiares por acreditarem em mim e por estarem do meu lado nessa jornada que é a vida. Especialmente minha irmã Grazyely e meu falecido avô José Bibica, vocês não deixaram eu desistir. Muito obrigada!

Aos meus amigos, àqueles que estão próximo e àqueles espalhados por aí, mas que sempre estiveram comigo.

Agradeço especialmente à Débora Munhoz, Ingrid Pontes, Lucas A. Chibli, Camila Capel e Jolindo Freitas, pelas horas de sorriso, pelo apoio incondicional e parceria eterna que começou durante o curso de doutorado e que agora levo para a vida. Gratidão por ter vocês meus amigos.

Ao Lucas A. Chibli devo um agradecimento especial pela parceria imensurável nesse trabalho.

Ao Renê Oliveira do Couto, grande filósofo, amigo, mentor, que me incentivou desde a iniciação científica. Gratidão sempre, meu irmão.

Ao Alexandre Junqueira pelo companheirismo, carinho e suporte nos momentos finais dessa tese. Obrigada!

Aos professores Doutores Leonardo Gobbo Neto (FCFRP-USP), Maique Weber Biavatti (UFSC) e José Pinto de Siqueira Júnior (UFPB) pelas amostras de substâncias isoladas cedidas e que fizeram parte na composição da biblioteca de produtos naturais (AsterLibl).

Ao pessoal do Laboratório de Farmacognosia da Faculdade de Ciências Farmacêuticas de Ribeirão Preto – USP pela convivência, pelos aprendizados, carinho e suporte técnico, nunca esquecerei desse Laboratório de pesquisa único! Obrigada professores, técnicos e colegas.

Obrigada especialmente ao técnico Dr. Luiz Fernando da Silva, não somente pelo suporte técnico em experimentos, mas pela troca de conhecimento e grande coração.

Obrigada aos membros do grupo de pesquisa AsterBioChem pelas horas de aprendizado nesse excitante campo de pesquisa onde se misturam o antigo e o novo na busca pelo desconhecido. Em especial à Daniela Chagas de Paula, à

Rosana Casotti, ao Gari Vidal e ao Federico Padilla pelo “grande punhado” de amostras que cederam a esse trabalho. Espero que o que foi desenvolvido aqui possa colaborar em retorno.

À Doutora Rejane Barbosa de Oliveira, à Doutora Daniela Chagas de Paula e ao Doutor Bruno Leite Sampaio pelos primeiros passos no laboratório de Farmacognosia.

À Coordenação do Programa de Pós-Graduação em Ciências Farmacêuticas da Faculdade de Ciências Farmacêuticas de Ribeirão Preto da Universidade de São Paulo.

À Faculdade de Ciências Farmacêuticas de Ribeirão Preto da Universidade de São Paulo, local onde esse trabalho foi desenvolvido.

À CAPES, À FAPESP e ao CNPQ pelo apoio financeiro. Em especial à CAPES e à FAPESP pelas bolsas de doutorado e mestrado (processo nº2013/03024-8). Ainda devo agradecer à FAPESP pelo financiamento do projeto temático (processo nº 2010/51454-3) e finalmente pelo projeto de auxílio regular (processo nº 2014/26866-7) do qual esse trabalho fez parte.

RESUMO

ROSA, A. L. **Aplicação de ferramentas computacionais e analíticas na construção de bibliotecas de produtos naturais da família Asteraceae**. 2018. Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2018.

Desde as primeiras civilizações, o ser humano se utiliza da diversidade química encontrada na natureza como parte do seu desenvolvimento e até mesmo sobrevivência. Dentre os produtos naturais, as plantas possuem destaque pela quantidade de candidatos a fármacos que chegam à fase final de estudos clínicos. Uma família de plantas de grande interesse é a Asteraceae, cujas plantas biossintetizam diferentes tipos de terpenos, poliacetilenos e substâncias fenólicas. Na busca por novas substâncias ativas, surgem as bibliotecas ou coleções de substâncias, extratos e frações que se tornaram uma das bases para a pesquisa das propriedades químicas e biológicas na indústria. Para a análise e conhecimento dessas bibliotecas são utilizadas técnicas analíticas no estado da arte, como a Cromatografia Líquida de Ultra Eficiência (do inglês, UHPLC), acoplada a Espectrômetros de Massas de Alta Resolução (do inglês, HRMS). Nesse trabalho foi desenvolvido um método analítico em CLAE-EM utilizando a abordagem metabolômica da impressão digital. Esse método foi otimizado utilizando ferramentas *in silico* e validado segundo padrões internacionais para validação bioanalítica, mostrando-se linear, seletivo, robusto, preciso e acurado para o que foi proposto. Esse método foi utilizado ao longo do processo de otimização para a extração de 244 diferentes espécies de Asteraceae, após a obtenção da melhor condição: 20 mg/mL de droga vegetal, proporção de etanol e água de 80% e tempo de extração de ultrassom de 15 min. A partir dessa condição, foram obtidos os extratos e construída assim a coleção denominada AsterLibII. Cento e vinte substâncias de origem natural isoladas pelo grupo de pesquisa AsterBioChem foram devidamente organizadas para compor a AsterLibI. As estruturas 2D obtidas a partir dessas substâncias, juntamente com mais de 8000 estruturas adicionais obtidas pela conversão de arquivos de levantamento bibliográfico, utilizando ferramentas computacionais, formaram a AsterLibIII. Os dados obtidos pela análise da AsterLibII por UHPLC-HRMS mostram a diversidade dos sinais que representam possíveis substâncias em uma faixa de polaridade de alta a média-baixa, não sendo facilmente interpretadas relações entre as amostras, devido à baixa sensibilidade de gráficos de dispersão e da correlação linear necessária para se obter a redução dimensional efetiva por PCA (*Principal Component Analysis*). Os modelos matemáticos utilizados tanto para a otimização do método analítico como para a obtenção da melhor condição para as bibliotecas se mostraram estatisticamente significativos e forneceram informações sobre as variáveis. Através das ferramentas analíticas e computacionais utilizadas, foi possível a construção de bibliotecas de produtos naturais promissoras para estudos biológicos e químicos, incluindo estudos voltados para modelos de desreplicação.

Palavras-chave: Bibliotecas de produtos naturais, Quimiometria, Asteraceae.

ABSTRACT

ROSA, A. L. **Applying analytical and computational tools to build natural products libraries from Asteraceae family**. 2018. Faculdade de Ciências Farmacêuticas de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2018.

