



**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE CIÊNCIA E TECNOLOGIA DE ALIMENTOS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE
ALIMENTOS (PPGCTA)**

Aplicação do MS-Finder, Sirius e GNPS no estudo de compostos fenólicos

Luana Peixoto Mallmann

PORTO ALEGRE

2024

Luana Peixoto Mallmann

**APLICAÇÃO DO MS-FINDER, SIRIUS E GNPS NO ESTUDO DE COMPOSTOS
FENÓLICOS**

Tese apresentada ao Programa de Pós-Graduação
em Ciência e Tecnologia de Alimentos como um
dos requisitos para obtenção do grau de Doutor em
Ciência e Tecnologia de Alimentos.

Orientador: Prof. Dr. Eliseu Rodrigues

Coorientador: Prof. Dr. Alessandro de O. Rios

Porto Alegre

2024

CIP - Catalogação na Publicação

Mallmann, Luana Peixoto

Aplicação do Sirius, MS-Finder e GNPS no estudo de compostos fenólicos / Luana Peixoto Mallmann. -- 2024.

188 f.

Orientador: Eliseu Rodrigues.

Coorientador: Alessandro de Oliveira Rios.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Instituto de Ciência e Tecnologia de Alimentos, Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, Porto Alegre, BR-RS, 2024.

1. compostos fenólicos. I. Rodrigues, Eliseu, orient. II. Rios, Alessandro de Oliveira, coorient. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

Autora: Luana Peixoto Mallmann (Engenheira de Alimentos/UFRGS)

Título da Tese: Aplicação do MS-Finder, Sirius e GNPS no estudo de compostos fenólicos

Submetida como parte dos requisitos para a obtenção do grau de
DOUTOR EM CIÊNCIA E TECNOLOGIA DE ALIMENTOS

Aprovada em: ___/___/___

Homologada em: ___/___/___

Pela banca examinadora:

Por:

Prof. Dr. Eliseu Rodrigues
Orientador
Doutor em Ciência de Alimentos
ICTA/UFRGS

Prof. Dr. Rafael Costa Rodrigues
Coordenador do PPGCTA
ICTA/UFRGS

Prof. Dr. Alessandro de Oliveira Rios
Coorientador
Doutor em Ciência de Alimentos
ICTA/UFRGS

Prof^a. Dra. Florência Cladera Oliveira
Diretora
ICTA/UFRGS

Banca: Prof^a. Dra. Giovana Domeneghini Mercali
Doutora em Engenharia Química
ICTA/UFRGS

Banca: Prof^a. Dra. Vivian Caetano Bochi
Doutora em Ciência de Alimentos
Dep. Nutrição/UFCSPA

Banca: Prof^a. Dra. Lilian Regina Barros Mariutti
Doutora em Ciência de Alimentos
DECAN - FEA- UNICAMP

AGRADECIMENTOS

A Deus, por todas as oportunidades que me deste nesta vida, por ter me guiado pelos melhores caminhos que me levaram à realização de sonhos.

Aos meus pais Paulo e Ligia, por estarem sempre ao meu lado, pelo amor, exemplo e educação.

Ao meu marido Anderson, pelo apoio, dedicação à nossa família e por todo amor. Obrigada por estar sempre ao meu lado e por ser meu grande parceiro nesta vida.

Ao meu filho Joaquim por me mostrar o verdadeiro sentido da vida e o maior amor do mundo. Obrigada por ser minha motivação para levantar da cama todos os dias.

Ao Professor orientador Eliseu Rodrigues, pela orientação, pelos ensinamentos e até mesmo pelas cobranças, obrigada por estar sempre disponível.

Ao Professor coorientador Alessandro Rios, pela disponibilidade, pela orientação e pelos ensinamentos.

Aos meus colegas do NTC-ICTA por todo apoio e ajuda, em especial a Andressa, Ana Raísa, Michele e Diogo pelos almoços, desabafos e agradáveis momentos que passamos juntos.

Aos bolsistas Fernanda, Gabriel e Tayná pela importante ajuda nas análises e preparação de amostras.

Ao Programa de Pós-graduação em Ciências e Tecnologia de Alimentos, da Universidade Federal do Rio Grande do Sul, pela oportunidade de qualificação profissional.

À banca examinadora, pelas valiosas contribuições.

A todos que, de alguma forma, contribuíram para realização deste trabalho, meus sinceros agradecimentos.

RESUMO

A anotação e identificação de compostos fenólicos fundamenta-se na análise de dados gerados por cromatografia líquida acoplada à espectrometria de massas (LC-MS). No entanto, a interpretação manual dos espectros de massas é uma tarefa não trivial e demorada, além disso, depende do conhecimento sobre a técnica e o perfil de fragmentação de compostos fenólicos. Ainda, esse sistema manual de anotação continua sendo a principal abordagem utilizada no estudo de compostos fenólicos na área de Ciência e Tecnologia de Alimentos. Os pesquisadores gastam muito tempo comparando os seus espectros de massas com aqueles disponibilizados em publicações. Assim, uma quantidade excessiva de tempo é utilizada na anotação de compostos já relatados, não permitindo muitas vezes o avanço na descoberta de novos compostos fenólicos. Neste contexto, a aplicação de ferramentas de metabolômica, nomeadamente, os *softwares* MS-Finder e Sirius, além da plataforma GNPS (Rede Social Molecular Global de Produtos Naturais) nos estudos de compostos fenólicos poderá impulsionar a descoberta de novos compostos fenólicos ou ao menos acelerar a anotação de moléculas já relatadas. Avaliamos ambos os *softwares* para anotar compostos fenólicos de 8 amostras de alimentos (café, chá verde, suco de *cranberry*, suco de uva, suco de laranja, suco de maçã, extrato de soja e extrato de salsa). O MS-Finder e Sirius juntos conseguiram anotar corretamente mais de 90% dos compostos fenólicos analisados por LC-MS. A aplicação da Rede Social Molecular Global de Produtos Naturais (GNPS) com os outros *softwares* acelerou a anotação dos compostos fenólicos extraíveis (CFE) e não-extraíveis (CFNE) de duas variedades de cada uma das 5 frutas mais consumidas no Brasil (banana, maçã, laranja, mamão e manga). Através do GNPS, foi possível anotar rapidamente mais de 33% dos compostos fenólicos encontrados. Além disso, foi possível a criação de redes moleculares que agruparam compostos com similaridades estruturais nas mesmas famílias moleculares, facilitando a interpretação das similaridades entre espécies, variedades e frações. As redes moleculares permitiram a anotação de moléculas por propagação e indicaram possíveis marcadores fenólicos de cada fruta e variedade. Diferenças marcantes na composição fenólica das frações CFE e CFNE também foram reveladas pela aplicação do GNPS. A maioria dos compostos fenólicos obtidos após as hidrólises ácida e alcalina eram diferentes dos extraídos apenas por solventes, além de se diferenciarem entre si. O GNPS proporcionou também a possibilidade de compartilhar os dados brutos de padrões analíticos e das amostras analisadas na plataforma com acesso para comunidade global, trazendo avanços no compartilhamento de dados e resultados de espectrometria de massas. Desta forma, a combinação da triagem inicial usando MS-Finder, Sirius e GNPS com análises

manuais de informações adicionais é uma abordagem poderosa e eficiente para anotar e identificar compostos fenólicos.

Palavras-chave: compostos fenólicos; anotação; identificação; MS-Finder; Sirius; GNPS.

ABSTRACT

The annotation and identification of phenolic compounds is based on the analysis of data generated by liquid chromatography coupled to mass spectrometry (LC-MS). However, manual interpretation of mass spectra is a non-trivial and time-consuming task, in addition, it depends on knowledge about the technique and the fragmentation profile of phenolic compounds. Still, this manual annotation system continues to be the main approach used in the study of phenolic compounds in Food Science and Technology field. Researchers spend a lot of time comparing their mass spectra with those available in publications. Thus, an excessive amount of time is spent on annotating already reported compounds, often not allowing progress in the discovery of new phenolic compounds. In this context, the application of metabolomics tools, namely the MS-Finder and Sirius software, in addition to the GNPS platform (Global Social Molecular Network of Natural Products) to studies of phenolic compounds could boost the discovery of new phenolic compounds or at least accelerate the annotation of molecules already reported. We evaluated both software to annotate phenolic compounds from 8 food samples (coffee, green tea, cranberry juice, grape juice, orange juice, apple juice, soy extract and parsley extract). MS-Finder and Sirius together managed to correctly annotate more than 90% of the phenolic compounds analyzed by LC-MS. The application of the Global Social Molecular Network of Natural Products (GNPS) with other software accelerated the annotation of extractable (CFE) and non-extractable (CFNE) phenolic compounds from two varieties of each of the 5 most consumed fruits in Brazil (banana, apple, orange, papaya, and mango). Using GNPS, it was possible to quickly record more than 33% of the phenolic compounds found. Furthermore, it was possible to create molecular networks that grouped compounds with structural similarities into the same molecular families, facilitating the interpretation of similarities between species, varieties, and fractions. Molecular networks allowed the annotation of molecules by propagation and indicated possible phenolic markers for each fruit and variety. Outstanding differences in the phenolic composition of the CFE and CFNE fractions were also revealed by the application of GNPS. Most of the phenolic compounds obtained after acid and alkaline hydrolysis were different from those extracted only by solvents, in addition to differing from each other. The GNPS also provided the possibility of sharing raw data of analytical standards and analyzed samples on the platform with access to the global community, bringing advances in sharing data and mass spectrometry results. Thus, combining initial screening using MS-

Finder, Sirius and GNPS with manual analysis of additional information is a powerful and efficient approach to annotating and identifying phenolic compounds.

Keywords: phenolic compounds; annotation; identification; MS-Finder; Sirius; GNPS.

SUMÁRIO

1	INTRODUÇÃO.....	11
2	OBJETIVOS.....	15
3	REVISÃO BIBLIOGRÁFICA.....	16
3.1.	Compostos fenólicos.....	16
3.1.1.	Estrutura química.....	16
3.1.2.	Compostos fenólicos extraíveis (CFE) e não extraíveis (CFNE).....	22
3.1.1.	Composição fenólica das frutas.....	24
3.2.	Análise de compostos fenólicos por LC-MS.....	26
3.2.1.	Cromatografia Líquida (LC).....	26
3.2.2.	Espectrometria de massas (MS).....	27
3.3.	Ferramentas de bioinformática.....	32
3.3.1.	MS-Finder.....	32
3.3.2.	Sirius.....	35
3.3.3.	GNPS.....	36
4	ARTIGO 1.....	40
5	ARTIGO 2.....	108
6	DISCUSSÃO GERAL.....	176
7	CONCLUSÃO GERAL.....	178
8	REFERÊNCIAS.....	179

1 INTRODUÇÃO

Os compostos fenólicos são metabólitos secundários muito estudados devido à sua abundância e função nas plantas e aos prováveis efeitos benéficos à saúde humana. Uma pesquisa no *Scopus* (palavra-chave: *phenolic compounds*) mostra que foram publicados mais de 100 mil estudos nas duas últimas décadas, os quais têm focos variando de diferentes técnicas de extração, identificação e quantificação, até estudos *in vitro* e *in vivo* com objetivos de compreender os mecanismos de ação dessas moléculas na saúde humana (Li et al., 2015; Soares et al., 2020).

Graças às diferentes pesquisas realizadas, atualmente, sabemos que há cerca de 10 mil compostos fenólicos na natureza, frequentemente extraídos com misturas de água e solventes orgânicos como metanol, etanol e acetona. Essas extrações podem ser realizadas por técnicas simples, como maceração e agitação, até a utilização de tecnologias emergentes, como ultrassom e micro-ondas, as quais aumentam a eficiência do processo. As técnicas de identificação permitiram saber que os compostos fenólicos possuem diferentes esqueletos carbônicos, variando desde simples fenóis de poucos Daltons até taninos condensados com elevado grau de polimerização (Garcia-salas et al., 2010; Shi et al., 2011; Farias et al., 2022).

Algumas dessas classes estão largamente distribuídas nas plantas, tais como os flavonóis, e outros estão presentes apenas em algumas espécies, tais como as isoflavonas. Especialmente nos últimos anos, alguns trabalhos indicaram que há uma fração de compostos fenólicos pouco investigada, os compostos fenólicos não-extraíveis (CFNE). Trata-se da fração que não é extraída diretamente das matrizes devido às suas fortes interações com os componentes dos alimentos. Para alguns alimentos, foi relatada que essa fração, a qual muitas vezes possui composição diferente da fração extraível, pode até ser a majoritária e aquela responsável pelos efeitos benéficos à saúde humana encontrados (Esparza-Martínez et al., 2016; Pérez-Ramírez et al., 2018; Reynoso-Camacho et al., 2018).

Ao longo de décadas de estudo sobre as propriedades biológicas dos compostos fenólicos, observa-se que eles foram testados exaustivamente quanto as suas propriedades antioxidantes, o que está diretamente relacionada a sua estrutura química de natureza fenólica, a qual é capaz de desativar espécies reativas de oxigênio e de nitrogênio através da doação de

hidrogênio, elétrons e formação de aduto. Assim, milhares de trabalhos verificaram através de métodos *in vitro* que os compostos fenólicos são capazes de desativar tanto radicais não-biológicos (ABTS e DPPH) quanto espécies reativas biológicas (ROO•, RO•). Os resultados desses estudos tornaram-se pouco relevante à medida que foi relatada reiteradamente na literatura a baixa biodisponibilidade dessas moléculas, o que é fundamental para a sua atuação como antioxidantes no organismo humano (Eseberri et al., 2022).

Além disso, os compostos fenólicos sofrem transformação durante a digestão gastrointestinal, o que é completamente desconsiderado quando se aplica métodos *in vitro* para a avaliação da capacidade antioxidante. Recentemente, trabalhos têm indicado que o principal mecanismo envolvido na atuação dos compostos fenólicos na saúde humana envolve a sua interação com a microbiota intestinal, onde essas moléculas são capazes de modular a microbiota, favorecendo a colonização do intestino por microrganismos benéficos e impedindo a colonização por bactérias patogênicas. Desta forma, há um avanço na compreensão do papel dessas moléculas no organismo humano, porém, há um longo caminho que deve ser percorrido para entendermos os mecanismos de ação dessas moléculas no organismo humano (Arranz & Calixto, 2010; Cardona et al., 2013; Domínguez-Rodríguez et al., 2022).

Tal avanço depende fundamentalmente de aprofundarmos o conhecimento sobre a identidade dos compostos fenólicos presentes nos alimentos bem como dos metabólicos gerados a partir da digestão gastrointestinal e pela metabolização da microbiota intestinal. Essa tarefa tem sido conduzida majoritariamente pela técnica de espectrometria de massas de alta resolução (MS), a qual está frequentemente acoplada à cromatografia líquida (LC). A execução dessa técnica é relativamente simples, sendo o grande desafio a interpretação dos dados gerados, o que está relacionado tanto a complexidade intrínseca dos compostos fenólicos e seus metabólitos derivados quanto aos dados gerados na aplicação das técnicas citadas.

Tradicionalmente na área de Ciência e Tecnologia de Alimentos, a análise dos dados gerados pela aplicação da espectrometria de massas no estudo de compostos fenólicos tem sido conduzida de forma manual, onde os pesquisadores comparam os dados de MS e MS² obtidos experimentalmente com dados provenientes de padrões, dados da literatura e bibliotecas. Há um número muito limitado de padrões disponíveis comercialmente, e parte desses é muito caro, sendo inviável a sua compra pela maioria dos grupos de pesquisa.

Assim, a interpretação depende muito da comparação com aqueles compostos fenólicos já identificados previamente na literatura e da interpretação química dos dados de massas pelos pesquisadores. Essa abordagem é cansativa e ineficiente, visto que é contraproducente e imprecisa a comparação de dados experimentais de massas com dados tabelados em publicações científicas. De outro lado, a interpretação química das informações de MS e MS² depende de conhecimentos que grande parte dos pesquisadores não possui, visto que a área que estuda compostos fenólicos é multidisciplinar, com pesquisadores de diferentes áreas, como engenharia, tecnologia de alimentos e até nutrição.

De acordo com a Metabolomics Standards Initiative (MSI), uma molécula é considerada “identificada” quando seus dados experimentais são comparados a um padrão analítico por pelo menos dois tipos de dados ortogonais (por exemplo, tempo de retenção e padrão de fragmentação), enquanto um composto seria considerado “anotado” se a identificação não fosse alcançada (Sumner et al., 2007). Neste contexto, a etapa de anotação (ou identificação) é claramente a principal dificuldade na análise de dados de compostos fenólicos analisados por LC-MS e uma grande barreira no avanço da compreensão do papel dessas moléculas no organismo humano. Desta forma, propomos neste trabalho a integração da análise manual com ferramentas largamente utilizadas na metabolômica não-direcionada, o MS-Finder, o Sirius e o GNPS.

A primeira etapa de desenvolvimento do presente trabalho (Artigo 1) utilizou ferramentas de fragmentação *in silico* para anotação de compostos fenólicos analisados por cromatografia líquida acoplada a espectrometria de massas (LC-MS) em oito amostras de alimentos (café, chá verde, suco de cranberry, suco de uva, suco de laranja, suco de maçã, extrato de soja e extrato de salsinha). Dois *softwares* gratuitos disponíveis na internet, o MS-Finder e o Sirius foram utilizados por serem os mais completos em termos de elucidação estrutural. Ambos os softwares já foram explorados para a anotação de metabólitos de plantas, fungos, esponjas marinhas, algas e microalgas e apresentaram ótimos resultados no concurso *Critical Assessment of Small Molecule Identification* (CASMI) (Dührkop et al., 2013; Vaniya et al., 2017). Nesse trabalho utilizou-se o termo “identificação” o que era usual para estudos na área de ciência dos alimentos. A partir disso, foram utilizados os termos “identificação” e “anotação”.

Na segunda etapa do trabalho (Artigo 2), foi utilizado a Rede Social Molecular Global de Produtos Naturais (GNPS; <http://gnps.ucsd.edu>), uma base de conhecimento de acesso aberto para a comunidade mundial para análise de moléculas. Os objetos de estudo foram as frações de compostos fenólicos extraíveis (CFE) e não extraíveis (CFNE) de duas variedades das 5 frutas comumente consumidas no Brasil (banana, maçã, laranja, mamão e manga). As relações e similaridades entre as frutas, variedades e processos de extração foram exploradas.

2 OBJETIVOS

O objetivo deste estudo foi a aplicação das ferramentas de bioinformática MS-Finder, Sirius e GNPS para melhorar a performance da análise dos compostos fenólicos presentes nos alimentos e nas frações de CFE e CFNE das frutas comumente consumidas no Brasil.

As seguintes etapas foram realizadas para atingir o objetivo proposto:

1. Otimização das condições do espectrômetro de massas (MS) para a elucidação estrutural de compostos fenólicos através dos *softwares* MS-Finder e Sirius.
2. Avaliação da capacidade de anotação de compostos fenólicos pelos *softwares* MS-Finder e Sirius com padrões analíticos e amostras de alimentos.
3. Extração de compostos extraíveis e não-extraíveis de frutas comumente consumidas no Brasil através de duas hidrólises (ácida e alcalina).
4. Compartilhamento de dados de espectrometria de massas em tandem (MS/MS) brutos.
5. Aplicação do GNPS para formação de redes moleculares para exploração das similaridades entre frutas, variedades e tratamentos.
6. Aplicação do GNPS com os *softwares* MS-Finder e Sirius para uma melhor performance da análise dos compostos fenólicos presentes nas frutas comumente consumidas no Brasil.

3 REVISÃO BIBLIOGRÁFICA

3.1. Compostos fenólicos

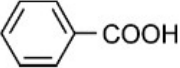
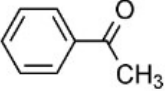
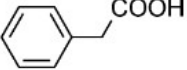
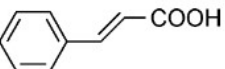
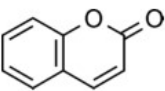
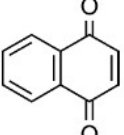
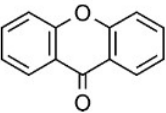
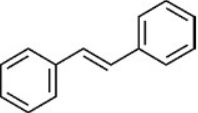
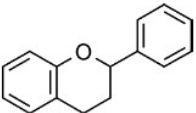
Os compostos fenólicos são metabólitos secundários das plantas sintetizados durante o seu desenvolvimento normal. A sua produção pode ser influenciada por situações de estresse, tais como injúrias, exposição à radiação ultravioleta (UV), entre outros fatores (Haminiuk et al., 2012). Aproximadamente 10.000 compostos fenólicos foram relatados na literatura, sendo que desse total, 500 estão presentes em plantas que são utilizadas na alimentação humana (Ashraf et al., 2018; Tahara, 2007).

3.1.1. Estrutura química

Os compostos fenólicos são caracterizados por terem pelo menos um anel aromático com um ou mais grupos hidroxila ligados. Eles variam de um único anel aromático, simples e de baixa massa molecular até grandes complexos de compostos fenólicos de elevada massa molecular. Uma fração significativa dos compostos fenólicos é encontrada glicosilada ou esterificada na matriz vegetal (Haminiuk et al., 2012).

Os compostos fenólicos constituem um grupo grande e diversificado de compostos que podem ser classificados de acordo com o número de carbonos na molécula (**Tabela 1**) e, posteriormente, pelo arranjo desses carbonos na sua estrutura. Usualmente, os compostos fenólicos tem sido divididos em dois grandes grupos: os flavonoides e os não flavonoides (Vermerris & Nicholson, 2006; Crozier et al., 2009).

Tabela 1. Estrutura básica dos compostos fenólicos classificados com base no número de carbonos na molécula.

Fórmula básica	Classificação	Estrutura básica
C_6-C_1	Ácidos fenólicos	
C_6-C_2	Acetofenonas	
C_6-C_2	Ácido fenilacético	
C_6-C_3	Ácidos hidroxicinâmicos	
C_6-C_3	Cumarinas	
C_6-C_3	Naftoquinonas	
C_6-C_3	Xantonas	
C_6-C_3	Estilbenos	
$C_6-C_3.C_6$	Flavonoides	

Fonte: própria autora.

Os ácidos fenólicos compõem o grupo mais estudado dentre os não-flavonoides, especialmente em frutas. Eles são caracterizados por possuírem um substituinte carboxílico na sua estrutura fenólica, podendo ter também uma ou mais hidroxilas, além de um ou mais grupos metoxila. Essa classe de compostos apresenta um esqueleto carbônico muito similar, porém, as posições das hidroxilas e das metoxilas no anel aromático variam bastante (Robbins, 2003).

Grande parte dos ácidos fenólicos não é encontrada no estado livre na natureza, mas sob forma de ésteres. Os taninos hidrolisáveis são constituídos de ésteres de ácidos gálicos e ácidos elágicos glicosilados, formados a partir do chiquimato, onde os grupos hidroxila do açúcar são esterificados com os ácidos fenólicos. Taninos elágicos são muito mais frequentes que os gálicos, embora ambos tenham como ponto de partida o composto pentagalactose. A variação estrutural entre esses compostos é causada por acoplamento oxidativo de ácidos gálicos vizinhos ou por oxidação de anéis aromáticos (Mueller-Harvey, 2001).

Os flavonoides são caracterizados por possuírem 15 carbonos (C6-C3-C6) em sua estrutura e dois anéis aromáticos (A e B) conectados por um heterocíclico de três átomos de carbono, anel C, (**Figura 1**). Os substituintes no anel C criam as variações de compostos dando origem a diversas classes (Moussa-Ayoub, El-Samahy, Kroh, & Rohn, 2011; Padilha et al., 2017).

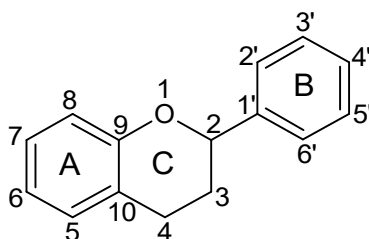
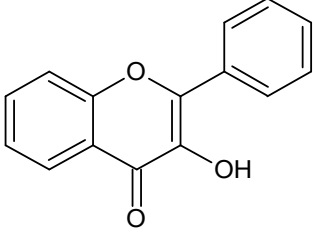
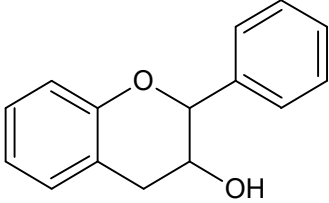
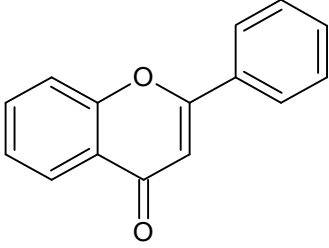
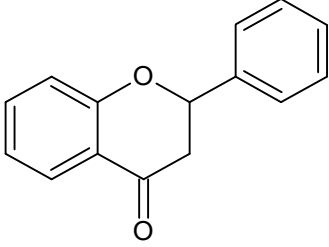


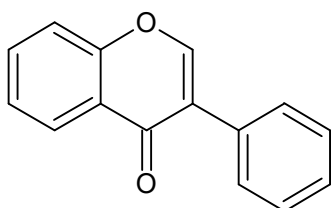
Figura 1. Estrutura química básica dos flavonoides.

Fonte: própria autora.

Hidroxilas e metoxilas também estão presentes em diferentes posições nos anéis A e B, o que propicia uma grande variedade de moléculas em cada uma dessas classes dos flavonoides (**Tabela 2**). O anel B é tipicamente hidroxilado podendo, também, ter éteres metílicos como substituintes. Além disso, os flavonoides podem estar ligados a açúcares, como ramnose, rutinose e glicose (Vermerris & Nicholson, 2006).

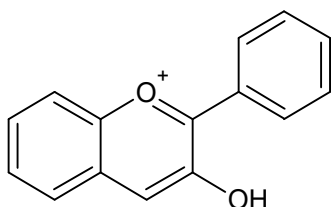
Tabela 2. Estrutura química das diferentes classes dos flavonoides.

Flavonoide	Substituição	Exemplos
<p>Flavonóis</p> 	<p>3,5,7,3',4' – OH</p> <p>5,7,3',4' – OH; 3 – rutinose</p>	<p>Quercetina</p> <p>Rutina</p>
<p>Flavanóis</p> 	<p>3,5,7,3',4' – OH</p> <p>3,5,7,3',4' – OH</p> <p>5,7,3',4' – OH; 3 – O-glicosídeo</p>	<p>Catequina</p> <p>Epicatequina</p> <p>Catequina 3-O-glicosídeo</p>
<p>Flavonas</p> 	<p>5,7,4' – OH</p>	<p>Apigenina</p>
<p>Flavanonas</p> 	<p>5,4' – OH; 7- rhamnoglucose</p> <p>3,5,7,3',4' – OH</p>	<p>Naringina</p> <p>Taxifolin<u>a</u></p>

Isoflavonas

5,7,4' – OH

Genisteína

Antocianidinas

3,5,7,3',4' – OH

Cianidina

Fonte: própria autora.

Os flavonóis são abundantes em pequenos frutos do tipo *berry* (*edelberry*, *cranberry*, *blueberry* ou mirtilo) e muito característicos de ervas e temperos (Rothwell et al., 2013). O anel C dos flavonóis contém um grupo carbonila no C4, uma insaturação entre C2-C3 e um grupo hidroxila ligado no C3. A quercetina, rutina, kaempferol e miricetina são os flavonóis mais amplamente distribuídas na natureza. As flavonas apresentam estrutura semelhante aos flavonóis, porém, sem a oxigenação no C-3. Aipo, salsa e ervas são boas fontes de flavonas, como a apigenina e luteolina (Moussa-Ayoub et al., 2011).

As antocianinas são flavonoides e constituem o maior e, provavelmente, o mais importante grupo de pigmentos vegetais solúveis em água, sendo que sua cor é determinada pelo pH da seiva. O azul e o vermelho de determinadas flores e frutos podem ser devidos ao mesmo glicosídeo, em pH diferente. As antocianinas, derivadas das antocianidinas, têm um cátion flavílium como estrutura central, com diferentes açúcares ligados na posição C3. Romãs, ameixas, uvas tintas e, conseqüentemente, o vinho tinto são fontes de antocianidinas (Sharma et al., 2016).

Os flavanóis formam a subclasse de flavonoides mais complexa abrangendo desde simples monômeros (catequina, epicatequina, epigallocatequina) até estruturas complexas incluindo os oligômeros e polímeros denominados proantocianidinas. Esses compostos são encontrados em chás, cacau, uvas e vinho tinto (Díaz-de-Cerio et al., 2017).

As proantocianidinas são produtos oligoméricos e poliméricos de catequina e epicatequina resultantes da biossíntese de flavonoides. Proantocianidinas do tipo A possuem uma ligação éter interflavanóis adicional entre C2, O e C7 se comparadas com proantocianidinas do tipo B (**Figura 2**) (Colletti et al., 2021; Rauf et al., 2019). A massa molecular das proantocianidinas varia de 500 a 3000 Da.

Todas as proantocianidinas acima de 3.000 Da são considerados polímeros cujas massas moleculares podem atingir até 20.000 Da. À medida que a massa molecular das proantocianidinas aumenta, tornam-se cada vez mais polares devido à presença de grupos hidroxila adicionais e, portanto, com mais potencial para formar ligações de hidrogênio. As proantocianidinas poliméricas são frequentemente chamadas também de taninos condensados tendem a formar agregados complexos. Elas possuem dificuldades significativas de solubilização em solventes orgânicos polares usados e são conhecidos por precipitar proteínas. Esses taninos têm massas moleculares na faixa de 1.000 a 5.000 Da (Bedran-Russo et al., 2014).

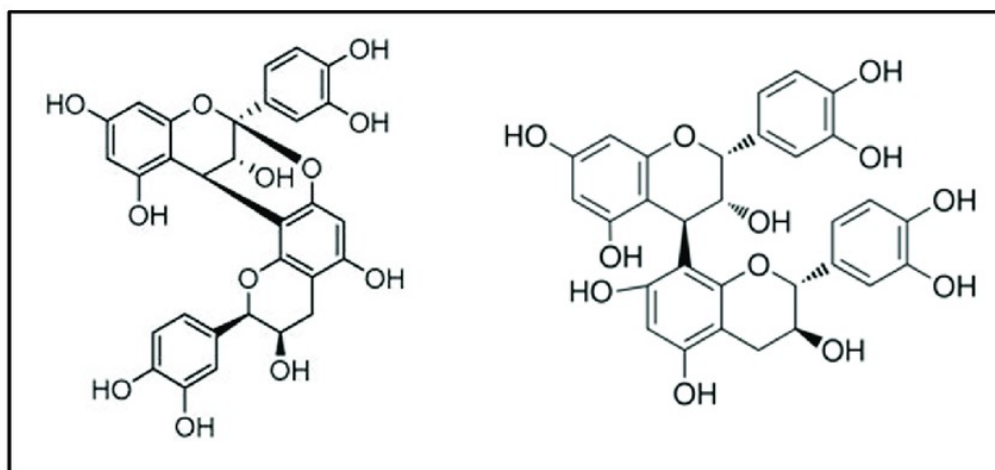


Figura 2. Proantocianidina tipo A (esquerda) e tipo B (direita).

Fonte: Colletti et al. (2021)

As flavanonas são caracterizadas pela ausência de ligação dupla no anel C e apresentam um centro quiral no C-2. As principais flavanonas encontrados em plantas são a naringina e hesperedina, comumente encontradas em frutas cítricas. As isoflavonas possuem o anel B ligado ao C3 ao invés C2. As isoflavonas são encontradas quase exclusivamente em leguminosas, sendo a soja amplamente estudada e conhecida como fonte de isoflavonas. A daidzeína e a genisteína são as principais representantes dessa classe (Rodríguez-Mateos et al., 2014).

3.1.2. Compostos fenólicos extraíveis (CFE) e não extraíveis (CFNE)

Grande parte dos estudos sobre os compostos fenólicos se concentra nos compostos extraídos de alimentos vegetais com solventes orgânicos ou mistura de água e solventes orgânicos. No entanto, uma parte considerável dos compostos fenólicos não é extraída pela aplicação de extração direta com solventes orgânicos e, portanto, é ignorada. Estudos recentes demonstraram que estes compostos fenólicos não extraíveis são uma parte importante da dieta total e que exibem uma atividade biológica significativa (**Tabela 3**) (Durazzo, 2018; Pérez-Ramírez et al., 2018).

Tabela 3. Teor de CFNE em frutas.

Matriz	% de CFNE	Referência
Maçã	86%	Arranz et al. (2010)
Pêssego	77%	Arranz et al. (2010)
Nectarina	80%	Arranz et al. (2010)
Tangerina	40%	Esparza-Martínez et al. (2016)
Uva	25%	Pérez-Ramírez et al. (2018)
Araçá vermelho	38%	Mallmann et al. (2020)
Araçá amarelo	33%	Mallmann et al. (2020)

Os compostos fenólicos extraíveis (CFE) são aqueles que, por definição, são extraídos da matriz pela aplicação direta de solventes orgânicos ou misturas desses com água. Na sua maioria, são moléculas de baixa massa molecular agrupados em diversas classes e subclasses (descritas previamente no item 3.1.1). Por outro lado, os compostos fenólicos não extraíveis (CFNE) permanecem nos resíduos das extrações e somente são liberados por hidrólise química ou enzimática. Os CFNE obtidos em diversos estudos são na sua maioria ácidos fenólicos, ésteres de ácidos fenólicos e taninos condensados (proantocianidinas) que encontram-se ligados por ligações fracas ou simplesmente imobilizados fisicamente pelas fibras da matriz (Durazzo, 2018; Mallmann et al., 2020; Pérez-Ramírez et al., 2018).

A extração de compostos fenólicos é afetada por fatores como o solvente utilizado, tempo e temperatura assim como pela matriz da qual se deseja extraí-los. Devido à sua natureza complexa, a extração dos CFNE deve seguir duas etapas. Primeiramente, a matriz é submetida a uma extração exaustiva com solvente (água e solventes orgânicos) para a retirada total dos compostos fenólicos extraíveis e de outros compostos orgânicos solúveis. O resíduo da amostra deve passar, então, por hidrólise química ou enzimática para liberação dos CFNE (Pérez-Jiménez et al., 2013).

A hidrólise química, por sua vez, pode ser ácida ou alcalina. Para a hidrólise ácida são utilizados ácidos inorgânicos como o ácido clorídrico e o ácido sulfúrico. Neste tratamento ácido, ocorre a hidrólise das ligações glicosídicas da celulose e hemicelulose. Neste método, o binômio tempo e temperatura é essencial, já que estudos reportaram perdas de flavonóis em tratamentos a 90 °C por uma hora. Metanol acidificado com HCl tem demonstrado bons resultados na extração dos compostos fenólicos ligados, sendo um método acessível e rápido (Arranz & Calixto, 2010; Moussa-Ayoub et al., 2011).

Na hidrólise alcalina acontece a quebra da ligação éster do composto fenólico ao tecido vegetal causando a liberação do composto fenólico da cadeia de polissacarídeo. As bases mais utilizadas para o tratamento alcalino são: hidróxido de sódio, hidróxido de amônio e hidróxido de cálcio. NaOH em concentrações de 2 a 4 mol L⁻¹ tem sido o método mais utilizado. Infelizmente, o uso da hidrólise alcalina pode causar degradação de alguns compostos fenólicos. Alguns estudos relataram perdas de até 67 % de ácido cafeico em sementes oleaginosas que

sofreram tratamento alcalino (Domínguez-Rodríguez et al., 2022; Krygier et al., 1982; Verma et al., 2009).

