中国科学: 化学 2022 年 第 52 卷 第 9 期: 1580 ~ 1591
SCIENTIA SINICA Chimica chemcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
评 述 庆祝汪尔康院士90华诞专刊

基于质谱数据的计算代谢组学方法学研究进展

杨军[1,2], 刘心昱[1,3*], 许国旺[1,2,3]

1. 中国科学院分离分析化学重点实验室, 中国科学院大连化学物理研究所, 大连 116023
2. 中国科学院大学, 北京 100049
3. 辽宁省代谢组学重点实验室, 大连 116023
*通讯作者, E-mail: liuxy2012@dicp.ac.cn

收稿日期: 2022-04-15; 接受日期: 2022-05-23; 网络版发表日期: 2022-08-29
国家重点研究计划(编号: 2019YFC1605100)、国家自然科学基金委(编号: 21876169, 21934006)和中国科学院青年创新促进会(编号: 2021186)资助项目

摘要 代谢组学旨在对代谢组进行全景的分析, 从而发现生物现象或规律. 由于代谢物的种类和数量都非常庞大, 对代谢组的分析高度依赖于分析仪器和方法. 质谱是代谢组分析最有用的工具, 超高效液相色谱-高分辨质谱可以从生物样品中获得大量的代谢物离子特征, 获得丰富的代谢组信息, 但如何对质谱数据进行挖掘和利用仍面临极大的挑战. 计算代谢组学可以充分利用质谱采集的数据, 结合统计、化学计量学、人工智能等方法实现对代谢组学数据的高效处理和分析, 推动代谢组学的发展. 本文在给出代谢组数据特点的基础上, 综述了数据驱动的计算代谢组学方法学进展, 包括特征提取、代谢物的注释和鉴定, 并简要介绍了知识辅助的计算代谢组学方法, 最后对计算代谢组学方法学下一步的发展进行了展望.

关键词 计算代谢组学, 质谱, 特征提取, 代谢物注释与鉴定

1 引言

代谢组学是继基因组学、转录组学、蛋白质组学等之后的一种新的“组学”技术, 旨在对代谢组进行全景的研究[1]. 代谢组是生物样本中小分子代谢物的集合, 人类代谢组由生物体中的内源性代谢物(如氨基酸、脂质、有机酸等)组成, 也受外源性化学物质(如食物、药物、环境污染物及其代谢产物等)影响[2]. 因此, 代谢组是基因和环境因素共同作用的结果, 对代谢组的分析可以全面揭示生物体在代谢和健康层面的信息. 代谢组学已广泛应用于疾病诊断和机理研究[3]、药物发现[4]、食品安全[5]、环境暴露[6]等众多领域. 代谢组的构成极为复杂, 根据基因和代谢反应的预测和估计, 人类代谢组大约有217920种代谢物[7]. 代谢物种类非常丰富, 理化性质各不相同. 代谢组的复杂性使得代谢组学与其他组学相比, 更加依赖于分离分析的仪器和方法[8].

质谱有着灵敏度高、检测范围宽、易于与分离仪器连接的特点, 高分辨质谱获得的精确离子质量更是有助于代谢物的鉴定, 因此质谱是代谢组学中使用最广泛的分析仪器[9]. 气相色谱-质谱联用(gas chromatography-mass spectrometry, GC-MS)与液相色谱-质谱联

引用格式: Yang J, Liu X, Xu G. New advances in mass spectrometry data-based computational metabolomics methods. *Sci Sin Chim*, 2022, 52: 1580–1591, doi: 10.1360/SSC-2022-0084

用(liquid chromatography-mass spectrometry, LC-MS)是常用的代谢组学分析技术. 代谢物在色谱中的保留时间(retention time, tr)、质谱中的*m*/*z*值及响应强度构成了LC-MS或GC-MS的谱图[10].

基于质谱的代谢组学分析流程包括样品采集和预处理、仪器分析、数据处理和分析等步骤[11], 数据处理和分析是从仪器采集的数据获得生物问题解释的重要过程, 包括原始数据预处理、代谢物鉴定、单变量和多变量统计模型构建等[12], 许多化学计量学方法被应用到其中. 从仪器上导出的原始数据包含各个样品在不同保留时间下的离子质荷比和强度, 它的预处理一般包括噪音过滤、基线校正、峰检测、去卷积、峰对齐等, 每个步骤均有多种算法和工具实现自动处理[13], 用户可通过相关软件或工具输入参数获得简洁、清晰的峰表. 对数据进行代谢物鉴定是至关重要的一步, 目前最可靠的代谢物鉴定方法仍是通过数据库中谱图检索完成[14]. 经过预处理和鉴定后得到的数据是关于样本和代谢物的矩阵, 因此构建统计或机器学习模型是解释数据、发现代谢模式的基本方法[15]. 由于代谢模式的复杂性, 一些非线性的机器学习方法也可用于代谢组学数据分析并提高模型的准确率[16]. 限于篇幅, 单变量统计分析和多变量分类分析部分不在此文展开.

根据分析对象和研究目的不同, 基于不同质谱仪器平台的众多分析方法不断被开发出来[17], 仪器性能的提高导致代谢组学的分析精准度和覆盖度提升, 也使得代谢组学数据的维度和复杂程度不断上升. 通过高分辨质谱, 一个生物样品就可能会产生数万到十几万个离子特征(feature)[18]. 为了充分挖掘其中的信息, 近年来, 许多基于质谱数据的计算代谢组学方法和工具被提出. 计算代谢组学方法通过统计、化学计量学或机器学习原理等, 充分利用质谱数据, 实现对数据自动或半自动的处理分析. 目前计算代谢组学的主要方向是从数据中分辨信号和噪音从而准确地提取代谢物的特征, 进一步根据谱图数据对代谢物进行鉴定或注释[19]. 因此计算代谢组学方法对推动代谢组学的发展与应用有着重要意义.

本文综述了近年来计算代谢组学方法的研究进展. 一方面, 代谢组学数据作为医学大数据的一种, 数据驱动的各种算法是挖掘其中信息的重要手段. 另一方面, 充分利用代谢组学数据库、代谢物在其他维度的化学信息也是分析代谢组学数据挖掘的方向. 因此下文将从数据驱动和知识辅助的计算代谢组学方法两方面分别展开介绍.