From the beginning, civilizations use the chemical diversity found in nature as part of their growing culture and knowledge, even to overcome disease. Among natural products, plants are most noticed because of their success in reaching the final stage of drug discovery. Asteraceae species are of great interest because of the chemical diversity and their dissemination in nature. The chemical compounds found in Asteraceae are terpenes, flavonoids, coumarins, polyacetylenes, benzofurans and phenylpropanoids. With regard to drug discovery, a new emerging approach is the construction of libraries, collections of extracts, compounds and fractions, which have become a fundamental tool for understanding biological and chemical properties, especially in industry. To perform library analysis towards knowledge, state-of-the-art techniques used are UHPLC and HRMS, mainly coupled. In this work, the development and validation of an analytical methodology using the metabolic fingerprint approach was carried out. After validation, the analytical method was considered linear, precise, robust, accurate and selective for the metabolomics analysis. The analytical method was applied together with response surface studies to obtain an efficient extraction condition that allowed to reach most of the chemical diversity. After that, 244 plant samples were extracted and AsterLibII was constructed. 120 natural compounds isolated by members of the AsterBioChem group were duly organized and became AsterLibI. Then, the 2D structures drawn from AsterLibI, along with about 8000 2D structures extracted and converted by computational means from the bibliographic search, became AsterLibIII. Graphical and mathematical analysis was performed on AsterLibII, although it was not easy to see patterns and continue the investigation of metabolites. This was due to the low sensitivity of the dispersion plots and the expected linear correlation to obtain a real dimensional reduction using Principal Component Analysis (PCA). The mathematical models used to understand the variables in the extraction and development of the method showed statistical significance and they were important to access information. Using analytical and computational tools, it was possible to construct libraries of natural products with great potential for biological and chemical studies, including dereplication studies.

Keywords: Natural Product Libraries, Chemometrics, Asteraceae.

LISTA DE FIGURAS

Figura 1. Estruturas 2D de substâncias isoladas pelo grupo de pesquisa AsterBioChem a partir de espécies de Asteraceae (a). Em (b), captura de tela mostrando dados obtidos a partir das 120 substâncias isoladas pelo grupo de pesquisa.....	19
Figura 2. Sobreposição dos cromatogramas de íons no modo positivo e negativo da análise das 12 substâncias de origem natural e o padrão interno (IS), a cafeína.....	20
Figura 3. Estruturas em 2D das substâncias utilizadas no desenvolvimento inicial do método analítico e avaliação das colunas cromatográficas. O números apresentados são os seus identificadores na biblioteca de substâncias isoladas (AsterLibI).....	21
Figura 4. Cromatogramas de íons no modo positivo da análise das 12 substâncias em mistura, (a) coluna Kinetex e (b) coluna Ace. Para essa análise foi utilizado o gradiente 2 mostrado na Tabela 2.....	22
Figura 5. Cromatogramas das doze substâncias utilizadas no desenvolvimento do método para cada gradiente proposto, vide Tabela 2. Figura 5a, gradiente 1; Figura 5b, gradiente 2 e Figura 5c, gradiente 3.....	23
Figura 6. Sobreposição das réplicas do ponto 1 do planejamento fatorial completo para otimização do método analítico. Foi utilizada como amostra a mistura dos extratos hidroalcóolicos de 41 espécies de plantas da família Asteraceae. Para esse cromatograma as variáveis estão nos seguintes níveis : gradiente = 3, porcentagem de ácido fórmico na fase móvel = 0,1; e temperatura = 30 °C.....	24
Figura 7. Sobreposição de cromatogramas da análise da mistura de 41 espécies da família Asteraceae em diferentes gradientes de eluição gradientes 2 (em verde) e 3 da Tabela 2.....	25
Figura 8. Gráficos para diagnóstico da regressão, onde são mostradas variáveis de interesse versus os resíduos da regressão e o comportamento dos dados modelados em um gráfico de probabilidade do tipo QQ (<i>quantil-quantil</i>) para a resposta 'número de sinais'.....	28
Figura 9. Gráficos para diagnóstico da regressão, onde é mostrada a variável de interesse versus os resíduos da regressão e o comportamento dos dados modelados em um gráfico de probabilidade do tipo QQ (<i>quantil-quantil</i>) para a resposta 'área dos sinais'.....	29
Figura 10. O cromatograma de íons no modo positivo para o extrato diclorometano da raiz de <i>Aldama arenaria</i> , ampliado na região de 16-30 minutos. É possível a visualização de sinais de intensidade entre 5×10^8 e 3×10^9	30
Figura 11. Curvas de calibração obtidas pela média das curvas de calibração variando-se o volume de injeção para cada padrão interno (cafeína, daidzeína, 3- α -acetóxi-copálico).....	31
Figura 12. Curvas de calibração obtidas pela média das curvas de calibração para cada padrão interno variando-se a concentração (cafeína, daidzeína, 3- α -acetóxi-copálico).....	32

Figura 13. O cromatograma do extrato de <i>T. diversifolia</i> contendo somente os padrões internos é mostrado em (a). Em (b) é possível ver o cromatograma do extrato adicionado com as seguintes substâncias: ácido clorogênico, cumarina, budleína A e ácido ent-caurenóico mostrados em análise isolada em 'c'. Os cromatogramas apresentados são análises no modo negativo de ionização.....	36
Figura 14. Os cromatogramas em (a) mostram (acima) a análise do ácido clorogênico, cumarina, budleína A, do ácido ent-caurenóico e do extrato de <i>T. diversifolia</i> adicionado dessas substâncias. Em (b) são mostrados os espectros para o ácido ent-caurenóico analisado isoladamente e adicionado ao extrato de <i>T. diversifolia</i>	37
Figura 15. Cromatogramas no modo de ionização negativo, mostrando as análises de repetitividade (precisão) realizadas com a espécie de <i>T. diversifolia</i> para validação do método analítico proposto.....	40
Figura 16. Ilustração mostrando curvas obtidas para as razões da área dos sinais do extrato de <i>T. diversifolia</i> pela a área do padrão interno p6 (daidzeína).....	45
Figura 17. Gráfico de contorno para a superfície de resposta mostrando a faixa de máximo para a resposta 'soma da área dos sinais' do modelo calculado com as variáveis x1 (massa de droga vegetal), x2 (proporção de H ₂ O: EtOH).....	49
Figura 18. Superfície de resposta mostrando a faixa de máximo para a resposta 'soma da área dos sinais' do modelo calculado com as variáveis x1 (massa de droga vegetal), x2 (proporção de H ₂ O: EtOH)	50
Figura 19. Gráficos para diagnóstico da regressão realizada para o cálculo da superfície de resposta. Onde são mostradas variáveis de interesse versus os resíduos da regressão e o comportamento dos dados modelados em um gráfico de probabilidade do tipo QQ (<i>quantil-quantil</i>) para a resposta 'área dos sinais'	51
Figura 20. Gráficos para diagnóstico da regressão, onde são mostradas variáveis de interesse versus os resíduos da regressão e o comportamento dos dados modelados em um gráfico de probabilidade do tipo QQ (<i>quantil-quantil</i>) para a resposta 'número de sinais'.....	54
Figura 21. Cromatogramas de íons do modo positivo da análise dos 244 extratos hidroalcóolicos de plantas da família Asteraceae.....	55
Figura 22. Frequência dos sinais nos cromatogramas de íons para o modo positivo (número de vezes que o sinal aparece por amostras) para os 244 extratos de espécies da família Asteraceae.....	56
Figura 23. Gráfico de dispersão dos sinais obtidos para a análise dos 244 extratos que compõem a biblioteca de produtos naturais AsterLibII no modo positivo de análise. Tempo de retenção (Rt) por massa acurada (accuratemass)	57
Figura 24. Gráfico de dispersão mostrando a m/z dos íons pelo tempo de retenção – Mapa de íons para as 120 substâncias da AsterLibI no modo positivo.....	58