Já a hidrólise enzimática é realizada através da aplicação de uma mistura de enzimas, tais como celulases, hemicelulases, pectinases, amilases e glucanases. A hidrólise por reação enzimática tem seu efeito dependente da ligação do composto à matriz e da enzima aplicada, além disso, é influenciada pelo tamanho de partícula da amostra e pelo solvente utilizado após a hidrólise. Como desvantagem, esse método tem custo elevado (Meyer et al., 1998; Moussa-Ayoub et al., 2011).

3.1.1. Composição fenólica das frutas

As frutas são as principais fontes de compostos fenólicos da dieta humana. As cinco frutas consumidas com mais frequência no Brasil são, nomeadamente, banana, maçã, laranja, mamão e manga. Diferentes tipos de frutas contêm diferentes perfis de compostos fenólicos e a combinação desses compostos no organismo podem gerar uma sinergia na diminuição do risco de desenvolvimento de algumas doenças e na manutenção da saúde (Rodríguez-Mateos et al., 2014).

Frutas de variedades distintas podem ter formação de diferentes compostos. O grau de maturação também influencia diretamente o perfil de compostos fenólicos. Frutas mais verdes tendem a ter maior quantidade de fenólicos totais quando comparadas com frutas mais maduras. O solo, condição climática e genética da planta são fatores que também influenciam a composição fenólica das frutas (Fateme et al., 2012).

Banco de dados abertos, como o Phenol-Explorer (<http://phenol-explorer.eu/>) apresentam de forma resumida a composição fenólica de alimentos e agrupam resultados de diferentes pesquisas. Porém, os resultados apresentados pelo Phenol-Explorer mostram uma ínfima parte de compostos fenólicos encontrados em artigos científicos, especialmente quando se trata de frutas tropicais.

A banana (*Musa acuminata* L.) é a fruta mais comumente consumida no Brasil (IBGE, 2020) e é conhecida por sua riqueza em ácido gálico, catequina, epicatequina, proantocianidinas até tetrâmeros (Gu et al., 2004; Pascual-Teresa et al., 2000; Sidhu & Zafar,

2018). Um estudo com variedades australianas de banana identificou também ácidos hidroxicinâmicos, hidroxibenzóicos e hidroxifenilacéticos, além de antocianinas, flavonóis, flavanóis, flavonas, flavanonas e hidroxicomarinas (Bashmil et al., 2021). O grau de maturação das bananas foi estudado e verificou-se que bananas mais verdes apresentam maior quantidade de fenólicos totais quando comparadas com bananas mais maduras (Fatemeh et al., 2012). Hellström et al. (2009) estudou a fração de CFNE de bananas e encontrou uma proporção significativa de proantocianidinas na fruta com grau médio de polimerização de 102.

A maçã (*Malus domestica* Borkh), segunda fruta mais consumida no Brasil, já foi muito estudada quanto ao seu perfil fenólico uma vez que representa uma cultura largamente explorada nos hemisférios norte e sul. Banco de dados e estudos de revisão dividem os dados por variedades ou por produtos (cidra, suco, purê, etc.) devido a quantidade de dados disponíveis sobre a fruta. Maçãs são ricas em flavonóis, como a quercetina, mas também são uma boa fonte de proantocianidinas (Gu et al., 2004; Hellström et al., 2009; Pascual-Teresa et al., 2000; Tsao et al., 2003). Ácidos fenólicos como derivados de ácido cafeico, cumárico e ferúlico fazem parte dos compostos majoritários dessas frutas (Schieber et al., 2001).

Maçãs da variedade ‘Golden Delicious’ com casca tiveram sua fração de CFNE estudada (Arranz et al., 2010). A fração obtida de CFE (20 mg 100 g⁻¹ peso fresco) foi caracterizada por ácido gálico, clorogênico, procianidinas e derivados de epicatequina, floretina e quercetina. Enquanto a fração de CFNE (126 mg 100 g⁻¹ peso fresco) teve seus compostos identificados como ácido gálico, hidroxibenzoico, protatecuico, elágico, vanílico, ferúlico, cinâmico, cafeico e sinápico, além de epicatequina, galocatequina e proantocianidinas. A fração de CFNE representou 86% dos compostos fenólicos totais da fruta.

As laranjas (*Citrus sinensis* L.) possuem na sua composição fenólica a presença de flavanonas como hesperidina e naringenina, mas também são fontes de ácido clorogênico e cumárico, luteolina, miricetina e quercetina (Freeman et al., 2010). O suco de laranja também tem sido explorado já que é mundialmente consumido e bastante apreciado; ele apresenta principalmente flavonas e flavanonas ligadas a diferentes substituintes (Mallmann et al., 2023).

Um estudo com laranja da variedade ‘Navelina’ investigou os efeitos da maturação nas alterações dinâmicas no conteúdo de CFE e CFNE. Os compostos fenólicos totais diminuíram gradualmente com o amadurecimento das frutas, com exceção ácido sinápico na fração de CFE

e dos ácidos ferúlico e cafeico na fração de CFNE. O ácido ferúlico foi o mais dominante de todos os ácidos fenólicos em todos os estágios de crescimento (Hou et al., 2021).

O mamão (*Carica papaya* L.) possui um número reduzido de estudos comparado com as frutas anteriores, visto que é cultivado apenas em regiões tropicais e subtropicais. A base de dados “Phenol-Explorer” apresenta somente um composto fenólico para essa fruta: a lignina Matairesinol. Poucos compostos fenólicos foram relatados para a polpa do mamão. Ácido cafeico, gálico, *p*-cumárico e siríngico foram encontrados no mamão inteiro (polpa, casca e semente) (Jadaun et al., 2023). Sete compostos fenólicos foram identificados em resíduo de mamão (casca e semente), nomeadamente ácidos gálico, ferúlico, clorogênico, cafeico e cumárico, além de rutina e quercetina (Fuentes et al., 2023).

A manga (*Mangifera indica* L.) foi mais estudada em relação a sua composição fenólica do que o mamão. Catequina foi o flavonoide encontrado na manga e o único composto registrado no “Phenol-Explorer” para essa fruta (Arts et al., 2000). Polímeros de proantocianidinas foram relatados também para a manga com grau de polimerização de até 6 (Gu et al., 2004). Apesar de não se ter o estudo da fração de CFNE da manga, um estudo encontrou 15 ácidos cinâmicos, 14 ácidos fenilpropanoicos, 12 ácidos fenilacéticos, 28 ácidos benzoicos, 2 ácidos mandélicos e 15 hidroxibenzenos na urina e plasma de indivíduos entre 0 e 24 horas após a ingestão de polpa de manga (Cáceres-Jiménez et al., 2023).

3.2. Análise de compostos fenólicos por LC-MS

3.2.1. Cromatografia Líquida (LC)

A cromatografia líquida de alta eficiência (HPLC) é o método dominante para a análise de compostos fenólicos. Nesta técnica acontece a separação dos componentes de uma amostra, os quais se distribuem em duas fases, uma estacionária (coluna) e a outra móvel (solventes). Colunas cromatográficas de fase reversa, essencialmente octadecilsilano (C18), são as mais utilizadas mesmo que sua habilidade seja limitada quando se trata de polifenóis de alta massa molecular. A separação dos analitos ocorre utilizando-se gradiente de eluição de água ultrapura e acetonitrila, ambos acidificados com ácido um ácido orgânico fraco, como o fórmico (Pereira et al., 2010; Liu et al., 2017).

Os compostos fenólicos possuem múltiplas ligações duplas conjugadas e, assim, possuem absorção no espectro UV-vis. Para uma ampla análise dos compostos fenólicos de diversas classes, o detector de arranjo de diodos (DAD) permite uma varredura completa da absorção dos compostos na região do UV-vis (Motilva et al., 2013). Compostos fenólicos tem absorção máxima por volta de 280, 320, 360 e 520 nm (Rohr et al., 2000).

Apesar dessas diferenças, diversos compostos de mesmo máximo de absorção coeluem (possuem o mesmo tempo de retenção), fazendo com que a identificação desses compostos seja dificultada. A complexidade de composição de uma matriz vegetal impossibilita a identificação dos compostos similares somente pelo detector DAD (Mezni et al., 2018). Além disso, a identificação de compostos fenólicos devidamente separados necessita do uso de padrões analíticos comerciais para fins de comparação e confirmação. Isso onera muito o custo da análise uma vez que uma amostra pode conter centenas de compostos fenólicos e não seria viável a aquisição de padrões analíticos para cada composto encontrado. O método HPLC-DAD ainda que sensível, apresenta muitas dificuldades na identificação de compostos fenólicos quando não há a comparação com padrões analíticos (Mesquita & Monteiro, 2018).

3.2.2. Espectrometria de massas (MS)

A utilização da espectrometria de massas permite a obtenção de informações de grande relevância para a determinação dos compostos fenólicos. A técnica permite a anotação e identificação dos compostos, pois fornece informação estrutural dos analitos, mesmo sem a comparação com padrões analíticos (Kruve et al., 2015).

A introdução da amostra no equipamento pode se dar de forma direta ou conectada em série após a saída do equipamento HPLC-DAD, formando o sistema conhecido como HPLC-DAD-MS. O espectrômetro de massas mede a concentração de moléculas ionizadas geradas na amostra. Dentre os métodos de ionização, as fontes ESI são muito versáteis e por isso as mais aplicadas para os compostos fenólicos.

Inicialmente, o analito é introduzido na fonte ESI para dessolvatação e geração de carga. Nesta etapa, moléculas desprotonadas são produzidas por reações ácido-base de Brønsted-Lowry (protonação/desprotonação). A maioria dos compostos fenólicos tendem a desprotonar

com mais facilidade, deste modo, as análises de compostos fenólicos são geralmente operadas no modo negativo. Esses “íons de elétrons pares” ($[M-H]^-$) normalmente exibem baixo conteúdo de energia interna e, conseqüentemente, pouca ou nenhuma fragmentação. Esses íons passam pelo analisador de massas e temos o espectro MS no formato razão massa-carga (m/z) que é igual a massa molecular do composto dividido pelo número de cargas na estrutura (La Barbera et al., 2017).

A célula de colisão promove a fragmentação da molécula para obtenção dos espectros resultantes: MS^2 (ou MS/MS) e os íons fragmentados gerados são detectados no detector de massas. Esses fragmentos podem ser usados para fins de elucidação estrutural. Normalmente, as reações de fragmentação ocorrem através de mecanismos combinados e podem ocorrer mesmo em baixas energias de colisão, dependendo da estrutura química, método de ionização, alvo de colisão em massa e outros parâmetros do equipamento (Cheng & Gross, 2000).

A fragmentação por rearranjo de hidrogênio são aquelas envolvidas em reações de desidratação em compostos contendo um grupo hidroxila em suas estruturas, como ácidos fenólicos. A eliminação do açúcar de compostos fenólicos glicosilados também ocorre através do rearranjo de hidrogênio (**Figura 3**). Embora estas reações sejam uma das fragmentações mais prevalentes em produtos naturais, as intensidades relativas dos íons do produto no espectro MS/MS nem sempre são altas. O fator de entropia coloca restrições estéricas na formação do produto. No entanto, essas reações podem ser favorecidas energeticamente quando a ligação π formada é adjacente ao local de carga (Demarque et al., 2016).

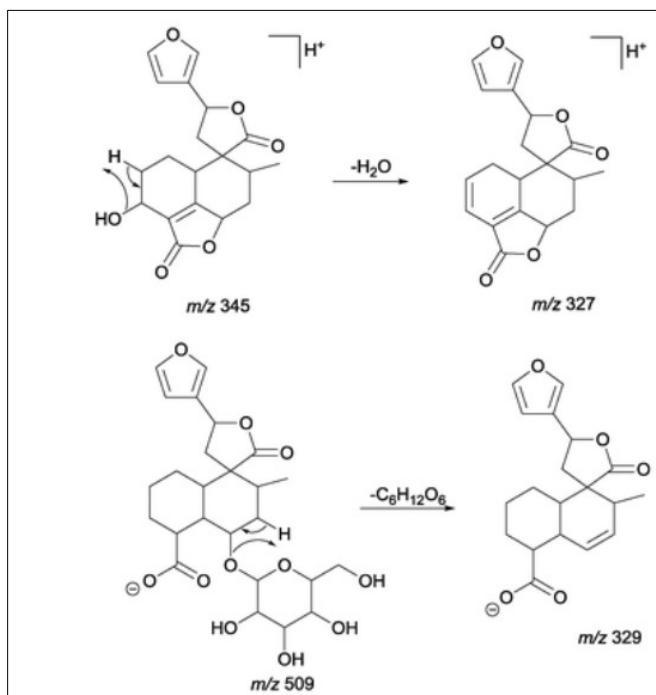


Figura 3. Mecanismo da fragmentação por rearranjo de hidrogênio remoto.

Fonte: adaptado de Demarque et al., 2016.

As reações Retro-Diels-Alder (RDA) desempenham um papel fundamental na anotação de flavonoides e seus derivados. A análise dos fragmentos resultantes permite a determinação dos grupos ligados ao núcleo estrutural dos flavonoides e a variação estrutural nesta classe de compostos (Delcambre & Saucier, 2012). Reações RDA são a reação de fragmentação mais importante para flavonas, dihidroflavonas, flavonóis e isoflavonas. A fragmentação no anel C pode ocorrer em todas as subclasses de flavonoides, e o mecanismo RDA pode seguir as vias a, b e/ou c (**Figura 4**), de acordo com a subclasse estrutural. A ligação dupla presente no O do anel C de flavonas, flavonóis, chalconas e isoflavonas leva a um mecanismo RDA adicional em comparação com dihidroflavonas e dihidroflavonóis (de Rijke et al., 2006).

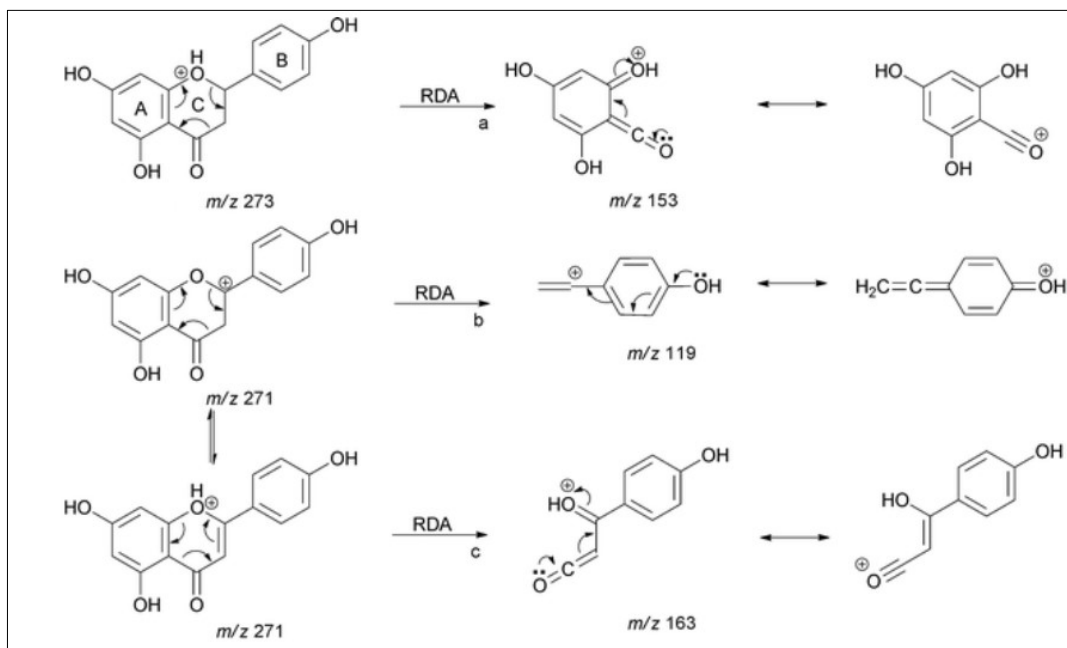


Figura 4. Mecanismo da fragmentação por reação Retro-Diels–Alder (RDA).

Fonte: adaptado de Demarque et al., 2016.

Nas proantocianidinas, o grupo ligado ao anel A é geralmente outra unidade catequina, que desempenha um papel importante na elucidação dos dímeros de proantocianidinas (**Figura 5a**) (Jaiswal et al., 2012). A unidade açúcar dos compostos fenólicos glicosilados também pode fragmentar por RDA. Normalmente, uma perda de H_2O precede a fragmentação RDA pelo rearranjo remoto do hidrogênio, formando uma porção insaturada na unidade de açúcar, que permite o processo RDA (**Figura 5b**). É importante observar a preferência de formação de insaturação entre os carbonos 4' e 5' para formar um alceno mais substituído. Ocasionalmente, o fragmento desidratado não é detectado (Hui et al., 2007).

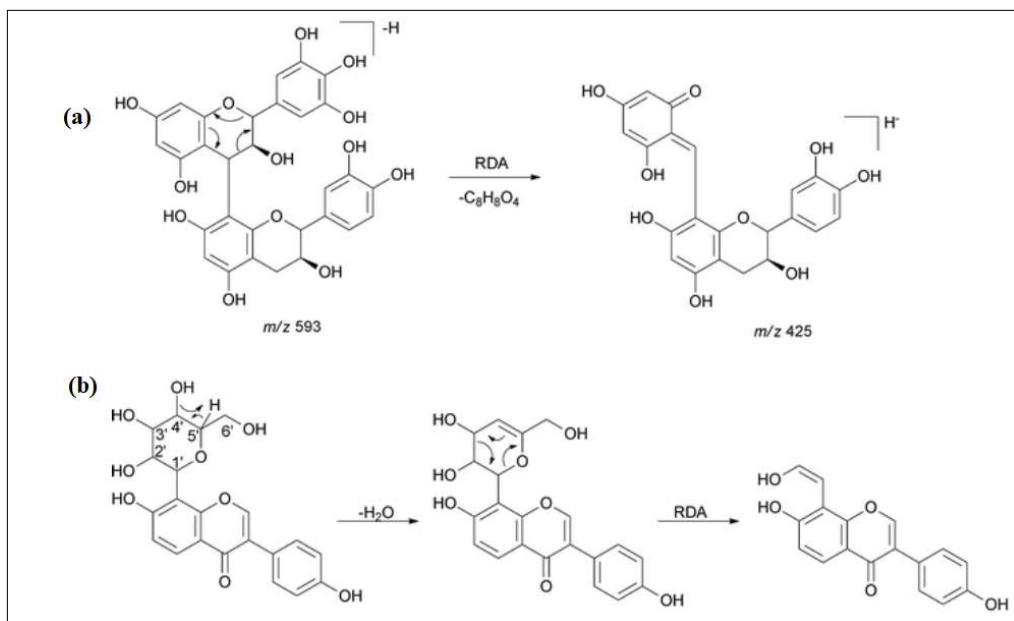


Figura 5. Mecanismo da fragmentação por reação Retro-Diels–Alder (RDA) em (a) procianidinas e (b) unidades de hexoses.

Fonte: adaptado de Demarque et al., 2016.

Um dos processos de fragmentação mais comuns em compostos fenólicos glicosilados é a eliminação da porção de açúcar. Quando a unidade aglicona possui uma carbonila próxima ao açúcar, espera-se que um rearranjo do tipo McLafferty esteja envolvido nesta eliminação (**Figura 6**).

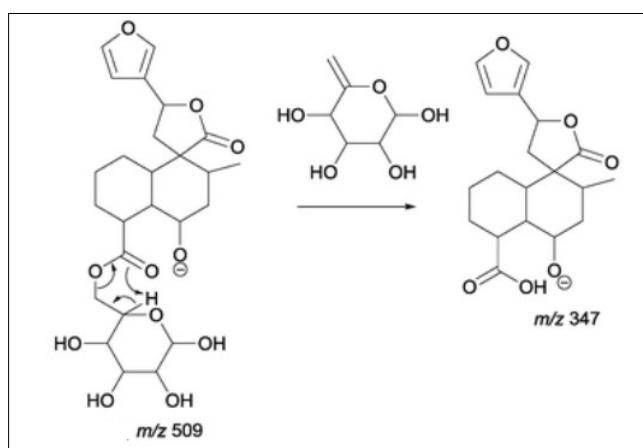


Figura 6. Mecanismo da fragmentação por rearranjo do tipo McLafferty.

Fonte: adaptado de Demarque et al., 2016.

A identificação de compostos fenólicos se dá pela combinação das informações obtidas da espectrometria de massas, como informações espectrais de MS, distribuição isotópica e MS², aliadas a informações ortogonais provenientes da técnica de cromatografia líquida (tempo de retenção, ordem de eluição e absorvância no espectro UV-Vis). O uso combinado dessas informações gera uma melhor capacidade de identificação.

3.3. Ferramentas de bioinformática

A identificação de moléculas permanece uma questão central na química analítica, em particular para pesquisa de produtos naturais, análises de metabolômica não-direcionada, pesquisa ambiental e descoberta de biomarcadores. Devido à sua alta sensibilidade, a espectrometria de massas é adequada para a caracterização de biomoléculas. A interpretação automatizada de dados de espectrometria de massas é muitas vezes limitada à pesquisa em bibliotecas espectrais, de modo que o usuário pode identificar apenas compostos para os quais um padrão analítico foi analisado e seu espectro armazenado.

A interpretação manual é trabalhosa, pois a atual tecnologia de espectrometria de massas pode produzir centenas de milhares de espectros de fragmentação (MS²) por dia em um único instrumento. Os softwares MS-Finder e Sirius utilizam ferramentas de fragmentação *in silico* (via simulação computacional) e são capazes de auxiliar na anotação e identificação dos compostos fenólicos a partir de dados de LC-MS.

3.3.1. MS-Finder

O *software* MS-Finder está disponível gratuitamente na página do Instituto Riken (<http://prime.psc.riken.jp/compms/msfinder/main.html>). O MS-Finder utiliza um algoritmo que prevê a fórmula do metabólito a partir de espectros experimentais de MS e MS² enquanto aplica uma série de regras heurísticas e um banco de dados interno. As fórmulas moleculares de íons precursores são determinadas a partir de informações precisas de massa, razão isotópica e padrão de fragmentação. Todas as estruturas de isômeros da fórmula prevista são recuperadas de bancos de dados próprio e as fragmentações de MS² são previstas por simulação computacional. As estruturas são classificadas por uma pontuação de ponderação combinada

considerando energias de dissociação de ligações, acurácia de massa, fragmentos e nove regras de rearranjo de hidrogênios (Vaniya et al., 2017).

O programa foi validado por sua capacidade de calcular corretamente fórmulas moleculares com 98,0% de precisão para 5.063 registros do banco de dados MS² e produzir o isômero estrutural correto com 82,1% de precisão. Em um teste com 936 espectros identificados manualmente de um conjunto de dados QTOF-MS de plasma humano, as fórmulas foram previstas corretamente em 90,4% dos casos, e a estrutura isomérica correta foi identificada com 80,4% de probabilidade, inclusive para compostos que estavam ausentes em bibliotecas de espectros de massas (Tsugawa et al., 2016). Além disso, no concurso CASMI 2016, O MS-Finder (versão 1.62) foi capaz de identificar corretamente 89% das fórmulas moleculares usando um banco de dados interno composto por 13 repositórios metabolômicos com 45.181 fórmulas. As amostras testadas foram produtos naturais originários de plantas, fungos, esponjas marinhas, algas ou microalgas (Vaniya et al., 2017).

O programa possui uma interface gráfica amigável onde o usuário entra com dados em forma de texto para MS e MS² (**Figura 7**). A partir do íon precursor m/z , modo de ionização e tipo de aduto, as fórmulas candidatas são geradas computacionalmente e filtradas usando regras de valência e razões elementares. As configurações padrão para tipos de adutos (perda de hidrogênio) e tolerância de massa podem ser ajustadas manualmente por meio da interface. As fórmulas candidatas são então classificadas por erros de massa, razão isotópica, fragmentos, perdas neutras e presença em um base de dados de fórmulas de compostos que atualmente contém 90.227 fórmulas (Tsugawa et al., 2016).

Para o ranking de anotação, as estruturas correspondentes são recuperadas de um banco de dados de estrutura que atualmente contém 224.663 estruturas únicas, representando dados de 14 bancos de dados de metabolomas disponíveis. Para fórmulas elementares que não são encontradas no banco de dados de estrutura do metaboloma, o MS-Finder pesquisa o candidato no banco de dados de expansão do metaboloma por simulação computacional MINE23 (atualmente com 643307 estruturas únicas) e no banco de dados PubChem. Finalmente, as estruturas candidatas são pontuadas e classificadas pelo resultado da anotação por simulação de MS/MS usando as regras rearranjo de hidrogênio (**Figura 8**).

Create a query...

Folder path: D:\identification of sample 1 Browse

File name: Compound 1

Meta data

Scan number: -1

Retention time [min]: 0.00

Precursor m/z [Da]: 179.035

Precursor type: [M-H]⁻

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

SMILES:

Intensity: 0

MS1 spectrum or EI spectrum

179.0350	100.0
180.0377	9.8
181.0389	1.2

MS/MS spectrum or EI spectrum

109.0265	2.8
117.0330	6.2
118.0347	0.6
132.0195	1.0
133.0265	3.0
134.0359	81.5
135.0434	100.0
136.0462	8.7

Finish Cancel

Figura 7. Ilustração da interface gráfica do MS-Finder.
Fonte: própria autora.

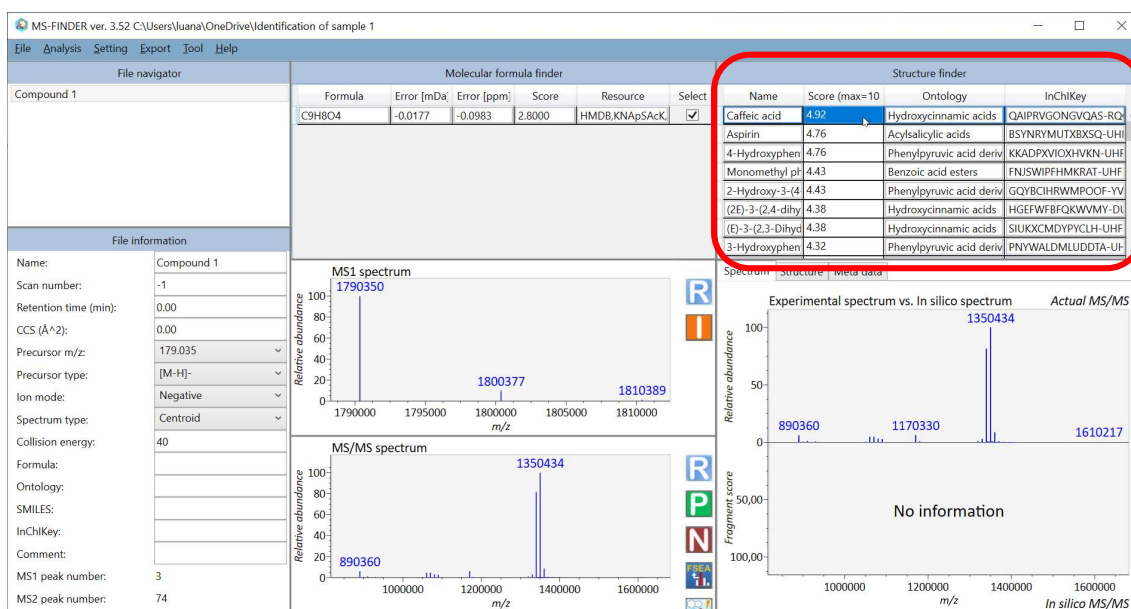


Figura 8. Ilustração da aplicação do MS-Finder para a identificação do padrão de ácido cafeico.

Fonte: própria autora.

3.3.2. Sirius

O *software* Sirius 5 (versão atual) é uma ferramenta especializada que aborda duas questões fundamentais: a fórmula molecular do composto e a estrutura molecular que melhor explica os dados experimentais usando para comparação bases de dados de estruturas moleculares utilizando um buscador da *web CSI:FingerID*, gerando um percentual de similaridade (**Figura 9**). O reconhecimento da fórmula molecular usando análise de padrão isotópico foi implementada pela primeira vez na versão *Sirius 1*, lançada em 2009. Em 2011, o *Sirius 2* acrescentou métodos para a análise de dados MS^2 usando árvores de fragmentação. A versão 3 do Sirius foi lançada em 2015 com uma releitura completa do software e melhorou as taxas de comprovação de fórmulas moleculares corretas dos dados de MS^2 em 2,5 vezes. Diferente das versões anteriores, o *Sirius 3* não tinha uma interface gráfica com o usuário.

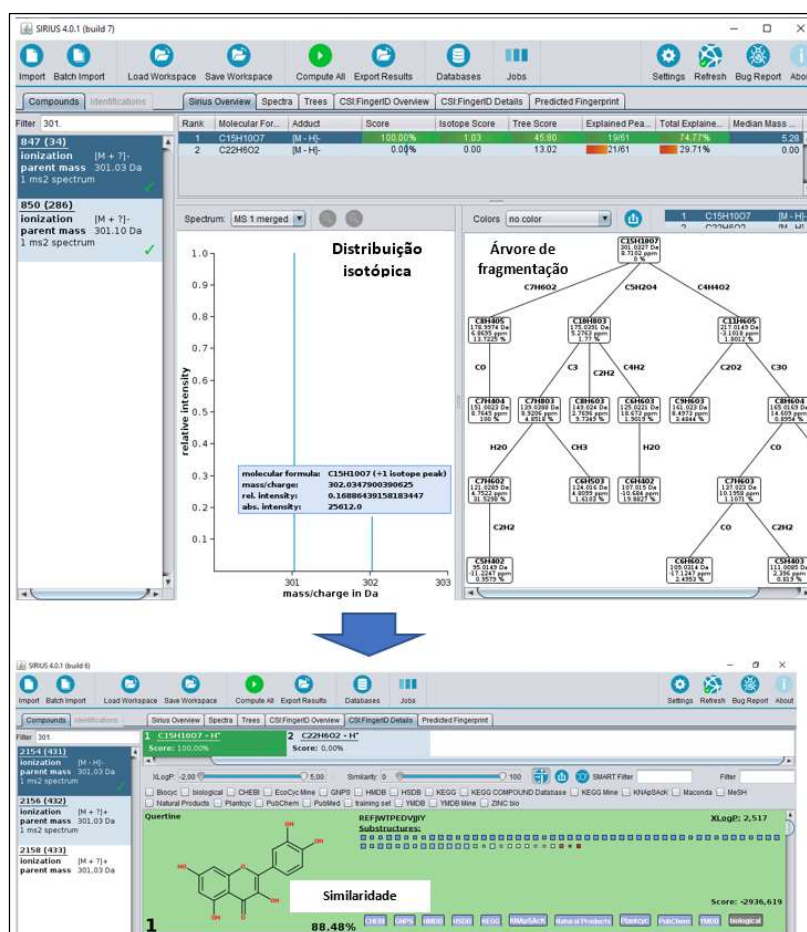


Figura 10. Ilustração da aplicação do Sirius para a identificação do padrão de quercetina.

Fonte: própria autora.

No nível conceitual, o software Sirius integra análise do padrão de distribuição isotópica em alta resolução e árvores de fragmentação com elucidação estrutural para fornecer uma avaliação combinada e coerente de estruturas moleculares de dados de MS² para grandes conjuntos de dados. Os usuários podem analisar conjuntos de dados de cromatografia líquida e espectrometria de massas (LC-MS), em vez de apenas um espectro de cada vez. Considera-se que o Sirius é uma ferramenta poderosa na identificação e anotação de compostos fenólicos em alimentos.

O Sirius identificou corretamente 61% das amostras teste no concurso CASMI 2016 usando fórmulas PubChem e as Sete Regras de Ouro identificaram corretamente 83% usando o Dicionário de Produtos Naturais como banco de dados direcionado. As amostras testadas foram produtos naturais originários de plantas, fungos, esponjas marinhas, algas ou microalgas (Vaniya et al., 2017).

No nível técnico, o software Sirius e seu serviço da Web incluem 213.071 linhas de código Java, das quais cerca de 94% são novas em comparação com a versão anterior. O software fornece uma interface de usuário gráfica intuitiva com seis visualizações. A engenharia extensiva de algoritmos e paralelização através de um sistema de agendamento de trabalho reduziram os tempos de execução do Sirius em mais de duas ordens de grandeza e, assim, reduziram o tempo necessário para analisar uma execução total de LC-MS de horas para minutos. Além disso, ele integra um modelo para a pontuação de padrões isotópicos que combina ruído absoluto e relativo para intensidades de pico e supera todas as pontuações anteriores (Dührkop et al., 2019).

3.3.3. GNPS

A Rede Social Molecular Global de Produtos Naturais (GNPS; <http://gnps.ucsd.edu>), é uma plataforma de acesso aberto que permite o compartilhamento e organização de dados de espectrometria de massas, tanto dados brutos, processados e anotados. O GNPS auxilia na anotação de moléculas a partir de dados de MS² pois tem a capacidade de analisar um conjunto de dados e compará-lo a todos os dados publicamente disponíveis.

As bibliotecas espectrais GNPS permitem a anotação de espectros em redes moleculares pelas correspondências aproximadas aos espectros de moléculas relacionadas e as espelha para

comparação (**Figura 11**). O GNPS coleta bibliotecas espectrais MS² disponíveis relevantes para produtos naturais (metabólitos e moléculas), incluindo MassBank, ReSpec e NIST. Ao todo, essas bibliotecas de terceiros totalizam 212.230 espectros MS² representando 12.694 compostos únicos. Embora esta coleção combinada de espectros de referência forneça um ponto de partida para a replicação, apenas 1,0 % de todos os espectros em conjuntos de dados GNPS públicos identificados (Horai et al., 2010; Sawada et al., 2012).

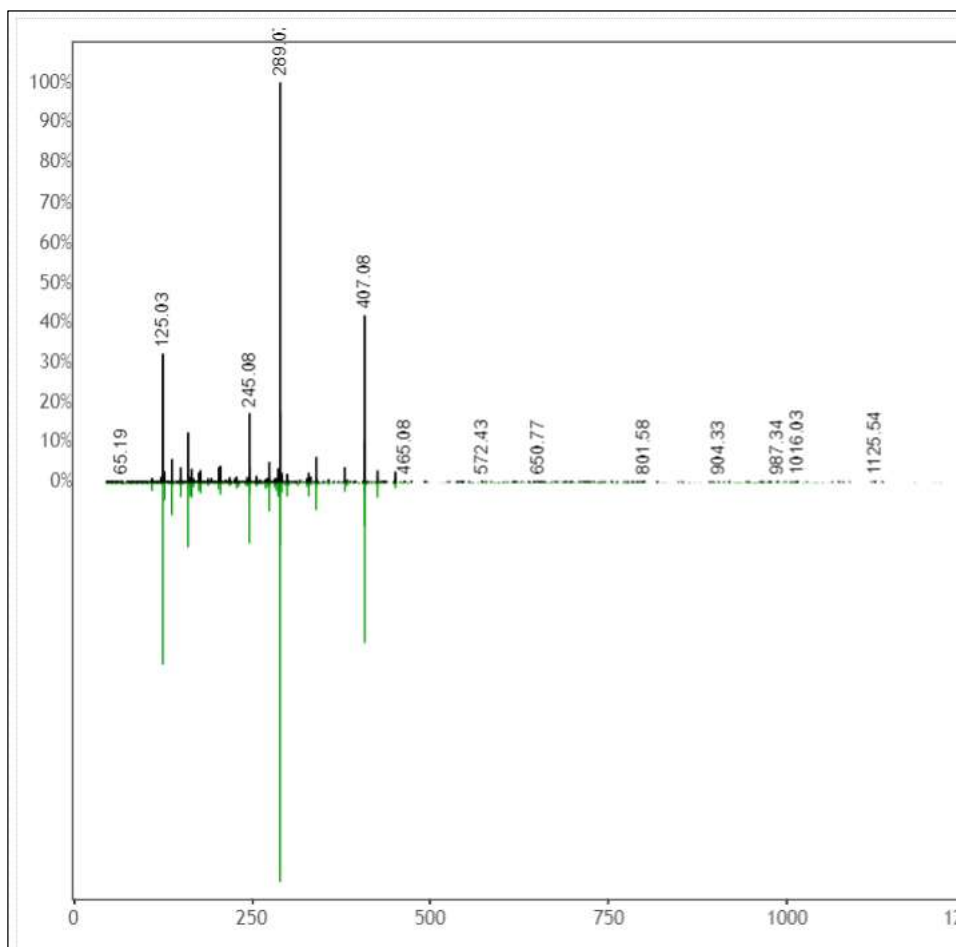


Figura 11. Comparação do espectro de fragmentação de uma amostra (preto) espelhada com a biblioteca espectral do GNPS (verde).