2 数据驱动的计算代谢组学方法

2.1 代谢物碎片特征提取

从GC/MS或LC/MS质谱数据中提取代谢物碎片特征(feature)是进行后续数据分析的前提. 数据中的质谱特征有三个来源, 分别是样本中实际的代谢物、试剂中杂质和仪器残留、噪音[20]. 目前特征提取的主要方法是根据信噪比的阈值去除强度较低的信号, 并通过各种算法实现对峰形的约束. 例如, XCMS的特征提取原理是根据密度确定感兴趣的质荷比区域(regions of interest, ROI), 再通过连续小波变换(continuous wavelet transform, CWT)解析色谱峰[21]. Mzmine 2则提供了多种算法用于不同类型的质谱数据, 如局部极值法(local maxima)、递归阈值法(recursive threshold)、精确质量法(exact mass)等[22]. 但是XCMS和Mzmine 2均会产生较多的假阳性特征, 且二者获得的峰表也存在显著差异[23]. Ju等[20]提出应用信息熵和统计相关性的方法去除假阳性特征, 在实际尿样的测试中, 将SIEVE软件提取的7182个特征减少到2522个, 而98%的已鉴定代谢物的特征则保留了下来. 如图1所示, 对于特定的提取离子流图(extracted ion chromatogram, EIC), 最清晰的代谢物特征峰形应该呈正态分布, 但随着噪音的增强, 正态分布逐渐转化为无规则的非正态分布. 因此该研究提出了熵指数(entropy index)的概念, 可根据EIC中相对最高点高度算出, 如果是理想的正态分布, 熵指数会等于0, 熵指数越大, 表示EIC谱越杂乱, 是噪音的可能性更大. 同时, 对于真的代谢物特征, 不同样品间的EIC谱应该具有较高的统计相关性, 在优化了计算参数和阈值后, 可以根据熵指数和相关系数依次去除假阳性峰. Samanipour等[24]提出一种自调整的特征提取方法, 与XCMS等对Centroid模式数据处理不同, 它对质谱数据进行三维高斯拟合获取特征, 可以充分利用数据中的每个点, 提高特征检测的准确度, 而且不需要进行像XCMS中的参数优化. 不足之处是由于该方法需要使用特征中的每个点并拟合, 所以它的时间复杂度要高于XCMS, 计算耗时更久.

提高低丰度信号覆盖度是特征提取的另一关键科

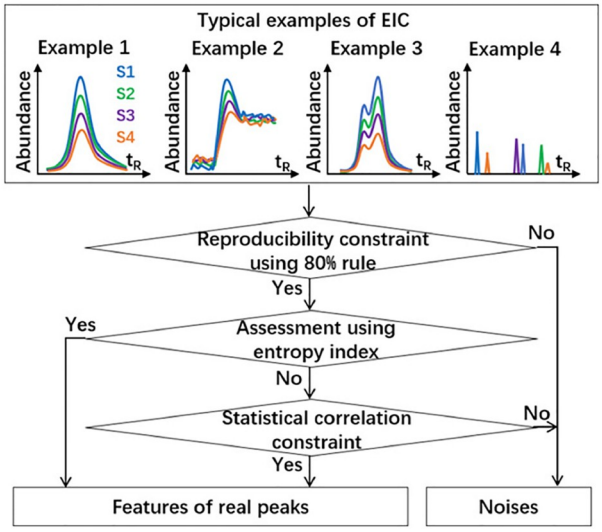

图 1 对不同情形EIC特征提取的流程[20] (网络版彩图)

Figure 1 Workflow of feature detection in different situations of EIC [20] (color online).

学问题. 不同算法从不同角度获得质谱数据中的代谢物特征, 因此, 把多个算法识别的特征进行整合有利于低丰度代谢物信息的获取. 但问题是这些不同算法识别的特征存在冗余, 去除冗余就成为此种方法获得较好效果的核心问题. 为此, Ju等[25]提出了一种基于图密度的方法整合不同峰检测软件的结果并去除冗余. 特征被看作图的节点, 将其在保留时间和*m*/*z*二维空间中的欧氏距离设为节点距离, 距离小于阈值的节点被连上一条边. 在构建了所有特征的图后, 检索图中的极大完全子图(maximal complete sub-graph)合并成一个特征, 从而实现了充分保留"真"信息、去除冗余的目的.

另一方面, 特征提取还需要考虑代谢物发生源内裂解(in-source fragmentation, ISF)、形成加合离子和同位素离子的问题, 由于这个原因, 一种代谢物可产生多个特征信号, 增加了代谢组学数据的复杂性, 所以要对加合离子等进行融合. Guo等[26]提出了基于LC-MS数据的ISF识别工具ISFrag, 它根据ISF离子的3个特点进行识别, 分别是ISF离子和母离子保留时间一致、会在母离子的二级谱图上出现、与母离子的二级谱图较为相似, ISFrag可以不借助任何数据库信息实现对多种获取模式下质谱数据的处理.

近年来, 深度学习在图像分类、物体识别等计算机视觉方面取得了很大的进步[27], 而在谱图中的特征检测也是一个与物体识别或分类较为类似的任务, 因此深度学习方法也能够应用到质谱代谢组学的特征检测中. Melnikov等[28]提出了Peakonly算法, 通过构建卷积神经网络(convolutional neural networks, CNNs)实现精确的峰检测和积分. 该算法分为三步, 分别是ROI的检测、ROI的分类、峰检测及积分. ROI的检测方法与XCMS方法类似, 随后根据峰形将ROI进行分类, 并通过约4000个人工分类的ROI训练了一个卷积神经网络. 由于网络构建前已将每个ROI的最大强度标度到相同强度, 因此ROI的分类只会根据峰形判断, 而不会考虑强度的因素, 这有助于发现低丰度的特征. 该分类在测试集中取得了87%的准确率, 误分类的情况主要是一些人工也难以区分的特征, 因此模型对ROI的分类效果是较好的. 在ROI分类后, 为了识别出ROI中的峰从而准确积分, 他们又应用图像分割的思路构建了一个类似于U-Net (一种可以快速准确的实现图像分割的卷积神经网络算法)的CNN模型, 实现了较好的峰检测和积分效果. Gloaguen等[29]提出了基于深度学习的特征提取工具(NeatMS), 通过预训练构建了初步的CNN模型, 通过迁移学习(transfer learning)可以扩展该模型, 从而使其适用于不同的LC-MS分析流程.

2.2 代谢物注释与鉴定

2.2.1 代谢物注释

在获得质谱代谢特征的峰表后, 下一步是要挖掘其中包含的代谢组信息. 从质谱数据中准确鉴定出代谢物(metabolite identification)是代谢组学研究的瓶颈[30]. 代谢物的注释(metabolite annotation)和鉴定并不完全相同, 代谢物鉴定的目标是得到化合物的准确结构, 而代谢物注释倾向于获得化合物的部分信息(如分子式、类别、官能团等)[19]. 虽然理论上代谢物注释要比代谢物鉴定简单, 但要对大量未知代谢物获得准确信息, 仍然是很大的挑战.

分子式的注释是代谢物结构鉴定的前提. 目前许多结构鉴定的软件都是通过SIRIUS的同位素模式匹配和二级质谱分析模块获得分子式, 从而进一步推导出结构[31]. SIRIUS的不足之处在于当化合物的尺寸增大时, 不同原子组合的可能性更多, 因此它对分子量大于500 Da的化合物的错误率较高[32]. 为了克服这一问题, 2020年, Ludwig等[33]提出了一种不依赖于数据库的分子式注释方法ZODIAC. 该方法通过贝叶斯统计

和Gibbs采样对SIRIUS给出的前50名候选分子式进行重新排名, 由于考虑了不同分子式同时出现的概率, 因此比SIRUS提升了注释的准确性.