Figura 25. Gráfico de dispersão mostrando a m/z dos íons pelo tempo de retenção – Mapa de íons para as 120 substâncias da AsterLibI no modo negativo.....	58
Figura 26. Quantidade de sinais encontrados no cromatograma de íons para cada amostra da AsterLibII.....	59
Figura 27. Análise de componentes principais (PCA) para a análise em CLAE-EM dos 244 extratos da AsterLibII no modo positivo de ionização.....	60

LISTA DE TABELAS

Tabela 1 – Substâncias utilizadas para a avaliação do método analítico e das colunas cromatográficas, seus ID's e suas massas exatas.....	11
Tabela 2 – Gradientes de eluição utilizados para avaliar as colunas ACE e Kinetex (Phenomenex); fase móvel A: água (0,1% de ácido fórmico), fase móvel B: acetonitrila (0,1% de ácido fórmico).....	12
Tabela 3. Parâmetros de análise utilizados para o espectrômetro de massas acoplado ao CLAE.....	12
Tabela 4.Os 45 gêneros da família Asteraceae utilizados na otimização do processo de extração.....	16
Tabela 5. Valores reais e normalizados para as variáveis utilizadas nos cálculos para a construção do modelo de superfície de resposta: x1 (proporção de droga vegetal), x2 (proporção de EtOH e H2O do solvente extrator) e x3 (tempo de extração em ultrassom). Os experimentos foram realizados em 5 níveis: -1,68; -1; 0; 1 e 1,68, com 5 réplicas no ponto central.....	17
Tabela 6. São mostrados os valores de número de pratos teóricos (N) e as resoluções (R) média e mínima para a injeção de 12 substâncias em três diferentes gradientes (Tabela 2) e em duas colunas: Kinetex e Ace.....	20
Tabela 7. Condições dos experimentos do planejamento fatorial completo. Os resultados analisados foram o número de sinais detectados e a soma da área desses sinais. Os resultados são expressos como a média das réplicas $\pm$ o desvio padrão.....	25
Tabela 8. Dados obtidos para a regressão linear a partir das variáveis estudadas no planejamento fatorial completo (v1: temperatura da coluna, v2: porcentagem de ácido fórmico na fase móvel e; v3: gradiente), tendo como resposta o 'número de sinais'. Os cálculos apresentados abaixo foram para o primeiro modelo.....	26
Tabela 9. Dados obtidos para a regressão linear a partir das variáveis estudadas no planejamento fatorial completo (v1: temperatura da coluna, v2: porcentagem de ácido fórmico na fase móvel e; v3: gradiente), tendo como resposta a 'área dos sinais'. Os cálculos apresentados abaixo se referem ao primeiro modelo obtido.....	26
Tabela 10. Dados obtidos para a regressão linear a partir das variáveis estudadas no planejamento fatorial completo (v1: temperatura da coluna, v2: porcentagem de ácido fórmico na fase móvel e; v3: gradiente), tendo como resposta a 'área dos sinais'. Os cálculos apresentados abaixo são referentes ao modelo recalculado utilizando somente a variável 2 (v2: porcentagem de ácido fórmico na fase móvel).....	27

Tabela 11. Valores de resolução (R) par a par obtidos para dezessete sinais da análise do extrato de <i>T. diversifolia</i> . Segue também na Tabela o tempo de retenção médio (Tr médio) e a largura para os sinais (L). Os sinais de p1-p12 referem-se ao modo negativo de ionização e de 14-17 ao modo positivo de ionização.....	33
Tabela 12. Valores de tempo de retenção para os sinais de p1 a p17 obtidos a partir da análise do extrato de <i>T. diversifolia</i> em oito condições diferentes no sistema cromatográfico de ultra alta eficiência, variando-se os lotes da coluna cromatográfica, o volume de injeção ($4 \mu\text{L} \pm 0,4$) e a temperatura do forno da coluna ($30 \text{ }^{\circ}\text{C} \pm 3$). São mostrados os valores de média, desvio padrão e coeficiente de variação, sendo esse último, o parâmetro avaliado para se verificar a robustez do método).....	34
Tabela 13. Valores de área para os sinais de p1 a p17 obtidos da análise do extrato de <i>T. diversifolia</i> em oito condições diferentes em sistema cromatográfico de ultra alta performance (do inglês UHPLC), variando-se os lotes da coluna cromatográfica, o volume de injeção ($4 \mu\text{L} \pm 0,4$) e a temperatura do forno da coluna ($30 \text{ }^{\circ}\text{C} \pm 3$). São mostrados os valores de média, desvio padrão e coeficiente de variação, sendo esse último o parâmetro avaliado para se verificar a robustez do método).....	35
Tabela 14. Resultados para a análise de exatidão por recuperação. São apresentados para cada substância utilizada no enriquecimento do extrato de <i>T. diversifolia</i> (ácido clorogênico, ácido ent-caurenóico, cumarina e budleína A) os valores finais após o procedimento de extração por ultrassom e análise em CLAE acoplado a EM, expressos em massa recuperada (μg) e porcentagem de recuperação.....	48
Tabela 15. A seguir são mostrados os valores para a média de massa recuperada dos padrões utilizados no teste de exatidão por recuperação e seus respectivos intervalos de confiança com probabilidade de 95%.....	39
Tabela 16. São mostrados os dados (médias, desvio padrão e CV) obtidos para o ensaio de repetibilidade dentro da avaliação da precisão do método analítico, tendo como resposta o tempo de retenção para os sinais de p1 a p17.	41
Tabela 17. São mostrados os dados (médias, desvios padrões e CV) obtidos para o ensaio de repetibilidade dentro da avaliação da precisão do método analítico, tendo como resposta a área dos sinais de p1 a p17.....	42
Tabela 18. São mostrados os dados (médias, desvios padrões e CV) obtidos para a precisão intermediária (interanalistas) dentro da avaliação da precisão do método analítico tendo como resposta o tempo de retenção e a área dos sinais de p1 a p17. Aqui são consideradas as médias para as seis réplicas obtidas pelos analistas.....	43
Tabela 19. São apresentadas as equações para as curvas obtidas a partir da regressão linear dos dados para a área dos sinais do extrato de <i>T. diversifolia</i>	44

e seus respectivos coeficientes de correlação (r).....