Fonte: obtida da plataforma GNPS (<http://gnps.ucsd.edu>)

Para criar uma biblioteca robusta, é preciso garantir que os envios sejam revisados por pares e, se necessário, as anotações serem corrigidas ou atualizadas conforme apropriado. Os espectros de referência submetidos à biblioteca comunitária GNPS são categorizados pela confiabilidade estimada das submissões propostas. Os espectros de referência de ouro devem

ser derivados de compostos sintéticos ou purificados estruturalmente caracterizados e podem ser submetidos apenas por usuários aprovados. A aprovação é dada aos colaboradores que passaram por treinamento. O treinamento é iniciado entrando em contato com os autores correspondentes ou administradores do programa. Os espectros de referência de prata precisam ser apoiados por uma publicação associada, e os espectros de referência de bronze compreendem todas as supostas anotações restantes (M. Wang et al., 2016).

Com base na infraestrutura computacional do Centro de Espectrometria de Massas Computacional (CCMS) da Universidade da Califórnia em San Diego (UCSD; <http://proteomics.ucsd.edu/>), o GNPS fornece deposição e/ou recuperação de conjuntos de dados públicos através do Repositório de dados do Ambiente Virtual Interativo de Espectrometria de Massas (MassIVE). Isso permite que investigadores de diversas localidades acessem dados brutos de seus pares que analisaram amostras semelhantes, impulsionando a velocidade de análise de dados (M. Wang et al., 2016).

Outra grande ferramenta do GNPS é a criação de redes moleculares. Redes moleculares são exibições visuais do espaço químico presente em experimentos de MS e auxiliam a descoberta de relação estrutural entre moléculas presentes nas amostras a partir de dados de MS². Isso possibilita a descoberta de relação de composição química entre diferentes amostras. O GNPS pode ser usado para redes moleculares, uma correlação espectral e abordagem de visualização que pode detectar conjuntos de espectros de moléculas relacionadas (as chamadas redes espectrais), mesmo quando os próprios espectros não correspondem a nenhum conhecido.

O alinhamento espectral detecta espectros semelhantes de moléculas estruturalmente relacionadas, assumindo que essas moléculas se fragmentam de maneiras semelhantes refletidas em seus padrões MS². A similaridade entre moléculas é calculada pelo cosseno do ângulo entre elas, isso porque o GNPS expressa os espectros MS² de cada molécula como um vetor. Moléculas idênticas apresentam ângulo de valor zero entre elas, ou seja, cosseno igual a 1. Ou seja, quanto mais próximo de 1 é o cosseno, maior a semelhança estrutural entre moléculas. Ao analisar dados, o valor de cosseno é um dos parâmetros a ser seguido. Na área de metabolômica, considera-se 0,7 um valor mínimo aceitável para similaridade entre espectros (Nguyen et al., 2013; Watrous et al., 2012).

O GNPS é atualmente a única infraestrutura pública que permite a criação de redes moleculares. A visualização de redes moleculares no GNPS representa cada espectro como um nodo, e relação entre espectros como arestas (conexões) entre nodos (**Figura 12**). Os nodos podem ser complementados com metadados, incluindo correspondências de replicação ou informações fornecidas pelo usuário, origem do produto, espécie ou tratamento, que podem ser refletidas no tamanho ou na cor de um nodo. É possível visualizar o mapa de moléculas relacionadas como uma rede molecular (Wang et al., 2016; Yang et al., 2013; Bittremieux et al., 2023).

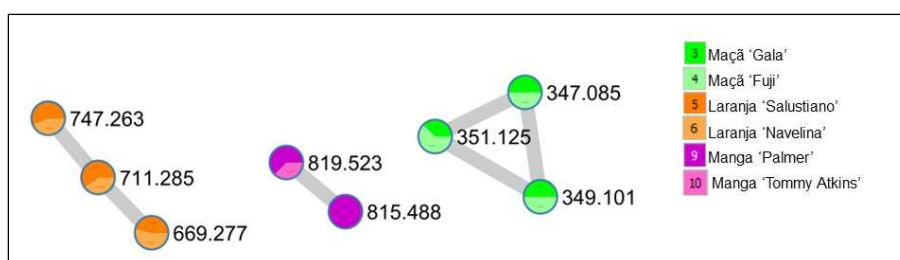


Figura 12. Ilustração de redes moleculares formadas por nodos e arestas.

Fonte: própria autora.

4 ARTIGO 1

(Publicado pela Food Research Internacional em janeiro de 2023;

DOI: <https://doi.org/10.1016/j.foodres.2022.112315>)

MS-FINDER and SIRIUS for phenolic compound identification from high-resolution mass spectrometry data

Luana P. Mallmann¹, Alessandro de Oliveira Rios¹, Eliseu Rodrigues^{1*}

¹Instituto de Ciência e Tecnologia de Alimentos, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, 91501-970, Porto Alegre, Brazil.

*Corresponding author: e-mail: eliseu.rodrigues@ufrgs.br; tel: +555133089786

Abstract

LC-HR-MS/MS is the predominant analytical technique in phenolic compound (PC) research. However, the manual interpretation of mass spectra is a heavy nontrivial time-consuming task and depends on mass spectrometry and phenolic compounds fragmentation deep knowledge. We think this manual approach should be partially translated into a practical software that allows users to perform such complicated analyses. In silico fragmentation software have been tested for small molecule identification, MS FINDER and SIRIUS stood out at identification contests and challenges. We evaluated both software to identify PC from two data categories: 1st MS/MS spectra from 18 phenolic compound standards (PCS) and 2nd phenolic compounds (FPC) from 8 samples (coffee, green tea, cranberry juice, grape juice, orange juice, apple juice, soy extract and parsley extract). MS FINDER and SIRIUS were able to correctly identify more than 90% of the PCS by LC HR-MS/MS. The main FPC were also correctly identified by MS-FINDER (70%) and SIRIUS (38%). We highlight that these software packages were unable to differentiate PC isomers. This task is only possible by using additional information, such as chromatographic behaviour and manual analysis of the relative intensity of fragments in the MS/MS spectra. Therefore, the combination of initial screening by using MS FINDER and SIRIUS with manual analyses of additional information is a powerful and efficient approach for identifying phenolic compounds.

Keywords: phenolic compounds, high-resolution mass spectrometry, identification, bioinformatics

1. Introduction

Phenolic compounds (PC) are the most studied plant secondary metabolites due to either their abundance and possible positive effects on human health. Currently, over 10,000 phenolic compounds have been reported in the literature, varying from phenolic acids and flavonoids to condensed and hydrolysable tannins (Truzzi et al., 2021).

The main technique utilized for their structural elucidation is high-resolution mass spectrometry, which is usually coupled to LC (HPLC or UPLC). Experimental conduction of this analysis is relatively easy, however, MS-data analysis is far away from trivial. Thus, no doubt identification is one of the major hindrances in the PC research field.

There are several major criteria for PC identification, such as elution order in the chromatographic column, UV-vis absorption characteristics, accurate mass, isotopic pattern and fragmentation (MS/MS) pattern. Among these criteria, MS/MS pattern is the most meaningful information to verify molecular substructures and distinguish isomers (Vaniya et al., 2017). PC identification has been performed by using a comparison with authentic compounds analysed under the same conditions, data available in the literature and libraries (Schymanski et al., 2021). However, these data sources are limited, incomplete and, in the case of standards, are expensive.

In silico fragmentation software tools have been utilized as an alternative strategy for compound identification of known compounds and are used to identify MS/MS spectra when the reference MS/MS spectrum is unavailable. Such software tools include MS-FINDER (Vaniya et al., 2017), SIRIUS (Dührkop et al., 2019), CFM ID (Djoumbou-Feunang et al., 2019), MetFrag (Schymanski et al., 2021), MIDAS (Wang et al., 2014), MAGMa (Verdegem et al., 2016) and Input output kernel regression (IOKR) (Brouard et al., 2019).

MS-FINDER algorithm simulates the alpha-cleavage of linear chains up to three chemical bonds and considers also bond dissociation energies. Multiple bonds (double, triple, or cycles) are modelled as penalized single bonds in which hydrogens are lost (hydrogen rearrangement rules). The total score also includes mass accuracy, isotopic ratio, product ion assignment, neutral loss assignment and existence of the compound in an internal structure database. MS-FINDER shows great capacity for unknown compound identification and the MS data can be directly input into the molecular networking analysis platform (Tsugawa et al., 2016; Vaniya et al., 2017; Wang et al., 2022).

SIRIUS is a Java software for analysing metabolites from tandem mass spectrometry data. It combines the analysis of high-resolution isotope patterns in MS spectra with the analysis of fragmentation trees with structural elucidation in MS/MS spectra. SIRIUS first collects all fragmentation spectra and their precursor information, then looks for resources associated with those fragmentation spectra (precursor ions, adduct ions, isotope peaks). The software uses CSI:FingerID as a web service database to search and provide an assessment of molecular structures from MS/MS data and is based on the prediction of a molecular fingerprint of a query compound from MS/MS data (Dührkop et al., 2019; Hoffmann et al., 2021; Vaniya & Fiehn, 2015).

Both software are explored for the identification of metabolites from plants, fungi, marine sponges, algae and micro-algae. MS-FINDER and SIRIUS presented great results at Critical Assessment of Small Molecule Identification (CASMI) contest (Dührkop et al., 2013; Vaniya et al., 2017). In this study, we aim to show that MS FINDER and SIRIUS software can be employed for phenolic compounds identification from data of LC-HR-MS/MS. MS-FINDER and SIRIUS were used for phenolic compound identification from high-resolution mass spectrometry data of LC-HR-MS/MS.

2. Material and methods

2.1. Chemicals

Standards of caffeic, 5-caffeoylquinic, ellagic, ferulic, gallic, hydroxybenzoic, *p*-coumaric and vanillic acids, apigenin, (+)-catechin, daidzein, (-)-epicatechin, kaempferol, naringenin, punicalagin A + B, quercetin, quercetin 3-rutinoside and resveratrol (97-99% purity) were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile and methanol both of HPLC grade were from Honeywell (Charlotte, North Carolina, USA). Formic acid was purchased from Merck (Darmstadt, Germany). Methanol (P.A.) was purchased from Neon Comercial (São Paulo, Brazil). Ultrapure water (Milli-Q) was generated by the Millipore System (Molsheim, FR).

2.2. Standards and samples preparation

Eighteen phenolic compound standards (PCS) from different classes and with different molecular weights were used to represent the complexity of the phenolic composition of different foods (**Figure 1**). PCS were diluted in HPLC-grade methanol to obtain a solution of 20 mg L⁻¹.

Eight food samples with different mixes of phenolic compound classes were used. Samples of juices (apple, grape, cranberry, and orange), green tea in sachets, fresh parsley, soluble coffee brew and soy extract powder were purchased in a local market in Porto Alegre, Rio Grande do Sul, Brazil.

The soluble coffee brew was prepared by dissolving 0.05 g of coffee in ultrapure water (23°C) in a 10 mL volumetric flask. Green tea was prepared using the sachet/water ratio recommended by the manufacturer: 1 sachet of green tea for 200 mL of hot water (80°C). After

3 minutes, the tea was diluted 4 times. Cranberry, grape and apple juices were only filtered. Orange juice was centrifuged at 10,000 g for 15 minutes (15 °C) and the supernatant was used for analysis.

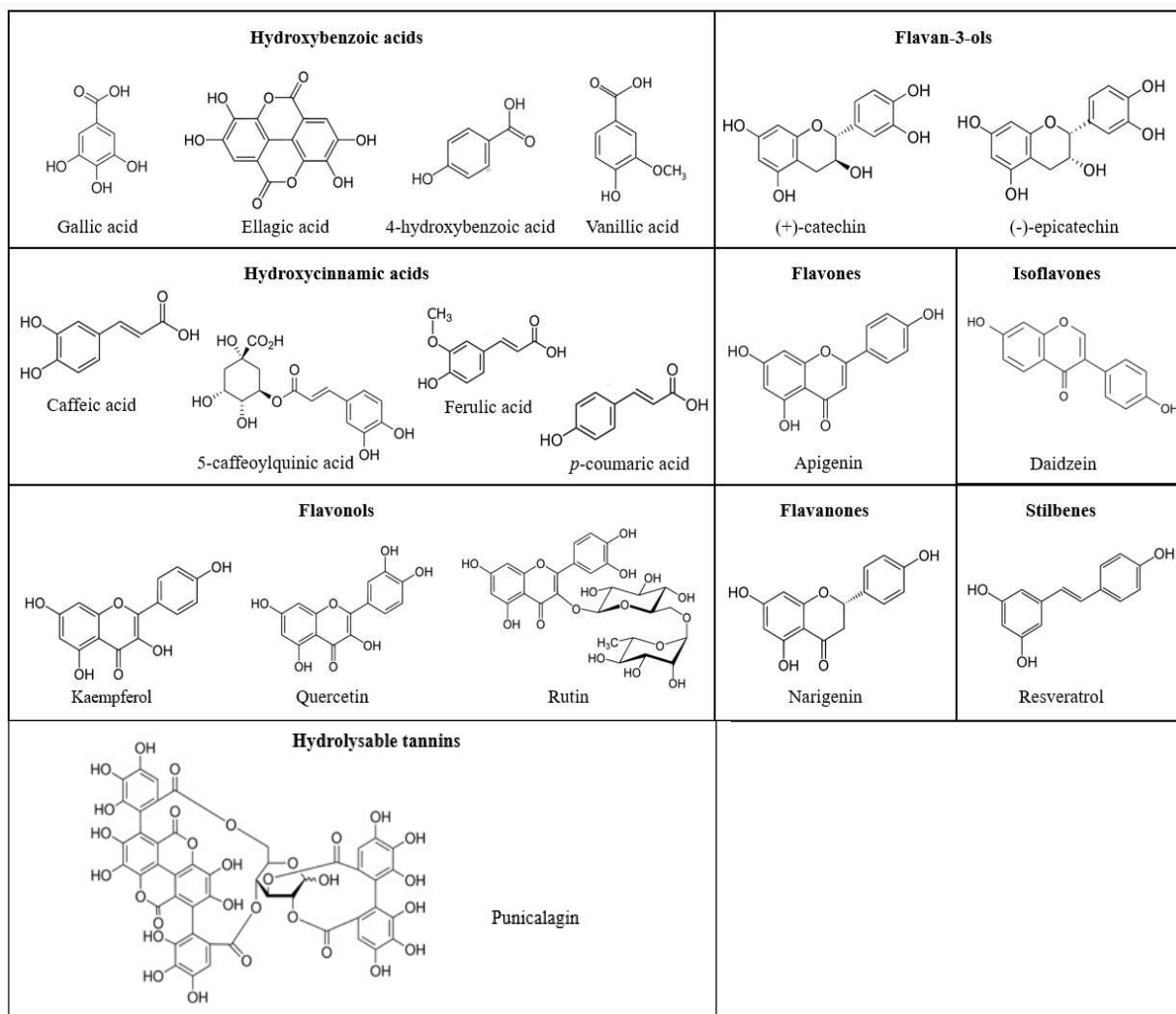


Figure 1. Phenolic compound standards from different classes used for identification tests.

Soy extract was prepared using the extract powder/water ratio recommended by the manufacturer: 2.5g of powder extract for 25 mL of water. The dissolved extract was centrifuged at 10,000 g for 15 minutes (15 °C) and the supernatant was diluted 4 times. Thereafter, the diluted supernatant was filtered.

Fresh parsley was minced to 0.5 cm² particle size. Phenolic compounds were extracted from the sample (2.0 g) using 10 mL of methanol:water mixture (8:2, v/v, acidified with 0.35% formic acid) and vortexing (USC-1400-Unique) for 3 minutes. The extract was centrifuged at 10,000 g for 15 minutes (15 °C) and the supernatant was used for analysis.

All samples were filtered in hydrophilic PTFE membranes (0.22 µm) directly to a vial before LC-HR-MS/MS analysis.

2.3. LC-DAD-ESI-MS/MS analysis

A Shimadzu (Kyoto, Japan) HPLC apparatus connected in series to a DAD detector (SPD-M20A) and a mass spectrometer (MS) with Quadrupole-Time-of Flight (QTOF) analyzer and an electrospray ionization source (ESI) (Bruker Daltonics, micrOTOF-Q III model, Bremen, Germany) were used to analyze the phenolic compounds.

The chromatographic conditions used for phenolic compounds separation were based on Mallmann et al. (2020) with some modifications. The separation was carried out with a Phenomenex C18 (250 mm × 4.6 mm × 4 µm) column. Solvents were classified as A (water acidified with 0.1% formic acid) and B (acetonitrile acidified with 0.1% formic acid). The samples were eluted according to the binary gradient, which began with 5% B as the initial condition and reached 50.2% B at 46 minutes. The flow rate was 0.5 mL min⁻¹ and the injection volume was 20 µL.

After the separation of the phenolic compounds in the column, the eluate entered the ESI interface. The ESI source was operated under the following conditions: negative mode, capillary voltage: 3500 V, scan range of *m/z* 50 to 1500, dry temperature and gas flow (N₂): 310 °C and 8 Lmin⁻¹, nebulizer gas pressure: 4 bar. MS/MS spectra were acquired in Auto-MS² mode (data-dependent acquisition) and the fragmentation energies (eV) were estimated

dynamically based on the set values for the following m/z ratios: 35 eV to m/z lower than 500 Da and calculated proportionally up to 50 eV for 1000 Da and up to 70 eV for 2000 Da.

Instrument calibration was performed using a syringe pump (kdScientific, Holliston, USA) directly connected to the system. A solution of 10 mM sodium formate cluster was used for a calibrant as recommended by Bruker Daltonics (Bremen, Germany) for molecules up to m/z 1500. This same solution was used as an internal calibrant in the LC-HR-MS/MS analysis, where 20 μ L was injected at the first 0.25 minutes of each analysis. This internal calibration was used to automatically calibrate each analysis in the equipment software by obtaining a calibration curve with 20 cluster masses each differing by 68 Da in the negative mode.

2.4. Phenolic compounds identification

The phenolic compounds were identified using two approaches: (1st) manual interpretation, and (2nd) software-aided identification by MS-FINDER and SIRIUS.

For manual identification, the following parameters were considered: elution order in the C₁₈ column, UV-vis absorption characteristics, accurate mass and fragmentation pattern in comparison to standards analyzed under the same conditions and/or data available in the literature and Mona library (<https://mona.fiehnlab.ucdavis.edu/>).

The MS-FINDER software (version 3.52) was downloaded from the Riken institute website (<http://prime.psc.riken.jp/compms/msfinder/main.html>). File name, precursor m/z (Da), MS and MS/MS spectra were typed in the “Create a query” tab. Both spectra contained mass-to-charge-ratio (m/z) of each compound and its relative intensity (%) extracted directly from the LC-HR-MS/MS analysis software Data Analysis 4.3 (Bruker, Germany) raw data were added. Compound identification was realized by a batch job set with formula and structure finder with a 5 top N hits automatically applied to the structure finder program.

The SIRIUS software (version 4.9.12), was downloaded from Lehrstuhl Bioinformatik Jena website (<https://bio.informatik.uni-jena.de/software/sirius/>). The text files of both MS and MS/MS spectra were used. The text files contained the m/z of each compound and its Relative Intensity (%) extracted directly from the LC-HR-MS/MS analysis software Data Analysis 4.3 (Bruker, Germany) raw data. To compute the molecular formulas, instrument type was set as Q-TOF, mass accuracy was set as 20 ppm and possible ionization was selected as $[M-H]^-$. C, H, and O were selected for element searches, and the number of candidates was set to 10. Structure elucidation by CSI:FingerID was set to search in Bio Database and $[M-H]^-$ as the only adduct. Canopus Class Prediction was also abled.

Both software were used on a standard personal computer with a 2.40 GHz Intel Core-i7 CPU and 16 GB of RAM under the Windows 10 operating system. We created an easy step-by-step procedure that will allow all analysts to perform the PC by using MS-FINDER and SIRIUS (**Supplementary material**).

3. Results and Discussion

3.1. Identification of phenolic compounds standards

Both software have shown excellent performance being able to correctly generate 100% of the chemical formulas. We considered as correctly identified when the correspondent PCS appeared ranked as 1st place in the software ranking. A total of 74% and 58% of PCS were correctly identified by MS-FINDER and SIRIUS respectively (**Table 1**). Although the remaining PCS were not considered as correctly identified, they were ranked within Top-7 place for both software. Thus, even in these cases, the software's results might help the researcher, giving a starting point for deeper exploration using its knowledge about the food matrix under

investigation and additional information (i.e. elution order in the C18 column and UV-vis absorption characteristics).

Excellent performance in the PCS identification was found even for those where the exact mass deviation (mass accuracy) was relatively high or the number of fragments was low. The official online documentation for the SIRIUS MS/MS software (<https://boecker-lab.github.io/docs.sirius.github.io>) states that the mass deviation of the MS spectra should be within 20 ppm. However, no molecular formula was found when the mass deviation was higher than 8.2 ppm. In that case, we guaranteed that all PCS had an exact mass deviation lower than 7.5 ppm.

Moreover, at least three representative fragments were imputed as MS/MS data for each compound with exception of hydroxybenzoic acid (HA), which had one fragment and vanillic acid (VA) which did not show any fragment. Both software were able to generate the correct HA molecular formula, however, only MS-FINDER was able to predict the molecular structure. It is important to highlight that the fragments hold structural information and so doing, a few fragment mass spectra will not have enough information to deduce the structure. It is a case of the HA, which is a low-molecular-weight phenolic acid that generates solely the fragment m/z 93 Da, corresponding to the neutral loss of carbon dioxide [HA-44].

Previous tests were conducted by direct infusion of each standard on the ESI source. Vanillic acid presented at least four fragments on those tests and was identified correctly by both software. Without any fragment generated by LC-HR-MS/MS analysis, neither software was able to suggest a molecular structure for this compound.

Table 1. Identification of phenolic compound standards analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Standard	t _R (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS-FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	Gallic acid	12.3	C ₇ H ₆ O ₅	169.0125	7.1	125(100), 124(13), 97(5), 95(4)	Gallic acid (1°)	4.67	Gallic acid (1°)	81.366%
2	5-caffeoylquinic acid	21.1	C ₁₆ H ₁₈ O ₉	353.0848	7.1	191(100), 127(1), 93(1), 85(1)	Chlorogenic acid (1°)	4.25	3-caffeoylquinic acid (1°)	73.311%
3	4-hydroxybenzoic acid	21.8	C ₇ H ₆ O ₃	137.0243	2.9	93(100)	4-Hydroxybenzoic acid (1°)	5.25	n.d	-
4	(+)-catechin	21.8	C ₁₅ H ₁₄ O ₆	289.0726	4.8	123(100), 109(80), 125(46), 151(45)	Epicatechin (1°)	4.63	Catechin (1°)	76.889%
5	(-)-epicatechin	23.5	C ₁₅ H ₁₄ O ₆	289.0727	5.2	123(100), 109(80), 125(37), 151(28)	Epicatechin (1°)	4.63	Catechin (1°)	78.509%
6	Caffeic acid	23.8	C ₉ H ₈ O ₄	179.0345	0.0	135(100), 134(17), 107(2), 117(1)	Caffeic acid (1°)	4.91	Caffeic acid (10°)	58.470%
7	Punicalagin A	24.3	C ₄₈ H ₂₈ O ₃₀	1083.0533	5.0	541(100), 301(64), 275(39), 542(24),	Punicalagin (2°)	3.13	Punicalagin (2°)	69.205%
8	Vanillic acid	24.4	C ₈ H ₈ O ₄	167.0343	1.2	n.d	n.d	-	n.d	-
9	Punicalagin B	25.8	C ₄₈ H ₂₈ O ₃₀	1083.0563	2.2	289(100), 541(74), 290(15), 325(15), 300(100), 301(63), 179(5)	Punicalagin (2°)	3.15	Punicalagin (2°)	59.459%
10	Rutin	27.1	C ₂₇ H ₃₀ O ₁₆	609.1406	8.2	119(100), 93(3), 117(1), 91(1)	Rutin (1°)	4.10	Rutin (6°)	84.498%
11	<i>p</i> -coumaric acid	29.4	C ₉ H ₈ O ₃	163.0393	1.2	229(100), 300(98), 284(74), 201(50)	<i>m</i> -coumaric acid (4°)	4.61	<i>p</i> -coumaric acid (1°)	80.769%
12	Ellagic acid	30.5	C ₁₄ H ₆ O ₈	300.9995	3.3	134(100), 133(22), 178(18), 137(15)	Ellagic acid (1°)	4.14	Ellagic acid (1°)	85.714%
13	Ferulic acid	31.4	C ₁₀ H ₁₀ O ₄	193.0495	3.1	224(100), 209(65), 133(64), 208(63)	Trans-ferulic acid (1°)	4.76	Ferulic acid (2°)	60.513%
14	Daidzein	38.9	C ₁₅ H ₁₀ O ₄	253.0511	4.0	143(100), 117(15), 185(12)	Daidzein (1°)	4.62	Daidzein (1°)	87.135%
15	Resveratrol	39.4	C ₁₄ H ₁₂ O ₃	227.0692	7.0	151(100), 121(34), 107(27), 179(12)	Resveratrol (1°)	4.65	Resveratrol (1°)	78.689%
16	Quercetin	40.6	C ₁₅ H ₁₀ O ₇	301.0354	1.7		Quercetin (1°)	4.73	Quercetin (1°)	82.833%

Table 1 Continued

17	Apigenin	45.4	C ₁₅ H ₁₀ O ₅	269.0459	3.3	117(100), 151(40), 149(25), 107(18)	Apigenin (2°)	4.75	Apigenin (1°)	92.896%
18	Kaempferol	45.8	C ₁₅ H ₁₀ O ₆	285.0419	7.0	185(100), 229(83), 239(75), 211(59)	Kaempferol (1°)	4.77	Kaempferol (7°)	64.407%
19	Naringenin	45.9	C ₁₅ H ₁₂ O ₅	271.0621	5.2	119(100), 117(75), 151(61), 107(34)	Naringenin (1°)	4.74	Naringenin (1°)	78.509%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined

3.2. Identification of phenolic compounds in food sample

The eight food samples (coffee, green tea, cranberry juice, grape juice, orange juice, apple juice, soy extract and parsley) presented a total of 149 identified PC varying from sixteen to twenty-two PC for each sample (**Table 2**). Here, we considered the correct identification when standard was available (level 1) and by manual identification as probable structure (level 2), according to Schymanski et al. (2014). The software correct identification was considered when the correspondent PC appeared ranked as 1st place in the software ranking. MS-FINDER showed better results by correctly identifying 70% of the PC, while SIRIUS correctly identified 38% of the compounds.

Cranberry juice was the sample with highest number of correctly identified compounds for both software (100% and 69%, respectively). MS-FINDER presented the lower identification with coffee PCs (40%) while SIRIUS showed more difficulties to identify PCs from soy extracts (13%). Those results can be related to the different compositions of PC in those samples. In the following sections, we discuss the characteristic PCs of each sample and present an overview of the results based on the phenolic compound classes.

Table 2. Identification of phenolic compounds in food samples analyzed by LC-HR-MS/MS using MS-FINDER and SIRIUS software.

Sample	N° of compound identified	Identified by MS-FINDER (%)	Identified by SIRIUS (%)
Coffee	20	40%	30%
Green tea	21	81%	43%
Cranberry juice	17	100%	69%
Grape juice	17	76%	53%
Orange juice	17	41%	24%
Apple juice	22	95%	32%
Soy extract	16	69%	13%
Parsley extract	19	68%	26%
Total	149	70%	38%

3.2.1. Coffee

Coffee's PC were mainly hydroxycinnamic acids and resulted in the worst results for MS-FINDER identification. Coffee has caffeoylquinic acids as one of its major PC, which can be mainly as a 5-, 4- or 3-caffeoylquinic acid isomers (5-CQA, 4-CQA and 3-CQA) (**Figure 2**). Moreover, other isomers of feruloylquinic acid, coumaroylquinic acid, caffeoyl-quinolactone, dicaffeoylquinic acid, feruloyl-quinolactone and caffeoyl-feruloylquinic acid were found. All these isomers were not differentiated by MS-FINDER and SIRIUS as shown in **Table 3**.

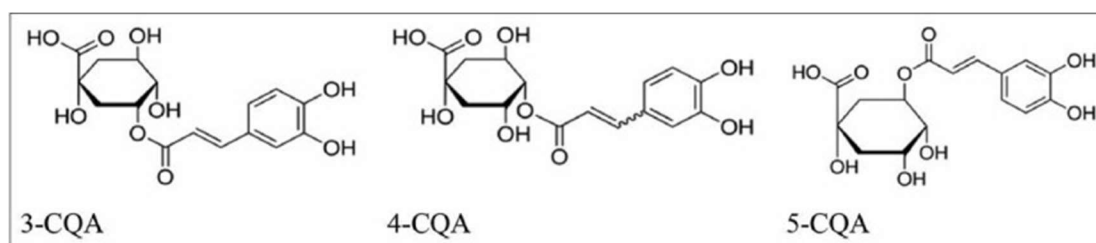


Figure 2. Molecular structure of 3-, 4- and 5-caffeoylquinic acid isomers.

The three main caffeoylquinic acids were identified as chlorogenic acid by MS-FINDER and as 3-caffeoylquinic acid by SIRIUS. Thus, the software's information tells us that these molecules are formed by a caffeic acid molecule linked to a quinic acid molecule. The discrimination among the three main isomers can be performed by using their additional information. 5-, 4- or 3-Caffeoylquinic acid isomers have the same UV-Vis absorption profile with a maximum absorption wavelength at around 326 nm. These isomers have different chromatographic behaviour with the following constant elution order in a C18 column: 3-CQA, 5-CQA and 4-CQA (Rodrigues & Bragagnolo, 2013).

Afterward, we applied the hierarchical scheme of identification described by Clifford et al.,(2003), which is based on fragmentation patterns. Although we utilized mass spectrometry equipment with different mass analysers, some information can be obtained from MS/MS spectra

(**Figure 3**). The 4-CQA was the only one to show an intense signal in MS/MS spectrum at m/z 173 ([quinic acid-H-H₂O]⁻) due to the dehydration of the quinic acid ion, a feature of cinnamoyl group bonded to the quinic acid moiety at the position 4. The 3-CQA and 5-CQA could be distinguished by comparing the relative intensity of the secondary ion from caffeoyl moiety ([caffeic acid-H]⁻) at m/z 179 (Clifford et al., 2003), which corresponded to 60% of the base peak in 3-CQA and 1% in 5-CQA.

3.2.2. Green tea

Green tea is known for its (+)-catechin and (-)-epicatechin abundance which is responsible for its astringent flavor (**Table 4**). Once stereoisomers have the same fragmentation profile, MS-FINDER and SIRIUS are unable to differentiate them. In this case, (+)-catechin and (-)-epicatechin were both randomly identified as (-)-epicatechin by SIRIUS and (+)-catechin by MS-FINDER. In this situation, the additional information is fundamental to differentiate them. Considering that these PC have the same UV-vis absorption profile ($\lambda_{\text{max.}} = 279$ nm), the elution order in the C₁₈ column is the only parameter that can help in this task. Under a conventional elution in the C₁₈ column, (+)-catechin elutes before (-)-epicatechin as reported by several works (Mariutti et al., 2014; Carvalho et al., 2015; De Freitas et al., 2018; Mallmann et al., 2020).

Quercetin 3-rutinoside-7-glucoside (MW = 771.1998 Da) (peak 14) was in the third position on the identification ranking for MS-FINDER while SIRIUS did not show any possible structure for this compound. One single fragment ($m/z = 301$ Da) was found in the MS/MS spectrum of this compound, which was generated by the neutral loss of 3-rutinoside-7-glucoside (C₁₅H₃₀O₁₄). MS-FINDER used this information, added to mass accuracy and isotopic ratio, to search for matching compounds in its local databases and quercetin 3-rutinoside-7-glucoside was in the top-3 ranking. For SIRIUS, one fragment was not sufficient for SIRIUS to draw a proper fragmentation tree and identify this compound.