在生命体中, 许多在相近通路上的代谢物会拥有共同的部分结构, 这样的共同结构被称为子结构(substructure)[34]. 获得子结构要比获得完整的分子结构容易, 并且在一些生物学问题上, 子结构信息就足以解决问题. 代谢物的特定化学修饰也会产生特定的子结构, 如乙酰化、硫酸化、葡萄糖醛酸化等, 因此可以辅助代谢物结构的鉴定[35,36], 从而大大提高了代谢物结构鉴定的准确度和覆盖度. 因此, 从质谱数据中挖掘代谢物的子结构是一种有前景的分析新视角.

Ma等[37]在2014年提出的MS2Analyzer就是一个对化合物子结构进行注释的工具, 它可以根据母离子和子离子m/z差(即中性丢失)得到对应的丢失基团(如NH_3、H_2O、CO等). 但MS2Analyzer的参数依赖于人工设定, 找到的基团在大多数化合物中均较为常见, 而部分样品中的特有基团则容易被忽视. Van der Hooft等[38]在2016年提出的MS2LDA是一种无监督的子结构分析方法, 将来源于文本挖掘的隐迪利克雷分配模型(latent Dirichlet allocation, LDA)算法用于处理代谢组学数据. 该方法的基础是代谢物子结构在碎裂谱图中有着较为保守的碎片离子和中性丢失. 在MS2LDA算法中, 首先通过XCMS和MzMatch进行峰检测, 通过RMassBank检测MS1-MS2离子对, 从而得到一个以MS1为列、以MS2特征(碎片和中性丢失)为行、以MS2特征相对强度为数值的矩阵. 通过LDA推断处理该矩阵就可以得到各MS1的Mass2Motifs, 根据Mass2Motifs在各样本中的分布情况就可以解释样本的代谢和通路变化. 为了对大量样本的非靶向代谢组学数据进行Mass2Motifs的同时分析, Van der Hooft等[39]又将MS2LDA扩展成MS2LDA+ (图2), 可以自动快速地获得多个样本中Mass2Motifs变化. 由于许多Mass2Motifs在实验中出现的概率较高, 但其对应的结构一般还需要人工判断, 因此MS2LDA又引入了MotifDB数据库, 用于存储已经注释的Mass2Motifs, 并整合MAGMa和ClassyFirePredict两种工具用于子结构的自动注释[40]. 除了MS2Analyzer和MS2LDA子结构分析工具外, Liu等[41]在2020年提出了MESSAR (metabolite sub-structure auto-recommender)工具, MESSAR基于关联规则挖掘(association rule mining, ARM)原理, 从GNPS谱图数据库中学习“从MS/MS到子结构”的规则, 进而给未知质谱谱图推荐可能的子结构.

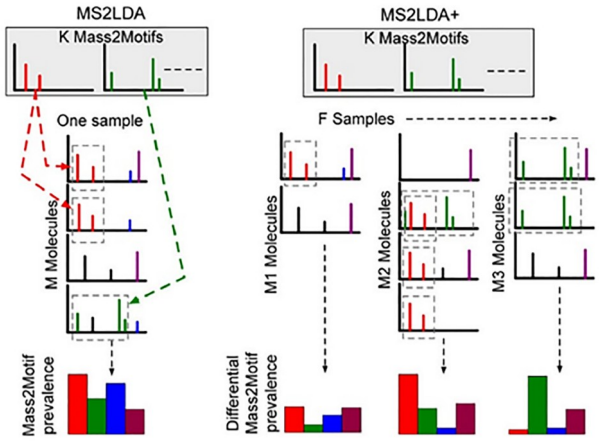

图 2　从MS2LDA到MS2LDA+的扩展[39] (网络版彩图)

Figure 2　Extension from MS2LDA to MS2LDA+ [39] (color online).

代谢物的分类往往可以预测其在人体的代谢命运, 因此对代谢物进行类别的注释也是一个较为关注的方向. 类别的注释与子结构注释较为类似, 因为同一类别的代谢物往往有着共同的特定子结构, 但化合物的分类需要依据一定的化学分类标准[34]. 由于代谢物种类和结构较为复杂, 手动分类较为困难, Feunang等[42]在2016年提出了ClassyFire分类工具, 可以根据结构给出代谢物依次分层的类别. 但ClassyFire并不能直接根据质谱谱图得到类别注释, 为了实现MS/MS数据的化合物类别注释, Dührkop等[43]提出了CANOPUS工具. 当输入一个MS/MS谱图, CANOPUS首先将该谱图通过核SVM计算出预测的分子指纹, 再将分子指纹输入到深度神经网络(DNN)模型中, 从而输出化合物的类别. 因此核SVM和DNN的训练是预测的关键. CANOPUS中SVM是通过MS/MS谱图数据库训练得到, DNN是通过庞大的化合物结构数据库训练得到, 两种模型的结合使得CANOPUS可以对2497种ClassyFire类别进行高质量的预测. 分子网络(molecular networking, MN)也是代谢物注释的一种方法, 由Watrous等[44]在2012年首次提出. 分子网络通过计算MS/MS谱图的相似性, 可以构建一个以谱图(或代谢物)为节点、相似程度为边的网络. 构建分子网络可以更加全面地考虑谱图之间的联系, 而不是单独地分析每一个谱图, 并且谱图的相似反映了化合物结构的相似. 构建大规模

的分子网络有助于质谱数据和信息的共享，推动代谢组学及其他天然产物质谱分析的研究.

全球天然产物分子网络(global natural products social molecular networking, GNPS, https://gnps.ucsd.edu/)公共平台[45]中有两种构建分子网络的方法，分别是经典分子网络(classical MN)和基于特征的分子网络(feature-based MN, FBMN). 经典分子网络先对样品中的二级谱图通过MS-Cluster算法合并来自同一化合物的谱图, 对于合并后的谱图, 需要根据各谱图母离子的质量差进行校正，从而可以正确计算谱图余弦相似性，并构建分子网络[46] (图3a). FBMN对经典方法进行了补充, 对LC-MS/MS中的特征构建分子网络, 可以区分不同保留时间的异构体，通过可视化的分子网络辅助代谢物的注释[47,48]. Yu等[49]将分子网络分析方法用于药物代谢研究，比较了经典分子网络、FBMN和其他代谢物检测工具对质谱数据的分析结果，分子网络方法的高效率和聚类可视化功能使其成为一种较有前景的药物代谢研究工具. 将分子网络与*in silico*谱图预测结合可以提升分子网络的注释效率. Da Silva等[50]在2018年提出了网络注释传播(network annotation propagation, NAP)方法(图3b~g). NAP首先对实验MS/MS数据构建分子网络，再通过谱图数据库搜索出可以确定结构的节点, 对于未知节点则通过MetFrag(一种*in silico*结构注释工具)给出候选结构及打分，最后通过网络中谱图的关联对打分进行调整，调整后的打分可提高注释的准确性，减少人工检查的负担. Ernst等[51]在

