生 物 工 程 学 报
Chinese Journal of Biotechnology
http://journals.im.ac.cn/cjbcn
DOI: 10.13345/j.cjb.210271

Mar. 25, 2022, 38(3): 961-975

·综 述·

蛋白质组学中色谱保留时间对齐算法的研究进展

刘祎[1,2,3]，常乘[2,3]，朱云平[2,3]

1 北京工业大学 环境与生命学部，北京 100124
2 北京蛋白质组研究中心，北京 102206
3 国家蛋白质科学中心 (北京)，北京 102206

刘祎, 常乘, 朱云平. 蛋白质组学中色谱保留时间对齐算法的研究进展. 生物工程学报, 2022, 38(3): 961-975.
LI Y, CHANG C, ZHU YP. Advances of chromatogram retention time alignment algorithms in proteomics. Chin J Biotech, 2022, 38(3): 961-975.

摘 要：色谱是目前蛋白质组学流程中的一个基本环节，而色谱的保留时间对齐是有效提高鉴定和定量准确性的重要步骤之一。经过多年的发展，目前已经产生了一系列保留时间对齐算法。文中主要从可用性角度对蛋白质组学分析中色谱保留时间对齐算法及工具进行了系统总结，并对其发展趋势及应用方向进行了展望。

关键词：蛋白质组学；色谱；保留时间对齐；算法；软件

Advances of chromatogram retention time alignment algorithms in proteomics

LIU Yi[1,2,3], CHANG Cheng[2,3], ZHU Yunping[2,3]

1 The Faculty of Environment and Life, Beijing University of Technology, Beijing 100124, China
2 Proteome Research Center, Beijing 102206, China
3 National Center for Protein Sciences (Beijing), Beijing 102206, China

Abstract: Chromatography is a basic process in the current proteomics workflow, and the retention time alignment of the chromatogram is one of the important steps to effectively improve the identification and quantification accuracy. After years of development, a series of algorithms for retention time alignment have been developed. This review summarizes the advances of

Received: April 1, 2021; **Accepted:** June 30, 2021; **Published online:** July 8, 2021
Supported by: National Key Research and Development Program of China (2020YFE0202200)
Corresponding authors: CHANG Cheng. Tel: +86-10-61777046; E-mail: changcheng@ncpsb.org.cn
ZHU Yunping. Tel: +86-10-61777058; E-mail: zhuyunping@ncpsb.org.cn
基金项目：国家重点研发计划 (2020YFE0202200)

chromatographic retention time alignment algorithms and tools for proteomics analysis from the perspective of proteomics users, and discusses the development and future application directions.

Keywords: proteomics; chromatography; retention time alignment; algorithm; software

色谱作为一种分离手段常常与质谱联合使用，在分析混合样品方面有着广泛的应用。目前，在蛋白质组学中，液相色谱-质谱联用是最常用的样本分析方案。色谱的保留时间(retention time, RT) 指被分离样品组分从进样开始到柱后出现该组分浓度极大值的时间。保留时间对齐指的是将多次实验中相同组分的保留时间校正到一致的过程。本文主要从可用性角度对蛋白质组分析中色谱保留时间对齐算法及工具进行回顾和总结。

基于质谱的蛋白质组学定量流程如图 1 所示，可以分为有标定量和无标定量两种[1]。相较于无标定量，有标定量需要稳定同位素标记，不同的同位素用于区别不同的样本。由于有标定量策略可以同时检测多个样本，因此，保留时间对齐对于有标定量来说并不是一个难题。但是受限于有限的同位素标签，有标定量方法能同时检测的样本数量是有限制的。无标定量则不同，每个样品都是单独上机，理论上可以进行大规模样品间的比较。

目前无标定量方法基本上都以重构离子流色谱峰 (extracted ion current, XIC) 为前提，其过程如图 2 所示，在时间维度上根据连续谱图中质荷比相同的质谱谱峰强度拟合出合适的色谱峰，即 XIC。最后一般以 XIC 的峰面积或最高强度等特征作为定量值。但是不同的样本在检测时，每个样本的保留时间会由于温度等差异产生不确定的偏移，如果不进行校正，我们往往会将不同样本间不同的 XIC 进行匹配，得到错误的比较结果。所以，为了准确地进行不同样本间的比较，保留时间对齐是一个关键步骤。本文后续讨论均以无标定量作为前提。