Table 3. Identification of phenolic compound in coffee sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	t _R (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS-FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	3-caffeoylquinic acid	16.9	C ₁₆ H ₁₈ O ₉	353.0883	2.8	191(100), 179(60), 135(49), 180(6)	Chlorogenic acid (1°)	4.30	3-caffeoylquinic acid (1°)	79.264%
2	4-caffeoylquinic acid	20.3	C ₁₆ H ₁₈ O ₉	353.0884	3.1	191(100), 173(98), 135(82), 179(78)	Chlorogenic acid (1°)	4.30	3-caffeoylquinic acid (1°)	87.713%
3	5-caffeoylquinic acid	21.2	C ₁₆ H ₁₈ O ₉	353.0893	5.7	191(100), 161(2), 179(1), 135(1)	Chlorogenic acid (1°)	4.07	3-caffeoylquinic acid (1°)	84.138%
4	3-Feruloylquinic acid	22.0	C ₁₇ H ₂₀ O ₉	367.1058	7.9	193(100), 134(85), 149(8), 117(8)	3- <i>O</i> -Feruloylquinic acid (3°)	3.81	3- <i>O</i> -Feruloylquinic acid (1°)	82.848%
5	5- <i>p</i> -Coumaroylquinic acid	23.9	C ₁₆ H ₁₈ O ₈	337.0932	2.4	173(100), 93(35), 163(22), 119(20)	3- <i>O</i> - <i>p</i> - Coumaroylquinic acid (1°)	4.16	4- <i>p</i> -Coumaroylquinic acid (1°)	82.796%
6	4- <i>p</i> -Coumaroylquinic acid	25.2	C ₁₆ H ₁₈ O ₈	337.0940	4.7	191(100), 93(26), 119(25), 163(12)	4- <i>p</i> -Coumaroylquinic acid (4°)	3.88	4- <i>p</i> -Coumaroylquinic acid (1°)	73.455%
7	5-Feruloylquinic acid	25.6	C ₁₇ H ₂₀ O ₉	367.1056	7.4	173(100), 134(27), 193(22), 93(16)	4- <i>O</i> -feruloylquinic acid (2°)	3.87	3- <i>O</i> -Feruloylquinic acid (2°)	83.819%
8	3-Caffeoyl-1,5- quinolactone	26.4	C ₁₆ H ₁₆ O ₈	335.0791	7.2	161(100), 173(55), 135(34), 179(32)	3-Caffeoyl-1,5- quinolactone (4°)	3.79	3-Caffeoyl-1,5- quinolactone (5°)	70.405%
9	4-Feruloylquinic acid	26.9	C ₁₇ H ₂₀ O ₉	367.1052	6.3	191(100), 93(23), 193(19), 134(12)	4-Feruloylquinic acid (2°)	3.88	4-Feruloylquinic acid (3°)	81.818%
10	5-Caffeoyl-1,5- quinolactone	27.1	C ₁₆ H ₁₆ O ₈	335.0759	2.4	161(100), 133(16), 179(3), 137(3)	3-Caffeoyl-1,5- quinolactone (4°)	3.80	3-Caffeoyl-1,5- quinolactone (6°)	53.055%
11	4-Caffeoyl-1,5- quinolactone	27.7	C ₁₆ H ₁₆ O ₈	335.0766	0.3	161(100), 133(15), 179(3), 134(3)	3-Caffeoyl-1,5- quinolactone (4°)	3.81	3-Caffeoyl-1,5- quinolactone (5°)	60.000%
12	3,4-Dicaffeoylquinic acid	31.2	C ₂₅ H ₂₄ O ₁₂	515.1216	5.1	179(100), 173(96), 191(45), 353(29)	Dicaffeoylquinic acid (1°)	3.94	3,4-Di- <i>O</i> - caffeoylquinic acid (1°)	81.433%
13	3-Feruloyl-1,5- quinolactone	31.9	C ₁₇ H ₁₈ O ₈	349.0939	4.4	175(100), 160(92), 134(72), 193(32)	3-Feruloyl-1,5- quinolactone (2°)	3.67	3-Feruloyl-1,5- quinolactone (6°)	52.353%
14	5-Feruloyl-1,5- quinolactone	32.5	C ₁₇ H ₁₈ O ₈	349.0942	5.3	175(100), 160(96), 93(15), 134(14)	3-Feruloyl-1,5- quinolactone (2°)	3.66	4-Feruloyl-1,5- quinolactone (3°)	62.577%

Table 3 Continued

15	3,5-Dicaffeoylquinic acid	33.1	C ₂₅ H ₂₄ O ₁₂	515.1225	6.8	191(100), 179(71), 353(18), 135(13)	Dicaffeoylquinic acid (1°)	3.92	Dicaffeoylquinic acid (4°)	73.826%
16	4-Feruloyl-1,5-quinolactone	33.4	C ₁₇ H ₁₈ O ₈	349.0931	2.1	175(100), 160(65), 134(15), 93(12)	4-Feruloyl-1,5-quinolactone (3°)	3.67	4-Feruloyl-1,5-quinolactone (9°)	54.762%
17	4,5-Dicaffeoylquinic acid	33.7	C ₂₅ H ₂₄ O ₁₂	515.1216	5.1	173(100), 179(84), 353(39), 191(30)	3,4-Di- <i>O</i> -caffeoylquinic acid (3°)	3.94	3,4-Di- <i>O</i> -caffeoylquinic acid (1°)	86.957%
18	3-Feruloyl,4-caffeoylquinic acid	35.0	C ₂₆ H ₂₆ O ₁₂	529.1353	1.3	193(100), 173(82), 179(53), 155(19)	1-Caffeoyl-5-feruloylquinic acid (3°)	3.54	1-Feruloyl-5-caffeoylquinic acid (10°)	62.356%
19	3-Caffeoyl,4-feruloylquinic acid	35.5	C ₂₆ H ₂₆ O ₁₂	529.1367	4.0	173(100), 193(22), 179(10), 155(8)	3-Caffeoyl-4-feruloylquinic acid (5°)	3.58	3-Caffeoyl-4-feruloylquinic acid (3°)	78.659%
20	4-Caffeoyl,5-feruloylquinic acid	38.0	C ₂₆ H ₂₆ O ₁₂	529.1386	7.6	191(100), 179(36), 193(29), 173(21)	3-Caffeoyl-4-feruloylquinic acid (5°)	3.56	3-Caffeoyl-4-feruloylquinic acid (4°)	66.667%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid.

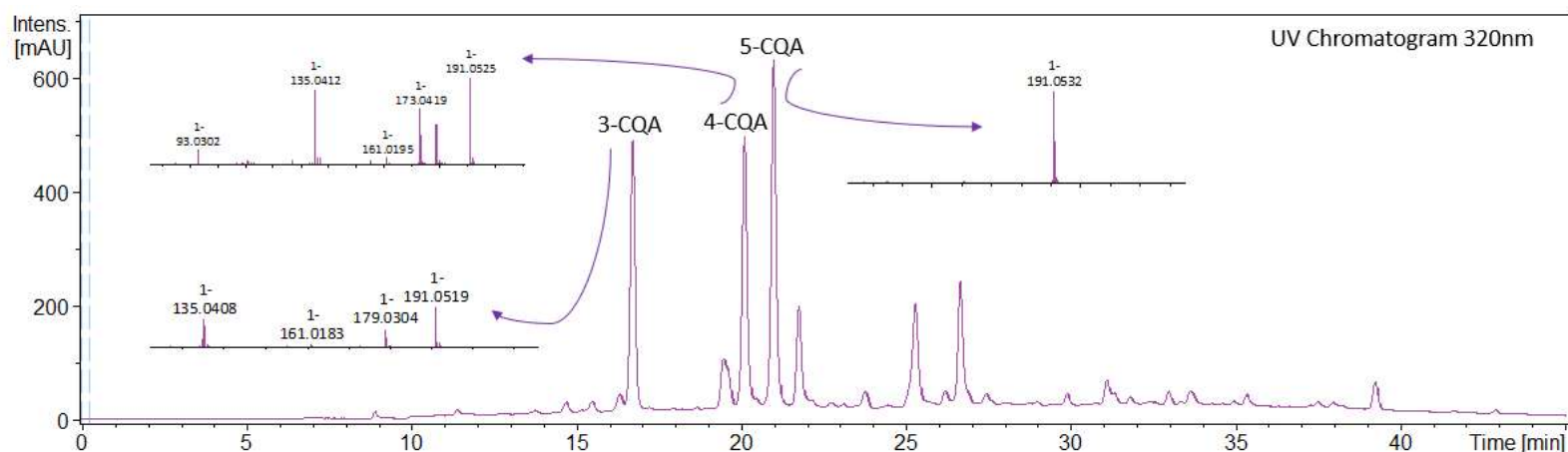


Figure 3. UV chromatogram and fragmentation pattern of 3-, 4- and 5-caffeoylquinic acid isomers.

Table 4. Identification of phenolic compound in green tea sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	tr (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS- FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	Gallic acid	12.4	C ₇ H ₆ O ₅	169.0135	1.2	125(100), 124(13), 97(5), 95(4)	Gallic acid (1°)	4.73	Gallic acid (1°)	81.366%
2	Gallocatechin	17.1	C ₁₅ H ₁₄ O ₇	305.0678	5.4	125(100), 137(41), 139(35), 167(28)	(-)-Epigallocatechin (1°)	4.46	Gallocatechin (1°)	80.631%
3	3-caffeoylquinic acid	16.9	C ₁₆ H ₁₈ O ₉	353.0867	1.7	191(100), 127(1), 93(1), 85(1)	Chlorogenic acid (1°)	4.30	3-caffeoylquinic acid (1°)	73.311%
4	Epigallocatechin	19.0	C ₁₅ H ₁₄ O ₇	305.0684	7.4	125(100), 109(34), 167(32), 165(26)	(-)-Epigallocatechin (1°)	4.44	Gallocatechin (1°)	81.567%
5	4-caffeoylquinic acid	20.3	C ₁₆ H ₁₈ O ₉	353.0870	0.8	191(100), 135(89), 173(68), 179(57)	Chlorogenic acid (1°)	4.30	n.d	-
6	5-caffeoylquinic acid	21.1	C ₁₆ H ₁₈ O ₉	353.0874	0.3	191(100), 127(1), 93(1), 85(1)	Chlorogenic acid (1°)	4.30	3-caffeoylquinic acid (1°)	73.311%
7	Isovitexin-7- <i>O</i> - glucoside (saponarin)	21.4	C ₂₇ H ₃₀ O ₁₅	593.1505	0.3	305(100), 125(92), 423(66), 407(43)	Neosaponarin (2°)	3.91	Isovitexin 8-C- glucoside (4°)	54.407%
8	(+)-catechin	21.8	C ₁₅ H ₁₄ O ₆	289.0727	5.2	123(100), 109(78), 125(42) 121(26)	Epicatechin (1°)	4.63	Catechin (1°)	76.923%
9	Quercetin 3-(2"- galoylrutinoside)	22.0	C ₃₄ H ₃₄ O ₂₀	761.1589	3.2	423(100), 125(61), 424(39), 301(29)	Quercetin 3-(2"- galoylrutinoside) (1°)	2.99	Quercetin 3-(2"- galoylrutinoside) (3°)	50.860%
10	(-)-epicatechin	23.4	C ₁₅ H ₁₄ O ₆	289.0734	7.6	123(100), 109(91), 137(43), 151(34)	Epicatechin (1°)	4.61	Catechin (1°)	82.805%
11	Naringin	23.5	C ₂₇ H ₃₂ O ₁₄	579.1686	4.8	289(100), 290(15), 125(8), 203(4)	Naringin (1°)	4.01	Naringin (1°)	62.963%
12	Fertaric acid	23.6	C ₁₄ H ₁₄ O ₉	325.0538	6.7	125(100), 109(100), 137(59), 161(30)	Fertaric acid (1°)	3.83	Fertaric acid (3°)	41.156%
13	Vitexin-2-rhamnoside	23.9	C ₂₇ H ₃₀ O ₁₄	577.1559	0.3	289(100), 407(64), 125(63), 290(19)	Vitexin 2"- <i>O</i> -rhamnoside (3°)	3.95	Isovitexin 2"- <i>O</i> - rhamnoside (7°)	49.730%
14	Quercetin 3-rutinoside- 7-glucoside	25.8	C ₃₃ H ₄₀ O ₂₁	771.1998	1.8	301(100)	Quercetin 3-rutinoside 7- galactoside (3°)	3.61	n.d	-

Table 4 Continued

15	Gallocatechin gallate	26.2	C ₂₂ H ₁₈ O ₁₁	457.0788	3.7	169(100), 125(30), 170(2), 305(4)	Epigallocatechin gallate (1°)	4.15	Gallocatechin 3"- <i>O</i> - gallate (4°)	65.018%
16	Rutin	27.2	C ₂₇ H ₃₀ O ₁₆	609.1498	7.1	300(100), 301(72), 302(12), 343(2)	Rutin (1°)	4.14	Rutin (6°)	87.798%
17	Epigallocatechin gallate	27.3	C ₂₂ H ₁₈ O ₁₁	457.0805	7.4	169(100), 125(33), 126(4), 171(3)	Epigallocatechin gallate (1°)	4.12	Gallocatechin 7- <i>O</i> - gallate (6°)	63.448%
18	Myricitrin	29.2	C ₂₁ H ₂₀ O ₁₂	463.0905	6.1	300(100), 301(46), 302(7), 271(3)	Myricitrin (4°)	4.05	Myricitrin (3°)	80.328%
19	Isovitexin-7- <i>O</i> - glucoside (saponarin)	29.5	C ₂₇ H ₃₀ O ₁₅	593.1529	3.8	285(100), 284(45), 286(15), 327(2)	Kaempferol-3-glucoside- 7-rhamnoside (1°)	3.96	Kaempferol-3- glucoside-7- rhamnoside (5°)	81.571%
20	Epicatechin-gallate	30.6	C ₂₂ H ₁₈ O ₁₁	441.0787	7.9	169(100), 125(34), 289(32), 245(19)	(-)-Catechin 3- <i>O</i> -gallate (1°)	4.13	(-)-Epicatechin gallate (1°)	65.862%
21	Quercitrin	31.5	C ₂₁ H ₂₀ O ₁₁	447.0948	4.6	284(100), 285(36), 255(12), 256(8)	Quercitrin (1°)	4.21	Quercitrin (3°)	77.451%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined

3.2.3. Cranberry juice

Besides the presence of caffeic acid derivatives and proanthocyanidins, cranberry juice's PC are mainly flavonols. This sample showed the best results of identification for both software. MS-FINDER and SIRIUS identified correctly 100% and 69% of the PC from cranberry juice (**Table 5**), respectively.

MS-FINDER had no trouble identifying the seventeen compounds of cranberry juice, while SIRIUS could not suggest any structure for β -1-caffeoylglucose and α -1-caffeoylglucose (peak 2 and 3; $[M-H]^- = 341.0847$ Da). The discrimination among these isomers can be performed by using their additional information. β - and α -1-caffeoylglucose isomers have the same UV-Vis absorption profile with a maximum absorption wavelength at around 278 nm and 324 nm, but β -1-caffeoylglucose elutes before α -1-caffeoylglucose in a reversed phase LC-MS method (Jaiswal et al., 2014).

Jaiswal et al. (2014) applied the hierarchical key scheme of identification in 10 different caffeoylglucose stereoisomers and obtained information about MS/MS spectra of those compounds. The β -1-caffeoylglucose was the only one to show an intense signal in MS/MS spectrum at m/z 161 ($[\text{caffeic acid-H-H}_2\text{O}]^-$) due to the dehydration of the caffeic acid ion. The two anomers could be distinguished from others stereoisomers of caffeoylglucose by comparing the relative intensity of the secondary ions from caffeoyl moiety ($[\text{caffeic acid-H}]^-$) at m/z 179 and ($[\text{caffeic acid-CO}_2\text{-H}]^-$) at m/z 135.

The fragmentation pattern information added to mass accuracy and isotopic ratio was sufficient for MS-FINDER to search for matching compounds in its local databases. Eventhough SIRIUS computed the correct molecular formula for both compounds, two fragments were not sufficient for a proper fragmentation tree drawing and identification of this compound. SIRIUS did not show any possible structure for both compounds.

Table 5. Identification of phenolic compound in cranberry juice sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	t _R (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS- FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	3-caffeoylquinic acid	17.0	C ₁₆ H ₁₈ O ₉	353.0879	1.7	191(100), 179(62), 135(51), 161(6)	Chlorogenic acid (1°)	4.31	3-caffeoylquinic acid (1°)	76.898%
2	β-1-caffeoylglucose	17.7	C ₁₅ H ₁₈ O ₉	341.0847	7.6	161(100), 135(50), 179(30)	1- <i>O</i> -Caffeoylglucose (1°)	3.82	n.d	-
3	α-1-caffeoylglucose	19.7	C ₁₅ H ₁₈ O ₉	341.0847	7.6	135(100), 179(71)	1- <i>O</i> -Caffeoylglucose (1°)	4.04	n.d	-
4	B-type Proanthocyanidin	20.3	C ₃₀ H ₂₆ O ₁₂	577.1314	5.5	289(100), 125(52), 407(44), 245(23)	Procyanidin (1°)	3.86	Procyanidin B5 (1°)	56.044%
5	5-caffeoylquinic acid	21.3	C ₁₆ H ₁₈ O ₉	353.0867	1.7	191(100), 161(2), 179(1)	Chlorogenic acid (1°)	4.30	1-caffeoylquinic acid (1°)	68.421%
6	Epicatechin	23.5	C ₁₅ H ₁₄ O ₆	289.0692	6.9	123(100), 109(64), 151(40), 125(38)	Epicatechin (1°)	4.58	Catechin (1°)	67.227%
7	B-type Proanthocyanidin	23.5	C ₃₀ H ₂₆ O ₁₂	577.1345	0.2	289(100), 407(47), 125(46), 245(19)	Procyanidin (1°)	3.90	Procyanidin B5 (4°)	56.491%
8	Myricetin-3- <i>O</i> - hexoside	25.8	C ₂₁ H ₂₀ O ₁₃	479.0852	5.4	316(100), 317(26), 271(3), 287(2)	Myricetin 3-galactoside (1°)	4.18	Myricetin 3-glucoside (1°)	91.304%
9	B-type Proanthocyanidin	23.5	C ₃₀ H ₂₆ O ₁₂	577.1302	7.6	289(100), 203(71), 407(60), 273(53)	Procyanidin (1°)	3.90	Procyanidin B5 (1°)	51.961%
10	Rutin	27.3	C ₂₇ H ₃₀ O ₁₆	609.1433	3.6	300(100), 301(67), 179(4)	Rutin (1°)	4.15	Rutin (9°)	79.942%
11	Myricetin-3- <i>O</i> - pentoside	27.6	C ₂₀ H ₁₈ O ₁₂	449.0705	3.3	316(100), 318(7), 271(5)	Myricetin 3- arabinoside (1°)	3.84	Myricetin 3-xyloside (1°)	80.731%
12	Myricetin-3- <i>O</i> - pentoside	28.5	C ₂₀ H ₁₈ O ₁₂	449.0736	3.6	316(100), 318(4), 179(3), 287(2)	Myricetin 3- arabinoside (1°)	3.85	Myricetin 3-xyloside (1°)	80.135%
13	Quercetin 3-galactoside	29.3	C ₂₁ H ₂₀ O ₁₃	463.0898	4.6	300(100), 301(38), 302(4), 271(3)	Quercetin 3-galactoside (1°)	4.27	Quercetin 3-glucoside (1°)	95.146%
14	Quercetin-3- <i>O</i> - pentoside	31.0	C ₂₀ H ₁₈ O ₁₁	433.0783	2.8	300(100), 301(45), 302(7), 255(5)	Quercetin 3-α-L- arabinofuranoside (1°)	3.84	Quercetin 3- <i>O</i> -β-D- xylopyranoside (1°)	81.311%
15	Quercetin-3- <i>O</i> - pentoside	31.7	C ₂₀ H ₁₈ O ₁₁	433.0796	5.8	300(100), 301(49), 302(7), 255(6)	Quercetin 3-α-L- arabinofuranoside (1°)	3.82	Quercetin 3- <i>O</i> -β-D- xylopyranoside (1°)	84.839%
16	A-type Proanthocyanidin	32.0	C ₃₀ H ₂₄ O ₁₂	575.1190	0.0	285(100), 289(65), 125(31), 423(37)	Proanthocyanidin A2 (1°)	3.72	Procyanidin A2 (2°)	62.583%

Table 5 Continued

17	Quercitrin	32.2	C ₂₁ H ₂₀ O ₁₁	447.0941	3.0	300(100), 301(70), 302(12), 151(3)	Quercitrin (1°)	4.21	Quercitrin (1°)	89.967%
----	------------	------	---	----------	-----	---------------------------------------	-----------------	------	-----------------	---------

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined.

Three isomers of B-type Proanthocyanidin were identified as Procyanidin by MS-FINDER and as Procyanidin B5 by SIRIUS. Information from these software packages tells us that these molecules are formed from a dimer of (+)-catechin and (-)-epicatechin with oxidative coupling occurring between the C4 of the heterocycle and the C6 or C8 positions. The discrimination among the five main isomers (B1, B2, B3, B4 and B5) can be performed by using their additional information. All B-type procyanidin isomers have the same UV-Vis absorption profile with a maximum absorption wavelength at around 280 nm. These isomers have different chromatographic behaviour with the following constant elution order: B3, B1, B4, B2 and B5 (Regos et al., 2009; Dixon et al., 2005).

Little information was obtained from MS/MS spectra since the isomers have very similar fragmentation patterns: intense signal at m/z 289 ($[(\text{epi})\text{catechin-H}]^-$) originated from a QM (quinone methide) fragment and m/z 407 ($[\text{M-H-170 Da}]^-$) by the loss of a RDA (retro Diels-Alder) fragment (152 Da) followed by the loss of a water molecule (18 Da) (Jaiswal et al., 2014). Therefore, we could not indicate the correct stereoisomer for B-type proanthocyanidins.

One isomer of A-type Proanthocyanidin was identified as Proanthocyanidin A2 by MS-FINDER and Procyanidin A2 by SIRIUS (the second compound in the identification ranking). This compound (peak 16) produced MS² base peaks at m/z 285 and 289 ($[(\text{epi})\text{catechin-H}]^-$) originated from a QM (quinone methide) and m/z 423 ($[\text{M-H-152 Da}]^-$) originated by the loss of a RDA (retro Diels–Alder) fragment (Dixon et al., 2005).

3.2.4 Grape juice

Tartaric acid esters of hydroxycinnamic acids are representative compounds of grape juice, mainly caffeoyltartaric acid (caftaric acid), p-coumaroyl tartaric acid (coutaric acid)

and feruloyltartaric acid (fertaric acid) (**Table 6**). These compounds are more abundant in their *trans*-molecular form with the following constant elution order: *trans*-caftaric acid, *trans*-coutaric acid and *trans*-fertaric acid (Stalmach et al., 2011).

Both software identified caftaric acid correctly. Three isomers of coutaric acid were identified, MS-FINDER identified all as caffeoylmalic acid and coutaric acid was second in the identification ranking. Although SIRIUS computed the correct molecular formula for the three isomers, it identified only the second isomer as coutaric acid. For the other two, SIRIUS did not generate any identification. Two isomers of fertaric acid were identified, MS-FINDER identified both isomers correctly. SIRIUS computed the correct molecular formula for both fertaric acid isomers but did not suggest any structure for the first isomer. For the second isomer, fertaric acid was second on the identification ranking of SIRIUS.

Resveratrol is a widely known compound and a reference in phenolic studies of grapes, grape juice and wine. Peak 16 was correctly identified by MS-FINDER as resveratrol 3-*O*-glucoside. SIRIUS computed the correct molecular formula for this compound, but it did not suggest any structure for identification. This may be due to the unique fragment presented for this compound ($m/z = 227$ Da) which is the deprotonated resveratrol, formed after the loss of glucose. One fragment was not sufficient for the fragmentation tree drawing and identification of this compound.

3.2.5. Orange juice

Orange juice presented mainly flavones and flavanones linked to different substituents (**Table 7**). MS-FINDER and SIRIUS suggested many other structures and the correct identification was found on lower positions up to 9th and 27th positions, respectively.

Table 6. Identification of phenolic compound in bordô grape juice sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	t _R (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS- FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	Caffeoyltartaric acid	17.6	C ₁₃ H ₁₂ O ₉	311.0384	6.1	135(100), 179(21), 149(17), 150(1)	Caftaric acid (1°)	3.85	Caftaric acid (1°)	44.358%
2	B type Proanthocyanidin	20.3	C ₃₀ H ₂₆ O ₁₂	577.1359	2.3	289(100), 407(50), 125(44), 161(20)	Procyanidin (1°)	3.73	Procyanidin B5 (2°)	61.372%
3	p-coumaroyl tartaric acid	20.7	C ₁₃ H ₁₂ O ₈	295.0437	5.8	119(100), 149(42), 163(23)	cis-Coutaric acid (2°)	3.91	n.d	-
4	Feruloyltartaric acid	21.3	C ₁₄ H ₁₄ O ₉	325.0561	0.3	145(100), 117(19), 146(10), 119 (6)	Fertaric acid (1°)	3.89	n.d	-
5	p-coumaroyl tartaric acid	21.5	C ₁₃ H ₁₂ O ₈	295.0450	1.4	119(100), 163(37), 164(4), 149(4)	cis-Coutaric acid (2°)	3.95	Coutaric acid (1°)	47.773%
6	Catechin	21.9	C ₁₅ H ₁₄ O ₆	289.0700	4.2	123(100), 109(64), 151(40), 125(38)	Epicatechin (1°)	4.61	Catechin (1°)	67.227%
7	Feruloyltartaric acid	22.7	C ₁₄ H ₁₄ O ₉	325.0560	0.2	145(100), 119(56), 163(36), 161 (31)	Fertaric acid (1°)	3.89	Fertaric acid (2°)	41.830%
8	B type Proanthocyanidin	23.5	C ₃₀ H ₂₆ O ₁₂	577.1371	4.3	289(100), 407(94), 125(79), 290(47)	Procyanidin B5 (2°)	3.73	Procyanidin B5 (1°)	65.441%
9	Epicatechin	23.5	C ₁₅ H ₁₄ O ₆	289.0713	0.3	123(100), 109(64), 151(40), 125(38)	Epicatechin (1°)	4.63	Catechin (1°)	67.227%
10	p-coumaroyl tartaric acid	24.0	C ₁₃ H ₁₂ O ₈	295.0442	4.1	119(100), 163(48)	cis-Coutaric acid (2°)	3.93	n.d	-
11	Myricetin 3- <i>O</i> - glucuronide	26.1	C ₂₁ H ₁₈ O ₁₄	493.0652	6.8	317(100), 289(20), 318(14), 179(9)	Myricetin 3- <i>O</i> - glucuronide (1°)	3.87	Myricetin 3- <i>O</i> - glucuronide (1°)	76.765%
12	Myricetin-3- <i>O</i> -glucoside	26.3	C ₂₁ H ₂₀ O ₁₃	479.0829	0.6	316(100), 317(26), 271(2), 287(1)	Myricetin 3- galactoside (1°)	4.19	Myricetin-3- <i>O</i> - glucoside (1°)	91.611%
13	Rutin	27.2	C ₂₇ H ₃₀ O ₁₆	609.1445	1.6	300(100), 301(67), 179(4)	Rutin (1°)	4.16	Rutin (9°)	80.175%
14	Quercetin 3-glucoronide	28.9	C ₂₁ H ₁₈ O ₁₃	477.0700	6.5	301(100), 179(6), 151(6), 273(2)	Quercetin 3- <i>O</i> - glucuronide (1°)	4.12	Quercetin 3- <i>O</i> - glucuronide (1°)	89.939%
15	Quercetin 3-galactoside	29.3	C ₂₁ H ₂₀ O ₁₂	463.0900	5.0	300(100), 301(49), 255(3), 179(2)	Quercetin 3- galactoside (1°)	4.27	Quercetin 3- galactoside (1°)	96.040%

Table 6 Continued

16	Resveratrol 3- <i>O</i> -glucoside	31.4	C ₂₀ H ₂₂ O ₈	389.1263	6.8	227(100)	Resveratrol 3- <i>O</i> -glucoside (1°)	3.72	n.d	-
17	Spinacetin 3-glucoside	31.6	C ₂₃ H ₂₄ O ₁₃	507.1150	2.2	344(100), 345(35), 329(8), 273(7)	Spinacetin 3-glucoside (1°)	3.74	Spinacetin 3-glucoside (9°)	71.726%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined.

Table 7. Identification of phenolic compound in orange juice sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	t _R (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS- FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	Myrciacitrin II	20.7	C ₂₄ H ₂₈ O ₁₁	491.1591	7.6	347(100), 125(88), 329(30)	Myrciacitrin II (2°)	3.53	Myrciacitrin II (27°)	36.735%
2	Myrciacitrin II	21.2	C ₂₄ H ₂₈ O ₁₁	491.1521	6.6	347(100), 161(92), 125(57), 329(30)	Myrciacitrin II (2°)	3.52	Myrciacitrin II (27°)	38.213%
3	Myrciacitrin II	22.2	C ₂₄ H ₂₈ O ₁₁	491.1536	3.6	347(100), 161(47), 125(42), 179(34)	Myrciacitrin II (2°)	3.55	Myrciacitrin II (26°)	36.567%
4	Myrciacitrin II	22.8	C ₂₄ H ₂₈ O ₁₁	491.1547	1.3	347(100), 125(70), 293(66), 161(47)	Myrciacitrin II (2°)	3.34	n.d	-
5	Hesperetin 7- <i>O</i> - glucoside	24.6	C ₂₂ H ₂₄ O ₁₁	463.1251	2.2	265(100), 266(16), 221(6), 247(5)	Hesperetin 7-glucoside (1°)	3.97	Hesperetin 7-glucoside (4°)	36.723%
6	Pedunculsumosi de D	23.5	C ₂₇ H ₃₀ O ₁₂	545.1681	4.0	221(100), 203(28), 99(11), 151(10)	Pedunculsumoside D (1°)	3.39	Pedunculsumoside D (14°)	40.462%
7	Naringenin 7- glucoside	26.4	C ₂₁ H ₂₂ O ₁₀	433.1161	6.0	99(100), 249(34), 125(30), 161(25)	Naringenin 4- <i>O</i> - glucoside (9°)	3.85	Naringenin 4- <i>O</i> - glucoside (7°)	51.479%
8	Myrciacitrin I	26.5	C ₂₃ H ₂₆ O ₁₁	477.1417	4.2	99(100), 125(33), 247(30), 316(20)	Myrciacitrin I (5°)	3.73	Myrciacitrin I (20°)	39.726%
9	Rutin	27.1	C ₂₇ H ₃₀ O ₁₆	609.1502	7.7	300(100), 301(72), 179(4)	Rutin (1°)	4.13	Rutin (5°)	78.979%
10	Quercetin 3-(2- galloylglucoside)	30.6	C ₂₈ H ₂₄ O ₁₆	615.0939	7.6	271(100), 579(15), 272(14), 151(2)	Quercetin 3-(2- galloylglucoside) (1°)	3.35	Quercetin 3-(2- galloylglucoside) (1°)	61.173%
11	Neokurarinol	32.0	C ₂₇ H ₃₄ O ₇	469.2252	5.4	115(100), 114(95), 145(65), 280(27)	Neokurarinol (5°)	3.57	Neokurarinol (10°)	36.158%
12	Quercetin 3-(3- <i>p</i> - coumaroylglucosi de)	32.1	C ₃₀ H ₂₆ O ₁₄	609.1209	5.8	301(100), 302(18), 303(3), 325(2)	Quercetin 3-(3- <i>p</i> - coumaroylglucoside) (2°)	3.66	Quercetin 3-(3- <i>p</i> - coumaroylglucoside) (1°)	67.663%
13	Hesperidin	32.5	C ₂₈ H ₃₄ O ₁₅	609.1828	1.4	301(100), 343(2), 286(2), 257(1),	Hesperidin (1°)	4.01	Neohesperidin (1°)	76.667%
14	Hesperidin	34.6	C ₂₈ H ₃₄ O ₁₅	609.1825	0.9	101(100), 99(71), 161(51), 125(51)	Hesperidin (1°)	2.34	n.d	-
15	Hesperidin	35.5	C ₂₈ H ₃₄ O ₁₅	609.1795	4.1	205(100), 161(45), 99(43), 187(27)	Hesperidin (1°)	3.99	Neohesperidin (1°)	53.827%

Table 7 Continued

16	Matterionate A	35.9	C ₃₀ H ₃₆ O ₁₄	619.2059	5.2	331(100), 99(81), 125(40), 101(25)	Matterionate A (2°)	3.19	Matterionate A (13°)	40.547%
17	Matterionate A	36.5	C ₃₀ H ₃₆ O ₁₄	619.2031	0.6	331(100), 99(77), 125(46), 101(34)	Matterionate A (2°)	3.21	Matterionate A (14°)	40.878%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined.

Neither software could identify the four isomers of myrciacitrin II ($[M-H]^- = 491.1547$ Da) correctly. Myrciacitrin II (peaks 1 to 4) appeared in the second position at MS-FINDER's identification ranking and positions 26th and 27th for SIRIUS. Although SIRIUS computed the correct molecular formula for all isomers, it did not suggest any structure for the fourth myrciacitrin II isomer. Although it presented more than four fragments in the MS/MS spectra, they did not match the fragmentation tree fingerprint for myrciacitrin II.

All myrciacitrin II isomers presented the base fragment at m/z 347 ($[M-H-144 Da]^-$) due to the loss of $C_9H_4O_2$. Peaks 1, 2 and 3 presented a secondary fragment at m/z 329 ($[M-H-144-18 Da]^-$) due to the dehydration of the first fragment. Other fragments in the MS/MS spectra of this compound were m/z 179 ($C_9H_8O_4$) followed by 161 ($C_9H_8O_4 - H_2O$) and m/z 125 ($C_2H_6O_6$).

Three isomers of hesperidin (peaks 13, 14 and 15) were correctly identified by MS-FINDER as hesperidin, while SIRIUS identified peaks 13 and 15 as neohesperidin. SIRIUS did not suggest any structure for peak 14 although it computed the correct molecular formula for this compound. The discrimination among these isomers can be performed by using their additional information. These isomers have the same UV-Vis absorption profile with a maximum absorption wavelength at 285 nm but hesperidin elutes before neohesperidin in a reversed phase LC-MS method (Li et al., 2004).

3.2.6. *Apple juice*

Apple juice has *p*-coumaroylquinic acid as one of its major PC, which can be mainly as a 5-, 4- or 3-*p*-coumaroylquinic acid isomers (5-*p*CoQA, 4-*p*CoQA and 3-*p*CoQA). Moreover, other isomers of caffeoylquinic acids, *p*-coumaroyl glucoses and feruloyl glucoses were found. All these isomers were not differentiated by MS-FINDER and SIRIUS as shown in **Table 8**.

Four isomers of *p*-coumaroylquinic acids (peaks 1, 9, 13 and 15; $[M-H]^- = 337.0924$ Da) were identified as 3-*p*-coumaroylquinic acid by MS-FINDER and as 4-*p*-coumaroylquinic acid by SIRIUS. Information from these software packages tells us that these molecules are formed by a *p*-coumaric acid linked to quinic acid. The discrimination of those isomers can be performed by using their additional information. 5-, 4- or 3-coumaroylquinic acid isomers have the same UV-Vis absorption profile with a maximum absorption wavelength at around 311 nm. These isomers have different chromatographic behaviour with the following constant elution order: 3-*p*CoQA, *cis*-3-*p*CoQA, 4-*p*CoQA and *cis*-5-*p*CoQA (Jaiswal et al., 2010).

Afterward, we applied the hierarchical scheme of identification described by Clifford et al., (2003). Information obtained from MS/MS spectra and elution order contributed to the manual identification of each isomer. The 4-*p*CoQA showed an intense signal in MS/MS spectrum at m/z 173 ($[quinic\ acid-H-H_2O]^-$) due to the dehydration of the quinic acid ion, a feature of cinnamoyl group bonded to the quinic acid moiety at the position 4.

The 3-*p*CoQA and *cis*-3-*p*CoQA showed an intense signal in MS/MS spectrum at m/z 119 ($[p\text{-coumaric}\ acid-H-CO_2]^-$), due to the *p*-coumaric acid after a loss of carbon dioxide. The isomers could be distinguished by elution order and by comparing the relative intensity of the secondary ion at m/z 191 ($[quinic\ acid-H]^-$) which corresponded to 29% of the base peak in 3-*p*CoQA and 74% in *cis*-3-*p*CoQA (Jaiswal et al., 2010). The *cis*-5-*p*CoQA was identified by comparing the relative intensity of the base peak at m/z 191 and the secondary ion at m/z 163 ($[p\text{-coumaric}\ acid-H]^-$) (Clifford et al., 2003; Jaiswal et al., 2010).