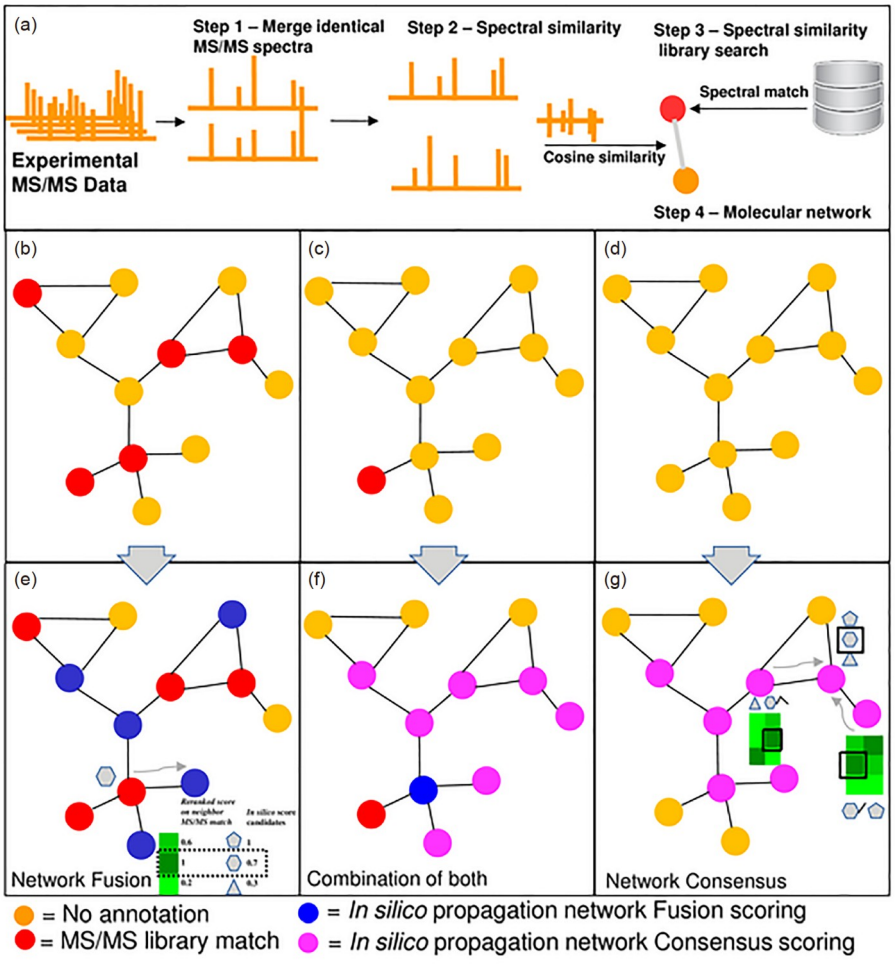

图 3 从质谱数据构建分子网络的过程(a)以及网络传播注释方法(b~g)示意图[50] (网络版彩图)

Figure 3 The process for creating a molecular network from MS/MS data (a) and network annotation propagation illustration (b–g) [50] (color online).

2019年提出了MolNetEnhancer工具，将分子网络、MS2LDA、NAP、Classify等多种质谱数据分析工具整合, 从而增加了代谢物化学信息的覆盖度. 这也使得MolNetEnhancer在实际样品中的代谢物注释能力大大提高[52]，对于天然产物的研究等领域有着广泛的应用前景[53].

2.2.2 代谢物鉴定

目前，代谢物的鉴定通常依赖于标准谱图库的检索，即将实验谱图与标准谱图比较并计算相似性分数, 根据分值确定代谢物结构[54]. GC-MS中代谢物鉴定一般通过检索电子电离(electron ionization, EI)的标准谱图, 由于EI的重复性好、标准化程度高, 因此有较为丰富的谱图数据库可以检索[55]. LC-MS以电喷雾电离(ESI)等软电离源为主，再通过碰撞诱导解离(collision-induced dissociation, CID)技术获得二级离子碎片或多级离子碎片，通过MS/MS或MS/MS″谱图的检索来鉴定代谢物[56]. 但通过实验方法构建谱图库较为复杂和耗时, 并且由于数据库内容的有限性, 目前, 在非靶向代谢组学实验中，只有1.8%的谱图可被注释[57]. 因此近年来许多研究都在关注通过对质谱数据的解析直接获得分子结构.

从质谱数据中推导出化合物结构通常也被叫做*de novo*分析[58]. 目前主要的思路是通过CID质谱数据计算碎片树(fragmentation tree), 树的节点是质谱图中的离子, 有向边代表离子之间的反应或转化过程, 因此可以通过碎片树的比较得到可能性最大的化合物结构[59]. 但从碎片树不能直接得到分子结构，一般是通过机器学习方法得到分子指纹，分子指纹是每一位代表一定的化合物子结构信息的矢量[60]. 通过分子指纹可以实现在化合物结构数据库中检索，而不是在数量非常有限的CID数据库中检索. 例如, Shen等[61]提出了多种核函数用于获得碎片树的相似性，并通过多核学习(multiple kernel learning, MKL)增加碎片树对分子指纹的预测能力(图4). 后来，该课题组[62]在此基础上又引入了多个核函数，并对分子指纹中每个分子属性训练单独的支持向量机(support vector machine, SVM), 从而更加准确地计算和比较分子指纹，发展出了CSI:FingerID结构鉴定工具. 为了对候选结构分子指纹打分过程进行优化，Ludwig等[63]又提出利用分子属性的联合概率计算打分，即不再假定分子指纹中各属性的独立性，通过训练数据集构建贝叶斯网络，根据各属性的边缘概率和协方差估计联合概率，从而计算打分，提升了结构检索的效果. 2019年，Dührkop等[31]将CSI:FingerID整合入SIRIUS 4，使得SIRIUS 4同时具有高分辨同位素模式分析和碎片树结构分析功能, 并能够实现对大规模LC-MS数据的自动化分析，用户可通过清晰的图形界面进行对应的操作. 由于多核学习方法预测分子指纹时是对不同的分子属性分别构建分类器[62]，这样可能会丢失分子指纹中隐藏的一些结构信息，因此，Fan等[64]提出了MetFID，利用二级质谱数据构建人工神经网络(artificial neural network, ANN)并预测分子指纹. 该ANN模型将二级质谱数据转化后得到的950维矢量作为输入层, 最终输出层可直接得到528维的分子指纹. 经过独立数据集的比较, MetFID的预测准确率比CSI:FingerID等工具提升了约5%. 类似地, Ji等[65]提出了基于深度神经网络的DeepEI工具, 可实现从EI质谱数据预测分子指纹，并根据分子指纹从分子结构数据库中检索出对应的结构.