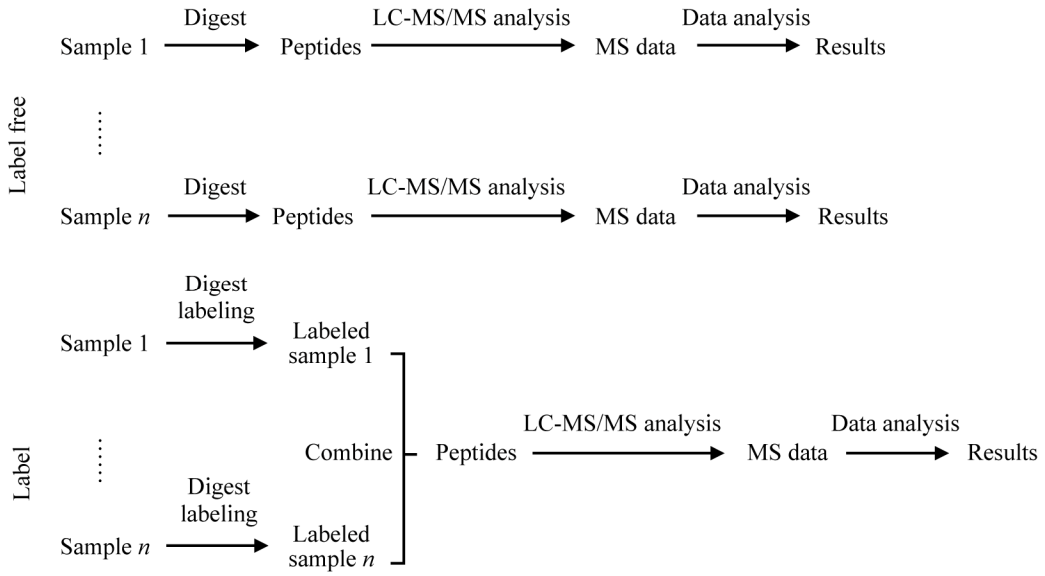

图 1　蛋白质组学基本定量流程

Figure 1　Basic quantitative workflow of proteomics.

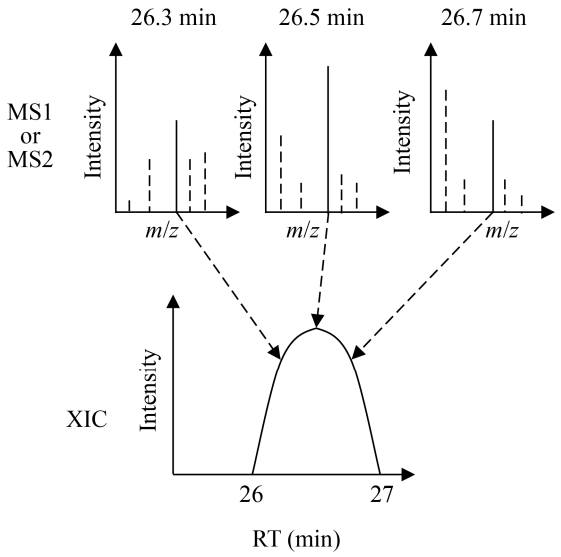

图 2　重构离子流色谱峰的基本过程　MS1 和 MS2 分别指质谱的一级谱图和二级谱图，RT 指色谱保留时间

Figure 2　Basic process of constructting XIC. MS1 and MS2 are indicate the MS and MS/MS spectrum, respectively. RT indicates the chromatographic retention time.

与保留时间对齐相关的另一个概念是保留时间预测。保留时间预测指的是根据序列预测该序列在实际样品中可能出现的保留时间，根据这个时间也可以将 XIC 进行一一匹配。保留时间预测本身也是一个重要的热门领域，有很多最新进展，本文不涉及保留时间预测相关的算法与研究。

另外，在蛋白质组分析中，有谱峰对齐、特征对齐等表述，指的是将不同样本中相同成分形成的谱峰或特征一一对应起来。保留时间对齐与谱峰对齐、特征对齐实际上是有细微区别的，保留时间对齐并不要求每个谱峰一一对齐，而只需要一些主要的峰对齐，再根据这些主要的峰校正每个样本的保留时间。但在蛋白质组分析中，保留时间对齐的最终目的同样是得到谱峰或特征的一一对应列表。所以本文并不刻意区分保留时间对齐与谱峰对齐及特征对齐。

1　算法

本章节中，虽有部分算法主要应用于代谢组数据分析，但其保留时间对齐部分仍然对蛋白质组数据分析有参考意义，故予以保留。

规整函数是一类可在某个维度上减小数据分布不均匀性的函数，而这个维度常常指时间维度。根据有没有使用规整函数，可以将对齐算法分为规整函数法与直接对齐法。其中，动态时间规整 (dynamic time warping, DTW) 和相关优化规整 (correlation-optimized warping, COW) 是两种最常用的规整函数。后续很多算法都是在这两种方法基础上的改进和提升。

直接对齐法则不需要引入规整函数。最简单的直接对齐就是定义一个保留时间阈值，不同样本中位于阈值之内的相似特征 (根据质荷比等信息确定) 就进行对齐。这种方法在质谱精度较高时较为有效，更多特征的加入也有利于对齐的准确性。通常，该方法并不会单独使用，而是作为最终得到一一对应特征的一个步骤。更复杂一些的方法会将该问题看作聚类问题，不同样品中相似的特征会被聚为一类，也就是被“对齐”了。有一些算法会采用 MS2 信息，比如 MSInspect[2]，而其他一些算法则会直接计算特征直接的相似性，根据相似性进行聚类。这些算法的主要区别就在于选取特征的不同和相似性计算方法的不同。

关于这两类算法，Smith 等[3]已经进行了比较全面的综述，本文不再赘述。

不管是对于试图开发新算法的科研人员，还是希望将已有算法引入自己流程的科研人员，算法的源代码都是重要的参考资料。但是部分文献着重于算法的论述，没有提供算法的

具体实现；另一些文献虽然提供了算法的实现，但是由于服务器的更改等原因，原文的链接已经无法使用。本文系统考察了 Smith 等综述中的所有算法，保留了迄今为止仍能使用的算法，并补充了于该文之后发表的保留时间对齐算法，总共有 16 种算法目前仍然能下载到相关软件。本文按引用次数由高到低对这 16 种算法进行介绍。