Tabela 20. Valores obtidos para as respostas “número de sinais” e a “soma da área dos sinais” para as condições do planejamento por composto central. As variáveis x1 (massa de droga vegetal em 2 mL de solvente extrator), x2 (porcentagem de etanol e água) e x3 (tempo de extração em ultrassom) são mostradas em seus valores reais e codificados..... 46

Tabela 21. Dados obtidos para o cálculo de modelo pela metodologia de superfície de resposta, tendo como variáveis x1 (massa de droga vegetal), x2 (proporção de H₂O e EtOH) e x3 (tempo de extração) e resposta a ‘soma da área dos sinais’ para a análise da mistura das 45 espécies de Asteraceae..... 47

Tabela 22. Análise de variância do modelo obtido para a resposta ‘soma da área dos sinais’, utilizando a técnica de superfície de resposta e tendo como variáveis x1 (massa de droga vegetal), x2 (proporção de H₂O e EtOH) e x3 (tempo de extração)..... 48

Tabela 23. Dados obtidos para o cálculo do modelo final pela metodologia de superfície de resposta, tendo como variáveis x1 (massa de droga vegetal) e x2 (proporção de H₂O e EtOH) e como resposta a ‘soma da área dos sinais’ para a análise da mistura das 45 espécies de Asteraceae..... 48

Tabela 24. Análise de variância do modelo final obtido para a resposta ‘soma da área dos sinais’, utilizando a técnica de superfície de resposta e tendo como variáveis x1 (massa de droga vegetal) e x2 (proporção de H₂O e EtOH). 49

Tabela 25. Dados obtidos para o cálculo de modelo pela metodologia de superfície de resposta, tendo como variáveis x1 (massa de droga vegetal), x2 (proporção de H₂O e EtOH) e x3 (tempo de extração) e como resposta o ‘número de sinais’ para a análise da mistura das 45 espécies de Asteraceae... 52

Tabela 26. Análise de variância do modelo obtido para a resposta ‘número de sinais’, utilizando a técnica de superfície de resposta e tendo como variáveis x1 (massa de droga vegetal) e x2 (proporção de H₂O e EtOH) e x3 (tempo de extração)..... 56

Tabela 27. Dados obtidos para o cálculo do modelo final pela metodologia de superfície de resposta, tendo como variável x1 (massa de droga vegetal) e resposta o ‘número de sinais’ para a análise da mistura das 45 espécies de Asteraceae..... 53

LISTA DE ABREVIATURAS E SIGLAS

UHPLC	Ultra-High Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometer
CLAE-EM	Cromatografia Líquida de Alta Eficiência – Espectrometria de Massas
2D	duas dimensões
PCA	Principal Component Analysis
CG	Cromatografia Gasosa
CLAE	Cromatografia Líquida de Alta Eficiência
EM	Espectrometria de Massas
RMN	Ressonância Magnética Nuclear
amu	Unified atomic mass unit
TOF	Time of flight
m/z	mass-to-charge ratio
HCA	Hierarchical Clustering Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
SVM	Support Vector Machine
CL	Cromatografia Líquida
CAS	Chemical Abstracts Service
SMILES	Simplified Molecular Input Line Entry System
IUPAC	International Union for Pure and Applied Chemistry
COX	Ciclo-oxigenase
LOX	Lipoxigenase
IC 50	Inhibitory concentration 50%
ESI	Electrospray Ionization
R	Resolução
N	Número de pratos teóricos

min	minutos
MeCN	Acetonitrila
LD	Limite de detecção
s	Desvio padrão
S	Coeficiente angular ou inclinação da curva analítica
RSM	Response Surface Methodology
TIC	Total Ion Chromatogram
FTMS	Fourier Transform Mass Spectrometer
Rt	Retention time

SUMÁRIO

Resumo	i
Abstract	ii
Lista de Figuras	iii
Lista de Tabelas	vi
Lista de abreviaturas e siglas	ix
1. INTRODUÇÃO.....	1
2. OBJETIVOS.....	7
2.1 Objetivos específicos.....	7
3. MATERIAIS E MÉTODOS.....	8
3.1 Biblioteca de substâncias isoladas e coleção de estruturas 2D da família Asteraceae.....	8
3.2 Desenvolvimento do método analítico em CLAE-EM.....	10
3.3 Validação do método analítico.....	13
3.3.1 Adequabilidade do sistema e robustez	13
3.3.2 Seletividade	14
3.3.3 Exatidão	14
3.3.4 Precisão	15
3.3.5 Linearidade	15
3.3.6 Limite de detecção	15
3.4 Uso da técnica de planejamento por composto central com superfície de resposta para otimização do processo de extração: construção da AsterLibII...	16
3.5 Processamento dos dados analíticos obtidos em CLAE-EM	17
3.6 Análises exploratórias das bibliotecas de produtos naturais	18
4. RESULTADOS.....	19
4.1 Biblioteca de substâncias isoladas e biblioteca de estruturas 2D da família	19

Asteraceae.....	
4.2 Desenvolvimento do método analítico em CLAE-EM.....	20
4.3 Validação do método analítico	32
4.3.1 Adequabilidade do sistema e robustez	32
4.3.2 Seletividade	38
4.3.3 Exatidão	39
4.3.4 Precisão	41
4.3.5 Linearidade	46
4.3.6 Limite de detecção	48
4.4 Uso da técnica de planejamento por composto central com superfície de resposta para otimização do processo de extração: construção da AsterLibII	49
4.5 Bibliotecas de produtos naturais da família Asteraceae (AsterLibI e II)	59
 5. DISCUSSÃO.....	 67
5.1 Desenvolvimento e validação do método analítico em CLAE-EM	67
5.2 Método de extração e Bibliotecas de produtos naturais	69
 6. CONCLUSÕES.....	 71
 7. REFERÊNCIAS.....	 72

1. INTRODUÇÃO

Desde as primeiras civilizações, o ser humano se utiliza da diversidade química encontrada na natureza como parte do seu desenvolvimento e até mesmo sobrevivência. Essa diversidade era utilizada na obtenção de aromas e corantes, no combate aos insetos e ervas daninhas e na cura de doenças (NEWMAN; CRAGG, 2016; ZHANEL et al., 2002). Ainda assim, com todas as aplicações e o conhecimento desenvolvido em consequência, a diversidade que conhecemos é considerada a camada mais superficial de milhares de anos de evolução (CRAGG; NEWMAN, 2013; LI; VEDERAS, 2009).

Ao longo da história, as fontes de origem natural, tais como plantas e micro-organismos, possibilitaram a descoberta de inúmeros fármacos, fosse através do isolamento de substâncias ativas ou pela obtenção de derivados e análogos (HARVEY, 2008; NEWMAN; CRAGG, 2016; PASCOLUTTI; QUINN, 2014). Exemplos do sucesso da utilização dessas fontes podem ser demonstrados por antibióticos como a penicilina, tetraciclina e a eritromicina, fármacos utilizados no combate à malária como a quinina e a artemisinina e imunossupressores como a ciclosporina, além de agentes anticâncer como o taxol (HARVEY, 2008).