除了从质谱数据推导化合物结构的*de novo*分析方法外，还可以利用化合物结构预测谱图并进一步用于结构鉴定. 直接从化合物结构预测其碎裂必然会涉及分子自身的热力学和动力学性质，因此量子化学的理论可以被用于预测化合物的质谱谱图，且不需要依赖于数据库[66]. 量子化学方法中比较有代表性的是Grimme等[67]提出的QCEIMS (quantum chemistry electron ionization mass spectrometry)方法, 该方法将波恩-奥本海默分子动力学和统计采样结合，用于预测分子的EI谱图，现在该方法也被进一步发展到可预测CID谱图[68]. 除了量子化学方法外，机器学习方法也被用于预测谱图上. 例如，2012年，Kangas等[69]就提出了MetISIS (*in silico* identification software, ISIS)用于预测脂质的碎裂谱图，该算法的原理是通过动力学蒙特卡罗(kinetic Monte Carlo, KMC)模拟离子在线性离子阱中的行为，并通过人工神经网络预测断键能量从而产生CID谱图，根据计算得到的谱图与实验谱图的相似性，通过遗传算法(genetic algorithm)不断优化神经网络的权重, 提升谱图的准确性.

另外一个目前较受欢迎的谱图预测工具是Wishart实验室在2014年提出的CFM-ID[70]. CFM-ID是一个界面简洁、便于使用的网络服务器，可快速实现化合物的QToF谱图预测、二级谱峰注释和化合物鉴

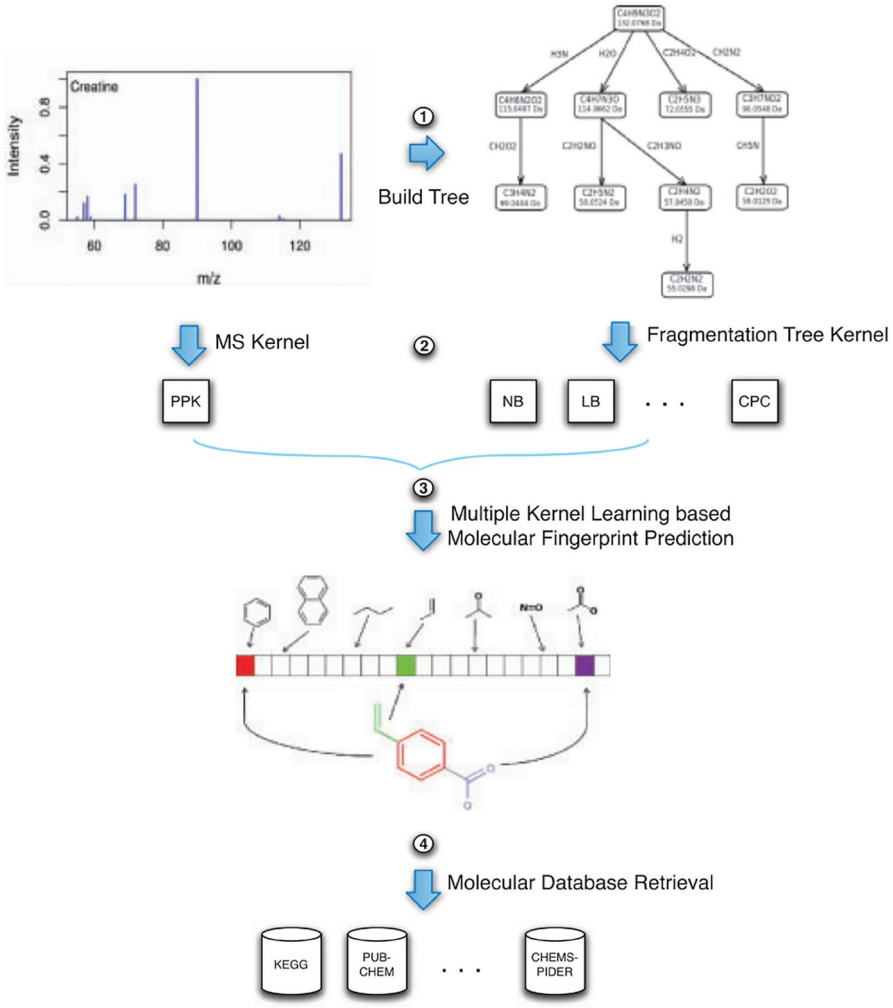

图 4 多核学习鉴定代谢物的流程示意图[61] (网络版彩图)
Figure 4 The metabolite identification framework through MKL [61] (color online).

定等功能. CFM-ID的谱图预测是基于一种竞争性碎裂模型(competitive fragmentation modeling, CFM), 这是一种基于概率计算的生成式模型[71]. 通过训练数据集可以学习CFM模型中的概率参数, 从而计算出离子出现概率, 预测出给定结构的CID谱图. Wishart实验室[72]仍在不断地对CFM-ID进行更新和升级, 2019年提出的CFM-ID 3.0补充了基于碎裂规则的脂质二级谱图预测方法, 提升了CFM-ID对脂质谱图的预测能力, 同时也在化合物鉴定和打分、分类算法上做出了改进. 最新的版本是2021年提出的CFM-ID 4.0 (https://cfmid.wishartlab.com/), 通过一种新颖的张量表示方法描述分子中化学键的拓扑结构, 引入拓扑信息后的机器学习模型可以有更好的预测效果, 同时他们还提出了处理环断裂的新方法, 使得对环状化合物的预测能力也显著提升[73]. 与量子化学方法不同的是, 机器学习方法的效果较为依赖训练数据集和碎裂过程的模型, 复杂的碎裂模型有利于提高准确度, 但会增加计算复杂性, 因此提升准确度和计算效率仍是谱图预测的发展方向.

3 知识辅助的计算代谢组学方法

计算代谢组学方法的发展推动了对代谢组学数据更加全面的解释, 从数据中获得的信息也更加全面. 但

是这些信息仍然存在偏差或错误的可能，准确定性代谢物是代谢组学中研究的瓶颈. 即使基于最小的质量误差去匹配，也不能保证定性结果的准确度，当对许多代谢物做这种匹配时，定性错误会传递到后面的分析中[74]. 因此，充分利用现有的知识是减少错误、提升信息质量的重要手段，这里简要介绍代谢组学数据库和代谢物其他维度的化学信息对代谢组学发展的推动的作用.