XCMS[4]是这些算法中引用次数最多的，但 XCMS 并不是一个单纯进行保留时间对齐的算法，而是一系列质谱原始数据预处理方法的集合，包括保留时间对齐、数据过滤、峰识别、峰提取等。XCMS 是一个 R 包，目前主要应用于代谢组学领域。XCMS 由于其开源的特性，是该领域最常用的原始数据处理软件之一。在 2018 年，该团队还在 *Nature Protocols* 发表了基于网页的工具 XCMS Online[5]。原文使用 XCMS 分别做了血浆和一个脂肪酸酰胺水解酶的研究，但是没有与别的软件进行比较。总的来说，XCMS 在代谢组学领域已得到了认可，使用 R 语言进行流程开发或数据分析的研究者可以方便进行整合。具体算法方面，XCMS 首先根据质荷比将谱峰进行粗分组，然后在这些粗分组中根据两条原则 (1：该分组包含几乎所有样本的谱峰；2：几乎所有样本都只有一个谱峰位于该组) 筛选出优质 (原文为 well-behaved) 分组，以这些分组作为参考，使用局部加权回归散点平滑法 (locally weighted scatterplot smoothing, LOESS) 对所有样本进行对齐。该方法最大的优点在于无需选择参考样本，但其效果受限于能否得到合适的优质分组。

MZMine 2[6]与 XCMS 类似，同样是一个主要服务于代谢组学领域的工具集，包括谱峰检测、保留时间对齐和数据可视化等一系列功能，由 Java 语言编写。MZMine[7]发表于 2006 年，目前已无法获取。相比于 MZMine，MZMine 2 的保留时间对齐提供了一种新的算法随机样本一致性对齐 (random sample consensus aligner, RANSAC)，这种算法会进行多次迭代，每次迭代会选取样本中的部分 XIC 构建一个 LOESS 模型，余下的 XIC 用于评估该模型。在多次迭代之后，选取效果最好的模型作为最终的对齐方案。与此同时，MZMine 2 保留了 MZMine 中根据固定窗口进行简单对齐的方案，称为 Join 方法。这种方法会根据用户定义的保留时间范围将不同样本中的 XIC 与一个参考样本进行对齐，在范围之外的 XIC 会被标记为未对齐。Lange 等[8]在 2008 年用 2 个蛋白质组数据集 P1、P2 与 2 个代谢组数据集 M1、M2 对 MSInspect[2]、MZmine[7]、OpenMS[9]、SpecArray[10]、Xalign[11]和 XCMS[4]这几个工具进行了评估，MZMine 2 在此基础上补充了 MZMine 2 的结果。P1、P2 数据集上 MZmine 2 的 RANSAC 算法与 OpenMS 表现最佳，MZmine、XCMS 与 Xalign 稍差一点，M1、M2 数据集上 MZmine 2、MZmine、OpenMS、XCMS、Xalign 差别不明显，都能得到比较准确的结果。总之，除了 MSInsepect 无法得到合理结果，其他软件各有优劣，其中 MZmine 2 的 RANSAC 算法与 OpenMS 在蛋白质组数据上稍胜一筹。MZmine 2 的算法虽然效果较好，但在使用时有相当多的参数需要用户定义，其中保留时间窗口范围等参数极其重要，会严重影响对齐结果的优劣。处理不同的样本集时如何找到最合适的参数对于用户来说是一个比较困难的问题。

DTW[12]与 COW[13]作为经典算法，由于时间较早，两篇相关文献中提供的代码链接都已失效。但是在 Tomasi 等的文章[14]中提供了目前仍可下载的 Matlab 代码，文中还探讨了 DTW 与 COW 的联系，比较有参考价值。

MetAlign[15]与 XCMS 及 MZMine 2 类似，是一个质谱原始数据预处理工具，由 C++语言编写，运行于 Windows 平台。目前其软件仍然可以下载使用，但并未开源。MetAlign 发表于 2009 年，据文章描述，经过了 8 年的开发，已得到一些实际研究的检验。MetAlign 的算法基础来源于对人工处理方式的模拟，该算法首先会根据用户设置的初始窗口对齐不同的样本，再根据参考样本对保留时间进行校正。迭代上述过程，每次迭代时采用更小的窗口，最终得到更精确的结果。MetAlign 已经被整合于另一个工具 TagFinder 中，但文献中并没有提供与其他工具的比较。

SpecArray[10]是较早期的一个定量工具，发表于 2005 年，由 C 语言编写，源代码仍可获取。其保留时间对齐主要由 PepMatch 和 PepArray 这两个模块完成，PepMatch 模块先构建一个多对多的肽段映射，PepArray 再根据距离最小原则对保留时间进行校准。SpecArray 在上文提到的 Lange 等的评估中表现一般，在 P1 数据集中稍稍优于表现最差的 MSInsepect，在 P2 数据集中效果与 MSInsepect 相当。同时，SpecArray 不适用于代谢组学数据，且无法调整参数，故其参考价值大于实际应用价值。

OBI-warp[16]发表于 2006 年，使用的对齐方法是改进过的 DTW。早期的 DTW 算法并不考虑质荷比维度，而 OBI-warp 进行了考虑，进而提高了精度。该软件由 C++编写，一直到 2010 年还在更新，同时也提供了源代码，是很好的参考资料。OBI-warp 在文章中与另一个对齐算法 ChAMS[17]进行了对比，在需要对齐的谱图相似度较高时两种算法的效果相当，但对于差别较大的谱图，OBI-warp 会有更好的结果。