Table 8. Identification of phenolic compound in apple juice sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	tr (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS-FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	3- <i>p</i> -Coumaroylquinic acid	20.1	C ₁₆ H ₁₈ O ₈	337.0909	4.4	119(100), 163(69), 191(29)	3- <i>O</i> - <i>p</i> -Coumaroylquinic acid (1°)	4.13	4- <i>O</i> - <i>p</i> -Coumaroylquinic acid (14°)	48.464%
2	3-caffeoylquinic acid	20.3	C ₂₃ H ₁₄ O ₄	353.0842	8.8	135(100), 191(88), 101(71), 179(43)	Chlorogenic acid (1°)	4.23	3-caffeoylquinic acid (1°)	39.015%
3	Procyanidin dimer B1	20.3	C ₃₀ H ₂₆ O ₁₂	577.1316	5.2	289(100), 407(55), 125(51), 245(21)	Procyanidin (1°)	3.87	Procyanidin B5 (1°)	48.789%
4	(-)-Epicatechin gallate	20.6	C ₂₂ H ₁₈ O ₁₀	441.0828	1.4	249(100), 161(26), 101(28), 143(3)	(-)-Catechin 3- <i>O</i> -gallate (1°)	4.19	(-)-Epicatechin gallate (1°)	53.268%
5	5-caffeoylquinic acid	21.1	C ₁₆ H ₁₈ O ₉	353.0898	7.1	191(100), 135(1), 133(1), 93(1)	Chlorogenic acid (1°)	4.28	1- <i>O</i> -Caffeoylquinic acid (29°)	36.712%
6	<i>p</i> -Coumaroyl glucose	21.8	C ₁₆ H ₁₈ O ₈	325.0933	2.9	145(100), 119(41), 161(27), 163(27)	1- <i>O</i> - <i>p</i> -Coumaroyl-beta-D-glucose (1°)	4.04	1- <i>p</i> -Coumaroyl glucose (5°)	47.464%
7	Catechin	21.8	C ₁₅ H ₁₄ O ₆	289.0710	0.7	109(100), 123(73)	Epicatechin (1°)	4.63	n.d	-
8	4-caffeoylquinic acid	22.2	C ₁₆ H ₁₈ O ₉	353.0901	7.9	191(100), 161(3), 127(2)	Chlorogenic acid (1°)	4.27	n.d	-
9	<i>cis</i> -3- <i>p</i> -Coumaroylquinic acid	23.0	C ₁₆ H ₁₈ O ₈	337.0949	7.5	119 (100), 191(74), 163(67)	3- <i>O</i> - <i>p</i> -Coumaroylquinic acid (1°)	4.13	4- <i>O</i> - <i>p</i> -Coumaroylquinic acid (27°)	36.391%
10	Feruloyl glucose	23.2	C ₁₆ H ₂₀ O ₉	355.1051	6.2	175(100), 160(97), 161(12), 176(11)	1- <i>O</i> -Feruloylglucose (1°)	4.02	1-Feruloyl-D-glucose (3°)	54.027%
11	Procyanidin dimer B2	23.5	C ₃₀ H ₂₆ O ₁₂	577.1355	1.6	289(100), 407(42), 125(41), 245(17)	Procyanidin (1°)	3.90	Procyanidin B1 (1°)	57.609%
12	Epicatechin	23.6	C ₁₅ H ₁₄ O ₆	289.0717	1.7	123(100), 109(65), 125(63), 221(34)	Epicatechin (1°)	4.64	Catechin (3°)	57.674%
13	4- <i>p</i> -Coumaroylquinic acid isomer	23.7	C ₁₆ H ₁₈ O ₈	337.0942	5.4	173(100), 93(34), 163(29), 111(11)	3- <i>O</i> - <i>p</i> -Coumaroylquinic acid (1°)	4.15	4- <i>p</i> -Coumaroylquinic acid (28°)	37.888%
14	Feruloyl glucose	24.6	C ₁₆ H ₂₀ O ₉	355.1047	5.1	175(100), 134(73), 193(68), 191(64)	1- <i>O</i> -Feruloylglucose (1°)	4.03	1-Feruloyl-D-glucose (5°)	46.405%
15	<i>cis</i> -5- <i>p</i> -Coumaroylquinic acid	25.2	C ₁₆ H ₁₈ O ₈	337.0909	4.4	191(100), 163(23), 173(10), 127(4)	3- <i>p</i> -Coumaroylquinic acid (1°)	3.86	4- <i>p</i> -Coumaroylquinic acid (26°)	38.776%
16	(-)-Epicatechin gallate	25.2	C ₂₂ H ₁₈ O ₁₀	441.0824	0.5	263(100), 101(39), 161(39), 131(18)	(-)-Catechin 3- <i>O</i> -gallate (1°)	4.19	n.d	-

Table 8 Continued

17	(-)-Epicatechin gallate	26.1	C ₂₂ H ₁₈ O ₁₀	441.0848	5.9	263(100), 101(33), 161(30), 113(14)	(-)-Catechin 3- <i>O</i> -gallate (1°)	4.17	n.d	-
18	Myricitrin	29.1	C ₂₁ H ₂₀ O ₁₂	463.0884	1.6	300(100), 302(9), 463(7)	Myricitrin (4°)	4.06	Myricitrin (1°)	54.799%
19	Mulberroside A	30.1	C ₂₆ H ₃₂ O ₁₄	567.1742	4.9	273(100), 274(16), 167(13), 125(4)	cis-Mulberroside A (1°)	3.64	Mulberroside A (3°)	49.481%
20	Avicularin	31.4	C ₂₀ H ₁₈ O ₁₁	433.0741	6.9	300(100), 301(51), 207(16), 271(11)	Avicularin (1°)	3.84	Avicularin (1°)	67.619%
21	Quercitrin	32.0	C ₂₁ H ₂₀ O ₁₁	447.0897	6.8	300(100), 301(54), 302(9)	Quercitrin (1°)	4.17	Quercitrin (9°)	53.030%
22	Phloretin 2'- <i>O</i> - glucoside	33.0	C ₂₁ H ₂₄ O ₁₀	435.1305	3.1	273(100), 167(97), 179(17), 274(12)	Phloretin 2'- <i>O</i> -glucoside (1°)	4.12	Phloretin 2'- <i>O</i> -glucoside (1°)	58.725%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined.

3.2.7. Soy extract

Isoflavones are the main PC in soy extracts (**Table 9**). SIRIUS had difficulties identifying those compounds, being the worst result for this software. Only seven compounds had structures suggested by SIRIUS and the correct identification was found on lower positions up to the 30th position.

MS-FINDER identified genistein (peak 10, $[M-H]^- = 269.0450$ Da) correctly eventhough the MS/MS spectrum showed only the parent ion ($m/z = 269$ Da). Beyond accurate mass, isotope ratio and product ion information, MS-FINDER considers bond dissociation energies, fragment linkages, and nine hydrogen rearrangement rules to predict the structures. SIRIUS predicts molecular fingerprints using the compound's fragments and identifies them by searching theoretical databases. Therefore, Genistein was not identified by SIRIUS because it did not present any fragmentation.

Similar behaviour was observed with 6-*O*-acetyldaizidin, 6-*O*-malonyldaizidin, 6-*O*-acetylglycitin and 6-malonylgenistin (peaks 6, 7, 9 and 14). 6-*O*-acetyldaizidin and 6-*O*-malonyldaizidin presented only one fragment at m/z 253 ($[daidzein-H]^-$) generated by the loss of the acetyl and malonyl group. 6-*O*-acetylglycitin presented only one fragment at m/z 283 ($[glycitein-H]^-$) generated by the loss of the acetyl group. And 6-malonylgenistin presented only one fragment at m/z 269 ($[genistein - H]^-$) generated by the loss of the malonyl group. MS-FINDER identified those four compounds correctly with a high score and ranked in the first position while SIRIUS did not compute any structure for those compounds. One fragment was not sufficient for a proper fragmentation tree fingerprint outline by SIRIUS.

Table 9. Identification of phenolic compound in soy extract sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	t _R (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS-FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	Genistin	20.2	C ₂₁ H ₂₀ O ₁₀	431.0989	2.4	125(100), 391(9), 99(9)	Genistin (1°)	4.25	n.d	-
2	Isoangustone A	21.2	C ₂₅ H ₂₆ O ₆	421.1647	1.0	389(100), 209(34)	Isoangustone A (1°)	3.84	n.d	-
3	(-)-Epigallocatechin	21.4	C ₁₅ H ₁₄ O ₇	305.0685	7.7	97(100), 225(19), 99(8), 147(4)	(-)-Epigallocatechin (1°)	4.24	n.d	-
4	Glycitein	27.3	C ₁₆ H ₁₂ O ₅	283.0616	3.3	240(100), 268(52), 196(23), 241(16)	Glycitein (2°)	4.47	Glycitein (30°)	50.000%
5	Daidzein	30.7	C ₁₅ H ₁₀ O ₄	253.0508	2.8	224(100), 209(27), 196(20), 133(23)	Daidzein (1°)	4.59	Daidzein (25°)	52.284%
6	6- <i>O</i> -Acetyldaidzin	30.8	C ₂₃ H ₂₂ O ₁₀	457.1149	3.1	253(100)	6''- <i>O</i> -Acetyldaidzin (1°)	3.78	n.d	-
7	6- <i>O</i> - Malonyldaidzin	30.7	C ₂₄ H ₂₂ O ₁₂	501.1044	2.2	253(100)	6''- <i>O</i> - Malonyldaidzin (1°)	3.92	n.d	-
8	Glycitein	31.1	C ₁₆ H ₁₂ O ₅	283.0613	2.2	240(100), 268(52), 196(23), 241(16)	Glycitein (2°)	4.47	Glycitein (30°)	50.000%
9	6- <i>O</i> -Acetylglycitin	31.1	C ₂₄ H ₂₄ O ₁₁	487.1202	7.9	283(100)	6''- <i>O</i> -Acetylglycitin (1°)	3.64	n.d	-
10	Genistein	31.3	C ₁₅ H ₁₀ O ₅	269.0435	5.6	269(100)	Genistein (1°)	4.81	n.d	-
11	Apigenin 7-di- <i>O</i> - xyloside	32.6	C ₂₅ H ₂₆ O ₁₅	565.1202	1.5	129(100), 147(62), 193(47), 197(38)	Apigenin 7-di- <i>O</i> - xyloside (6°)	3.34	Apigenin 7-di- <i>O</i> - xyloside (1°)	39.362%
12	Genistein	35.0	C ₁₅ H ₁₀ O ₅	269.0465	5.6	133(100), 159(31), 135(30), 225(29)	Genistein (1°)	4.85	Genistein (2°)	73.846%
13	6- <i>O</i> -Acetylgenistin	35.0	C ₂₃ H ₂₂ O ₁₁	473.1071	2.7	269(100), 268(52), 270(19), 335(9)	6''- <i>O</i> -Acetylgenistin (4°)	3.75	6''- <i>O</i> -Acetylgenistin (1°)	52.198%
14	6-Malonylgenistin	35.1	C ₂₄ H ₂₂ O ₁₃	517.0992	1.9	269(100)	6''-Malonylgenistin (1°)	3.96	n.d	-
15	Daidzein	38.9	C ₁₅ H ₁₀ O ₄	253.0497	1.6	223(100), 224(92), 132(23)	Daidzein (1°)	4.61	n.d	-
16	Apigenin	45.4	C ₁₅ H ₁₀ O ₅	269.0442	3.0	269(100), 133(38), 159(20), 135(17)	Apigenin (2°)	4.73	Apigenin (5°)	60.952%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined.

3.2.8. Parsley extract

Parsley is known for its richness of flavones, especially apigenin and its derivatives, likewise apiin (apigenin-7-apioglucoside) (**Table 10**). MS-FINDER and SIRIUS correctly identified apiin ($[M-H]^- = 563.1401$, peak 7), even though it presented only two fragments in the MS/MS spectra. The base peak was $[apigenin - H]^-$ at m/z 269 generated by the loss of apioglucoside and the secondary ion was at m/z 431.

Two isomers of malonylapiin ($[M-H]^- = 649.1410$, peaks 10 and 13) and two isomers of acetylapiin ($[M-H]^- = 605.1512$, peaks 11 and 12) were identified by both software as 6''-malonylapiin and 6''-acetylapiin, respectively. Differentiation between the isomers can be achieved through additional information, likewise chromatographic behavior. All compounds showed a fragment with an intense signal at m/z 269 and a secondary ion at m/z 545 ($[apiin-H-H_2O]^-$) due to the loss of the acetyl and malonyl group respectively.

Both isomers of diosmetin apiosylglucoside (peaks 8 and 9, $[M-H]^- = 593.1512$ Da) were identified by MS-FINDER as astragalin 7-rhamnoside and SIRIUS could not identify this compound. Both compounds presented only one fragment at m/z 299 Da ($[diosmetin-H]^-$) due to the loss of apiosylglucoside.

3.2.9. Overview

The 149 identified PC were divided into twelve phenolic compound classes (**Table 11**). Phenolic acid was the class with the highest number of examples (47 compounds), while homoflavonoids, chalcones and cumarins had only one compound in each class. For a representative discussion, only classes with more than five examples were considered.

Table 10. Identification of phenolic compound in parsley extract sample analyzed by LC-ESI-MS using SIRIUS 4.9.12 and MS-FINDER ver. 3.52.

Peak	Compound	tr (min) ^a	Molecular Formula	Exp [M-H] ⁻	Error (ppm)	MS/MS	Identification by MS-FINDER (Ranking)	Score by MS-FINDER	Identification by SIRIUS (Ranking)	Similarity by SIRIUS
1	Coumaric acid glucoside	19.4	C ₁₅ H ₁₈ O ₈	325.0936	3.8	119(100), 163(39)	trans- <i>o</i> -Coumaric acid 2-glucoside (2°)	4.03	trans- <i>o</i> -Coumaric acid 2-glucoside (1°)	76.744%
2	Kotanin	20.1	C ₂₄ H ₂₂ O ₈	437.1212	5.6	269(100), 161(69)	Kotanin (2°)	3.55	n.d	-
3	Asphodoside A	21.7	C ₂₈ H ₃₂ O ₁₇	639.1479	3.7	315(100), 477(95), 476(89), 313(33)	asphodoside A (1°)	3.16	asphodoside A (1°)	54.599%
4	Isovitexin-8-C-glucoside (Vicenin-2)	22.3	C ₂₇ H ₃₀ O ₁₅	593.1505	0.3	473(100), 353(53), 383(40), 503(22)	Vicenin-2 (3°)	3.91	Vicenin-2 (1°)	60.323%
5	Coumaric acid glucoside	24.8	C ₁₅ H ₁₈ O ₈	325.0943	5.9	119(100), 101(4), 117(2)	trans- <i>o</i> -Coumaric acid 2-glucoside (2°)	3.99	trans- <i>o</i> -Coumaric acid 2-glucoside (1°)	65.299%
6	(Apiosyl-malonyl)astragalinalin	31.5	C ₂₉ H ₃₀ O ₁₈	665.1378	2.9	97(100), 621(97), 563(28)	(3"-Apiosyl-6"-malonyl)astragalinalin (1°)	3.26	(3"-Apiosyl-6"-malonyl)astragalinalin (6°)	38.386%
7	Apigenin-7-apioglucoside (apiin)	31.6	C ₂₆ H ₂₈ O ₁₄	563.1392	1.6	269(100), 431(2)	Apiin (1°)	3.97	Apiin (1°)	70.487%
8	Diosmetin apiosylglucoside	32.4	C ₂₇ H ₃₀ O ₁₅	593.1488	4.0	299(100)	Astragalinalin 7-rhamnoside (1°)	3.95	n.d	-
9	Diosmetin apiosylglucoside	32.9	C ₂₇ H ₃₀ O ₁₅	593.1544	5.4	299(100)	Astragalinalin 7-rhamnoside (1°)	3.94	n.d	-
10	Malonylapiin	33.3	C ₂₉ H ₃₀ O ₁₇	649.1377	5.1	545(100), 269(44), 246(29), 605(11)	6"-Malonylapiin (1°)	3.31	6"-Malonylapiin (6°)	44.828%
11	Acetylapiin	33.9	C ₂₈ H ₃₀ O ₁₅	605.1477	5.8	269(100), 563(54), 564(19), 545(16)	6"-Acetylapiin (2°)	3.50	6"-Acetylapiin (1°)	55.039%
12	Acetylapiin	34.2	C ₂₈ H ₃₀ O ₁₅	605.1503	1.5	269(100), 563(47), 545(22), 270(20)	6"-Acetylapiin (2°)	3.53	6"-Acetylapiin (2°)	54.731%
13	Malonylapiin	34.4	C ₂₉ H ₃₀ O ₁₇	649.1396	2.2	269(100), 545(55), 270(15), 563(9)	6"-Malonylapiin (1°)	3.32	6"-Malonylapiin (6°)	51.117%
14	Kaempferol acetylgalactoside rhamnoside	35.0	C ₂₉ H ₃₂ O ₁₆	635.1618	0.0	299(100), 593(34), 300(17), 575(17)	Kaempferol 3-(6"-acetylgalactoside) 7-rhamnoside (1°)	3.58	Kaempferol 3-(6"-acetylgalactoside) 7-rhamnoside (11°)	55.789%
15	Kaempferol acetylgalactoside rhamnoside	35.4	C ₂₉ H ₃₂ O ₁₆	635.1630	1.9	299(100), 300(15), 593(11), 575(4)	Kaempferol 3-(6"-acetylgalactoside) 7-rhamnoside (1°)	3.58	n.d	-

Table 10 Continued

16	Di-acetylononin	41.3	C ₂₆ H ₂₆ O ₁₁	513.1416	3.7	163(100), 119(42), 145(27)	2",6"-Di- <i>O</i> -acetylononin (1°)	3.56	2",6"-Di- <i>O</i> -acetylononin (5°)	49.429%
17	Piceatannol galloylglucoside	41.3	C ₂₇ H ₂₆ O ₁₃	557.1294	0.3	163(100), 119(26), 145(21), 349(2)	Piceatannol 4'- galloylglucoside (1°)	3.22	Piceatannol 4'- galloylglucoside (8°)	51.233%
18	Piceatannol galloylglucoside	42.1	C ₂₇ H ₂₆ O ₁₃	557.1291	0.8	163(100), 145(31), 119(26), 349(4)	Piceatannol 4'- galloylglucoside (1°)	3.22	Piceatannol 4'- galloylglucoside (5°)	53.678%
19	Di-acetylononin	42.4	C ₂₆ H ₂₆ O ₁₁	513.1392	1.0	163(100), 119(34), 145(17)	2",6"-Di- <i>O</i> -acetylononin (1°)	3.56	2",6"-Di- <i>O</i> -acetylononin (6°)	46.703%

^aRetention time on the C₁₈ Synergi (4 μm) column and solvent: gradient of 0.1% formic acid in water and acetonitrile with 0.1% formic acid. n.d.: not determined.

MS-FINDER performed better than SIRIUS and correctly identified all 16 examples of flavanols and 91% of the 23 flavonols in the food samples. The worst identification rate of MS-FINDER was with flavones (42%). SIRIUS presented great difficulties identifying isoflavones (8%) once these compounds presented a poor fragmentation pattern with one or two fragments. The highest performance of SIRIUS identification was observed with proanthocyanidins (63%) and flavanols (56%).

Table 11. Identification of phenolic compound classes in food samples using MS-FINDER and SIRIUS.

Phenolic compound Class	N° of compounds	Identification by MS-FINDER	Identification by SIRIUS
Phenolic acids	47	60%	38%
Flavones	24	42%	42%
Flavonols	23	91%	39%
Flavanols	16	100%	56%
Isoflavones	12	75%	8%
Flavanones	10	60%	30%
Proanthocyanidins	8	88%	63%
Stilbenes	4	100%	0%
Isoflavonoids	2	100%	0%
Homoflavonoids	1	100%	0%
Chalcones	1	100%	100%
Cumarins	1	0%	0%
Total	149	70%	38%

4. Conclusion

The software SIRIUS and MS-FINDER are capable to assist in the identification of the phenolic compounds from LC-HR-MS/MS data. MS-FINDER showed a higher capacity to correctly identify the PC even in cases where MS/MS is poor in fragments. Identification capacity varies among PC classes and between the software. SIRIUS had difficulty identifying

isoflavones and derivatives. SIRIUS could not identify some phenolic acids with low molecular mass and poor fragmentation pattern. Still, flavanols were easily identified by both software although they could not distinguish the isomers. MS-FINDER showed great results with flavonols too and SIRIUS showed its better results with proanthocyanidins.

Researchers will be able to apply this easy software approach as an initial step to identify PC analysed by high-resolution mass spectrometry. From these software packages, it is possible to obtain a starting point for deeper exploration using its knowledge about the food matrix under investigation and other additional information (i.e. elution order in the C18 column and UV-vis absorption characteristics).

References

- Ali, A., Bashmil, Y. M., Cottrell, J. J., Suleria, H. A. R., & Dunshea, F. R. (2021). Lc-ms/ms-qtof screening and identification of phenolic compounds from australian grown herbs and their antioxidant potential. *Antioxidants*, *10*(11), 1–20. <https://doi.org/10.3390/antiox10111770>
- Arranz, S., & Calixto, S. (2010). Analysis of polyphenols in cereals may be improved performing acidic hydrolysis: A study in wheat flour and wheat bran and cereals of the diet. *Journal of Cereal Science*, *51*, 313–318. <https://doi.org/10.1016/j.jcs.2010.01.006>
- Arranz, S., Silván, J. M., & Saura-Calixto, F. (2010). Nonextractable polyphenols, usually ignored, are the major part of dietary polyphenols: A study on the Spanish diet. *Molecular Nutrition and Food Research*, *54*(11), 1646–1658. <https://doi.org/10.1002/mnfr.200900580>
- Arts, I. C. W., Putte, B. Van De, & Hollman, P. C. H. (2000). Catechin Contents of Foods Commonly Consumed in The Netherlands . 1 . Fruits , Vegetables , Staple Foods , and Processed Foods. *J. Agric. Food Chem*, *48*, 1746–1751.
- Ashraf, M. A., Iqbal, M., Rasheed, R., Hussain, I., Riaz, M., & Arif, M. S. (2018). Environmental Stress and Secondary Metabolites in Plants: An Overview. In *Plant Metabolites and Regulation under Environmental Stress*. <https://doi.org/10.1016/B978-0-12-812689-9.00008-X>
- Baquer, G., Sementé, L., Mahamdi, T., Correig, X., Ràfols, P., & García-Altres, M. (2023). What are we imaging? Software tools and experimental strategies for annotation and identification of small molecules in mass spectrometry imaging. *Mass Spectrometry Reviews*, *42*(5), 1927–1964. <https://doi.org/10.1002/mas.21794>
- Bedran-Russo, A. K., Pauli, G. F., Chen, S. N., McAlpine, J., Castellan, C. S., Phansalkar, R. S., ... Leme, A. A. (2014). Dentin biomodification: Strategies, renewable resources and clinical applications. *Dental Materials*, *30*(1), 62–76. <https://doi.org/10.1016/j.dental.2013.10.012>
- Bittremieux, W., Avalon, N. E., Thomas, S. P., Kakhkhorov, S. A., Gauglitz, J. M., Gerwick, W. H., ... Rima, F. (2023). Open Access Repository-Scale Propagated Nearest Neighbor Suspect Spectral Library for Untargeted Metabolomics. *BioRxiv*, 1–32. <https://doi.org/10.1038/s41467-023-44035-y>
- Cáceres-Jiménez, S., Rodríguez-Solana, R., Dobani, S., Pourshahidi, K., Gill, C., Moreno-

- Rojas, J. M., ... Pereira-Caro, G. (2023). UHPLC-HRMS Spectrometric Analysis: Method Validation and Plasma and Urinary Metabolite Identification after Mango Pulp Intake. *Journal of Agricultural and Food Chemistry*, 71(30), 11520–11533. <https://doi.org/10.1021/acs.jafc.3c03846>
- Cardona, F., Andrés-Lacueva, C., Tulipani, S., Tinahones, F. J., & Queipo-Ortuño, M. I. (2013). Benefits of polyphenols on gut microbiota and implications in human health. *Journal of Nutritional Biochemistry*, 24(8), 1415–1422. <https://doi.org/10.1016/j.jnutbio.2013.05.001>
- Carvalho, D. O., Curto, A. F., & Guido, L. F. (2015). Determination of Phenolic Content in Different Barley Varieties and Corresponding Malts by Liquid Chromatography-diode Array Detection-Electrospray Ionization Tandem Mass Spectrometry. *Antioxidants (Basel)*, 4(3), 563–576. <https://doi.org/10.3390/antiox4030563>
- Cheng, C., & Gross, M. L. (2000). Applications and mechanisms of charge-remote fragmentation. *Mass Spectrometry Reviews*, 19(6), 398–420. [https://doi.org/10.1002/1098-2787\(2000\)19:6<398::AID-MAS3>3.0.CO;2-B](https://doi.org/10.1002/1098-2787(2000)19:6<398::AID-MAS3>3.0.CO;2-B)
- Chung, T. Te, Chuang, C. Y., Teng, Y. H., Hsieh, M. J., Lai, J. C., Chuang, Y. T., ... Yang, S. F. (2017). Tricetin suppresses human oral cancer cell migration by reducing matrix metalloproteinase-9 expression through the mitogen-activated protein kinase signaling pathway. *Environmental Toxicology*, 32(11), 2392–2399. <https://doi.org/10.1002/tox.22452>
- Clifford, M. N., Johnston, K. L., Knight, S., & Kuhnert, N. (2003). Hierarchical scheme for LC-MSn identification of chlorogenic acids. *Journal of Agricultural and Food Chemistry*, 51(10), 2900–2911. <https://doi.org/10.1021/jf026187q>
- Colletti, A., Sangiorgio, L., Martelli, A., Testai, L., Cicero, A. F. G., & Cravotto, G. (2021). Highly active cranberry's polyphenolic fraction: New advances in processing and clinical applications. *Nutrients*, 13(8). <https://doi.org/10.3390/nu13082546>
- Crozier, A., Jaganath, I. B., & Clifford, M. N. (2009). Dietary phenolics: Chemistry, bioavailability and effects on health. *Natural Product Reports*, 26(8), 1001–1043. <https://doi.org/10.1039/b802662a>
- De Freitas, T. B., Santos, C. H. K., da Silva, M. V., Shirai, M. A., Dias, M. I., Barros, L., ... Leimann, F. V. (2018). Antioxidants extraction from Pinhão (*Araucaria angustifolia* (Bertol.) Kuntze) coats and application to zein films. *Food Packaging and Shelf Life*,

- 15(July), 28–34. <https://doi.org/10.1016/j.fpsl.2017.10.006>
- de Rijke, E., Out, P., Niessen, W. M. A., Ariese, F., Gooijer, C., & Brinkman, U. A. T. (2006). Analytical separation and detection methods for flavonoids. *Journal of Chromatography A*, *1112*(1–2), 31–63. <https://doi.org/10.1016/j.chroma.2006.01.019>
- Delcambre, A., & Saucier, C. (2012). Identification of new flavan-3-ol monoglycosides by UHPLC-ESI-Q-TOF in grapes and wine. *Journal of Mass Spectrometry*, *47*(6), 727–736. <https://doi.org/10.1002/jms.3007>
- Demarque, D. P., Crotti, A. E. M., Vessecchi, R., Lopes, J. L. C., & Lopes, N. P. (2016). Fragmentation reactions using electrospray ionization mass spectrometry: An important tool for the structural elucidation and characterization of synthetic and natural products. *Natural Product Reports*, *33*(3), 432–455. <https://doi.org/10.1039/c5np00073d>
- Díaz-de-Cerio, E., Tylewicz, U., Verardo, V., Fernández-Gutiérrez, A., Segura-Carretero, A., & Romani, S. (2017). Design of Sonotrode Ultrasound-Assisted Extraction of Phenolic Compounds from *Psidium guajava* L. Leaves. *Food Analytical Methods*, *10*(8), 2781–2791. <https://doi.org/10.1007/s12161-017-0836-z>
- Dixon, R. A., Xie, D. Y., & Sharma, S. B. (2005). Proanthocyanidins - A final frontier in flavonoid research? *New Phytologist*, *165*(1), 9–28. <https://doi.org/10.1111/j.1469-8137.2004.01217.x>
- Djoumbou-Feunang, Y., Pon, A., Karu, N., Zheng, J., Li, C., Arndt, D., ... Wishart, D. S. (2019). Cfm-id 3.0: Significantly improved esi-ms/ms prediction and compound identification. *Metabolites*, *9*(4), 1–23. <https://doi.org/10.3390/metabo9040072>
- Domínguez-Rodríguez, G., Ramón Vidal, D., Martorell, P., Plaza, M., & Marina, M. L. (2022). Composition of Nonextractable Polyphenols from Sweet Cherry Pomace Determined by DART-Orbitrap-HRMS and Their in Vitro and in Vivo Potential Antioxidant, Antiaging, and Neuroprotective Activities. *Journal of Agricultural and Food Chemistry*, *70*(26), 7993–8009. <https://doi.org/10.1021/acs.jafc.2c03346>
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., ... Böcker, S. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, *16*(4), 299–302. <https://doi.org/10.1038/s41592-019-0344-8>
- Dührkop, K., Scheubert, K., & Böcker, S. (2013). Molecular Formula Identification with SIRIUS. *Metabolites*, *3*(2), 506–516. <https://doi.org/10.3390/metabo3020506>

- Durazzo, A. (2018). Extractable and Non-extractable Polyphenols: an Overview. In *Non-extractable Polyphenols and Carotenoids: Importance in Human Nutrition and Health* (p. 382). Retrieved from <https://pubs.rsc.org/en/content/chapterpdf/2018/9781788013208-00037?isbn=978-1-78801-106-8&sercode=bk>
- Eseberri, I., Trepiana, J., Léniz, A., Gómez-García, I., Carr-Ugarte, H., González, M., & Portillo, M. P. (2022). Variability in the Beneficial Effects of Phenolic Compounds: A Review. *Nutrients*, *14*(9), 1–19. <https://doi.org/10.3390/nu14091925>
- Esparza-Martínez, F. J., Miranda-López, R., Mata-Sánchez, S. M., & Guzmán-Maldonado, S. H. (2016). Extractable and Non-Extractable Phenolics and Antioxidant Capacity of Mandarin Waste Dried at Different Temperatures. *Plant Foods Hum Nutr*, *71*, 294–300. <https://doi.org/10.1007/s11130-016-0559-0>
- Fatemeh, S. R., Saifullah, R., Abbas, F. M. A., & Azhar, M. E. (2012). Total phenolics, flavonoids and antioxidant activity of banana pulp and peel flours: Influence of variety and stage of ripeness. *International Food Research Journal*, *19*(3), 1041–1046.
- Freeman, B. L., Eggett, D. L., & Parker, T. L. (2010). Synergistic and antagonistic interactions of phenolic compounds found in navel oranges. *Journal of Food Science*, *75*(6), 570–576. <https://doi.org/10.1111/j.1750-3841.2010.01717.x>
- Fuentes, Y., Giovagnoli-Vicuña, C., Faúndez, M., & Giordano, A. (2023). Microencapsulation of Chilean Papaya Waste Extract and Its Impact on Physicochemical and Bioactive Properties. *Antioxidants*, *12*(10), 1–16. <https://doi.org/10.3390/antiox12101900>
- García-salas, P., Morales-soto, A., Segura-carretero, A., & Fernández-gutiérrez, A. (2010). *Phenolic-Compound-Extraction Systems for Fruit and Vegetable Samples*. 8813–8826. <https://doi.org/10.3390/molecules15128813>
- Gazolla, M. C., Marques, L. M. M., e Silva, M. G., Araújo, M. T. M. F., Mendes, R. L., da Silva Almeida, J. R. G., ... Lopes, N. P. (2020). Characterization of 3-aminospirostane alkaloids from roots of *Solanum paniculatum* L. with hepatoprotective activity. *Rapid Communications in Mass Spectrometry*, *34*(S3), 1–9. <https://doi.org/10.1002/rcm.8705>
- Gonzalez Ibarra, A. A., Wrobel, K., Yanez Barrientos, E., Corrales Escobosa, A. R., Gutierrez Corona, J. F., Enciso Donis, I., & Wrobel, K. (2017). Changes of metabolomic profile in *helianthus annuus* under exposure to Chromium(VI) Studied by capHPLC-ESI-QTOF-MS and MS/MS. *Journal of Analytical Methods in Chemistry*, *2017*(Vi). <https://doi.org/10.1155/2017/3568621>

- Gu, L., Kelm, M. A., Hammerstone, J. F., Beecher, G., Holden, J., Haytowitz, D., ... Prior, R. L. (2004). Concentrations of Proanthocyanidins in Common Foods and Estimations of Normal Consumption. *Journal of Nutrition*, 134(3), 613–617. <https://doi.org/10.1093/jn/134.3.613>
- Haminiuk, C. W. I., Maciel, G. M., Plata-Oviedo, M. S. V., & Peralta, R. M. (2012). Phenolic compounds in fruits - an overview. *International Journal of Food Science and Technology*, 47(10), 2023–2044. <https://doi.org/10.1111/j.1365-2621.2012.03067.x>
- He, M., Wu, H., Nie, J., Yan, P., Yang, T. B., Yang, Z. Y., & Pei, R. (2017). Accurate recognition and feature qualify for flavonoid extracts from Liang-wai Gan Cao by liquid chromatography-high resolution-mass spectrometry and computational MS/MS fragmentation. *Journal of Pharmaceutical and Biomedical Analysis*, 146, 37–47. <https://doi.org/10.1016/j.jpba.2017.07.065>
- Hellström, J. K., Törrönen, A. R., & Mattila, P. H. (2009). Proanthocyanidins in common food products of plant origin. *Journal of Agricultural and Food Chemistry*, 57(17), 7899–7906. <https://doi.org/10.1021/jf901434d>
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., ... Nishioka, T. (2010). MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703–714. <https://doi.org/10.1002/jms.1777>
- Hou, J., Liang, L., Su, M., Yang, T., Mao, X., & Wang, Y. (2021). Variations in phenolic acids and antioxidant activity of navel orange at different growth stages. *Food Chemistry*, 360(October 2020), 129980. <https://doi.org/10.1016/j.foodchem.2021.129980>
- Hui Li, Leren Wan, Y. H. and S. C. (2007). Fragmentation study of a 8-C-glycosyl isoflavone, puerarin, using electrospray ion trap time-of-flight mass spectrometry at high resolution. *Rapid Communications in Mass Spectrometry*, 21, 2497–2504. <https://doi.org/10.1002/rcm>
- Huo, C., Nguyen, Q. N., Alishir, A., Ra, M., Jung, S., & Yu, J. (2023). Global Natural Products Social (GNPS) -Based Molecular-Networking-Guided Isolation of Phenolic Compounds from Ginkgo biloba Fruits and the Identification of. *Plants*, 12(3970), 1–13.
- IBGE Instituto Brasileiro de Geografia e Estatística. (2020). *Pesquisa de Orçamentos Familiares 2017 - 2018*. Retrieved from <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101742.pdf>
- Jadaun, V., Prateeksha, P., Nailwal, T., & Singh, B. N. (2023). Antioxidant activity and