“组学”的发展和应用需要有信息全面的数据库支撑，基因组学和蛋白质组学、转录组学等都早已形成了较为完整的数据库供研究者使用，但由于代谢物的复杂性, 代谢组学的数据库仍在不断完善[8]. 目前代谢组学数据库主要包括人类代谢组数据库(如HMDB)、质谱谱图数据库(如NIST、METLIN、MassBank和mzCloud)和代谢通路数据库(如KEGG、MetaCyc和Reactome)[14], 这些数据库中的信息对代谢组学中各个步骤都存在一定的价值. 例如, Stancliffe等[75]提出一种基于谱图数据库的去卷积方法DecoID, 通过LASSO回归将混合谱图解卷积成数据库谱图的线性组合，该方法提高了去卷积的通量，也克服了传统去卷积方法对谱图获取方式的依赖. 代谢通路数据库中大量的生化反应可以帮助完成代谢物的注释. 例如，朱正江等[76]提出将基于代谢反应网络(metabolic reaction network, MRN)的递归算法(metabolite annotation and dysregulated network analysis, MetDNA)用于非靶向代谢组学的大规模代谢物注释，在一次实验中可以注释约2000种代谢物. 该方法首先根据KEGG构建MRN，即代谢物节点通过反应相互连接的网络，然后通过实验MS/MS谱图检索样品中的代谢物, 在网络中找到这些代谢物的位置, 再根据m/z、预测保留时间、二级谱图等检验其相邻节点代谢物是否存在, 这样递归检索, 直到没有新的代谢物被注释为止.

在代谢物的鉴定方面，代谢物在其他维度的化学信息也可以提高鉴定效率和可靠度. 在LC-MS中，预测代谢物在特定色谱分离条件下的保留时间可以减少质谱谱图检索中假阳性的结果[77]. 由于同一物质在不同色谱分离条件下tr并不相同，并且已知tr的代谢物数量较少，因此深度学习、迁移学习等方法也被用于tr预测[78,79]. 例如, Yang等[80]通过图神经网络(graph neural network, GNN)预测保留时间, 将小分子的结构转化为分子图后进行GNN的学习，分子图的预测性能比传统的分子描述符方法更好. 此外，预训练(pre-training)深度学习模型、再使用其他条件下数据微调(fine-tuning)的迁移学习策略也是一种较为有效的预测方法[81]. 除了保留时间，一些新型化学测量仪器可以向代谢组学引入新的测量维度，辅助代谢物的鉴定，但新维度的引入也要求建立相应的标准品数据库或*in silico*计算方法. 例如, 离子淌度质谱(ion mobility-mass spectrometry, IM-MS)可以根据代谢物离子的碰撞截面积(collision cross-section, CCS)区分生物样品中普遍存在的异构体, CCS、tr、多级质谱数据等多维数据整合后可以有效提高代谢物鉴定覆盖度和准确度[82]. 红外离子谱(infrared ion spectroscopy)可以提供经质谱质量选择后离子的红外振动指纹谱，从而反映离子的结构信息[83].

4 总结与展望

作为系统生物学的一部分，代谢组学和基因组学、转录组学、蛋白质组学等共同反映着生物体的生命健康状况，组学数据的充分挖掘将极大地推动精准医学的发展. 然而代谢组的复杂性限制了其发展和应用. 计算代谢组学方法可以优化代谢组学分析流程，实现数据的自动化处理和信息的充分挖掘，是代谢组学向大规模临床应用的重要基础.

计算代谢组学方法早期以化学计量学方法为主，后来一些机器学习方法也被用于降维、聚类或构建预测模型等. 近年来深度神经网络的发展为基于大规模数据的计算代谢组学方法学研究带来了新机遇，使数据处理速度和准确度都有了很大提升. 本文将计算代谢组学方法分为数据驱动和知识辅助两类，并根据目的或方法进行了进一步的分类，但这些类别并不是完全独立的. 由于特定的原理或模型往往只有在特定的数据中才有最优效果，因此大多数已有的方法都会存在一定的缺点. 此外，代谢组中有大量的低丰度代谢物，在生命过程中也起着重要作用，但检测出低丰度的离子特征、对这些特征进行结构的鉴定仍然具有较大的困难. 期望后续先进的计算代谢组学方法能解决这些问题, 并推动生物学研究.

今后计算代谢组学方法的发展主要有三个方向: 一是对现有方法进行改进, 不断克服其缺点; 二是通过集成现有方法结合各方法的优点，从而扩展其适用问

题的范围; 三是基于新的原理提出新方法. 近年来大数据分析越来越受到人们的关注, 大数据领域的发展也必然会推动计算代谢组学方法的发展. 随着计算方法性能的提升, 代谢组学数据也将给人们提供更加丰富、可靠的代谢信息, 从而有力地推动生物学研究和精准医学的发展.

参考文献

1 Nicholson JK, Lindon JC, Holmes E. *Xenobiotica*, 1999, 29: 1181–1189
2 Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. *Nucl Acids Res*, 2013, 41: D801–D807
3 Ghanem HZ, Kadry MO, Abdel-Megeed RM, Abdel-Hamid AHZ. *Egypt Pharmaceut J*, 2019, 18: 290–295
4 Wishart DS. *Nat Rev Drug Discov*, 2016, 15: 473–484
5 Li S, Tian Y, Jiang P, Lin Y, Liu X, Yang H. *Crit Rev Food Sci Nutr*, 2021, 61: 1448–1469
6 You L, Zheng F, Su C, Wang L, Li X, Chen Q, Kou J, Wang X, Wang Y, Wang Y, Mei S, Zhang B, Liu X, Xu G. *Environ Int*, 2022, 158: 106919
7 Wishart DS, Guo AC, Oler E, Wang F, Anjum A, Peters H, Dizon R, Sayeeda Z, Tian S, Lee BL, Berjanskii M, Mah R, Yamamoto M, Jovel J, Torres-Calzada C, Hiebert-Giesbrecht M, Lui VW, Varshavi D, Varshavi D, Allen D, Arndt D, Khetarpal N, Sivakumaran A, Harford K, Sanford S, Yee K, Cao X, Budinski Z, Liigand J, Zhang L, Zheng J, Mandal R, Karu N, Dambrova M, Schiöth HB, Greiner R, Gautam V. *Nucl Acids Res*, 2022, 50: D622–D631
8 Wishart DS. *Briefings Bioinf*, 2007, 8: 279–293
9 Roca M, Alcoriza MI, Garcia-Cañaveras JC, Lahoz A. *Anal Chim Acta*, 2021, 1147: 38–55
10 Alonso A, Marsal S, JuliÃ A. *Front Bioeng Biotechnol*, 2015, 3: 23
11 Pezzatti J, Boccard J, Codesido S, Gagnebin Y, Joshi A, Picard D, González-Ruiz V, Rudaz S. *Anal Chim Acta*, 2020, 1105: 28–44
12 Yi L, Dong N, Yun Y, Deng B, Ren D, Liu S, Liang Y. *Anal Chim Acta*, 2016, 914: 17–34
13 Castillo S, Gopalacharyulu P, Yetukuri L, Orešič M. *Chemometrics Intelligent Laboratory Syst*, 2011, 108: 23–32
14 Yi Z, Zhu Z. *Computational Methods and Data Analysis for Metabolomics*. Totowa: Humana Press Inc., 2020
15 Blaise BJ, Correia GDS, Haggart GA, Surowiec I, Sands C, Lewis MR, Pearce JTM, Trygg J, Nicholson JK, Holmes E, Ebbels TMD. *Nat Protoc*, 2021, https://doi.org/10.1038/s41596-021-00579-1
16 Chen N, Wang HB, Wu BQ, Jiang JH, Yang JT, Tang LJ, He HQ, Linghu DD. *Talanta*, 2021, 235: 122720
17 González-Riano C, Dudzik D, Garcia A, Gil-de-la-Fuente A, Gradillas A, Godzien J, López-Gonzálvez Á, Rey-Stolle F, Rojo D, Ruperez FJ, Saiz J, Barbas C. *Anal Chem*, 2020, 92: 203–226
18 Jones DP. *Toxicol Rep*, 2016, 3: 29–45
19 Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. *Chem Res Toxicol*, 2016, 29: 1956–1975
20 Ju R, Liu X, Zheng F, Zhao X, Lu X, Zeng Z, Lin X, Xu G. *Anal Chim Acta*, 2019, 1067: 79–87
21 Tautenhahn R, Böttcher C, Neumann S. *BMC BioInf*, 2008, 9: 504
22 Pluskal T, Castillo S, Villar-Briones A, Oresic M. *BMC BioInf*, 2010, 11: 11
23 Myers OD, Sumner SJ, Li S, Barnes S, Du X. *Anal Chem*, 2017, 89: 8689–8695
24 Samanipour S, O′Brien JW, Reid MJ, Thomas KV. *Anal Chem*, 2019, 91: 10800–10807
25 Ju R, Liu X, Zheng F, Zhao X, Lu X, Lin X, Zeng Z, Xu G. *Anal Chim Acta*, 2020, 1139: 8–14
26 Guo J, Shen S, Xing S, Yu H, Huan T. *Anal Chem*, 2021, 93: 10243–10250
27 Liu Y, Cheng Y, Wang W. A Survey of the Application of Deep Learning in Computer Vision. In: *Global Intelligent Industry Conference (GIIC)*. Beijing: Spie-Int Soc Optical Engineering, 2018
28 Melnikov AD, Tsentalovich YP, Yanshole VV. *Anal Chem*, 2020, 92: 588–592
29 Gloaguen Y, Kirwan JA, Beule D. *Anal Chem*, 2022, 94: 4930–4937
30 Neumann S, Böcker S. *Anal Bioanal Chem*, 2010, 398: 2779–2788
31 Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, Meusel M, Dorrestein PC, Rousu J, Böcker S. *Nat Methods*, 2019, 16: 299–302
32 Böcker S, Dührkop K. *J Cheminform*, 2016, 8: 26