IDEAL-Q[18]是 2010 年发表的一个定量工具，由 C#与 C++编写，目前能下载到软件并使用，但是软件并没有开源。IDEAL-Q 中的保留时间对齐算法依赖于鉴定结果，根据两个样品中的相同鉴定结果，IDEAL-Q 会拟合一个线性回归方程，再根据此方程对保留时间进行校准，该算法被称为 IDEAL。IDEAL-Q 在文章中并没有与别的工具进行比较，而只是比较了 IDEAL 算法与单纯的线性回归的效果，证明 IDEAL 强于单纯的线性回归。

OpenMS[9]是一个柏林自由大学的 Knut Reinert 团队、图宾根大学的 Oliver Kohlbacher 团队及苏黎世联邦理工学院的 Ruedi Aebersold 团队主导的开源项目，主要由 C++编写，应用于蛋白质组学领域及代谢组学领域质谱数据的分析和处理，它支持 Windows、macOS 和 Linux。OpenMS 包含了蛋白质组学中许多常见数据分析管道的工具，提供用于信号处理、特征查找、可视化，保留时间对齐和肽段鉴定等一系列算法。其中的保留时间对齐算法[19]使用的是一种几何方法，先拟合出一个仿射函数进行粗对齐，再用线性回归进行修正调整。原文中虽然没有给出与其他对齐软件的对比，但 Lange 等的测试中，OpenMS 在两个蛋白质组数据集上是表现效果最佳的软件之一，在两个代谢组数据集上的精确度和召回率也只是比最佳软件 MZmine 稍低。加上 OpenMS 开源的特性，该算法是开发新方法很好的参考，也是流程整合的最佳选择之一。

PTW-I[20]最早于 2010 年作为一个 R 包发表，是对原始 PTW[21]算法的优化。原始的 PTW 算法对信号强度过于敏感，对于信号强度较低的特征表现不佳。PTW-I 将 PTW 使用的距离度量方法由均方根差值 (root mean square difference, RMS) 替换为加权互相关 (weighted cross-correlation, WCC)。原文主要探讨了 WCC 相对于 RMS 的提升，并没有与其他工具进行

对比。作为开源 R 包，PTW-I 也是非常适合整合或是参考的工具。

msalign[22]是一个C语言编写的专用于保留时间对齐的工具，完全开源，发表于 2007 年。而 msalign 2[23]是其为了处理毛细管电泳质谱数据的修改版本，发表于 2009 年。虽然是专门为毛细管电泳质谱数据做了优化，但作者强调其同样适用于多种串联质谱数据。msalign 2 主要使用了一种遗传算法进行对齐，作者与 XCMS 做了比较，在作者提供的数据上，msalign 2 的对齐结果优于 XCMS。

LaCyTools[24]是 2016 年 Jansen 等发表的串联质谱数据处理工具，由 Python 编写，完全开源。LaCyTools 的对齐算法用到了信噪比和基于正态分布的背景及噪声判定 (normal-distribution based background and noise determination, NOBAN) 算法作为质控手段，再用最小二乘法 (least squares method) 拟合出一个幂律分布函数来对保留时间进行校准。文中作者与 msalign 2 及 MZmine 2 进行了对比，但只是描述了对比结果，没有提供直接对比数据。据作者描述，msalign 无法对提供的数据集进行有效对齐，而 MZmine 2 与 LaCyTools 均能完成有效对齐，且后两种方法结果没有显著的性能差异。所以，对于熟悉 Python 的研究者，LaCyTools 也是一种易于整合的方法。

AMSRPM[25]是一个开源 R 包，将鲁棒的点匹配算法 (robust point matching, RPM) 应用于保留时间对齐。AMSRPM 原文中并没有与其他工具进行比较，但作为一个开源的 R 包，AMSRPM 提供了 XCMS 之外的参考。

PMRM[26]是 2013 年 Tsai 等发表的保留时间对齐算法，提供了 Matlab 实现。PMRM 在算法层面上的意义更大一些，对于想要整合进自己流程的研究人员，别的算法可能是更好的选择。PMRM 主要使用了一个贝叶斯对齐模型 (Bayesian alignment model, BAM) 来对 2011 年发表的 SIMA 算法进行优化。经作者测试，PMRM 算法结果在精确度和召回率上都优于 Lange 等[8]文章中的几个工具，但需要注意的是测试所用的蛋白质组数据并不相同。另外，SIMA 的链接目前已经失效，所以 PMRM 的参考价值大于实际应用价值。

MassUntangler[27]是一个用 Python 编写的工具，专用于保留时间对齐。MassUntangler 在对齐时没有使用规整函数，是一种依赖于特征相似性的直接对齐算法，使用保留时间、质荷比以及二级谱图判断特征的相似性。作者分别在一个简单数据集与复杂数据集上与 Lange 等的文章中的几个工具做了对比，这几个工具中 OpenMS 表现最好。而 MassUntangler 效果比 OpenMS 稍差，与 MZmine 及 XAlign 效果相近。所以，对于使用 Python 作为编程语言的研究人员，MassUntangler 是一个不错的选择，否则，OpenMS 可能是更好的选择。但总的来说，MassUntangler 提供了一种新的思路，值得参考。