- simultaneous estimation of four polyphenolics in different parts of *Carica papaya* L. by a validated high-performance thin-layer chromatography method. *Journal of Planar Chromatography - Modern TLC*, 36(4), 211–221. <https://doi.org/10.1007/s00764-023-00251-2>
- Jaiswal, R., Jayasinghe, L., & Kuhnert, N. (2012). Identification and characterization of proanthocyanidins of 16 members of the *Rhododendron* genus (Ericaceae) by tandem LC-MS. *Journal of Mass Spectrometry*, 47(4), 502–515. <https://doi.org/10.1002/jms.2954>
- Jaiswal, R., Matei, M. F., Glembockyte, V., Patras, M. A., & Kuhnert, N. (2014). Hierarchical Key for the LC – MS. *Journal of Agricultural and Food Chemistry*, 62, 9252–9265.
- Jaiswal, R., Sovdat, T., Vivan, F., & Kuhnert, N. (2010). Profiling and Characterization by LC-MS n of the Chlorogenic Acids and Hydroxycinnamoylshikimate Esters in Mate (*Ilex paraguariensis*). *J. Agric. Food Chem.* 2010, 58, 5471–5484. <https://doi.org/10.1021/jf904537z>
- Kind, T., & Fiehn, O. (2017). Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. *Phytochemistry Letters*, 21, 313–319. <https://doi.org/10.1016/j.phytol.2016.11.006>
- Kruve, A., Rebane, R., Kipper, K., Oldekop, M. L., Evard, H., Herodes, K., ... Leito, I. (2015). Tutorial review on validation of liquid chromatography-mass spectrometry methods: Part I. *Analytica Chimica Acta*, 870(1), 29–44. <https://doi.org/10.1016/j.aca.2015.02.017>
- Krygier, K., Sosulski, F., & Hogge, L. (1982). *Food Chem* (Vol. 30). Retrieved from UTC website: <https://pubs.acs.org/sharingguidelines>
- La Barbera, G., Capriotti, A. L., Cavaliere, C., Montone, C. M., Piovesana, S., Samperi, R., ... Laganà, A. (2017). Liquid chromatography-high resolution mass spectrometry for the analysis of phytochemicals in vegetal-derived food and beverages. *Food Research International*, 100, 28–52. <https://doi.org/10.1016/j.foodres.2017.07.080>
- Li, C. W., Dong, H. J., & Cui, C. Bin. (2015). The synthesis and antitumor activity of twelve galloyl glucosides. *Molecules*, 20(2), 2034–2060. <https://doi.org/10.3390/molecules20022034>
- Li, X., Xiao, H., Liang, X., Shi, D., & Liu, J. (2004). LC – MS / MS determination of naringin , hesperidin and neohesperidin in rat serum after orally administrating the decoction of *Bulpleurum falcatum* L . and *Fractus aurantii*. *Journal of Pharmaceutical and Biomedical Analysis*, 34, 159–166. <https://doi.org/10.1016/j.japna.2003.08.002>

- Liu, Z., Wu, H. L., Xie, L. X., Hu, Y., Fang, H., Sun, X. D., ... Yu, R. Q. (2017). Direct and interference-free determination of thirteen phenolic compounds in red wines using a chemometrics-assisted HPLC-DAD strategy for authentication of vintage year. *Analytical Methods*, 9(22), 3361–3374. <https://doi.org/10.1039/c7ay00415j>
- Mallmann, L.P., Tischer, B., Vizzotto, M., Rodrigues, E., & Manfroi, V. (2020). Comprehensive identification and quantification of unexploited phenolic compounds from red and yellow araçá (*Psidium cattleianum* Sabine) by LC-DAD-ESI-MS/MS. *Food Research International*, 131. <https://doi.org/10.1016/j.foodres.2020.108978>
- Mallmann, Luana P., O. Rios, A., & Rodrigues, E. (2023). MS-FINDER and SIRIUS for phenolic compound identification from high-resolution mass spectrometry data. *Food Research International*, 163(December 2022), 112315. <https://doi.org/10.1016/j.foodres.2022.112315>
- Mariutti, L. R. B., Rodrigues, E., Chisté, R. C., Fernandes, E., & Mercadante, A. Z. (2014). The Amazonian fruit *Byrsonima crassifolia* effectively scavenges reactive oxygen and nitrogen species and protects human erythrocytes against oxidative damage. *Food Research International*, 64, 618–625. <https://doi.org/10.1016/j.foodres.2014.07.032>
- Martins, C. C., Kahmann, A., Anzanello, M. J., Rodrigues, R. C., Rodrigues, E., & Mercali, G. D. (2023). Acid hydrolysis conditions do affect the non-extractable phenolic compounds composition from grape peel and seed. *Food Research International*, 174(October). <https://doi.org/10.1016/j.foodres.2023.113636>
- Mesquita, E., & Monteiro, M. (2018). Simultaneous HPLC determination of flavonoids and phenolic acids profile in Pêra-Rio orange juice. *Food Research International*, 106(December 2017), 54–63. <https://doi.org/10.1016/j.foodres.2017.12.025>
- Meyer, A. S., Jepsen, S. M., & Sørensen, N. S. (1998). Enzymatic Release of Antioxidants for Human Low-Density Lipoprotein from Grape Pomace. *Journal of Agricultural and Food Chemistry*, 46(7), 2439–2446. <https://doi.org/10.1021/jf971012f>
- Motilva, M. J., Serra, A., & Macià, A. (2013). Analysis of food polyphenols by ultra high-performance liquid chromatography coupled to mass spectrometry: An overview. *Journal of Chromatography A*, 1292, 66–82. <https://doi.org/10.1016/j.chroma.2013.01.012>
- Moussa-Ayoub, E., El-Samahy, S. K., Kroh, L. W., & Rohn, S. (2011). *Analytical Methods Identification and quantification of flavonol aglycons in cactus pear (Opuntia ficus indica) fruit using a commercial pectinase and cellulase preparation.*

<https://doi.org/10.1016/j.foodchem.2010.07.032>

- Mueller-Harvey, I. (2001). Analysis of hydrolysable tannins. *Animal Feed Science and Technology*, *91*, 3–20.
- Nguyen, D. D., Wu, C. H., Moree, W. J., Lamsa, A., Medema, M. H., Zhao, X., ... Dorrestein, P. C. (2013). MS/MS networking guided analysis of molecule and gene cluster families. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28). <https://doi.org/10.1073/pnas.1303471110>
- Padilha, C. V. da S., Miskinis, G. A., de Souza, M. E. A. O., Pereira, G. E., de Oliveira, D., Bordignon-Luiz, M. T., & Lima, M. dos S. (2017). Rapid determination of flavonoids and phenolic acids in grape juices and wines by RP-HPLC/DAD: Method validation and characterization of commercial products of the new Brazilian varieties of grape. *Food Chemistry*, *228*, 106–115. <https://doi.org/10.1016/j.foodchem.2017.01.137>
- Pascual-Teresa, S., Santos-Buelga, C., & Rivas-Gonzalo, J. G. (2000). Quantitative analysis of flavan-3-ols in Spanish foodstuffs and beverages. *Journal of Agricultural and Food Chemistry*, *48*(11), 5331–5337. <https://doi.org/10.1021/jf000549h>
- Pereira, V., Câmara, J. S., Cacho, J., & Marques, J. C. (2010). HPLC-DAD methodology for the quantification of organic acids, furans and polyphenols by direct injection of wine samples. *Journal of Separation Science*, *33*(9), 1204–1215. <https://doi.org/10.1002/jssc.200900784>
- Pérez-Jiménez, J., Díaz-Rubio, M. E., & Saura-Calixto, F. (2013). Non-extractable polyphenols, a major dietary antioxidant: Occurrence, metabolic fate and health effects. *Nutrition Research Reviews*, *26*(2), 118–129. <https://doi.org/10.1017/S0954422413000097>
- Pérez-Ramírez, I. F., Reynoso-Camacho, R., Saura-Calixto, F., & Pérez-Jiménez, J. (2018). Comprehensive Characterization of Extractable and Nonextractable Phenolic Compounds by High-Performance Liquid Chromatography-Electrospray Ionization-Quadrupole Time-of-Flight of a Grape/Pomegranate Pomace Dietary Supplement. *Journal of Agricultural and Food Chemistry*, *66*(3), 661–673. <https://doi.org/10.1021/acs.jafc.7b05901>
- Rauf, A., Imran, M., Abu-Izneid, T., Iahisham-Ul-Haq, Patel, S., Pan, X., ... Rasul Suleria, H. A. (2019). Proanthocyanidins: A comprehensive review. *Biomedicine and Pharmacotherapy*, *116*(February). <https://doi.org/10.1016/j.biopha.2019.108999>
- Regos, I., Urbanella, A., & Treutter, D. (2009). Identification and quantification of phenolic

- compounds from the forage legume sainfoin (*Onobrychis viciifolia*). *Journal of Agricultural and Food Chemistry*, 57(13), 5843–5852. <https://doi.org/10.1021/jf900625r>
- Reynoso-Camacho, R., & M. S. M. Rufino, D. M. Amaya-Cruz, Pérez, A. M. (2018). Non-extractable Polyphenols in Tropical Fruits: Occurrence and Health-related Properties. In *Non-extractable Polyphenols and Carotenoids: Importance in Human Nutrition and Health* (p. 382).
- Robbins, R. J. (2003). Phenolic Acids in Foods: An Overview of Analytical Methodology. *J. Agric. Food Chem*, 51, 2866–2887. <https://doi.org/10.1021/jf026182t>
- Rodrigues, N. P., & Bragagnolo, N. (2013). Identification and quantification of bioactive compounds in coffee brews by HPLC-DAD-MSn. *Journal of Food Composition and Analysis*, 32(2), 105–115. <https://doi.org/10.1016/j.jfca.2013.09.002>
- Rodriguez-Mateos, A., Vauzour, D., Krueger, C. G., Shanmuganayagam, D., Reed, J., Calani, L., ... Crozier, A. (2014). Bioavailability, bioactivity and impact on health of dietary flavonoids and related compounds: an update. *Arch Toxicol*, 3, 1803–1853. <https://doi.org/10.1007/s00204-014-1330-7>
- Rohr, G. E., Meier, B., & Sticher, O. (2000). Analysis of procyanidins. *Studies in Natural Products Chemistry*, 21(1), 497–570. [https://doi.org/10.1016/S1572-5995\(00\)80013-7](https://doi.org/10.1016/S1572-5995(00)80013-7)
- Rothwell, J. A., Perez-Jimenez, J., Neveu, V., Medina-Remón, A., M'Hiri, N., García-Lobato, P., ... Scalbert, A. (2013). Phenol-Explorer 3.0: A major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database*, 2013, 1–8. <https://doi.org/10.1093/database/bat070>
- Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., ... Saito, K. (2012). RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database. *Phytochemistry*, 82, 38–45. <https://doi.org/10.1016/j.phytochem.2012.07.007>
- Schieber, A., Keller, P., & Carle, R. (2001). Determination of phenolic acids and flavonoids of apple and pear by high-performance liquid chromatography. *Journal of Chromatography. A*, 910(2), 265–273. [https://doi.org/10.1016/S0021-9673\(00\)01217-6](https://doi.org/10.1016/S0021-9673(00)01217-6)
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ru, M., Singer, H. P., & Hollender, J. (2014). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environmental Science and Technology*, 48, 2097–2098.
- Schymanski, E. L., Kondić, T., Neumann, S., Thiessen, P. A., Zhang, J., & Bolton, E. E. (2021).

- Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *Journal of Cheminformatics*, *13*(1), 1–15. <https://doi.org/10.1186/s13321-021-00489-0>
- Sharma, N., Saini, D., Kesharwani, R. K., Gupta, P. C., & Keservani, R. K. (2024). Advances in flavonoids for human health and prevention of diseases. In *Advances in Flavonoids for Human Health and Prevention of Diseases*. <https://doi.org/10.1201/9781003369813>
- Sharma, R. J., Gupta, R. C., Singh, S., Bansal, A. K., & Singh, I. P. (2016). Stability of anthocyanins- and anthocyanidins-enriched extracts, and formulations of fruit pulp of *Eugenia jambolana* ('jamun'). *Food Chemistry*, *190*, 808–817. <https://doi.org/10.1016/j.foodchem.2015.06.029>
- Shi, J., Li, G., Zhang, R., Zheng, J., Suo, Y., You, J., & Liu, Y. J. (2011). A validated HPLC-DAD-MS method for identifying and determining the bioactive components of two kinds of Luobuma. *Journal of Liquid Chromatography and Related Technologies*, *34*(7), 537–547. <https://doi.org/10.1080/10826076.2011.546173>
- Sidhu, J. S., & Zafar, T. A. (2018). Bioactive compounds in banana fruits and their health benefits. *Food Quality and Safety*, *2*(4), 183–188. <https://doi.org/10.1093/fqsafe/fyy019>
- Soares, J. C., Rosalen, P. L., Lazarini, J. G., Sardi, J. D. C. O., Massarioli, A. P., Nani, B. D., ... De Alencar, S. M. (2020). Phenolic profile and potential beneficial effects of underutilized Brazilian native fruits on scavenging of ROS and RNS and anti-inflammatory and antimicrobial properties. *Food and Function*, *11*(10), 8905–8917. <https://doi.org/10.1039/d0fo01763a>
- Stalmach, A., Edwards, C. A., Wightman, J. D., & Crozier, A. (2011). Identification of (Poly)phenolic compounds in concord grape juice and their metabolites in human plasma and urine after juice consumption. *Journal of Agricultural and Food Chemistry*, *59*(17), 9512–9522. <https://doi.org/10.1021/jf2015039>
- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., ... Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, *3*(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Tahara, S. (2007). A Journey of Twenty-Five Years through the Ecological Biochemistry of Flavonoids. *Bioscience, Biotechnology, and Biochemistry*, *71*(6), 1387–1404. <https://doi.org/10.1271/bbb.70028>

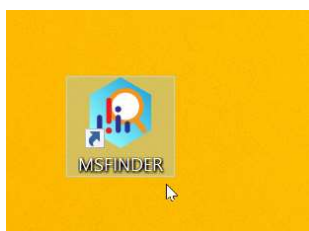
- Ticiane Carvalho Farias, Carolina Thomaz dos, Santos D’Almeida, Thaiza Serrano de Souza, Talita Pimenta do Nascimento, Fernanda de Sousa Bezerra, Roberta Nogueira Pereira da Silva, Mariana Simões Larráz Ferreira, Andrea Furtado Macedo, Maria Gabriela Bell, M. G. B. K. (2022). Application of Metabolomic Tools to Survey the Phenolic Composition of Food, Medicinal Plants, and Agro-Industrial Residues. In *Nutriomics* (p. 52).
- Tsao, R., Yang, R., Young, J. C., & Zhu, H. (2003). Polyphenolic profiles in eight apple cultivars using high-performance liquid chromatography (HPLC). *Journal of Agricultural and Food Chemistry*, *51*(21), 6347–6353. <https://doi.org/10.1021/jf0346298>
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., ... Arita, M. (2016). Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Analytical Chemistry*, *88*(16), 7946–7958. <https://doi.org/10.1021/acs.analchem.6b00770>
- Vaniya, A., Samra, S. N., Palazoglu, M., Tsugawa, H., & Fiehn, O. (2017). Using MS-FINDER for identifying 19 natural products in the CASMI 2016 contest. *Phytochemistry Letters*, *21*, 306–312. <https://doi.org/10.1016/j.phytol.2016.12.008>
- Verma, B., Hucl, P., & Chibbar, R. N. (2009). Phenolic acid composition and antioxidant capacity of acid and alkali hydrolysed wheat bran fractions. *Food Chemistry*, *116*, 947–954. <https://doi.org/10.1016/j.foodchem.2009.03.060>
- Vermerris, W., & Nicholson, R. (2006). *ISOLATION AND IDENTIFICATION OF PHENOLIC COMPOUNDS 151 ISOLATION AND IDENTIFICATION OF PHENOLIC COMPOUNDS A practical guide*. Retrieved from https://link.springer.com/content/pdf/10.1007/978-1-4020-5164-7_4.pdf
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., ... Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, *34*(8), 828–837. <https://doi.org/10.1038/nbt.3597>
- Wang, X., Li, N., Chen, S., Ge, Y., Xiao, Y., Zhao, M., & Wu, J. (2022). *MS-FINDER Assisted in Understanding the Profile of Flavonoids in Temporal Dimension during the Fermentation of Pu-erh Tea*. <https://doi.org/10.1021/acs.jafc.2c01595>
- Wang, Y., Kora, G., Bowen, B. P., & Pan, C. (2014). MIDAS: A database-searching algorithm for metabolite identification in metabolomics. *Analytical Chemistry*, *86*(19), 9496–9503. <https://doi.org/10.1021/ac5014783>

- Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., ... Dorrestein, P. C. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), 1743–1752. <https://doi.org/10.1073/pnas.1203689109>
- Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., ... Dorrestein, P. C. (2013). Molecular networking as a dereplication strategy. *Journal of Natural Products*, *76*(9), 1686–1699. <https://doi.org/10.1021/np400413s>

Supplementary material

Step-by-step procedure for identification using MS-Finder.

1. Open MS-FINDER



2. Click on File > Create a query...

A screenshot of the MS-FINDER ver. 3.52 software interface. The window title is 'MS-FINDER ver. 3.52'. The menu bar includes 'File', 'Analysis', 'Setting', 'Export', 'Tool', and 'Help'. The 'File' menu is open, showing 'Import...', 'Create a query...', and 'Create a query for MS-FINDER program...'. The main interface is divided into several panels. On the left is a 'File information' panel with input fields for Name, Scan number, Retention time (min), CCS (Å²), Precursor m/z, Precursor type, Ion mode, Spectrum type, Collision energy, Formula, Ontology, SMILES, InChIKey, Comment, MS1 peak number, and MS2 peak number. The central area contains two 'Mass chromatogram' plots showing 'Relative abundance' vs 'm/z' (0-100). The right side has a 'Structure finder' table with columns: Name, Score (max=1), Ontology, and InChIKey. Below the table are three plots: 'Spectrum' (Structure | Meta data), 'Measurement vs. Reference' (Actual MS/MS), and 'In silico MS/MS'. The 'Measurement vs. Reference' plot shows 'Relative abundance' vs 'm/z' (0-100). The 'In silico MS/MS' plot shows 'Fragment score' vs 'm/z' (0-100). There are also several icons on the right side of the interface, including a blue 'R', an orange 'I', a green 'P', a red 'N', and a blue 'S'.

3. In the “Create a query” tab, click on “Browse” to select a folder where the results should be saved.

Create a query...

Folder path:

File name:

Meta data

Scan number:

Retention time [min]:

Precursor m/z [Da]:

Precursor type:

Ion mode:

Spectrum type:

Collision energy:

Formula:

SMILES:

Intensity:

MS1 spectrum or EI spectrum

MS/MS spectrum or EI spectrum

4. In “File name”, type the name of the file. For example: Compound 1.

Create a query...

Folder path:

File name:

Meta data

Scan number:

Retention time [min]:

Precursor m/z [Da]:

Precursor type:

Ion mode:

Spectrum type:

Collision energy:

Formula:

SMILES:

Intensity:

MS1 spectrum or EI spectrum

MS/MS spectrum or EI spectrum

5. In “Ion mode”, select the ionization mode. For PC, we used negative mode.

The screenshot shows the 'Create a query...' dialog box with the following fields and settings:

- Folder path: D:\identification of sample 1
- File name: Compound 1
- Meta data:
 - Scan number: -1
 - Retention time [min]: 0.00
 - Precursor m/z [Da]: 0
 - Precursor type: [M+H]⁺
 - Ion mode: Positive (dropdown menu is open, showing options: Positive, Negative, Both)
 - Spectrum type: Positive
 - Collision energy: 0
 - Formula: (empty)
 - SMILES: (empty)
 - Intensity: 0
- MS1 spectrum or EI spectrum: (empty)
- MS/MS spectrum or EI spectrum: (empty)

Buttons: Finish, Cancel

6. In MS1 spectrum, type or paste m/z and Relatively intensity (I%) of the compound directly from raw data of the MS software.

The screenshot shows the 'Create a query...' dialog box with the following fields and settings:

- Folder path: D:\identification of sample 1
- File name: Compound 1
- Meta data:
 - Scan number: -1
 - Retention time [min]: 0.00
 - Precursor m/z [Da]: 0
 - Precursor type: [M-H]⁻
 - Ion mode: Negative
 - Spectrum type: Centroid
 - Collision energy: 40
 - Formula: (empty)
 - SMILES: (empty)
 - Intensity: 0
- MS1 spectrum or EI spectrum:

```
179.0350 100.0
180.0377 9.8
181.0389 1.2
```
- MS/MS spectrum or EI spectrum: (empty)

Buttons: Finish, Cancel

7. In Precursor m/z [Da], type or paste de m/z of the compound.

Create a query...

Folder path: D:\identification of sample 1 Browse

File name: Compound 1

Meta data

Scan number: -1

Retention time [min]: 0.00

Precursor m/z [Da]: 179.0350

Precursor type: [M-H]-

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

SMILES:

Intensity: 0

MS1 spectrum or EI spectrum

```
179.0350 100.0
180.0377 9.8
181.0389 1.2
```

MS/MS spectrum or EI spectrum

Finish Cancel

8. In MS/MS spectrum, type or paste m/z and Relatively intensity (I%) of the compound's fragments directly from raw data of the MS software.

Create a query...

Folder path: D:\identification of sample 1 Browse

File name: Compound 1

Meta data

Scan number: -1

Retention time [min]: 0.00

Precursor m/z [Da]: 179.035

Precursor type: [M-H]-

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

SMILES:

Intensity: 0

MS1 spectrum or EI spectrum

```
179.0350 100.0
180.0377 9.8
181.0389 1.2
```

MS/MS spectrum or EI spectrum

```
106.0411 4.6
107.0483 4.6
108.0217 3.1
109.0265 2.8
117.0330 6.2
133.0265 3.0
134.0359 81.5
135.0434 100.0
```

Finish Cancel

9. Then, click on “Finish”.

Create a query...

Folder path: D:\Identification of sample 1 Browse

File name: Compound 1

Meta data

Scan number: -1

Retention time [min]: 0.00

Precursor m/z [Da]: 179.035

Precursor type: [M-H]-

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

SMILES:

Intensity: 0

MS1 spectrum or EI spectrum

179.0350 100.0
180.0377 9.8
181.0389 1.2

MS/MS spectrum or EI spectrum

109.0265 2.8
117.0330 6.2
118.0347 0.6
132.0195 1.0
133.0265 3.0
134.0359 81.5
135.0434 100.0
136.0462 8.7

Finish Cancel

10. Compounds imputed will appear in the “File navigator”. Repeat steps 2 to 9 for all compounds.

MS-FINDER ver. 3.52 D:\Identification of sample 1

File Analysis Setting Export Tool Help

File navigator	Molecular formula finder						Structure finder			
	Formula	Error [mDa]	Error [ppm]	Score	Resource	Select	Name	Score (max=10)	Ontology	InChIKey
Compound 1										

File information

Name:

Scan number:

Retention time (min):

CCS (Å²):

Precursor m/z:

Precursor type:

Ion mode:

Spectrum type:

Collision energy:

Formula:

Ontology:

SMILES:

InChIKey:

Comment:

MS1 peak number:

MS2 peak number:

Mass chromatogram

Relative abundance vs m/z

Mass chromatogram

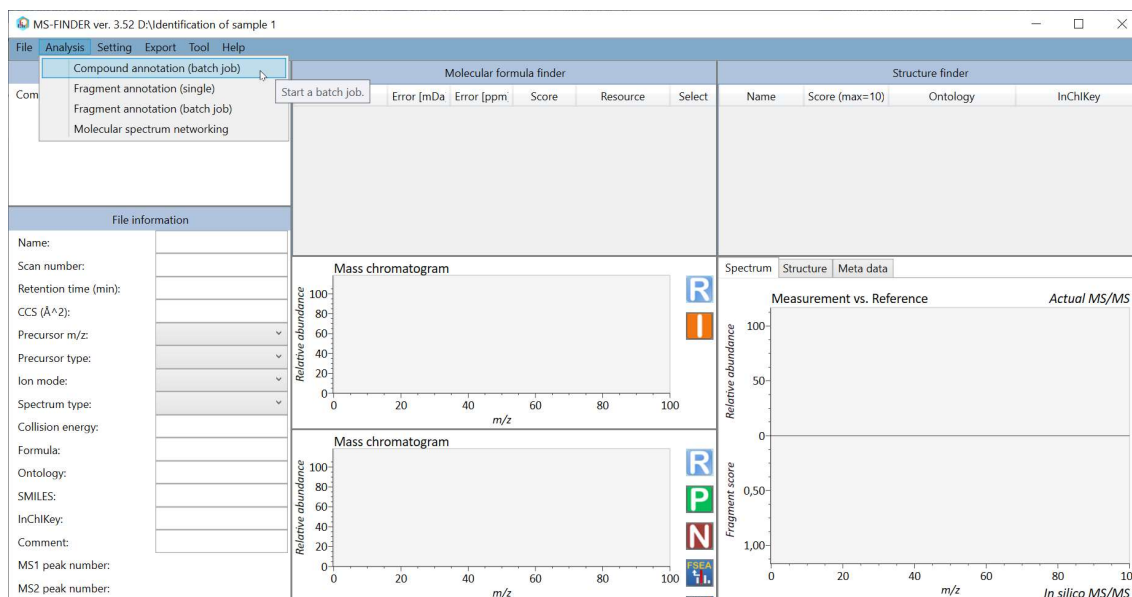
Relative abundance vs m/z

Measurement vs. Reference

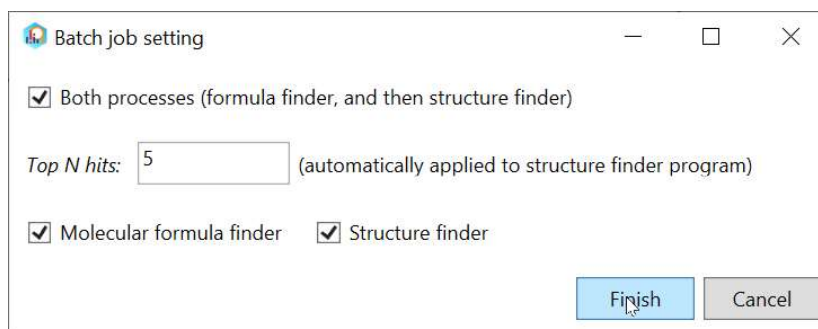
Actual MS/MS vs In silico MS/MS

Fragment score vs m/z

11. For identification, click on “Analysis and Compound annotation (batch job)”.



12. In the Batch job setting, check the information below and click on “Finish”.



13. A message indicating that identification succeeded will appear, click on “ok”.



14. Click in the compound of interest to see the information.

MS-FINDER ver. 3.52 C:\Users\luana\OneDrive\Identification of sample 1

File Analysis Setting Export Tool Help

File navigator

Compound 1

Molecular formula finder

Formula	Error [mDa]	Error [ppm]	Score	Resource	Select
C ₉ H ₈ O ₄	-0.0177	-0.0983	2.8000	HMDB,KNAPsAcK	<input checked="" type="checkbox"/>

Structure finder

Name	Score (max=10)	Ontology	InChIKey
Caffeic acid	4.92	Hydroxycinnamic acids	QAIPRVGONGVQAS-RQ
Aspirin	4.76	Acetylsalicylic acids	BSYNYRMUTXBXSQ-UHF
4-Hydroxyphen	4.76	Phenylpyruvic acid deriv	KKADPXVIOXHVKN-UHF
Monomethyl ph	4.43	Benzoic acid esters	FNJSWIPFHMKRAT-UHF
2-Hydroxy-3-(4	4.43	Phenylpyruvic acid deriv	GQYBCHRWMPPOOF-YV
(2E)-3-(2,4-dihy	4.38	Hydroxycinnamic acids	HGEPWFBFQKWMY-DL
(E)-3-(2,3-Dihyd	4.38	Hydroxycinnamic acids	SIUKXCMIDPYCLH-UHF
3-Hydroxyphen	4.32	Phenylpyruvic acid deriv	PNYWALDMLUDDTA-UH

File information

Name: Compound 1

Scan number: -1

Retention time (min): 0.00

CCS (Å²): 0.00

Precursor m/z: 179.035

Precursor type: [M-H]⁻

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

Ontology:

SMILES:

InChIKey:

Comment:

MS1 peak number: 3

MS2 peak number: 74

MS1 spectrum

MS/MS spectrum

Experimental spectrum vs. In silico spectrum Actual MS/MS

15. Possible molecular formulas will be indicated in the “Molecular formula finder”.

MS-FINDER ver. 3.52 C:\Users\luana\OneDrive\Identification of sample 1

File Analysis Setting Export Tool Help

File navigator

Compound 1

Molecular formula finder

Formula	Error [mDa]	Error [ppm]	Score	Resource	Select
C ₉ H ₈ O ₄	-0.0177	-0.0983	2.8000	HMDB,KNAPsAcK	<input checked="" type="checkbox"/>

Structure finder

Name	Score (max=10)	Ontology	InChIKey
Caffeic acid	4.92	Hydroxycinnamic acids	QAIPRVGONGVQAS-RQ
Aspirin	4.76	Acetylsalicylic acids	BSYNYRMUTXBXSQ-UHF
4-Hydroxyphen	4.76	Phenylpyruvic acid deriv	KKADPXVIOXHVKN-UHF
Monomethyl ph	4.43	Benzoic acid esters	FNJSWIPFHMKRAT-UHF
2-Hydroxy-3-(4	4.43	Phenylpyruvic acid deriv	GQYBCHRWMPPOOF-YV
(2E)-3-(2,4-dihy	4.38	Hydroxycinnamic acids	HGEPWFBFQKWMY-DL
(E)-3-(2,3-Dihyd	4.38	Hydroxycinnamic acids	SIUKXCMIDPYCLH-UHF
3-Hydroxyphen	4.32	Phenylpyruvic acid deriv	PNYWALDMLUDDTA-UH

File information

Name: Compound 1

Scan number: -1

Retention time (min): 0.00

CCS (Å²): 0.00

Precursor m/z: 179.035

Precursor type: [M-H]⁻

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

Ontology:

SMILES:

InChIKey:

Comment:

MS1 peak number: 3

MS2 peak number: 74

MS1 spectrum

MS/MS spectrum

Experimental spectrum vs. In silico spectrum Actual MS/MS

16. The ranking of identification is found in the up-right corner of the screen.

MS-FINDER ver. 3.52 C:\Users\luana\OneDrive\Identification of sample 1

File Analysis Setting Export Tool Help

File navigator

Compound 1

Molecular formula finder

Formula	Error [mDa]	Error [ppm]	Score	Resource	Select
C ₉ H ₈ O ₄	-0.0177	-0.0983	2.8000	HMDB,KNAP5AcK	<input checked="" type="checkbox"/>

Structure finder

Name	Score (max=10)	Ontology	InChIKey
Caffeic acid	4.92	Hydroxycinnamic acids	QAIPRVGONGVQAS-RQ
Aspirin	4.76	Acylsalicylic acids	BSYNRYMUTXBXSQ-UHF
4-Hydroxyphen	4.76	Phenylpyruvic acid deriv	KKADPXVOXHVKN-UHF
Monomethyl pff	4.43	Benzoic acid esters	FNJSWIPFHMKRAT-UHF
2-Hydroxy-3-(4	4.43	Phenylpyruvic acid deriv	GQYBCIHRWMPPOF-YV
(2E)-3-(2,4-dihy	4.38	Hydroxycinnamic acids	HGEPWFBFQKWMY-DL
(E)-3-(2,3-Dihyd	4.38	Hydroxycinnamic acids	SIUKXCMYPCYCLH-UHF
3-Hydroxyphen	4.32	Phenylpyruvic acid deriv	PNYWALDMLUDDTA-UH

File information

Name: Compound 1

Scan number: -1

Retention time (min): 0.00

CCS (Å²): 0.00

Precursor m/z: 179.035

Precursor type: [M-H]⁻

Ion mode: Negative

Spectrum type: Centroid

Collision energy: 40

Formula:

Ontology:

SMILES:

InChIKey:

Comment:

MS1 peak number: 3

MS2 peak number: 74

MS1 spectrum

MS/MS spectrum

Experimental spectrum vs. In silico spectrum

Actual MS/MS

Fragment score

Fragment score

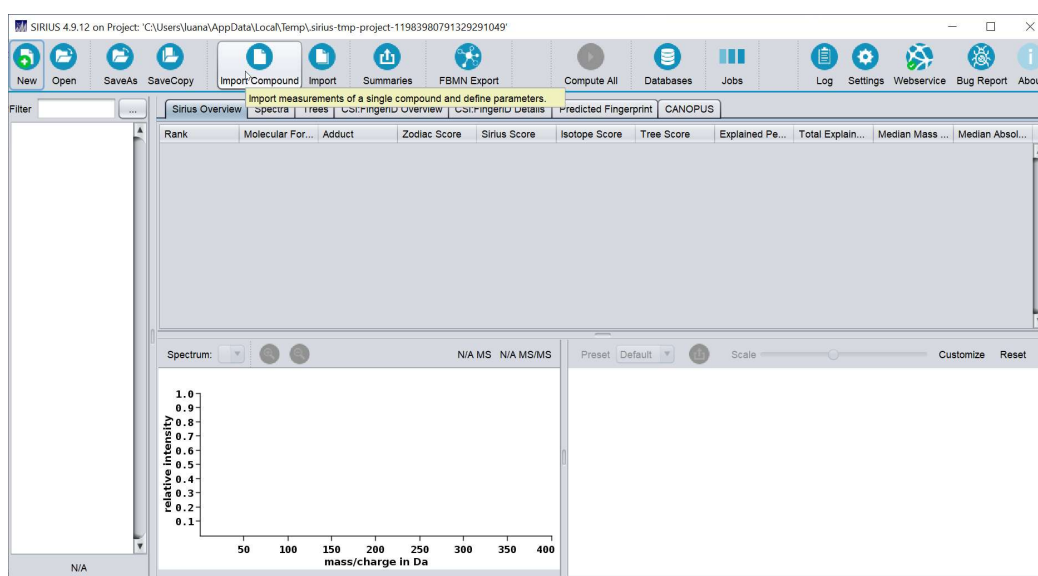
Fragment score

Step-by-step procedure for identification using SIRIUS.

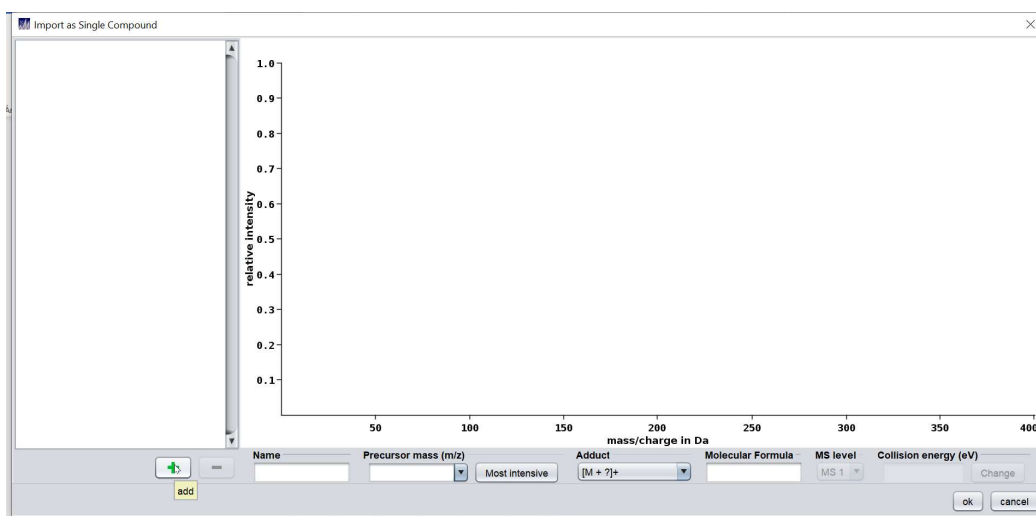
1. Open Sirius.



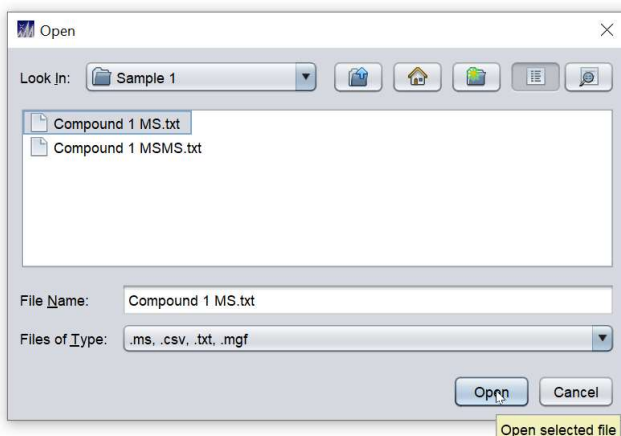
2. Click on "Import Compound".