33 Ludwig M, Nothias LF, Dührkop K, Koester I, Fleischauer M, Hoffmann MA, Petras D, Vargas F, Morsy M, Aluwihare L, Dorrestein PC, Böcker S. *Nat Mach Intell*, 2020, 2: 629–641

34 Beniddir MA, Kang KB, Genta-Jouve G, Huber F, Rogers S, van der Hooft JJJ. *Nat Prod Rep*, 2021, 38: 1967–1993

35 Dai W, Yin P, Zeng Z, Kong H, Tong H, Xu Z, Lu X, Lehmann R, Xu G. *Anal Chem*, 2014, 86: 9146–9153

36 Zheng S, Zhang X, Li Z, Hoene M, Fritsche L, Zheng F, Li Q, Fritsche A, Peter A, Lehmann R, Zhao X, Xu G. *Anal Chem*, 2021, 93: 10916–10924

37 Ma Y, Kind T, Yang D, Leon C, Fiehn O. *Anal Chem*, 2014, 86: 10724–10731

38 van der Hooft JJJ, Wandy J, Barrett MP, Burgess KEV, Rogers S. *Proc Natl Acad Sci USA*, 2016, 113: 13738–13743

39 van der Hooft JJJ, Wandy J, Young F, Padmanabhan S, Gerasimidis K, Burgess KEV, Barrett MP, Rogers S. *Anal Chem*, 2017, 89: 7569–7577

40 Rogers S, Ong CW, Wandy J, Ernst M, Ridder L, van der Hooft JJJ. *Faraday Discuss*, 2019, 218: 284–302

41 Liu Y, Mrzic A, Meysman P, De Vijlder T, Romijn EP, Valkenborg D, Bittremieux W, Laukens K. *PLoS ONE*, 2020, 15: e0226770

42 Feunang Y D, Eisner R, Knox C, Chepelev L, Hastings J, Owen G, Fahy E, Steinbeck C, Subramanian S, Bolton E, Greiner R, Wishart DS. *J Cheminform*, 2016, 8: 20

43 Dührkop K, Nothias LF, Fleischauer M, Reher R, Ludwig M, Hoffmann MA, Petras D, Gerwick WH, Rousu J, Dorrestein PC, Böcker S. *Nat Biotechnol*, 2021, 39: 462–471

44 Watrous J, Roach P, Alexandrov T, Heath BS, Yang JY, Kersten RD, van der Voort M, Pogliano K, Gross H, Raaijmakers JM, Moore BS, Laskin J, Bandeira N, Dorrestein PC. *Proc Natl Acad Sci USA*, 2012, 109: E1743–E1752

45 Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu WT, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, Boya P CA, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard PM, Phapale P, Nothias LF, Alexandrov T, Litaudon M, Wolfender JL, Kyle JE, Metz TO, Peryea T, Nguyen DT, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N. *Nat Biotechnol*, 2016, 34: 828–837

46 Aron AT, Gentry EC, McPhail KL, Nothias LF, Nothias-Esposito M, Bouslimani A, Petras D, Gauglitz JM, Sikora N, Vargas F, van der Hooft JJJ, Ernst M, Kang KB, Aceves CM, Caraballo-Rodríguez AM, Koester I, Weldon KC, Bertrand S, Roullier C, Sun K, Tehan RM, Boya P. CA, Christian MH, Gutiérrez M, Ulloa AM, Tejeda Mora JA, Mojica-Flores R, Lakey-Beitia J, Vásquez-Chaves V, Zhang Y, Calderón AI, Tayler N, Keyzers RA, Tugizimana F, Ndlovu N, Aksenov AA, Jarmusch AK, Schmid R, Truman AW, Bandeira N, Wang M, Dorrestein PC. *Nat Protoc*, 2020, 15: 1954–1991

47 Nothias LF, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A, Protsyuk I, Ernst M, Tsugawa H, Fleischauer M, Aicheler F, Aksenov AA, Alka O, Allard PM, Barsch A, Cachet X, Caraballo-Rodriguez AM, Da Silva RR, Dang T, Garg N, Gauglitz JM, Gurevich A, Isaac G, Jarmusch AK, Kameník Z, Kang KB, Kessler N, Koester I, Korf A, Le Gouellec A, Ludwig M, Martin H. C, McCall LI, McSayles J, Meyer SW, Mohimani H, Morsy M, Moyne O, Neumann S, Neuweger H, Nguyen NH, Nothias-Esposito M, Paolini J, Phelan VV, Pluskal T, Quinn RA, Rogers S, Shrestha B, Tripathi A, van der Hooft JJJ, Vargas F, Weldon KC, Witting M, Yang H, Zhang Z, Zubeil F, Kohlbacher O, Böcker S, Alexandrov T, Bandeira N, Wang M, Dorrestein PC. *Nat Methods*, 2020, 17: 905–908