Dentre os produtos naturais, as plantas possuem destaque pela quantidade de candidatos a fármacos que chegam à fase final de estudos clínicos (VILLAS-BÔAS; RASMUSSEN; LANE, 2005; WOLFENDER et al., 2013). Além disso, estudos mostram que o uso regulamentado de suplementos à base de plantas, plantas medicinais e fitoterápicos tem aumentado consideravelmente nos últimos anos (IZZO et al., 2016; OSHIRO et al., 2016).

Nesse contexto, uma família de plantas de grande interesse e com potencial para revelar novas e promissoras substâncias é a Asteraceae (Compositae). Essa família que possui cerca de 23.000 espécies e 1.535 gêneros distribuídos pelo mundo todo, correspondendo a 10% de toda a diversidade vegetal. No Brasil, essas plantas são encontradas principalmente na região Centro-sul do país (BREMER, 1994; NAKAJIMA; SEMIR, 2001; ZDERO; BOHLMANN, 1990).

Do ponto de vista morfológico, as espécies pertencentes à família Asteraceae possuem como característica principal as suas inflorescências em capítulo e flores reduzidas. Com relação ao hábito, podem ser desde ervas, arbustos até árvores, sendo conhecidos exemplos diversos de plantas com uso medicinal, ornamental e alimentício (BREMER, 1994; “The internacional Compositae Alliance”, [s.d.]). Como exemplo, é possível citar plantas medicinais nativas como o guaco (*Mikania laevigata* Sch. Bip.), o assa-peixe (*Vernonanthura ferruginea* Less.), a macela (*Achyrocline satureioides* (Lam.) DC), o mentrasto (*Ageratum conyzoides* L.), plantas exóticas que aqui foram introduzidas, como a calêndula (*Calendula officinalis* L.), a camomila (*Matricaria chamomilla* L.), a arnica (*Arnica montana* L.), além de plantas alimentícias como a alface (*Lactuca sativa* L.), a serralha (*Sonchus oleraceus* L.), o yacón (*Smallanthus sonchifolius* H. Rob.) e a alcachofra (*Cynara scolymus* L.).

Em relação à sua diversidade química, são encontrados em Asteraceae diferentes tipos de terpenos, poliacetilenos e substâncias fenólicas (ALVARENGA et al., 2001). Dentre essas substâncias, as lactonas sesquiterpênicas chamam a atenção pela diversidade do seu esqueleto carbônico e por sua distribuição taxonômica mais restrita, além de suas atividades biológicas (PICMAN, 1986; ZDERO; BOHLMANN, 1990). Por todas essas características levantadas, as plantas dessa família despertam o interesse de pesquisadores de diversas áreas da ciência, dentre elas a Química de Produtos Naturais, Farmacognosia e Farmacologia.

Uma abordagem recente tendo em vista a descoberta de novas substâncias ativas e o melhor entendimento de matrizes orgânicas é a construção de bibliotecas ou coleções de substâncias, extratos e frações (BOBZIN; YANG; KASTEN, 2000; FÜLLBECK et al., 2006; PASCOLUTTI; QUINN, 2014). Essas bibliotecas se tornaram a base para a pesquisa das propriedades químicas e biológicas de substâncias de interesse, principalmente para as pesquisas em escala industrial (LANG et al., 2008; LI; VEDERAS, 2009). Bibliotecas podem ser construídas a partir de substâncias sintéticas, produtos de origem natural, derivados semi sintéticos e mesmo de estruturas virtuais. (ABEL et al., 2002; FÜLLBECK et al., 2006; QUINN et al., 2008).

Existem diversas abordagens para a construção de bibliotecas, essas estão diretamente relacionadas com a origem das substâncias que serão utilizadas na sua composição (NTIE-KANG et al., 2013). Assim, com relação às bibliotecas de produtos naturais, as abordagens atualmente utilizadas levam principalmente em consideração as regras de Lipinski e utilizam técnicas de extrações seletivas (CAMP et al., 2012).

Essas extrações são realizadas em amostras advindas de bactérias do solo, fungos, plantas superiores e, mais recentemente, organismos de origem marinha e de invertebrados (GAUDÊNCIO; PEREIRA, 2015; LI; VEDERAS, 2009). No trabalho aqui descrito, a abordagem utilizada para a obtenção da biblioteca de extratos foi o uso de ferramentas analíticas e computacionais (*in silico*) para a extração e análise da maior diversidade de metabólitos possíveis, logo que vários fármacos e substâncias de interesse não seguem a regra de Lipinski (GIMÉNEZ et al., 2010).

Todavia, independente da abordagem utilizada, as bibliotecas de produtos naturais buscam a padronização na obtenção de seus constituintes, de modo a possibilitar posterior replicação e reposição da biblioteca frente ao seu consumo em estudos químicos e biológicos. Após a composição da biblioteca, seus elementos são geralmente analisados para a obtenção de informações analíticas e físico-químicas que possibilitem o conhecimento da diversidade química que esses oferecem para a investigação de atividades (GAUDÊNCIO; PEREIRA, 2015).

Atualmente, a investigação de produtos naturais se tornou mais rápida e eficiente devido ao avanço constante das técnicas analíticas, tais como Cromatografia Gasosa (CG), Cromatografia Líquida de Alta Eficiência (CLAE), Espectrometria de Massas (EM), Ressonância Magnética Nuclear (RMN), dentre outras, bem como suas combinações (técnicas hífenadas). Desta forma, não há a necessidade inicial do isolamento das substâncias, diminuindo a quantidade de tempo e recursos gastos nesse processo e evitando o re-isolamento de estruturas já conhecidas (BUTLER, 2004; LANG et al., 2008).

Em meio aos avanços dessas técnicas analíticas, surgiu a metabolômica, um campo de pesquisa que tem como principal objetivo a investigação de todos os metabólitos existentes em um organismo, ou seja, do seu metaboloma

(KUEHNBAUM; BRITZ-MCKIBBIN, 2013). Embora nenhuma técnica sozinha seja capaz de tal façanha, a combinação de técnicas como a CLAE e EM, ou a CLAE e RMN tem apresentado sucesso em análises de alta abrangência desses metabólitos (análises holísticas), sendo possível a visualização de milhares de sinais de uma vez (WOLFENDER et al., 2013).

Esse fato possibilita o estudo simultâneo de um maior número de substâncias com baixo peso molecular (< 1000 amu) e em matrizes orgânicas complexas, seja de forma qualitativa (técnica da impressão digital) ou quantitativa (técnica do perfil metabólico) (VILLAS-BÔAS; RASMUSSEN; LANE, 2005). São assim verificadas as diferenças na produção de metabólitos, com foco principalmente em metabolitos secundários – aqueles que não participam diretamente no processo de crescimento e desenvolvimento dos organismos –, ainda que em um determinado momento e em determinadas condições, com limitações intrínsecas de cada técnica (KIM; VERPOORTE, 2010; VERPOORTE et al., 2008).