LWBMatch[28]也是一种直接对齐方法，由 C++编写。LWBMatch 重点关注异质性问题，也就是差别较大的样本间的对齐，如不同色谱柱、不同的仪器或不同的设定下得到的数据。作者在异质性数据上对 LWBMatch、OpenMS、SuperHirn 以及 DTW 算法进行了测试，发现 LWBMatch 优于其他几种方法，尤其是在召回率上。OpenMS 及 SuperHirn 完全无法在异质性数据上得到有效的结果，作者认为这一点是由于这两个软件假设保留时间不会存在大的偏移，特征也不会存在顺序交换造成的，这两个假设在异质性数据中都不成立。所以，对于差异较大的样品，LWBMatch 会是一个不错的选择。

DIAlignR[29]则是一个专用于 DIA 数据的 R

包。上述工具都是在 MS1 层面上进行保留时间的对齐，但 DIAlignR 则是在 MS2 的层面上进行对齐。DIAlignR 在算法设计中也考虑到了不同样本中特征的顺序可能不同。文中作者通过比较，证明了 DIAlignR 优于采用 LOESS[30]的方法。对于 DIA 数据的 MS2 有对齐需求的研究者，虽然也可以把 MS2 当作 MS1 用传统工具进行对齐，但是 DIAlignR 提供了新的选择。

值得一提的是，笔者课题组长期致力于蛋白质组数据质量控制算法及定量算法等方向的研究，于 2011 年基于鲁棒局部回归方法建立了一个可逆的、非线性的色谱保留时间对齐方法[31]。该方法会选择一次实验作为参考实验，对于任意一个待对齐的目标实验，使用分段线性模型来描述肽段色谱保留时间在目标实验与参考实验之间的关系。该方法由 C++编写，最显著的优势在于可以大大节省计算时间，经测试，MSInspect、msalign 和 XCMS 在相同计算条件下的运行耗时至少是该方法的 2 倍。目前，该方法被整合于本实验室开发的 DDA 数据定量软件 LFQuant[32]与 PANDA[33]中。LFQuant 专用于无标数据的定量。而 PANDA 则可处理无标数据和有标数据。LFQuant 与 PANDA 能在保证定量准确性的前提下比 MaxQuant 更快完成定量过程。本文系统考察的保留时间对齐算法汇总为表 1。

总之，从对齐准确性上看，XCMS、MZmine 2 及 OpenMS 的效果较好，但没有一个方法存在明显的优势。值得注意的是，XCMS、MZmine 2 以及其他多种算法都选择 LOESS 作为算法的核心步骤之一，说明 LOESS 在保留时间对齐问题上的效果受到了大部分开发者的肯定。XCMS、MZmine 2 及 OpenMS 的对齐算法均基于拟合好的规整函数，这也是目前最主流的方法。规整函数方法是对全局进行对齐，相较之下，MassUntangler 所采用的则可以看作局部对齐，在特征数量较大的情况下，局部对齐所需的时间往往更长。另外，笔者认为目前的保留时间对齐算法主要存在 3 个问题：一是需要用户凭经验设置的参数过多。MZmine 2 等效果较

表 1　保留时间对齐算法统计

Table 1　Overview of the current available retention time alignment algorithms

Algorithm names	Year	Citation numbers*	Software availability	Source code availability	Programming language	References
COW	1998	818	No	No	C	[13]
DTW	1998	489	No	No	–	[12]
Bylund et al.	2002	255	No	No	Matlab	[34]
Wang et al.	2003	757	No	No	–	[35]
Peakmatch	2003	230	No	No	Matlab	[36]
RTAlign	2003	225	No	No	–	[37]
PARS	2003	116	No	No	Matlab	[38]
DTW and COW	2004	665	Yes	Yes	Matlab	[14]
PTW	2004	583	No	No	R	[21]
Radulovic et al.	2004	257	No	No	–	[39]
CPM	2004	165	No	No	–	[40]
Higgs et al.	2005	260	No	No	Perl and R	[41]
SpecArray	2005	230	Yes	Yes	C	[10]
Walczak et al.	2005	110	No	No	–	[42]

(待续)

(续表 1)

Algorithm names	Year	Citation numbers*	Software availability	Source code availability	Programming language	References
Xalign	2005	100	No	No	–	[11]
XCMS	2006	3 347	Yes	Yes	R	[4]
MZMine	2006	732	No	No	Java	[7]
MSInspect	2006	302	No	No	Java and R	[2]
Skov et al.	2006	284	No	No	Matlab	[43]
OBI-warp	2006	204	Yes	Yes	C++	[16]
STW	2006	191	No	No	–	[44]
LCMSWARP	2006	176	No	No	Matlab and C++	[45]
PEPPeR	2006	167	No	No	–	[46]
ChAMS	2006	135	No	No	–	[17]
Fischer et al.	2006	90	No	No	–	[47]
Chromalign	2006	84	No	No	C++	[48]
DeSouza et al.	2006	70	No	No	–	[49]
SuperHirn	2007	713	No	No	C++	[50]
OpenMS	2007	102	Yes	Yes	C++	[9]
PETAL	2007	89	No	No	R	[51]
msalign	2007	50	Yes	Yes	C	[22]
Auto-PABS	2007	49	No	No	Matlab	[52]
AMSRPM	2007	31	Yes	Yes	R	[25]
MCCA	2007	28	No	No	–	[53]
COW-CODA	2008	107	No	No	–	[54]
Suits et al.	2008	55	No	No	–	[55]
MetAlign	2009	646	Yes	No	C and C++	[15]
Podwojski et al.	2009	102	No	No	R	[56]
msalign 2	2009	65	Yes	Yes	C and R	[23]
Valkenborg et al.	2009	11	No	No	Matlab	[57]
MZMine 2	2010	1 988	Yes	Yes	Java	[6]
IDEAL-Q	2010	128	Yes	No	C# and C++	[18]
PTW-I	2010	86	Yes	Yes	R	[20]
Quality threshold clustering	2010	11	No	No	–	[58]
SIMA	2011	39	No	No	C++	[59]
MassUntangler	2011	17	Yes	Yes	Python	[27]
Zhang et al.	2012	29	No	No	C++	[60]
Supervised alignment	2012	13	No	No	–	[61]
PMRM	2013	31	Yes	Yes	Matlab	[26]
LWBMatch	2013	16	Yes	Yes	C++	[28]
SAGA	2013	11	No	No	–	[62]
PeakLink	2014	16	No	No	Matlab	[63]
LaCyTools	2016	56	Yes	Yes	Python	[24]
Li et al.	2017	11	No	No	–	[64]
DIAlignR	2019	10	Yes	Yes	R	[29]
Wang et al.	2019	3	No	No	–	[65]