48 Phelan VV. *Computational Methods and Data Analysis for Metabolomics*. Totowa: Humana Press Inc., 2020

49 Yu JS, Nothias LF, Wang M, Kim DH, Dorrestein PC, Kang KB, Yoo HH. *Anal Chem*, 2022, 94: 1456–1464

50 da Silva RR, Wang MX, Nothias LF, van der Hooft JJJ, Caraballo-Rodriguez AM, Fox E, Balunas MJ, Klassen JL, Lopes NP, Dorrestein PC. *PLoS Comput Biol*, 2018, 14: 26

51 Ernst M, Kang KB, Caraballo-Rodríguez AM, Nothias LF, Wandy J, Chen C, Wang M, Rogers S, Medema MH, Dorrestein PC, van der Hooft JJJ. *Metabolites*, 2019, 9: 144

52 Neto FC, Raftery D. *Anal Chem*, 2021, 93: 12001–12010

53 Ramabulana AT, Petras D, Madala NE, Tugizimana F. *Metabolites*, 2021, 11: 763

54 Mylonas R, Mauron Y, Masselot A, Binz PA, Budin N, Fathi M, Viette V, Hochstrasser DF, Lisacek F. *Anal Chem*, 2009, 81: 7604–7610

55 Wishart DS. *Bioanalysis*, 2009, 1: 1579–1596

56 Krettler CA, Thallinger GG. *Briefings BioInf*, 2021, 22:

57 da Silva RR, Dorrestein PC, Quinn RA. *Proc Natl Acad Sci USA*, 2015, 112: 12549–12550

58 Böcker S, Rasche F. *Bioinformatics*, 2008, 24: i49–i55

59 Rasche F, Svatos A, Maddula RK, Böttcher C, Böcker S. *Anal Chem*, 2011, 83: 1243–1251

60 Heinonen M, Shen H, Zamboni N, Rousu J. *Bioinformatics*, 2012, 28: 2333–2341

61 Shen H, Dührkop K, Böcker S, Rousu J. *Bioinformatics*, 2014, 30: i157–i164

62 Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. *Proc Natl Acad Sci USA*, 2015, 112: 12580–12585

63 Ludwig M, Dührkop K, Böcker S. *Bioinformatics*, 2018, 34: i333–i340

64 Fan Z, Alley A, Ghaffari K, Ressom HW. *Metabolomics*, 2020, 16: 11

65 Ji H, Deng H, Lu H, Zhang Z. *Anal Chem*, 2020, 92: 8649–8653

66 Borges RM, Colby SM, Das S, Edison AS, Fiehn O, Kind T, Lee J, Merrill AT, Merz Jr. KM, Metz TO, Nunez JR, Tantillo DJ, Wang LP, Wang S, Renslow RS. *Chem Rev*, 2021, 121: 5633–5670

67 Bauer CA, Grimme S. *Org Biomol Chem*, 2014, 12: 8737–8744

68 Koopman J, Grimme S. *J Am Soc Mass Spectrom*, 2021, 32: 1735–1751

69 Kangas LJ, Metz TO, Isaac G, Schrom BT, Ginovska-Pangovska B, Wang L, Tan L, Lewis RR, Miller JH. *Bioinformatics*, 2012, 28: 1705–1713

70 Allen F, Pon A, Wilson M, Greiner R, Wishart D. *Nucl Acids Res*, 2014, 42: W94–W99

71 Allen F, Greiner R, Wishart D. *Metabolomics*, 2014, 11: 98–110

72 Djoumbou-Feunang Y, Pon A, Karu N, Zheng J, Li C, Arndt D, Gautam M, Allen F, Wishart DS. *Metabolites*, 2019, 9: 72

73 Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart DS. *Anal Chem*, 2021, 93: 11692–11700

74 Karnovsky A, Li S Z. *Computational Methods and Data Analysis for Metabolomics*. Totowa: Humana Press Inc., 2020

75 Stancliffe E, Schwaiger-Haber M, Sindelar M, Patti GJ. *Nat Methods*, 2021, 18: 779–787

76 Shen X, Wang R, Xiong X, Yin Y, Cai Y, Ma Z, Liu N, Zhu Z. *Nat Commun*, 2019, 10: 14

77 Witting M, Böcker S. *J Sep Sci*, 2020, 43: 1746–1754

78 Domingo-Almenara X, Guijas C, Billings E, Montenegro-Burke JR, Uritboonthai W, Aisporna AE, Chen E, Benton HP, Siuzdak G. *Nat Commun*, 2019, 10: 5811

79 Kensert A, Bouwmeester R, Efthymiadis K, Van Broeck P, Desmet G, Cabooter D. *Anal Chem*, 2021, 93: 15633–15641

80 Yang Q, Ji H, Lu H, Zhang Z. *Anal Chem*, 2021, 93: 2200–2206

81 Ju R, Liu X, Zheng F, Lu X, Xu G, Lin X. *Anal Chem*, 2021, 93: 15651–15658

82 Zhou Z, Luo M, Chen X, Yin Y, Xiong X, Wang R, Zhu Z. *Nat Commun*, 2020, 11: 13

83 Martens J, van Outersterp RE, Vreeken RJ, Cuyckens F, Coene KLM, Engelke UF, Kluijtmans LAJ, Wevers RA, Buydens LMC, Redlich B, Berden G, Oomens J. *Anal Chim Acta*, 2020, 1093: 1–15

New advances in mass spectrometry data-based computational metabolomics methods

Jun Yang[1,2], Xinyu Liu[1,3*], Guowang Xu[1,2,3]

[1] *CAS Key Laboratory of Separation and Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *Liaoning Province Key Laboratory of Metabolomics, Dalian 116023, China*
**Corresponding author (email: liuxy2012@dicp.ac.cn)*

Abstract: Metabolomics aims at comprehensive analysis of the metabolome to discover biological phenomenon and mechanisms. Due to the complexity of metabolites in the biological samples, it is necessary to detect the metabolites as many as possible with advanced analytical techniques. Mass spectrometry is the most popular instrument in metabolomics research. The abundant ion signals from mass spectrometry are reflections of metabolome. However, it is still a huge challenge to process data and mine information from mass spectrometry data. Computational metabolomics fully exploits data from instruments, then combines methods from statics, chemometrics or artificial intelligence to make data processing and analysis efficiently, and therefore promotes the development of metabolomics. In this review, we firstly introduce general metabolomics data analysis procedure on the basis of the characteristic of data. Then data-driven computational metabolomics methods are reviewed, including feature detection, metabolites identification and annotation. Finally, knowledge-assisted computational metabolomics methods and future perspectives are briefly presented.

Keywords: computational metabolomics, mass spectrometry, feature detection, metabolites annotation and identification

doi: 10.1360/SSC-2022-0084