Em relação aos estudos metabolômicos, as técnicas de EM que se destacam são aquelas com uma maior sensibilidade e resolução, capazes de separar íons com erros na casa dos ppm, como os analisadores do tipo TOF (do inglês, *Time of Flight*) e Orbitrap (GLAUSER et al., 2013; XIE et al., 2008). Todavia, essa alta resolução e sensibilidade tornam os dados gerados em CLAE-EM complexos, possuindo esses mais que uma dimensão (por ex. tempo, intensidade, m/z), além da existência de ruídos e redundâncias (SMITH et al., 2006; VAN DEN BERG et al., 2006).

Devido a esse fato, há a necessidade da realização de tratamentos matemáticos para a obtenção de dados que possam ser interpretados pelos analistas (informação química) (PLUSKAL et al., 2010; SMITH et al., 2006; VAN DEN BERG et al., 2006). A obtenção desses dados inclui a utilização de filtros para a diminuição do ruído, suavização da linha de base, desconvolução dos sinais, o alinhamento dos sinais dos cromatogramas de íons, o agrupamento de isótopos para os espectros de massas, dentre outros (BRERETON, 2003; CASTILLO et al., 2011).

Após o tratamento inicial, é realizada a análise dos dados através de métodos computacionais. Algumas das técnicas utilizadas são para a observação do comportamento dos dados, portanto nenhuma informação inicial é fornecida ao algoritmo matemático. Tais análises são denominadas exploratórias e não-supervisionadas, são exemplos a PCA (do inglês, *Principal Component Analysis*) e a HCA (*Hierarchical Clustering Analysis*) (ABDELMOHSEN et al., 2014; TRYGG; HOLMES; LUNDSTEDT, 2007).

Outras técnicas utilizam de regressões, ou cálculos de similaridade com algoritmos de inteligência artificial para relacionar os dados, tendo o conhecimento prévio daquilo que se estuda, tal como uma propriedade química ou atividade biológica. São exemplos dessas técnicas, o PCR (*Principal Component Regression*), o PLS (*Partial Least Squares*) e técnicas de aprendizado (do inglês, *machine learning*), como SVM (*Support Vector Machine*) e redes neurais artificiais. Essas técnicas são portanto chamadas de supervisionadas (BRERETON, 2003; DÜHRKOP et al., 2015).

Outras características fundamentais dos estudos metabolômicos e que delimitam aquilo que será analisado pelo espectrômetro de massas, são os processos de extração e separação realizados para a amostra (KIM; VERPOORTE, 2010; LUQUE DE CASTRO; DELGADO-POVEDANO, 2013; SHIN et al., 2010). Esses processos selecionam os metabólitos secundários de acordo com suas características físico-químicas, sendo portanto fundamental o planejamento e análise das condições utilizadas nessas etapas (ASADOLLAHZADEH et al., 2014; TENG; CHOI, 2013) .

A técnica de separação que mais se destaca no estudo do metabolismo secundário, isso devido às características físico-químicas dessas substâncias é a Cromatografia Líquida (CL), sendo seu estado da arte, a Cromatografia Líquida de Ultra Alta Eficiência (do inglês UHPLC, *Ultra High Performance Liquid Chromatography*) e a modalidade mais utilizada a CL em fase reversa (FEKETE; OLÁH; FEKETE, 2012; GLAUSER et al., 2013).

Para a extração das amostras, são geralmente usadas misturas de solventes de média a alta polaridade que sejam compatíveis com o método de separação utilizado. Misturas de etanol-água e metanol-água são frequentes, sendo possível a extração não exaustiva de pequenas quantidades de amostras em metabolômica, devido à sensibilidade das técnicas de separação e detecção utilizadas. Para a técnica de CLAE-EM, por exemplo, a amostra utilizada pode ser da grandeza de miligramas (KIM; VERPOORTE, 2010; VAN DER KOOY et al., 2009).

Para o desenvolvimento dos métodos de extração e de análise, a abordagem geralmente utilizada é o uso da própria ferramenta analítica (por ex. EM) para a detecção sensível e resolvida dos metabólitos ao longo do processo. Contudo, para que esse processo não seja dispendioso em tempo e recursos, e ainda garanta de forma matemática a obtenção da melhor condição possível, podem ser utilizados modelos *in silico* para que ajudem a prever a melhor condição a partir de dados experimentais (GOODARZI; RUSSELL; VANDER HEYDEN, 2013).

Dentre tais técnicas computacionais estão a regressão e a geração de superfícies de resposta, tendo como fundamento para a experimentação, técnicas de planejamento como o planejamento fatorial completo e por composto central (ASADOLLAHZADEH et al., 2014; FERREIRA et al., 2007; JOSHI; KUMAR; RATHORE, 2017).

Na era em que a química de produtos naturais tem acesso à ferramentas analíticas e computacionais de ponta, ainda são considerados desafios a otimização de métodos para a obtenção de amostras, o desenvolvimento de métodos analíticos, bem como, o tratamento e análise dos dados químicos. Tudo isso, visando à diminuição do tempo da descoberta de novas entidades moleculares, com a confiança em se obter informação química que gere conhecimento (LANGE et al., 2008; SUGIMOTO et al., 2012; TCHOUMTCHOUA et al., 2013; YANG et al., 2013).

a inferir que a variabilidade de sinais encontrada para a AsterlibII é tão alta que pode não ser descrita linearmente, demonstrando assim uma grande diversidade de sinais obtidos, que pode se traduzir em uma diversidade de substâncias.

6. CONCLUSÃO

Nesse trabalho foram construídas bibliotecas de produtos naturais, físicas e virtuais (AsterLibI, AsterLibII e AsterLibIII), utilizando-se uma abordagem holística baseada na metabolômica, tanto para o método analítico como na etapa de extração. Visando, assim a obtenção de diversidade química.

Foi possível por meio de modelos matemáticos e baseado em técnicas analíticas, desenvolver um método analítico com grande abrangência na avaliação de produtos naturais. Além disso, esse método se mostrou robusto, preciso, acurado, linear e seletivo, mesmo sendo utilizado em uma abordagem qualitativa.

Assim, os dados obtidos, tanto no processo de construção das bibliotecas de produtos naturais, como sua posterior análise, não somente possibilitaram o acesso à biodiversidade como também garantiram a transformação de informação em conhecimento, criando-se assim uma plataforma para estudos químicos e biológicos tendo como modelo a família Asteraceae.