* Citation numbers: data from Google Scholar.

好的算法均需要用户指定保留时间阈值等参数，对于缺乏经验的用户很难使对齐算法达到最优效果。二是大部分算法需要指定一个参考样本。参考样本的不同会直接影响最终对齐的效果，但在对齐之前很难确定最优的参考样本。这一点上，XCMS 的算法无需选择参考样本，就显得尤为独特。三是目前的算法普遍在复杂样品上表现不佳。大部分算法均假设样品之间不存在洗脱顺序的变化，这一点在复杂样品中是不成立的。这也导致了目前大部分算法在进行复杂样品的保留时间对齐时效果下降。最新的一些算法已经开始考虑这个问题，如 DIAlignR 就不再严格要求肽段洗脱顺序不变，而是允许对齐的谱峰在一定时间范围内浮动。

2 应用

在这一部分，本文主要回顾近年来蛋白质组学领域一些主要工具对保留时间对齐算法的应用，按照数据类型介绍保留时间对齐在近年来一些蛋白质组分析工具中的应用。根据数据采集方式的不同，蛋白质组学数据可以分为数据依赖性采集 (data-dependent acquisition, DDA) 和数据非依赖性采集 (data independent acquisition, DIA)，其主要区别在于二级谱的采集策略上。DIA 定量准确性和可重复性更高，但其二级谱图的复杂性也更高，解析难度大于 DDA 数据。

2.1 DDA 数据

DDA 数据的二级谱与 DIA 数据相比易于解析，但由于数据采集机制的原因，并非每个母离子都有机会得到二级谱，同时，也并非所有的二级谱都能得到正确解析。所以，往往在 DDA 数据分析流程中，保留时间对齐会作为一种补充手段，将没有二级谱的母离子补充到定量结果中，以得到尽可能精确的最终结果。

MaxQuant 软件[66-68]的基本流程如图 3 所示。MaxQuant 首先会从质谱数据中提取信息，构建 XIC 峰。然后用自带的 Andromeda 搜索引擎进行数据库搜索。MaxQuant 支持有标定量与无标定量，对于无标定量，MaxQuant 采用的 match between runs (MBR) 机制就是一种保留时间对齐。该方法先通过二维的高斯平滑对不同的样品进行对齐，这一过程中为了节省计算时间，MaxQuant 会先对样品进行层次聚类，优先对齐相似的样本。在对齐之后，再根据保留时间阈值对 MS1 特征进行对齐。Lim 等使用人和酵母的混合样品评估了 MBR 机制[69]，虽然单独的 MBR 方法会引入大量的错误匹配，但配合 MaxQuant 自带的蛋白装配方法，可以有效减少最终定量结果的缺失值。

IonStar[70]包含从样本制备，到数据上机，再到数据分析的整个蛋白质组分析全流程，这里只总结其数据分析流程部分，如图 3 所示。保留时间对齐是该部分的关键步骤之一。

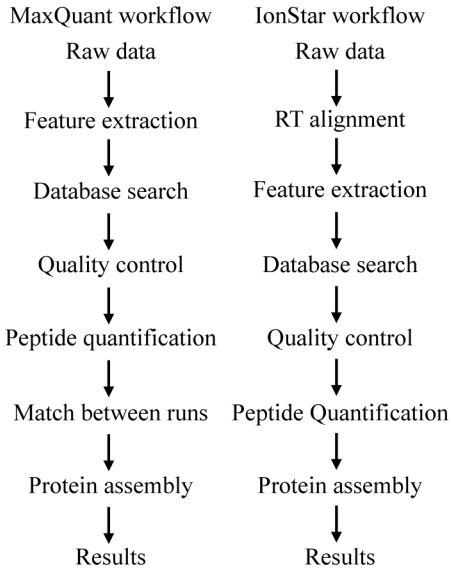

图 3 MaxQuant 及 IonStar 的基本流程

Figure 3 Basic workflows of MaxQuant and IonStar.