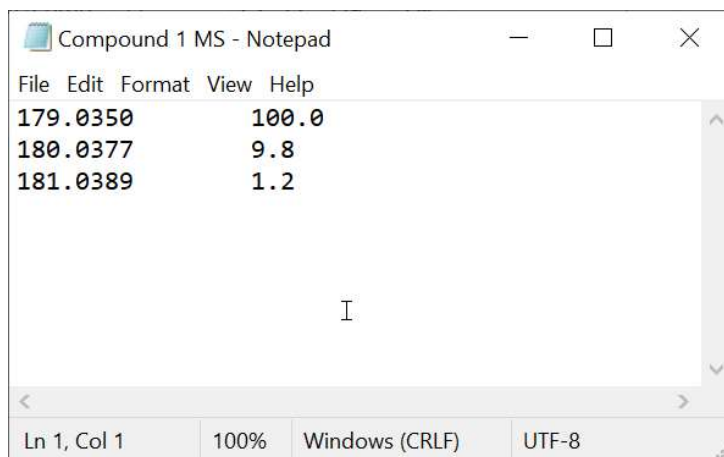
3. In the "Import as Single Compound" tab, click on the "add" sign.



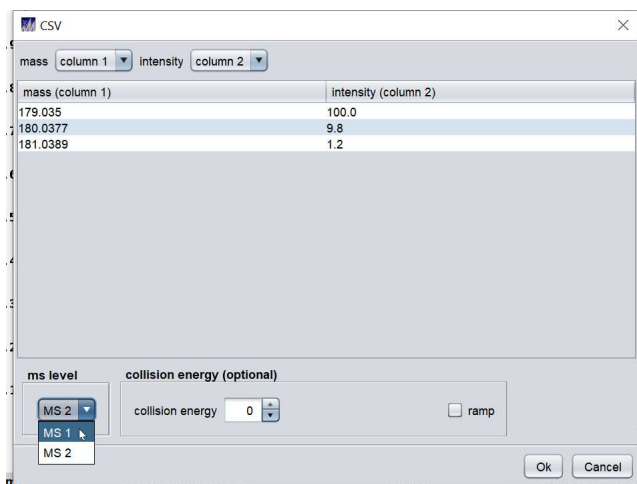
4. Select the archive in .txt format containing m/z and Relatively intensity (I%) of the compound.



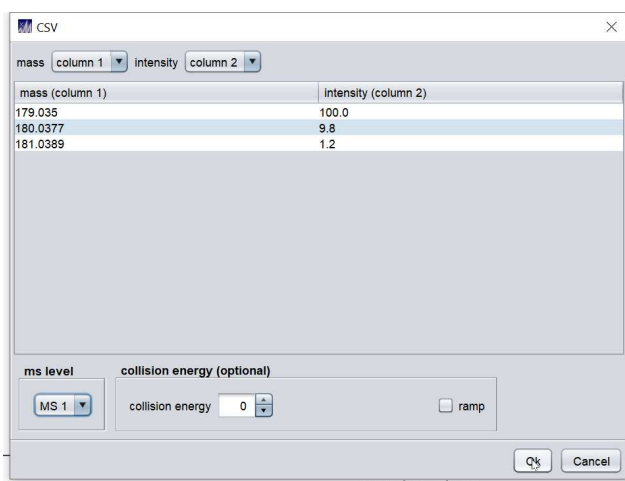
5. The information should be in this format, m/z and Relatively intensity (I%) of the compound from raw data in MS software to a .txt archive.



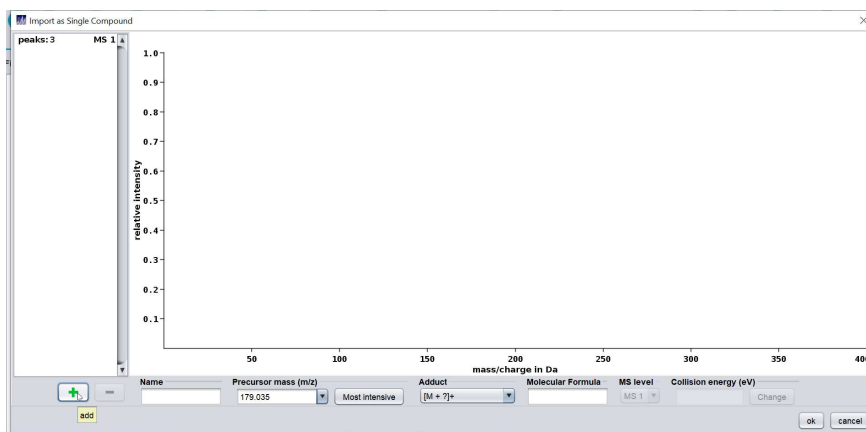
6. Select the "ms level", MS 1 for the m/z of the compound.



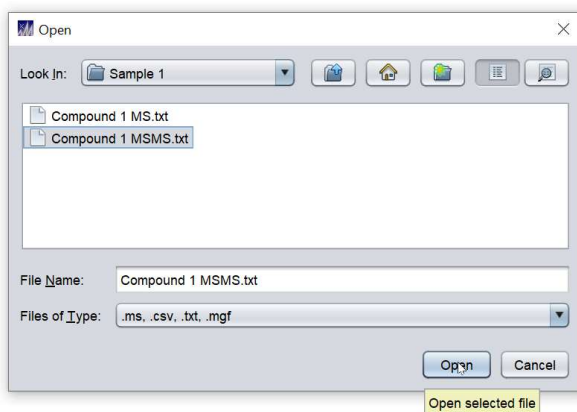
7. Click on “ok”.



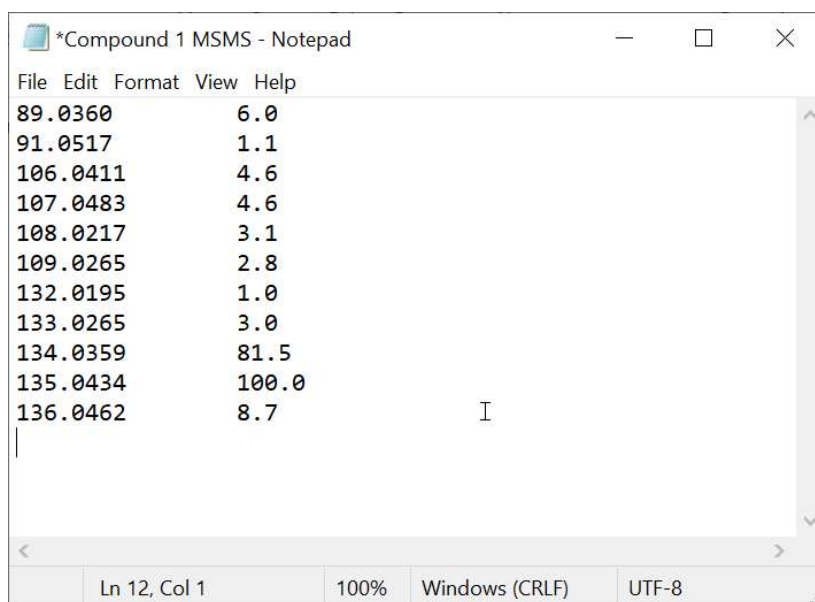
8. Click on the “add” sign again to add the MS/MS spectra.



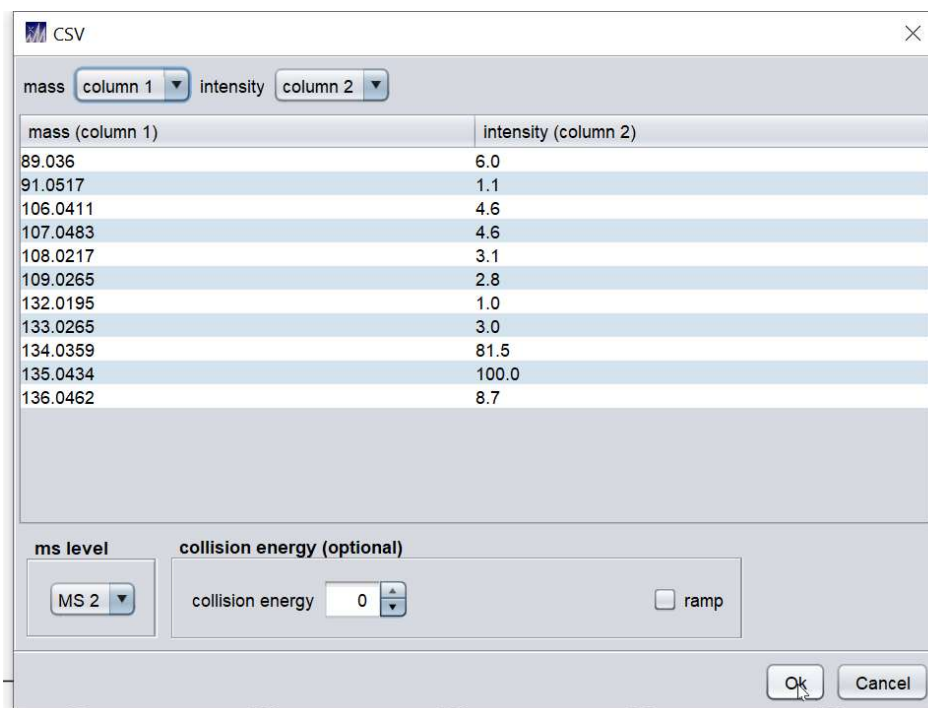
9. Select the archive in a .txt format containing m/z and Relatively intensity (I%) of the fragments.



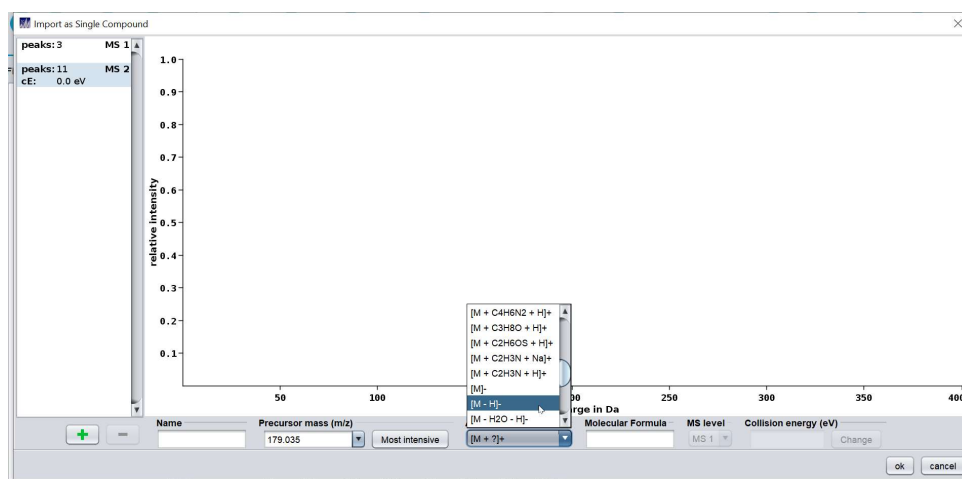
10. The information should be in this format, m/z and Relatively intensity (I%) of the compound's fragments from raw data in the MS software to a .txt archive.



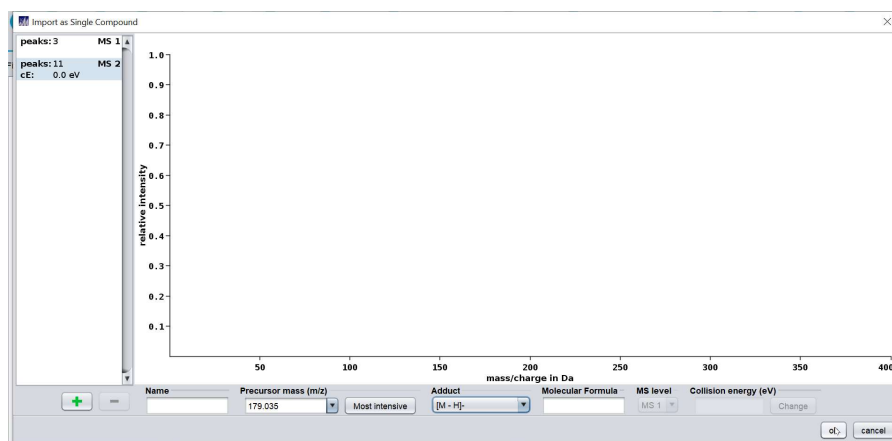
11. Check on "ms level" and click on "ok".



12. Select ionization mode. For PC, negative mode was used [M-H]⁻.



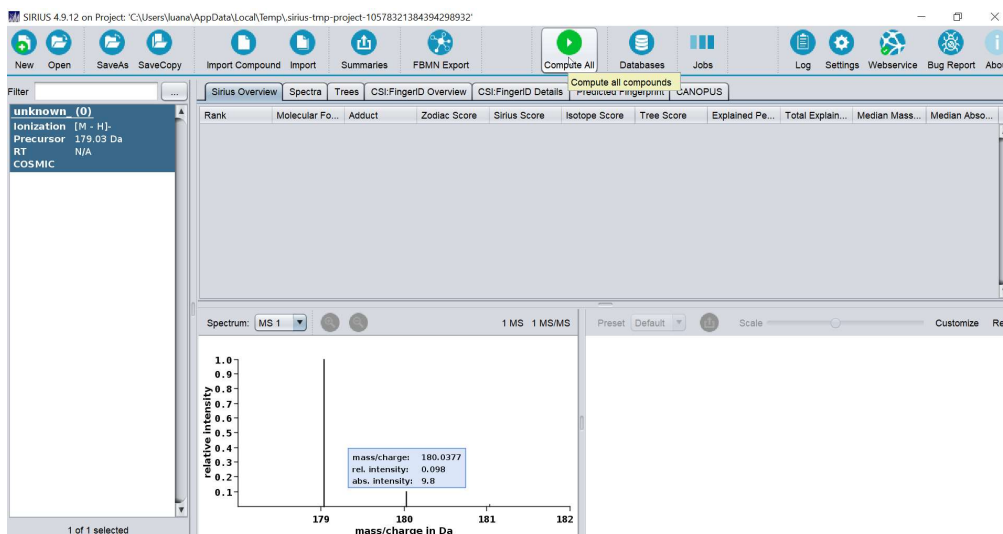
13. Click on “ok”.



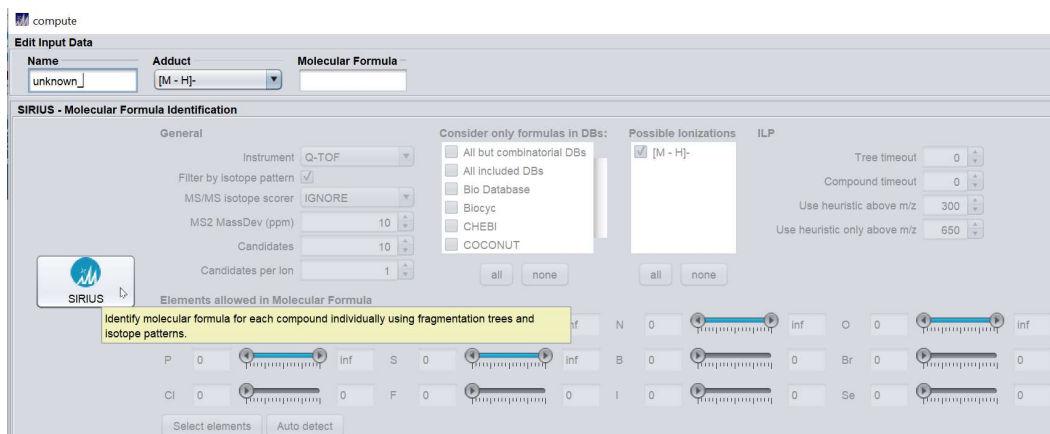
14. Compounds imputed will appear in the left of the screen. Repeat steps 2 to 13 for all compounds.



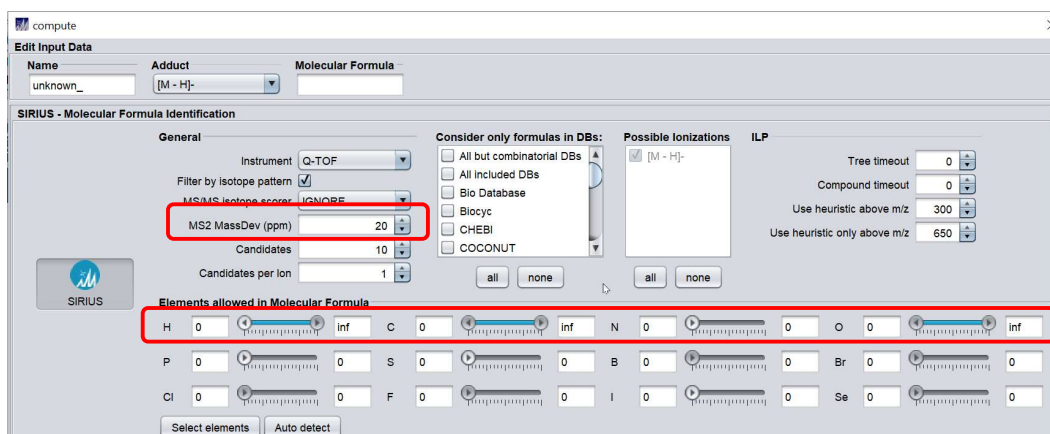
15. For identification, click on “Compute all”.



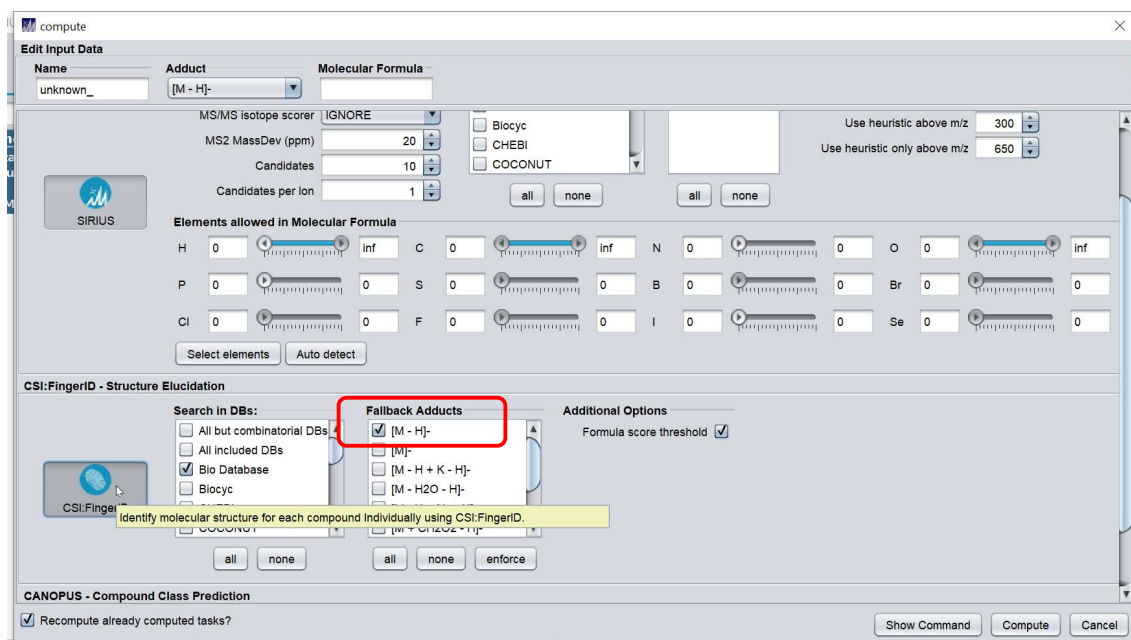
16. Click on the SIRIUS sign, this will enable the Sirius configuration.



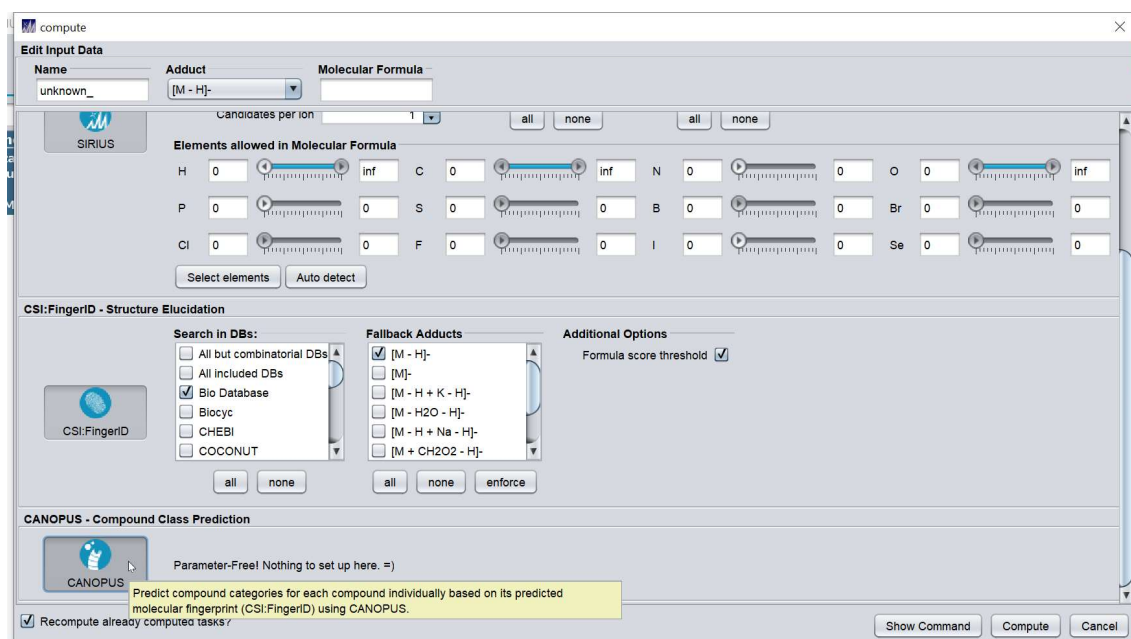
17. For PC identification, it was set 20 ppm of mass deviation and only H, C and O elements should be enabled, while the rest of the possible elements should be disabled.



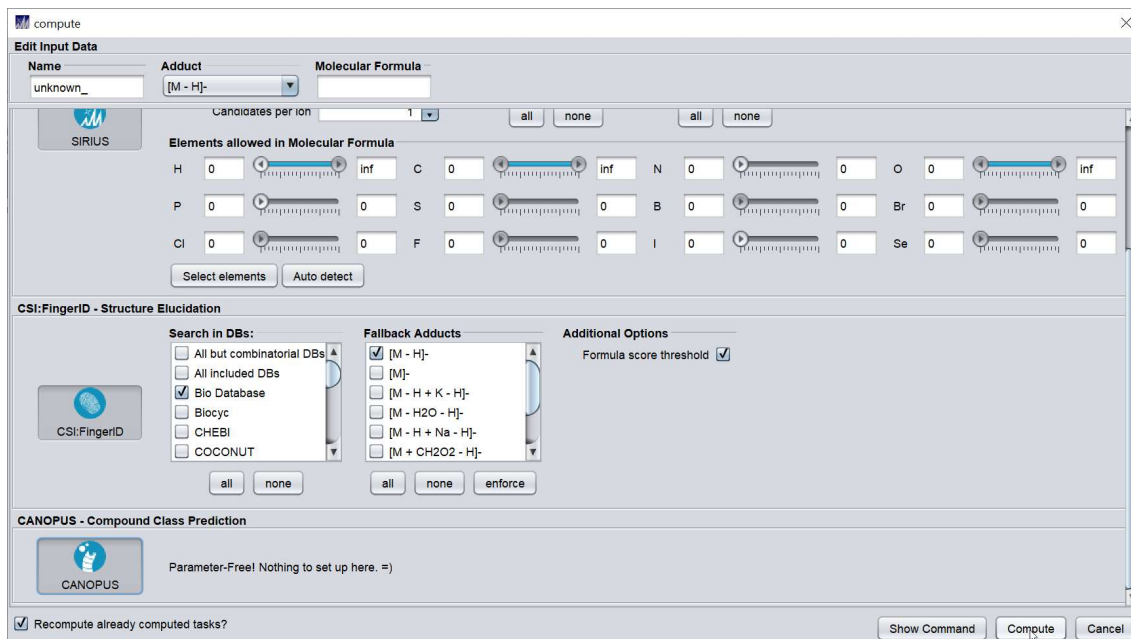
18. Click on the CSI:FingerID sign, this will enable the structure identification configuration. For PC, only [M-H]⁻ should be enabled.



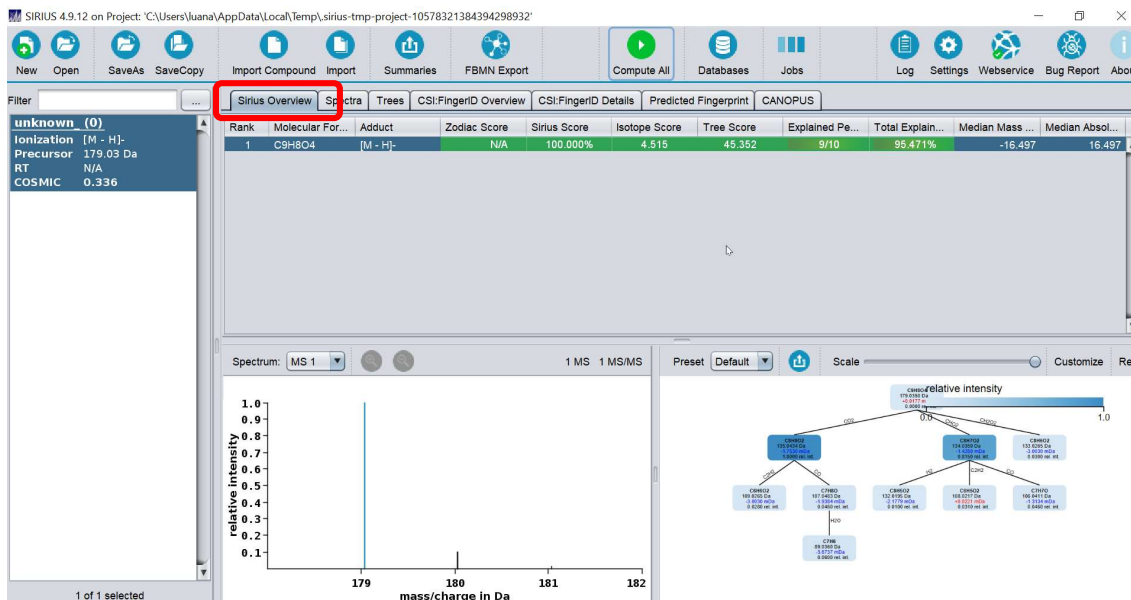
19. Click on the CANOPUS sign, this will enable the prediction of categories based on the compound fingerprint. There are no parameters to be set in this field.



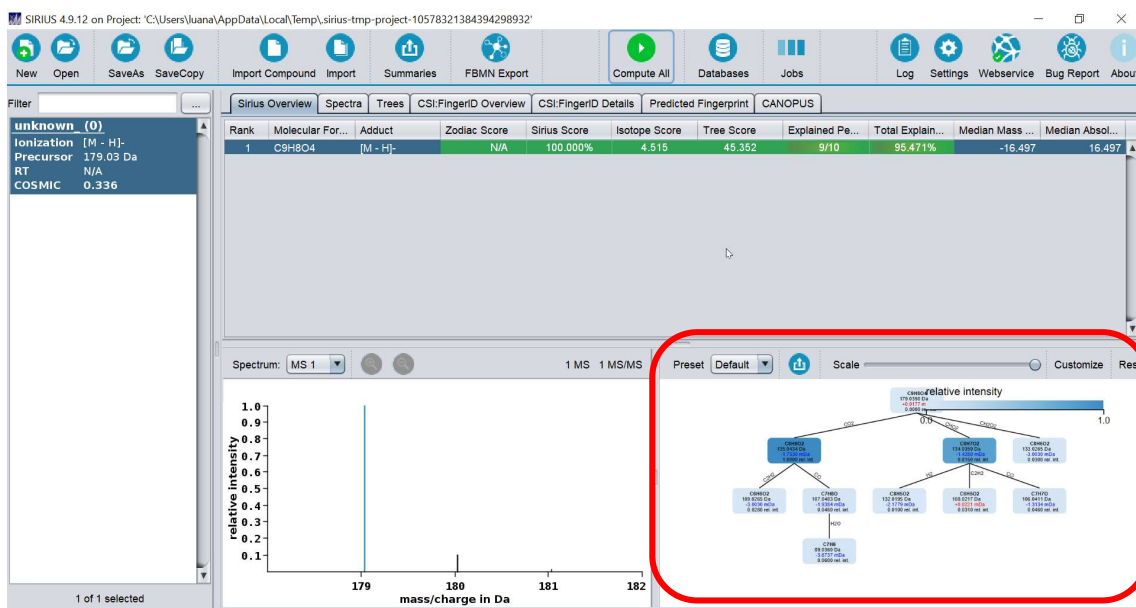
20. Click on “Compute”.



17. Information of the identification will appear in the screen. Possible molecular formulas will be indicated in the “Sirius Overview” tab.



18. A fragmentation tree will appear in the bottom right corner of the screen.



19. The ranking of identification will be shown in the “CSI:FingerID Overview” tab.

The screenshot shows the SIRIUS 4.9.12 interface with the 'CSI:FingerID Overview' tab selected. The table displays a list of potential identifications ranked by similarity. The top entry is Caffeate with a similarity of 71.429%. The table has columns: Rank, Name, SMILES, Molecular Form..., Adduct, CSI:FingerID Sc..., Similarity, #PubMed IDs, XLogP, and InChIKey. The table is filtered for XLogP: 1.00 and Similarity: 0. The table shows 27 rows of data, with the top row being Caffeate (Rank 1, SMILES C1=CC(=C(C=C...), Molecular Form... C9H8O4, Adduct [M - H]-, CSI:FingerID Sc... -85.557, Similarity 71.429%, #PubMed IDs 84.1280712.13..., XLogP 0.970, InChIKey GAIPIRVGONGV...).

Rank	Name	SMILES	Molecular Form...	Adduct	CSI:FingerID Sc...	Similarity	#PubMed IDs	XLogP	InChIKey
1	Caffeate	C1=CC(=C(C=C...	C9H8O4	[M - H]-	-85.557	71.429%	84.1280712.13...	0.970	GAIPIRVGONGV...
2	2,5-Dihydroxyl...	C1=CC(=C(C=...	C9H8O4	[M - H]-	-89.084	63.085%	1479375.11137...	1.200	JXIPYQZBOMJU...
3	(E)-3-(2,6-dihyd...	C1=CC(=C(C(=...	C9H8O4	[M - H]-	-112.920	60.119%		1.200	WFBVGDQDYL...
4	ACMC-20aocu	C1=CC(=C(C(=...	C9H8O4	[M - H]-	-118.830	61.850%	11826954.1903...	1.200	HGEFWFBFQK...
5	2,3-dihydroxyl...	C1=CC(=C(C=...	C9H8O4	[M - H]-	-135.594	54.072%	3158357.17454...	1.200	SIUKXCMQDPY...
6	3,5-Dihydroxyl...	C1=C(C=C(C=...	C9H8O4	[M - H]-	-141.489	58.480%	3158357.26712...	1.200	MFFCZSWTGM...
7	Methyl 5-Formyl...	COC(=O)C1=C...	C9H8O4	[M - H]-	-154.912	52.041%		1.037	XJFWPNWHD...
8	4-methylphthalate	CC1=CC(=C(C...	C9H8O4	[M - H]-	-155.243	41.361%	2170232.16158...	1.277	QWJAFQCTXF...
9	4-Hydroxy-enol...	C1=CC(=CC=C...	C9H8O4	[M - H]-	-156.758	54.651%		1.400	GOYBCHRWVM...
10	Methylterephth...	CC1=C(C=C(C...	C9H8O4	[M - H]-	-159.281	42.021%	15113539	1.300	UFMBQFGKHIX...
11	5-Acetylsalicylic...	CC(=O)C1=CC...	C9H8O4	[M - H]-	-161.254	51.500%	11699619.1117...	0.717	NZRDKNBIPL...
12	Benzoic acid, 2...	COC(=O)C1=C...	C9H8O4	[M - H]-	-166.283	49.744%		1.400	HDRLLQFTZ...
13	benzylmalonate	C1=CC=C(C=C...	C9H8O4	[M - H]-	-168.146	41.714%	7259141.11341...	0.846	WVYDZMNF...
14	ACMC-20adzr	COC(=O)C1=C...	C9H8O4	[M - H]-	-171.814	48.756%		1.400	ADSJWQKOKY...
15	Oprea1_539863	CC1=CC(=CC...	C9H8O4	[M - H]-	-178.798	52.105%		1.111	AAOLKUONXG...
16	D 3 (ester)	COC(=O)C1=C...	C9H8O4	[M - H]-	-180.473	39.896%	11853168.1266...	1.204	FNJSWPPFMK...
17	3-Methylphthalate	CC1=C(C(=CC...	C9H8O4	[M - H]-	-181.279	41.398%	3448154.11783...	1.277	IFJDBNISQJR...
18	4-hydroxybenzo...	C1=CC(=CC=C...	C9H8O4	[M - H]-	-188.978	47.208%		0.900	HRXAQVGNCW...
19	methyl 2-formyl...	COC(=O)C1=C...	C9H8O4	[M - H]-	-189.845	48.405%		1.400	WDYYLHVXCS...
20	Homoterephthal...	COC(=O)C1=C...	C9H8O4	[M - H]-	-194.589	38.021%	3723290.94065...	1.204	REDAIBAPLIA...
21	3-acetyl-4-hydr...	CC(=O)C1=C(C...	C9H8O4	[M - H]-	-195.781	49.500%		0.717	FPBXWQFOQ...
22	3,5-Dicarboxyto...	CC1=CC(=CC...	C9H8O4	[M - H]-	-199.212	40.223%	18154338.1913...	1.300	PMZBHPUNGN...
23	4-Carboxybenzo...	C1=CC(=CC=C...	C9H8O4	[M - H]-	-199.789	40.212%	21836970	0.956	DMEDOWYVXH...
24	Oprea1_251985	C1CO2C(C=C...	C9H8O4	[M - H]-	-200.700	46.009%	20824169	1.200	VCLSWKVAHAJ...
25	null	CC1=C(C=C(C...	C9H8O4	[M - H]-	-201.638	46.117%		1.100	CVXRFXBUIV...
26	Benzoic acid, 2...	COC1=CC=CC...	C9H8O4	[M - H]-	-203.928	48.500%		1.037	YIZSFEAVNICE...
27	Homophthalic a...	C1=CC=C(C(=...	C9H8O4	[M - H]-	-205.776	42.021%	2744421.83198...	0.956	ZHLQTKAVLK...

6 DISCUSSÃO GERAL

A determinação do perfil de compostos fenólicos presentes em alimentos tem sido objeto de estudo na área de Ciência e Tecnologia de Alimentos, onde a técnica de espectrometria de massas de alta resolução exibe um papel relevante na análise desses compostos. Não há dúvidas que o principal desafio não está na aplicação da técnica, mas sim, na interpretação dos dados de MS e MS² gerados. Neste sentido, o motivador dessa tese é fomentar a aplicação de ferramentas utilizadas na metabolômica não-direcionada no estudo de compostos fenólicos analisados por espectrometria de massas. Isso se justifica, visto que, apesar de estarmos nos referindo a uma classe específica de compostos, tal classe pode ser considerada um “mundo químico” único, possuindo grande complexidade estrutural, a qual é aumentada pelas suas interações com outras classes químicas.