7. REFERÊNCIAS

- ABDELMOHSEN, U. et al. Dereplication Strategies for Targeted Isolation of New Antitrypanosomal Actinosporins A and B from a Marine Sponge Associated-Actinokineospora sp. EG49. **Marine Drugs**, v. 12, n. 3, p. 1220–1244, 6 mar. 2014.
- ABEL, U. et al. Modern methods to produce natural-product libraries. **Curr Opin Chem Biol**, v. 6, p. 453–458, 2002.
- ALVARENGA, S. A. V. et al. Chemosystematic studies of natural compounds isolated from Asteraceae: Characterization of tribes by principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 56, n. 1, p. 27–37, 2001.
- AMBRÓSIO, S. R. et al. Trypanocidal Activity of Pimarane Diterpenes from *Viguiera arenaria* (Asteraceae). v. 1415, n. August, p. 1413–1415, 2008.
- AMBROSIO, S. R.; SCHORR, K.; DA COSTA, F. B. Terpenoids of *Viguiera arenaria* (Asteraceae). **Biochemical Systematics and Ecology**, v. 32, n. 2, p. 221–224, fev. 2004.
- APARECIDA, D.; PAULA, C. DE. Estudos metabolômicos de Astedraceae por UPLC - UV - HRFTMS , avaliação do potencial anti-inflamatório in vitro e suas correlações através de métodos in silico Estudos metabolômicos de Asteraceae por UPLC - UV - HRFTMS , avaliação do potencial anti-inflamatório (tese de doutorado). [s.l: s.n.].
- ASADOLLAHZADEH, M. et al. Response surface methodology based on central composite design as a chemometric tool for optimization of dispersive-solidification liquid-liquid microextraction for speciation of inorganic arsenic in environmental water samples. **Talanta**, v. 123, p. 25–31, 2014.
- BOBZIN, S. C.; YANG, S.; KASTEN, T. P. LC-NMR: a new tool to expedite the dereplication and identification of natural products. **Journal of Industrial Microbiology and Biotechnology**, v. 25, n. 6, p. 342–345, 1 dez. 2000.
- BREMER, K. **Asteraceae: cladistics and classification**. [s.l: s.n.].
- BRERETON, R. G. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. Bristol, UK: British Library, 2003. v. 8
- BUTLER, M. S. The role of natural product chemistry in drug discovery. **Journal of natural products**, v. 67, n. 12, p. 2141–53, dez. 2004.
- CAMP, D. et al. Drug-like Properties: Guiding Principles for the Design of Natural Product Libraries. **Journal of Natural Products**, 2012.

CASTILLO, S. et al. Algorithms and tools for the preprocessing of LC–MS metabolomics data. **Chemometrics and Intelligent Laboratory Systems**, v. 108, n. 1, p. 23–32, ago. 2011.

CHAMBERS, M. C. et al. A cross-platform toolkit for mass spectrometry and proteomics. **Nature biotechnology**, v. 30, n. 10, p. 918–20, out. 2012.

CHEMAXON. Marvin SketchChemaxon, , 2013. Disponível em: <<http://www.chemaxon.com/products/marvin/>>

CHEMAXON. **Jchem**, 2017. Disponível em: <<http://www.chemaxon.com>>

CRAGG, G. M.; NEWMAN, D. J. Natural products: a continuing source of novel drug leads. **Biochimica et biophysica acta**, v. 1830, n. 6, p. 3670–95, jun. 2013.

CREEK, D. J. et al. Metabolite identification: are you sure? And how do your peers gauge your confidence? **Metabolomics**, v. 10, n. 3, p. 350–353, 8 abr. 2014.

DA COSTA, F. B.; ALBUQUERQUE, S. Diterpenes and Synthetic Derivatives from *Viguiera aspilioides* with Trypanomicidal Activity. v. 62, p. 557–559, 1996.

DE VOS, R. C. H. et al. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. **Nature protocols**, v. 2, n. 4, p. 778–91, jan. 2007.

DÜHRKOP, K. et al. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. **Proceedings of the National Academy of Sciences**, v. 112, n. 41, p. 12580–12585, 2015.

FDA, F. AND D. A.; FOOD AND DRUG ADMINISTRATION. **Guidance for Industry: Bioanalytical method validation**. [s.l: s.n.].

FEKETE, S.; OLÁH, E.; FEKETE, J. Fast liquid chromatography: the domination of core-shell and very fine particles. **Journal of chromatography. A**, v. 1228, p. 57–71, 9 mar. 2012.

FERREIRA, S. L. C. et al. Box-Behnken design: An alternative for the optimization of analytical methods. **Analytica Chimica Acta**, v. 597, n. 2, p. 179–186, 2007.

FÜLLBECK, M. et al. Natural products: sources and databases. **Natural product reports**, v. 23, n. 3, p. 347–56, jun. 2006.

GAUDÊNCIO, S. P.; PEREIRA, F. Dereplication: racing to speed up the natural products discovery process. **Natural Product Reports**, v. 32, n. 6, p. 779–810, 2015.

GIMÉNEZ, B. G. et al. Evaluation of blockbuster drugs under the rule-of-five. **Pharmazie**, v. 65, n. 2, p. 148–152, 2010.

GLAUSER, G. et al. Ultra-high pressure liquid chromatography-mass spectrometry for plant metabolomics: a systematic comparison of high-resolution quadrupole-time-of-flight and single stage Orbitrap mass spectrometers. **Journal of chromatography. A**, v. 1292, p. 151–9, 31 maio 2013.

GOODARZI, M.; RUSSELL, P. J.; VANDER HEYDEN, Y. Similarity analyses of chromatographic herbal fingerprints: a review. **Analytica chimica acta**, v. 804, p. 16–28, 4 dez. 2013.

HARVEY, A. L. Natural products in drug discovery. **Drug discovery today**, v. 13, n. 19–20, p. 894–901, out. 2008.

ICH. **ICH harmonised tripartite guideline validation of analytical procedures: text and methodology Q2(R1)**. Guidance. Anais 2005. Disponível em: <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf>

INMETRO. Orientação Sobre Validação de Métodos de Ensaio Químicos - DOQ-CGCRE-008. **Instituto Nacional de Metrologia e Normalização e Qualidade Industrial**, p. 25, 2007.

IZZO, A. A. et al. A Critical Approach to Evaluating Clinical Efficacy, Adverse Events and Drug Interactions of Herbal Remedies. **Phytotherapy research : PTR**, v. 30, n. 5, p. 691–700, 2016.

JOLLIFFE, I. T.; CADIMA, J.; CADIMA, J. Principal component analysis : a review and recent developments Subject Areas : Author for correspondence : **Philosophical Transactions A**, v. 374, p. 1–16, 2016.

JOSHI, V. S.; KUMAR, V.; RATHORE, A. S. Optimization of ion exchange sigmoidal gradients using hybrid models: Implementation of quality by design in analytical method development. **Journal of Chromatography A**, v. 1491, p. 145–152, 2017.

KIM, H. K.; VERPOORTE, R. Sample preparation for plant metabolomics. **Phytochemical analysis : PCA**, v. 21, n. 1, p. 4–13, 2010.

KUEHNBAUM, N. L.; BRITZ-MCKIBBIN, P. New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. **Chemical reviews**, v. 113, n. 4, p. 2437–68, 10 abr. 2013.

LANG, G. et al. Evolving Trends in the Dereplication of Natural Product Extracts : New Methodology for Rapid , Small-Scale Investigation of Natural Product Extracts. p. 1595–1599, 2008.