IonStar 采用的是整合于商业软件 SIEVE 中的 ChromAlign[48]方法，在对齐时不会使用 MS2 信息，而是基于保留时间、质量及相对强度的三维信息基于一个参考样品对 MS1 特征进行对齐。原文中作者把 IonStar 与 MaxQuant 做了对比，能得到更少的缺失值。

Quandenser[71]是另一个试图减少定量缺失值的流程，其流程如图 4 所示。同样，保留时间的对齐是其重要步骤。Quandenser 首先采用阈值方法对 MS1 特征进行了初步筛选，接下来用一个训练好的机器学习方法判断特征对齐的 FDR。训练时，用正确的匹配结果作为 target 集，将这些正确的匹配结果在保留时间维度进行平移作为 decoy 集。Quandenser 在各个步骤中均提供了严格的质量控制，最终相较于 MaxQuant 能得到更多的差异蛋白。

2.2 DIA 数据

DIA 数据由于二级谱图更为复杂，往往更加依赖于保留时间对齐从谱图中提取有效信息。很多 DIA 方法都依赖于内标肽如 iRT 方法对保留时间进行对齐，除此之外也有一些对齐算法的应用。

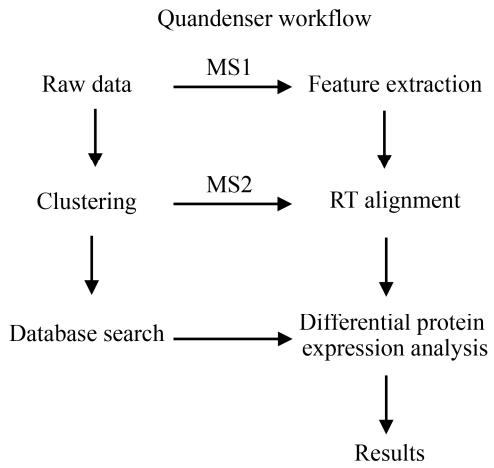

图 4 Quandenser 的基本流程

Figure 4 Basic workflow of Quandenser.

OpenSWATH 本身是一个 DIA 鉴定软件，与 DDA 数据处理中的数据库搜索类似，其鉴定过程依赖于一个预先构建的数据库，故被称为数据库依赖方法。OpenSWATH 也提供了从鉴定到定量的完整流程，如图 5 所示，在定量流程中定量过程使用了 TRIC 方法[72]，该方法是一个基于图的非线性对齐算法。利用最小生成树算法，TRIC 将相似的特征进行高效对齐。

DIA-Umpire[73]流程则是一个非数据库依赖 (library-free) 的 DIA 数据处理软件，其核心在于根据谱峰的相关性生成伪二级谱 (pseudo MS2 spectrum)。在鉴定过程之前，DIA-Umpire 并不会使用保留时间对齐，但是在鉴定过程之后使用了 IDEAL-Q[18]方法。该方法使用两个样

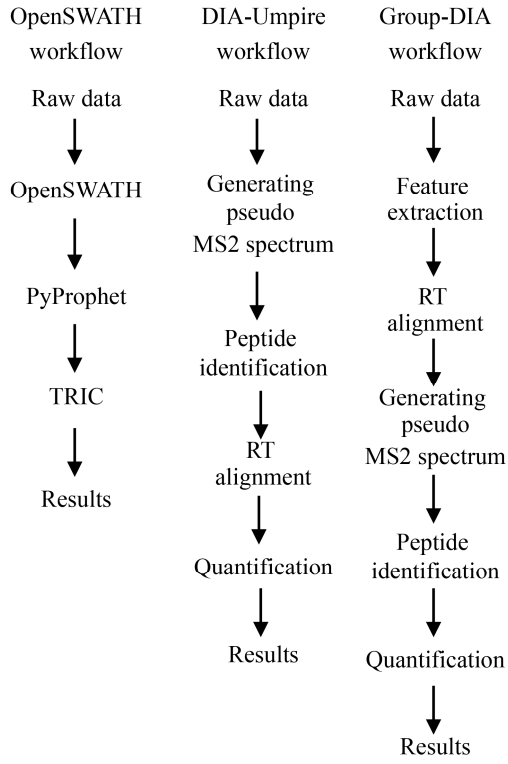

图 5 OpenSWATH、DIA-Umpire 和 Group-DIA 的基本流程

Figure 5 Basic workflows of OpenSWATH, DIA-Umpire and Group-DIA.

本中共同鉴定到的肽段构建线性回归方程，再结合一个校正函数，得到最终的对齐结果。作者与OpenSWATH 进行了比较，但发现 OpenSWATH 的效果更好,作者把这一点归结于 OpenSWATH 使用的数据库较伪二级谱搜库使用的数据库更小。在缩小了伪二级谱搜库所用的数据库后，DIA-Umpire 的效果有所上升，但仍然不如 OpenSWATH。

Group-DIA[74]是另一个非数据库依赖的 DIA 数据分析方法，但与 DIA-Umpire 不同，Group-DIA 会在鉴定之前使用 ChromAlign 方法对保留时间进行对齐,再根据相似性提取伪二级谱进行鉴定及后续步骤。据作者评估，所测试的样品越多,Group-DIA 对 DIA-Umpire 的优势越大。DIA-Umpire 检测到的肽段中大约 90%能被 Group-DIA 检测到，反过来 Group-DIA 检测到的肽段中只有 60%左右 DIA-Umpire 能被检测到。Group-DIA 与 DIA-Umpire 在生成伪二级谱时都使用的是相关性计算，最大的差别反而是计算相关性之前是否使用了保留时间对齐。从这点来看，保留时间对齐是相当重要的步骤，也是 Group-DIA 能在样品越多时效果越好的关键所在。