Os resultados apresentados nesta tese indicam fortemente que a aplicação do GNPS aos dados de espectrometria de massas de compostos fenólicos de frutas, sem pré-processamento prévio, permite uma anotação rápida de um número relevante de compostos fenólicos. Quando esses compostos não foram anotados, eles foram agrupados, através das redes moleculares, em famílias moleculares. Esse agrupamento indica que há similaridade estrutural, possivelmente, com as moléculas pertencendo a mesma classe de compostos fenólicos. A partir desse orientador, pode-se anotar a molécula por propagação, preferencialmente, utilizando os softwares MS-Finder e Sirius. Tal procedimento foi largamente aplicado no desenvolvimento do segundo trabalho (Artigo 2).

Cabe destacar que o GNPS, apesar de possuir cerca de 700 mil espectros de MS², ainda é muito limitado em número de compostos fenólicos em suas bibliotecas. Visando aumentar esse número, adicionamos informações espectrais de todos os padrões de compostos fenólicos de nosso laboratório, o que impulsionou o número de anotações nas amostras de frutas. Assim, fica evidente que o fomento para que outros pesquisadores da área façam o mesmo com os seus padrões e com as suas anotações manuais terá consequências significativas no avanço no estudo de compostos fenólicos.

A partir dos resultados obtidos nos dois artigos, sugerimos que a anotação e identificação de compostos fenólicos inicie sempre com a plataforma GNPS. Fazendo upload

de dados brutos da análise de padrões analíticos se tem um ótimo ponto de partida. A partir disso, a formação de redes moleculares com a inclusão de dados brutos das amostras no GNPS trará uma análise preliminar rápida e bastante acertiva. A partir das redes moleculares, a anotação por propagação poderá ser feita manualmente e confirmada com o uso dos *softwares* MS-Finder e Sirius.

Como perspectivas futuras, a aplicação de ferramentas do GNPS como NAP (*Network Annotation Propagation*) permitirá realizar a anotação por propagação automaticamente. O ponto fraco dessa ferramenta é o uso do *software* MetFrag para realizar as fragmentações *in silico* e posterior anotação. Esse *software* apresenta resultados inferiores quando comparados ao MS-Finder e Sirius que são campeões em anotação. Outra ferramenta do GNPS que poderá diminuir umas das maiores dificuldades de anotação é o *Feature-based Molecular Network* (FBMN), essa ferramenta utiliza dados ortogonais imputados previamente para anotação de isômeros por diferentes tempos de retenção, por exemplo. Ainda, futuramente pode-se prospectar que a união dessas ferramentas junto com a Inteligência Artificial pode substituir em parte a anotação manual.

Por fim, esse estudo não teve a pretensão de explorar todas as ferramentas de software e plataformas disponíveis para a análise de compostos fenólicos. Da mesma forma, não há uma discussão profunda dos algoritmos das ferramentas aplicadas, visto que o objetivo é mostrar para pesquisadores que não dominam inteiramente a área de espectrometria de massas e de bioinformática que é possível realizar a anotação de composto fenólicos de maneira sistemática e com eficiência.

7 CONCLUSÃO GERAL

Neste trabalho avaliamos as ferramentas de bioinformática MS-Finder, Sirius e GNPS para anotar e identificar compostos fenólicos de alimentos. Ainda não podemos dizer que a anotação e identificação de compostos fenólicos é inteiramente automática. Porém, avançamos grandemente com uso de softwares e de uma plataforma onde partimos de dados brutos e complexos para uma classificação, agrupamento e indicação de possíveis estruturas e fórmulas moleculares.

A partir da utilização dessas ferramentas de bioinformática, é possível obter um ponto de partida para uma exploração mais profunda utilizando o seu conhecimento sobre a matriz alimentar sob investigação e outras informações adicionais (ou seja, ordem de eluição na coluna C18 e características de absorção UV-vis). Por fim, ainda continua um desafio a anotação de isômeros, o qual depende do desenvolvimento de novas ferramentas que integrem informações ortogonais.

8 REFERÊNCIAS

- Ali, A., Bashmil, Y. M., Cottrell, J. J., Suleria, H. A. R., & Dunshea, F. R. (2021). Lc-ms/ms-qtof screening and identification of phenolic compounds from australian grown herbs and their antioxidant potential. *Antioxidants*, *10*(11), 1–20. <https://doi.org/10.3390/antiox10111770>
- Arranz, S., & Calixto, S. (2010). Analysis of polyphenols in cereals may be improved performing acidic hydrolysis: A study in wheat flour and wheat bran and cereals of the diet. *Journal of Cereal Science*, *51*, 313–318. <https://doi.org/10.1016/j.jcs.2010.01.006>
- Arranz, S., Silván, J. M., & Saura-Calixto, F. (2010). Nonextractable polyphenols, usually ignored, are the major part of dietary polyphenols: A study on the Spanish diet. *Molecular Nutrition and Food Research*, *54*(11), 1646–1658. <https://doi.org/10.1002/mnfr.200900580>
- Arts, I. C. W., Putte, B. Van De, & Hollman, P. C. H. (2000). Catechin Contents of Foods Commonly Consumed in The Netherlands . 1 . Fruits , Vegetables , Staple Foods , and Processed Foods. *J. Agric. Food Chem*, *48*, 1746–1751.
- Ashraf, M. A., Iqbal, M., Rasheed, R., Hussain, I., Riaz, M., & Arif, M. S. (2018). Environmental Stress and Secondary Metabolites in Plants: An Overview. In *Plant Metabolites and Regulation under Environmental Stress*. <https://doi.org/10.1016/B978-0-12-812689-9.00008-X>
- Baquer, G., Sementé, L., Mahamdi, T., Correig, X., Ràfols, P., & García-Altres, M. (2023). What are we imaging? Software tools and experimental strategies for annotation and identification of small molecules in mass spectrometry imaging. *Mass Spectrometry Reviews*, *42*(5), 1927–1964. <https://doi.org/10.1002/mas.21794>
- Bedran-Russo, A. K., Pauli, G. F., Chen, S. N., McAlpine, J., Castellan, C. S., Phansalkar, R. S., ... Leme, A. A. (2014). Dentin biomodification: Strategies, renewable resources and clinical applications. *Dental Materials*, *30*(1), 62–76. <https://doi.org/10.1016/j.dental.2013.10.012>
- Bittremieux, W., Avalon, N. E., Thomas, S. P., Kakhkhorov, S. A., Gauglitz, J. M., Gerwick, W. H., ... Rima, F. (2023). Open Access Repository-Scale Propagated Nearest Neighbor Suspect Spectral Library for Untargeted Metabolomics. *BioRxiv*, 1–32. <https://doi.org/10.1038/s41467-023-44035-y>
- Cáceres-Jiménez, S., Rodríguez-Solana, R., Dobani, S., Pourshahidi, K., Gill, C., Moreno-Rojas, J. M., ... Pereira-Caro, G. (2023). UHPLC-HRMS Spectrometric Analysis: Method Validation and Plasma and Urinary Metabolite Identification after Mango Pulp Intake. *Journal of Agricultural and Food Chemistry*, *71*(30), 11520–11533. <https://doi.org/10.1021/acs.jafc.3c03846>
- Cardona, F., Andrés-Lacueva, C., Tulipani, S., Tinahones, F. J., & Queipo-Ortuño, M. I. (2013). Benefits of polyphenols on gut microbiota and implications in human health. *Journal of*

- Nutritional Biochemistry*, 24(8), 1415–1422.
<https://doi.org/10.1016/j.jnutbio.2013.05.001>
- Carvalho, D. O., Curto, A. F., & Guido, L. F. (2015). Determination of Phenolic Content in Different Barley Varieties and Corresponding Malts by Liquid Chromatography-diode Array Detection-Electrospray Ionization Tandem Mass Spectrometry. *Antioxidants (Basel)*, 4(3), 563–576. <https://doi.org/10.3390/antiox4030563>
- Cheng, C., & Gross, M. L. (2000). Applications and mechanisms of charge-remote fragmentation. *Mass Spectrometry Reviews*, 19(6), 398–420. [https://doi.org/10.1002/1098-2787\(2000\)19:6<398::AID-MAS3>3.0.CO;2-B](https://doi.org/10.1002/1098-2787(2000)19:6<398::AID-MAS3>3.0.CO;2-B)
- Chung, T. Te, Chuang, C. Y., Teng, Y. H., Hsieh, M. J., Lai, J. C., Chuang, Y. T., ... Yang, S. F. (2017). Tricetin suppresses human oral cancer cell migration by reducing matrix metalloproteinase-9 expression through the mitogen-activated protein kinase signaling pathway. *Environmental Toxicology*, 32(11), 2392–2399. <https://doi.org/10.1002/tox.22452>
- Clifford, M. N., Johnston, K. L., Knight, S., & Kuhnert, N. (2003). Hierarchical scheme for LC-MSn identification of chlorogenic acids. *Journal of Agricultural and Food Chemistry*, 51(10), 2900–2911. <https://doi.org/10.1021/jf026187q>
- Colletti, A., Sangiorgio, L., Martelli, A., Testai, L., Cicero, A. F. G., & Cravotto, G. (2021). Highly active cranberry's polyphenolic fraction: New advances in processing and clinical applications. *Nutrients*, 13(8). <https://doi.org/10.3390/nu13082546>
- Crozier, A., Jaganath, I. B., & Clifford, M. N. (2009). Dietary phenolics: Chemistry, bioavailability and effects on health. *Natural Product Reports*, 26(8), 1001–1043. <https://doi.org/10.1039/b802662a>
- De Freitas, T. B., Santos, C. H. K., da Silva, M. V., Shirai, M. A., Dias, M. I., Barros, L., ... Leimann, F. V. (2018). Antioxidants extraction from Pinhão (*Araucaria angustifolia* (Bertol.) Kuntze) coats and application to zein films. *Food Packaging and Shelf Life*, 15(July), 28–34. <https://doi.org/10.1016/j.fpsl.2017.10.006>
- de Rijke, E., Out, P., Niessen, W. M. A., Ariese, F., Gooijer, C., & Brinkman, U. A. T. (2006). Analytical separation and detection methods for flavonoids. *Journal of Chromatography A*, 1112(1–2), 31–63. <https://doi.org/10.1016/j.chroma.2006.01.019>
- Delcambre, A., & Saucier, C. (2012). Identification of new flavan-3-ol monoglycosides by UHPLC-ESI-Q-TOF in grapes and wine. *Journal of Mass Spectrometry*, 47(6), 727–736. <https://doi.org/10.1002/jms.3007>
- Demarque, D. P., Crotti, A. E. M., Vessecchi, R., Lopes, J. L. C., & Lopes, N. P. (2016). Fragmentation reactions using electrospray ionization mass spectrometry: An important tool for the structural elucidation and characterization of synthetic and natural products. *Natural Product Reports*, 33(3), 432–455. <https://doi.org/10.1039/c5np00073d>
- Díaz-de-Cerio, E., Tylewicz, U., Verardo, V., Fernández-Gutiérrez, A., Segura-Carretero, A., & Romani, S. (2017). Design of Sonotrode Ultrasound-Assisted Extraction of Phenolic

- Compounds from *Psidium guajava* L. Leaves. *Food Analytical Methods*, 10(8), 2781–2791. <https://doi.org/10.1007/s12161-017-0836-z>
- Dixon, R. A., Xie, D. Y., & Sharma, S. B. (2005). Proanthocyanidins - A final frontier in flavonoid research? *New Phytologist*, 165(1), 9–28. <https://doi.org/10.1111/j.1469-8137.2004.01217.x>
- Djoumbou-Feunang, Y., Pon, A., Karu, N., Zheng, J., Li, C., Arndt, D., ... Wishart, D. S. (2019). Cfm-id 3.0: Significantly improved esi-ms/ms prediction and compound identification. *Metabolites*, 9(4), 1–23. <https://doi.org/10.3390/metabo9040072>
- Domínguez-Rodríguez, G., Ramón Vidal, D., Martorell, P., Plaza, M., & Marina, M. L. (2022). Composition of Nonextractable Polyphenols from Sweet Cherry Pomace Determined by DART-Orbitrap-HRMS and Their in Vitro and in Vivo Potential Antioxidant, Antiaging, and Neuroprotective Activities. *Journal of Agricultural and Food Chemistry*, 70(26), 7993–8009. <https://doi.org/10.1021/acs.jafc.2c03346>
- Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A. A., Melnik, A. V., Meusel, M., ... Böcker, S. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nature Methods*, 16(4), 299–302. <https://doi.org/10.1038/s41592-019-0344-8>
- Dührkop, K., Scheubert, K., & Böcker, S. (2013). Molecular Formula Identification with SIRIUS. *Metabolites*, 3(2), 506–516. <https://doi.org/10.3390/metabo3020506>
- Durazzo, A. (2018). Extractable and Non-extractable Polyphenols: an Overview. In *Non-extractable Polyphenols and Carotenoids: Importance in Human Nutrition and Health* (p. 382). Retrieved from <https://pubs.rsc.org/en/content/chapterpdf/2018/9781788013208-00037?isbn=978-1-78801-106-8&sercode=bk>
- Eseberri, I., Trepiana, J., Léniz, A., Gómez-García, I., Carr-Ugarte, H., González, M., & Portillo, M. P. (2022). Variability in the Beneficial Effects of Phenolic Compounds: A Review. *Nutrients*, 14(9), 1–19. <https://doi.org/10.3390/nu14091925>
- Esparza-Martínez, F. J., Miranda-López, R., Mata-Sánchez, S. M., & Guzmán-Maldonado, S. H. (2016). Extractable and Non-Extractable Phenolics and Antioxidant Capacity of Mandarin Waste Dried at Different Temperatures. *Plant Foods Hum Nutr*, 71, 294–300. <https://doi.org/10.1007/s11130-016-0559-0>
- Fatemeh, S. R., Saifullah, R., Abbas, F. M. A., & Azhar, M. E. (2012). Total phenolics, flavonoids and antioxidant activity of banana pulp and peel flours: Influence of variety and stage of ripeness. *International Food Research Journal*, 19(3), 1041–1046.
- Freeman, B. L., Eggett, D. L., & Parker, T. L. (2010). Synergistic and antagonistic interactions of phenolic compounds found in navel oranges. *Journal of Food Science*, 75(6), 570–576. <https://doi.org/10.1111/j.1750-3841.2010.01717.x>
- Fuentes, Y., Giovagnoli-Vicuña, C., Faúndez, M., & Giordano, A. (2023). Microencapsulation of Chilean Papaya Waste Extract and Its Impact on Physicochemical and Bioactive Properties. *Antioxidants*, 12(10), 1–16. <https://doi.org/10.3390/antiox12101900>

- Garcia-salas, P., Morales-soto, A., Segura-carretero, A., & Fernández-gutiérrez, A. (2010). *Phenolic-Compound-Extraction Systems for Fruit and Vegetable Samples*. 8813–8826. <https://doi.org/10.3390/molecules15128813>
- Gazolla, M. C., Marques, L. M. M., e Silva, M. G., Araújo, M. T. M. F., Mendes, R. L., da Silva Almeida, J. R. G., ... Lopes, N. P. (2020). Characterization of 3-aminospirostane alkaloids from roots of *Solanum paniculatum* L. with hepatoprotective activity. *Rapid Communications in Mass Spectrometry*, 34(S3), 1–9. <https://doi.org/10.1002/rcm.8705>
- Gonzalez Ibarra, A. A., Wrobel, K., Yanez Barrientos, E., Corrales Escobosa, A. R., Gutierrez Corona, J. F., Enciso Donis, I., & Wrobel, K. (2017). Changes of metabolomic profile in *helianthus annuus* under exposure to Chromium(VI) Studied by capHPLC-ESI-QTOF-MS and MS/MS. *Journal of Analytical Methods in Chemistry*, 2017(Vi). <https://doi.org/10.1155/2017/3568621>
- Gu, L., Kelm, M. A., Hammerstone, J. F., Beecher, G., Holden, J., Haytowitz, D., ... Prior, R. L. (2004). Concentrations of Proanthocyanidins in Common Foods and Estimations of Normal Consumption. *Journal of Nutrition*, 134(3), 613–617. <https://doi.org/10.1093/jn/134.3.613>
- Haminiuk, C. W. I., Maciel, G. M., Plata-Oviedo, M. S. V., & Peralta, R. M. (2012). Phenolic compounds in fruits - an overview. *International Journal of Food Science and Technology*, 47(10), 2023–2044. <https://doi.org/10.1111/j.1365-2621.2012.03067.x>
- He, M., Wu, H., Nie, J., Yan, P., Yang, T. B., Yang, Z. Y., & Pei, R. (2017). Accurate recognition and feature qualify for flavonoid extracts from Liang-wai Gan Cao by liquid chromatography-high resolution-mass spectrometry and computational MS/MS fragmentation. *Journal of Pharmaceutical and Biomedical Analysis*, 146, 37–47. <https://doi.org/10.1016/j.jpba.2017.07.065>
- Hellström, J. K., Törrönen, A. R., & Mattila, P. H. (2009). Proanthocyanidins in common food products of plant origin. *Journal of Agricultural and Food Chemistry*, 57(17), 7899–7906. <https://doi.org/10.1021/jf901434d>
- Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., ... Nishioka, T. (2010). MassBank: A public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7), 703–714. <https://doi.org/10.1002/jms.1777>
- Hou, J., Liang, L., Su, M., Yang, T., Mao, X., & Wang, Y. (2021). Variations in phenolic acids and antioxidant activity of navel orange at different growth stages. *Food Chemistry*, 360(October 2020), 129980. <https://doi.org/10.1016/j.foodchem.2021.129980>
- Hui Li, Leren Wan, Y. H. and S. C. (2007). Fragmentation study of a 8-C-glycosyl isoflavone, puerarin, using electrospray ion trap time-of-flight mass spectrometry at high resolution. *Rapid Communications in Mass Spectrometry*, 21, 2497–2504. <https://doi.org/10.1002/rcm>
- Huo, C., Nguyen, Q. N., Alishir, A., Ra, M., Jung, S., & Yu, J. (2023). Global Natural Products Social (GNPS) -Based Molecular-Networking-Guided Isolation of Phenolic Compounds from *Ginkgo biloba* Fruits and the Identification of. *Plants*, 12(3970), 1–13.

- IBGE Instituto Brasileiro de Geografia e Estatística. (2020). *Pesquisa de Orçamentos Familiares 2017 - 2018*. Retrieved from <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101742.pdf>
- Jadaun, V., Prateeksha, P., Nailwal, T., & Singh, B. N. (2023). Antioxidant activity and simultaneous estimation of four polyphenolics in different parts of *Carica papaya* L. by a validated high-performance thin-layer chromatography method. *Journal of Planar Chromatography - Modern TLC*, 36(4), 211–221. <https://doi.org/10.1007/s00764-023-00251-2>
- Jaiswal, R., Jayasinghe, L., & Kuhnert, N. (2012). Identification and characterization of proanthocyanidins of 16 members of the *Rhododendron* genus (Ericaceae) by tandem LC-MS. *Journal of Mass Spectrometry*, 47(4), 502–515. <https://doi.org/10.1002/jms.2954>
- Jaiswal, R., Matei, M. F., Glembockyte, V., Patras, M. A., & Kuhnert, N. (2014). Hierarchical Key for the LC – MS. *Journal of Agricultural and Food Chemistry*, 62, 9252–9265.
- Jaiswal, R., Sovdat, T., Vivan, F., & Kuhnert, N. (2010). Profiling and Characterization by LC-MS n of the Chlorogenic Acids and Hydroxycinnamoylshikimate Esters in Mate (*Ilex paraguariensis*). *J. Agric. Food Chem.* 2010, 58, 5471–5484. <https://doi.org/10.1021/jf904537z>
- Kind, T., & Fiehn, O. (2017). Strategies for dereplication of natural compounds using high-resolution tandem mass spectrometry. *Phytochemistry Letters*, 21, 313–319. <https://doi.org/10.1016/j.phytol.2016.11.006>
- Kruve, A., Rebane, R., Kipper, K., Oldekop, M. L., Evard, H., Herodes, K., ... Leito, I. (2015). Tutorial review on validation of liquid chromatography-mass spectrometry methods: Part I. *Analytica Chimica Acta*, 870(1), 29–44. <https://doi.org/10.1016/j.aca.2015.02.017>
- Krygier, K., Sosulski, F., & Hogge, L. (1982). *Food Chem* (Vol. 30). Retrieved from UTC website: <https://pubs.acs.org/sharingguidelines>
- La Barbera, G., Capriotti, A. L., Cavaliere, C., Montone, C. M., Piovesana, S., Samperi, R., ... Laganà, A. (2017). Liquid chromatography-high resolution mass spectrometry for the analysis of phytochemicals in vegetal-derived food and beverages. *Food Research International*, 100, 28–52. <https://doi.org/10.1016/j.foodres.2017.07.080>
- Li, C. W., Dong, H. J., & Cui, C. Bin. (2015). The synthesis and antitumor activity of twelve galloyl glucosides. *Molecules*, 20(2), 2034–2060. <https://doi.org/10.3390/molecules20022034>
- Li, X., Xiao, H., Liang, X., Shi, D., & Liu, J. (2004). LC – MS / MS determination of naringin, hesperidin and neohesperidin in rat serum after orally administrating the decoction of *Bulpleurum falcatum* L. and *Fractus aurantii*. *Journal of Pharmaceutical and Biomedical Analysis*, 34, 159–166. <https://doi.org/10.1016/j.japna.2003.08.002>
- Liu, Z., Wu, H. L., Xie, L. X., Hu, Y., Fang, H., Sun, X. D., ... Yu, R. Q. (2017). Direct and interference-free determination of thirteen phenolic compounds in red wines using a chemometrics-assisted HPLC-DAD strategy for authentication of vintage year. *Analytical*

- Methods*, 9(22), 3361–3374. <https://doi.org/10.1039/c7ay00415j>
- Mallmann, L.P., Tischer, B., Vizzotto, M., Rodrigues, E., & Manfroi, V. (2020). Comprehensive identification and quantification of unexploited phenolic compounds from red and yellow araçá (*Psidium cattleianum* Sabine) by LC-DAD-ESI-MS/MS. *Food Research International*, 131. <https://doi.org/10.1016/j.foodres.2020.108978>
- Mallmann, Luana P., O. Rios, A., & Rodrigues, E. (2023). MS-FINDER and SIRIUS for phenolic compound identification from high-resolution mass spectrometry data. *Food Research International*, 163(December 2022), 112315. <https://doi.org/10.1016/j.foodres.2022.112315>
- Mariutti, L. R. B., Rodrigues, E., Chisté, R. C., Fernandes, E., & Mercadante, A. Z. (2014). The Amazonian fruit *Byrsonima crassifolia* effectively scavenges reactive oxygen and nitrogen species and protects human erythrocytes against oxidative damage. *Food Research International*, 64, 618–625. <https://doi.org/10.1016/j.foodres.2014.07.032>
- Martins, C. C., Kahmann, A., Anzanello, M. J., Rodrigues, R. C., Rodrigues, E., & Mercali, G. D. (2023). Acid hydrolysis conditions do affect the non-extractable phenolic compounds composition from grape peel and seed. *Food Research International*, 174(October). <https://doi.org/10.1016/j.foodres.2023.113636>
- Mesquita, E., & Monteiro, M. (2018). Simultaneous HPLC determination of flavonoids and phenolic acids profile in Pêra-Rio orange juice. *Food Research International*, 106(December 2017), 54–63. <https://doi.org/10.1016/j.foodres.2017.12.025>
- Meyer, A. S., Jepsen, S. M., & Sørensen, N. S. (1998). Enzymatic Release of Antioxidants for Human Low-Density Lipoprotein from Grape Pomace. *Journal of Agricultural and Food Chemistry*, 46(7), 2439–2446. <https://doi.org/10.1021/jf971012f>
- Motilva, M. J., Serra, A., & Macià, A. (2013). Analysis of food polyphenols by ultra high-performance liquid chromatography coupled to mass spectrometry: An overview. *Journal of Chromatography A*, 1292, 66–82. <https://doi.org/10.1016/j.chroma.2013.01.012>
- Moussa-Ayoub, E., El-Samahy, S. K., Kroh, L. W., & Rohn, S. (2011). *Analytical Methods Identification and quantification of flavonol aglycons in cactus pear (Opuntia ficus indica) fruit using a commercial pectinase and cellulase preparation.* <https://doi.org/10.1016/j.foodchem.2010.07.032>
- Mueller-Harvey, I. (2001). Analysis of hydrolysable tannins. *Animal Feed Science and Technology*, 91, 3–20.
- Nguyen, D. D., Wu, C. H., Moree, W. J., Lamsa, A., Medema, M. H., Zhao, X., ... Dorrestein, P. C. (2013). MS/MS networking guided analysis of molecule and gene cluster families. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28). <https://doi.org/10.1073/pnas.1303471110>
- Padilha, C. V. da S., Miskinis, G. A., de Souza, M. E. A. O., Pereira, G. E., de Oliveira, D., Bordignon-Luiz, M. T., & Lima, M. dos S. (2017). Rapid determination of flavonoids and phenolic acids in grape juices and wines by RP-HPLC/DAD: Method validation and

- characterization of commercial products of the new Brazilian varieties of grape. *Food Chemistry*, 228, 106–115. <https://doi.org/10.1016/j.foodchem.2017.01.137>
- Pascual-Teresa, S., Santos-Buelga, C., & Rivas-Gonzalo, J. G. (2000). Quantitative analysis of flavan-3-ols in Spanish foodstuffs and beverages. *Journal of Agricultural and Food Chemistry*, 48(11), 5331–5337. <https://doi.org/10.1021/jf000549h>
- Pereira, V., Câmara, J. S., Cacho, J., & Marques, J. C. (2010). HPLC-DAD methodology for the quantification of organic acids, furans and polyphenols by direct injection of wine samples. *Journal of Separation Science*, 33(9), 1204–1215. <https://doi.org/10.1002/jssc.200900784>
- Pérez-Jiménez, J., Díaz-Rubio, M. E., & Saura-Calixto, F. (2013). Non-extractable polyphenols, a major dietary antioxidant: Occurrence, metabolic fate and health effects. *Nutrition Research Reviews*, 26(2), 118–129. <https://doi.org/10.1017/S0954422413000097>
- Pérez-Ramírez, I. F., Reynoso-Camacho, R., Saura-Calixto, F., & Pérez-Jiménez, J. (2018). Comprehensive Characterization of Extractable and Nonextractable Phenolic Compounds by High-Performance Liquid Chromatography-Electrospray Ionization-Quadrupole Time-of-Flight of a Grape/Pomegranate Pomace Dietary Supplement. *Journal of Agricultural and Food Chemistry*, 66(3), 661–673. <https://doi.org/10.1021/acs.jafc.7b05901>
- Rauf, A., Imran, M., Abu-Izneid, T., Iahitsham-Ul-Haq, Patel, S., Pan, X., ... Rasul Suleria, H. A. (2019). Proanthocyanidins: A comprehensive review. *Biomedicine and Pharmacotherapy*, 116(February). <https://doi.org/10.1016/j.biopha.2019.108999>
- Regos, I., Urbanella, A., & Treutter, D. (2009). Identification and quantification of phenolic compounds from the forage legume sainfoin (*Onobrychis viciifolia*). *Journal of Agricultural and Food Chemistry*, 57(13), 5843–5852. <https://doi.org/10.1021/jf900625r>
- Reynoso-Camacho, R., & M. S. M. Rufino, D. M. Amaya-Cruz, Pérez, A. M. (2018). Non-extractable Polyphenols in Tropical Fruits: Occurrence and Health-related Properties. In *Non-extractable Polyphenols and Carotenoids: Importance in Human Nutrition and Health* (p. 382).
- Robbins, R. J. (2003). Phenolic Acids in Foods: An Overview of Analytical Methodology. *J. Agric. Food Chem*, 51, 2866–2887. <https://doi.org/10.1021/jf026182t>
- Rodrigues, N. P., & Bragagnolo, N. (2013). Identification and quantification of bioactive compounds in coffee brews by HPLC-DAD-MSn. *Journal of Food Composition and Analysis*, 32(2), 105–115. <https://doi.org/10.1016/j.jfca.2013.09.002>
- Rodriguez-Mateos, A., Vauzour, D., Krueger, C. G., Shanmuganayagam, D., Reed, J., Calani, L., ... Crozier, A. (2014). Bioavailability, bioactivity and impact on health of dietary flavonoids and related compounds: an update. *Arch Toxicol*, 3, 1803–1853. <https://doi.org/10.1007/s00204-014-1330-7>
- Rohr, G. E., Meier, B., & Sticher, O. (2000). Analysis of procyanidins. *Studies in Natural Products Chemistry*, 21(1), 497–570. [https://doi.org/10.1016/S1572-5995\(00\)80013-7](https://doi.org/10.1016/S1572-5995(00)80013-7)

- Rothwell, J. A., Perez-Jimenez, J., Neveu, V., Medina-Remón, A., M'Hiri, N., García-Lobato, P., ... Scalbert, A. (2013). Phenol-Explorer 3.0: A major update of the Phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. *Database*, 2013, 1–8. <https://doi.org/10.1093/database/bat070>
- Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M., Sakata, A., ... Saito, K. (2012). RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database. *Phytochemistry*, 82, 38–45. <https://doi.org/10.1016/j.phytochem.2012.07.007>
- Schieber, A., Keller, P., & Carle, R. (2001). Determination of phenolic acids and flavonoids of apple and pear by high-performance liquid chromatography. *Journal of Chromatography A*, 910(2), 265–273. [https://doi.org/10.1016/S0021-9673\(00\)01217-6](https://doi.org/10.1016/S0021-9673(00)01217-6)
- Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ru, M., Singer, H. P., & Hollender, J. (2014). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environmental Science and Technology*, 48, 2097–2098.
- Schymanski, E. L., Kondić, T., Neumann, S., Thiessen, P. A., Zhang, J., & Bolton, E. E. (2021). Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. *Journal of Cheminformatics*, 13(1), 1–15. <https://doi.org/10.1186/s13321-021-00489-0>
- Sharma, N., Saini, D., Kesharwani, R. K., Gupta, P. C., & Keservani, R. K. (2024). Advances in flavonoids for human health and prevention of diseases. In *Advances in Flavonoids for Human Health and Prevention of Diseases*. <https://doi.org/10.1201/9781003369813>
- Sharma, R. J., Gupta, R. C., Singh, S., Bansal, A. K., & Singh, I. P. (2016). Stability of anthocyanins- and anthocyanidins-enriched extracts, and formulations of fruit pulp of *Eugenia jambolana* ('jamun'). *Food Chemistry*, 190, 808–817. <https://doi.org/10.1016/j.foodchem.2015.06.029>
- Shi, J., Li, G., Zhang, R., Zheng, J., Suo, Y., You, J., & Liu, Y. J. (2011). A validated HPLC-DAD-MS method for identifying and determining the bioactive components of two kinds of Luobuma. *Journal of Liquid Chromatography and Related Technologies*, 34(7), 537–547. <https://doi.org/10.1080/10826076.2011.546173>
- Sidhu, J. S., & Zafar, T. A. (2018). Bioactive compounds in banana fruits and their health benefits. *Food Quality and Safety*, 2(4), 183–188. <https://doi.org/10.1093/fqsafe/fyy019>
- Soares, J. C., Rosalen, P. L., Lazarini, J. G., Sardi, J. D. C. O., Massarioli, A. P., Nani, B. D., ... De Alencar, S. M. (2020). Phenolic profile and potential beneficial effects of underutilized Brazilian native fruits on scavenging of ROS and RNS and anti-inflammatory and antimicrobial properties. *Food and Function*, 11(10), 8905–8917. <https://doi.org/10.1039/d0fo01763a>
- Stalmach, A., Edwards, C. A., Wightman, J. D., & Crozier, A. (2011). Identification of (Poly)phenolic compounds in concord grape juice and their metabolites in human plasma and urine after juice consumption. *Journal of Agricultural and Food Chemistry*, 59(17), 9512–9522. <https://doi.org/10.1021/jf2015039>

- Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., ... Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, 3(3), 211–221. <https://doi.org/10.1007/s11306-007-0082-2>
- Tahara, S. (2007). A Journey of Twenty-Five Years through the Ecological Biochemistry of Flavonoids. *Bioscience, Biotechnology, and Biochemistry*, 71(6), 1387–1404. <https://doi.org/10.1271/bbb.70028>
- Ticiane Carvalho Farias, Carolina Thomaz dos, Santos D’Almeida, Thaiza Serrano de Souza, Talita Pimenta do Nascimento, Fernanda de Sousa Bezerra, Roberta Nogueira Pereira da Silva, Mariana Simões Larráz Ferreira, Andrea Furtado Macedo, Maria Gabriela Bell, M. G. B. K. (2022). Application of Metabolomic Tools to Survey the Phenolic Composition of Food, Medicinal Plants, and Agro-Industrial Residues. In *Nutriomics* (p. 52).
- Tsao, R., Yang, R., Young, J. C., & Zhu, H. (2003). Polyphenolic profiles in eight apple cultivars using high-performance liquid chromatography (HPLC). *Journal of Agricultural and Food Chemistry*, 51(21), 6347–6353. <https://doi.org/10.1021/jf0346298>
- Tsugawa, H., Kind, T., Nakabayashi, R., Yukihiro, D., Tanaka, W., Cajka, T., ... Arita, M. (2016). Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Analytical Chemistry*, 88(16), 7946–7958. <https://doi.org/10.1021/acs.analchem.6b00770>
- Vaniya, A., Samra, S. N., Palazoglu, M., Tsugawa, H., & Fiehn, O. (2017). Using MS-FINDER for identifying 19 natural products in the CASMI 2016 contest. *Phytochemistry Letters*, 21, 306–312. <https://doi.org/10.1016/j.phytol.2016.12.008>
- Verma, B., Hucl, P., & Chibbar, R. N. (2009). Phenolic acid composition and antioxidant capacity of acid and alkali hydrolysed wheat bran fractions. *Food Chemistry*, 116, 947–954. <https://doi.org/10.1016/j.foodchem.2009.03.060>
- Vermerris, W., & Nicholson, R. (2006). *ISOLATION AND IDENTIFICATION OF PHENOLIC COMPOUNDS 151 ISOLATION AND IDENTIFICATION OF PHENOLIC COMPOUNDS A practical guide*. Retrieved from https://link.springer.com/content/pdf/10.1007/978-1-4020-5164-7_4.pdf
- Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., ... Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8), 828–837. <https://doi.org/10.1038/nbt.3597>
- Wang, X., Li, N., Chen, S., Ge, Y., Xiao, Y., Zhao, M., & Wu, J. (2022). *MS-FINDER Assisted in Understanding the Profile of Flavonoids in Temporal Dimension during the Fermentation of Pu-erh Tea*. <https://doi.org/10.1021/acs.jafc.2c01595>
- Wang, Y., Kora, G., Bowen, B. P., & Pan, C. (2014). MIDAS: A database-searching algorithm for metabolite identification in metabolomics. *Analytical Chemistry*, 86(19), 9496–9503. <https://doi.org/10.1021/ac5014783>

- Watrous, J., Roach, P., Alexandrov, T., Heath, B. S., Yang, J. Y., Kersten, R. D., ... Dorrestein, P. C. (2012). Mass spectral molecular networking of living microbial colonies. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), 1743–1752. <https://doi.org/10.1073/pnas.1203689109>
- Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., ... Dorrestein, P. C. (2013). Molecular networking as a dereplication strategy. *Journal of Natural Products*, *76*(9), 1686–1699. <https://doi.org/10.1021/np400413s>