LANGE, E. et al. Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. v. 19, p. 1–19, 2008.

- LI, J. W.-H.; VEDERAS, J. C. Drug discovery and natural products: end of an era or an endless frontier? **Science (New York, N.Y.)**, v. 325, n. 5937, p. 161–5, 10 jul. 2009.
- LUQUE DE CASTRO, M. D.; DELGADO-POVEDANO, M. M. Ultrasound: A subexploited tool for sample preparation in metabolomics. **Analytica Chimica Acta**, v. 806, p. 74–84, 2 jan. 2013.
- NAEEM, S.; HYLANDS, P.; BARLOW, D. Construction of an Indonesian herbal constituents database and its use in Random Forest modelling in a search for inhibitors of aldose reductase. **Bioorganic & Medicinal Chemistry**, v. 20, n. 3, p. 1251–1258, 2012.
- NAKAJIMA, J. N.; SEMIR, J. Asteraceae do Parque Nacional da Serra da Canastra, Minas Gerais, Brasil. **Revta brasil. Bot.**, v. 4, n. 24, p. 471–478, 2001.
- NEWMAN, D. J.; CRAGG, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. **Journal of Natural Products**, v. 79, n. 3, p. 629–661, 2016.
- NTIE-KANG, F. et al. ConMedNP: A natural product library from Central African medicinal plants for drug discovery. **RSC. Adv.**, v. 0, p. 1–6, 2013.
- OSHIRO, M. C. et al. A evolução do registro e prescrição de fitoterápicos no Brasil sob a perspectiva legal e sanitária. **Vigilância Sanitária em Debate**, v. 4, n. 4, p. 116–122, 2016.
- PASCOLUTTI, M.; QUINN, R. J. Natural products as lead structures: chemical transformations to create lead-like libraries. **Drug discovery today**, v. 19, n. 3, p. 215–21, mar. 2014.
- PICMAN, A. K. Review Article Number 7 Biological Activities of Sesquiterpene Lactones. **Biochemical Systematics and Ecology**, v. 14, n. 3, p. 255–281, 1986.
- PLUSKAL, T. et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. **BMC bioinformatics**, v. 11, p. 395–405, jan. 2010.
- QUINN, R. J. et al. Developing a Drug-like Natural Product Library $\perp$. **J. Nat. Prod.**, v. 71, p. 464–468, 2008.
- R CORE TEAM (R FOUNDATION FOR STATISTICAL COMPUTING). R: A Language and Environment for Statistical Computing. 2017.
- RIBANI, M. et al. Validação em métodos cromatográficos e eletroforéticos. **Quim. Nova**, v. 27, n. 5, p. 771–780, 2004.

SEAMAN, F. C. SESQUITERPENE LACTONES AS TAXONOMIC. v. 48, n. 2, p. 121–592, 1982.

SEGER, C.; STURM, S.; STUPPNER, H. Mass spectrometry and NMR spectroscopy: modern high-end detectors for high resolution separation techniques--state of the art in natural product HPLC-MS, HPLC-NMR, and CE-MS hyphenations. **Natural product reports**, v. 30, n. 7, p. 970–87, jul. 2013.

SHIN, M. H. et al. Evaluation of Sampling and Extraction Methodologies for the Global Metabolic Profiling of *Saccharophagus degradans*. v. 82, n. 15, p. 6660–6666, 2010.

SISTEMA, N. O. et al. a Implantação Do Programa De Plantas Medicinais E Fitoterápicos No Sistema Público De Saúde No Brasil: Uma Revisão De Literatura. v. 1, p. 1–11, 2015.

SMITH, C. A et al. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. **Analytical chemistry**, v. 78, n. 3, p. 779–87, 1 fev. 2006.

SNYDER R., LLOYD, KIRKLAND, JOSEPH, DOLAN, J. W. **INTRODUCTION TO MODERN LIQUID CHROMATOGRAPHY**. Third Edit ed. [s.l.] Wiley, 2010.

SUGIMOTO, M. et al. Bioinformatics Tools for Mass Spectroscopy-Based Metabolomic Data Processing and Analysis. **Current bioinformatics**, v. 7, n. 1, p. 96–108, mar. 2012.

TCHOUMTCHOUA, J. et al. Structure-oriented UHPLC-LTQ Orbitrap-based approach as a dereplication strategy for the identification of isoflavonoids from *Amphimas pterocarpoides* crude extract. **J. Mass Spectrom.**, v. 48, p. 561–575, 2013.

TENG, H.; CHOI, Y. H. Optimization of microwave-assisted extraction of bioactive alkaloid compounds from *Rhizoma Coptidis* (*Coptis chinensis* Franch.). **Food Science and Biotechnology**, v. 22, n. 5, p. 1–8, 2013.

The internacional Compositae Alliance. Disponível em: <<http://www.vkcomp.org/>>. Acesso em: 29 out. 2014.

TRYGG, J.; HOLMES, E.; LUNDSTEDT, T. Chemometrics in metabonomics. **Journal of proteome research**, v. 6, n. 2, p. 469–79, fev. 2007.

VAN DEN BERG, R. A et al. Centering, scaling, and transformations: improving the biological information content of metabolomics data. **BMC genomics**, v. 7, p. 142, jan. 2006.

VAN DER KOOY, F. et al. Quality control of herbal material and phytopharmaceuticals with MS and NMR based metabolic fingerprinting. **Planta medica**, v. 75, n. 7, p. 763–75, jun. 2009.

VERPOORTE, R. et al. Metabolomics: Back to basics. **Phytochemistry Reviews**, v. 7, n. 3, p. 525–537, 2008.

VILLAS-BÔAS, S. G.; RASMUSSEN, S.; LANE, G. A. Metabolomics or metabolite profiles? **Trends in biotechnology**, v. 23, n. 8, p. 385–6, ago. 2005.

WENIG, P.; ODERMATT, J. OpenChrom: a cross-platform open source software for the mass spectrometric analysis of chromatographic data. **BMC bioinformatics**, v. 11, p. 405, jan. 2010.

WOLFENDER, J. et al. Plant Metabolomics : From Holistic Data to Relevant Biomarkers. **Current Medicinal Chemistry**, v. 20, p. 1056–1090, 2013.

XIE, G. X. et al. Application of ultra-performance LC-TOF MS metabolite profiling techniques to the analysis of medicinal Panax herbs. **Metabolomics**, v. 4, n. 3, p. 248–260, 2008.

YANG, J. Y. et al. Molecular networking as a dereplication strategy. **Journal of natural products**, v. 76, n. 9, p. 1686–99, 27 set. 2013.

ZDERO, C.; BOHLMANN, F. Systematics and evolution within the Compositae, seen with the eyes of a chemist. **Pl. Syst. Evol.**, v. 171, p. 1–14, 1990.

ZHANEL, G. G. et al. Ketolides: A Critical Review. v. 62, n. 12, p. 1771–1804, 2002.