3 总结与展望

不论在 DDA 流程中，还是在 DIA 流程中，保留时间对齐都是有效提高无标定量方法定量准确性的重要手段。保留时间对齐算法的发展最迅速的是 2000 年到 2010 年这 10 年，除了经典的 DTW、COW 及最新的一些算法外，现有的大部分算法都出现在这个时间段。从引用数量上看，最受欢迎的是综合性软件。将常用功能整合到一起且提供可视化界面是引用量最高的几个软件(除开经典算法) 的共性，功能单一的软件往往很难得到广泛的应用。

在研究相关算法时笔者发现，保留时间对齐算法的效果与数据的关联很大，但目前缺乏在多个数据集上对已有算法进行评估的工作。Lange 等[8]虽然在两个蛋白质组数据集上和两个代谢组数据集上评估了 6 种工具，但仅凭如此少的数据集很难对所有工具作出一个全面合理的评估。以蛋白质组数据集中综合表现最好的 OpenMS 为例，其在蛋白质组数据集 1 上的表现明显优于蛋白质组数据集 2，MZmine 在蛋白质组数据集 2 上的效果并不比 OpenMS 差。所以各种算法可能有其适合的应用场景，目前并不存在优势特别明显的算法，但还没有工作系统地统计和总结各个算法适用的情况。未来该领域需要在更多的数据集上系统地比较各个算法的优劣。最新出现的算法一是对应于新的数据类型，如 DIA 数据，另一个发展趋势就是允许待对齐样品间有更大的差异。早期的算法往往假设待对齐样品是相似的，因为我们在设计实验时一般会尽量保证仪器相同、仪器参数相同、运行时间接近，避免引入更多的误差。但随着蛋白质组学技术的发展,检测样品的数量越来越多,很难保证所有的数据均来自同一台仪器。而且随着质谱数据共享平台的发展，我们往往希望整合多个来源的数据来得到更多的提示或更好地验证我们的假设。在这种情况下，待对齐样品相似的假设是不成立的，如何在这种情况下进行更准确地对齐可能是未来该领域所关注的 一个重要方向。类似的问题有一种更规范的划分方法，研究人员将更加相似的样品之间的时间偏移称为单调性偏移(monotonic shift)[75],这种偏移在所有样品上不会发生顺序的变化。但实际上，由于色谱柱的不同以及色谱参数的不同,不同的样品往往会发生洗脱顺序的改变，这种偏移称为非单调性偏移 (non-monotonic shift)。色谱条件差别越大，

非单调性偏移的比例也会更多。所以，非单调性偏移的对齐可能会是未来的研究重点。这一点上笔者认为直接对齐会是比规整函数更好的方法，或者是用规整函数作粗略的对齐，再用合适的方法进行对齐。

另一方面，仪器和技术的进步也在推动着这个领域的发展。一是目前质谱的精度越来越高，依靠质荷比维度可以区分开更多的肽段，简化了共洗脱肽段的复杂程度，减少了对齐算法的压力。二是质谱在试图引入更多的特征，比如离子淌度质谱，增加一个维度同样可以区分开更多的肽段，同样会减小对齐算法所面对数据的复杂性。三是各种新算法的应用，以深度学习为代表的技术随着计算成本的降低及框架的完善正走入更多的领域，实际上也出现了将深度学习方法用于气相色谱-质谱联用数据的保留时间对齐[76]。理论上，气相色谱-质谱联用与液相质谱-质谱联用在数据层面上没有本质的区别，在合适的蛋白质组数据上对相应的方法进行优化，或许我们就能得到一个更为准确的保留时间对齐工具。

当前的研究大多集中于保留时间预测，就算是最新的工具，使用的保留时间对齐方法也还是多年前的经典方法。这说明保留时间对齐算法的研究仍然是一个很有潜力的方向，结合一些最新的技术，有希望能提升 XIC 匹配的准确性。随着蛋白质组学的发展，其在精准医疗等领域已经渐渐展现出巨大的潜力[77]，“蛋白质组学驱动的精准医学”研究新模式呼之欲出。基于蛋白质组学的临床研究越来越多，其中涉及到成百上千例样本之间的比较。然而，更大的样本规模使错误匹配的风险更高，可能导致整个研究的失败。因此，发展更精准的保留时间对齐算法将为蛋白质组学走向临床应用提供重要的方法学支撑。

REFERENCES

[1] 常乘, 朱云平. 基于质谱的定量蛋白质组学策略和方法研究进展. 中国科学: 生命科学, 2015, 45(5): 425-438.
Chang C, Zhu YP. Strategies and algorithms for quantitative proteomics based on mass spectrometry. Sci Sin Vitae, 2015, 45(5): 425-438 (in Chinese).

[2] Bellew M, Coram M, Fitzgibbon M, et al. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. Bioinformatics, 2006, 22(15): 1902-1909.

[3] Smith R, Ventura D, Prince JT. LC-MS alignment in theory and practice: a comprehensive algorithmic review. Brief Bioinform, 2013, 16(1): 104-117.

[4] Smith CA, Want EJ, O'Maille G, et al. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem, 2006, 78(3): 779-787.

[5] Forsberg EM, Huan T, Rinehart D, et al. Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. Nat Protoc, 2018, 13(4): 633-651.

[6] Pluskal T, Castillo S, Villar-Briones A, et al. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics, 2010, 11: 395.

[7] Katajamaa M, Miettinen J, Oresic M. MZmine: toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics, 2006, 22(5): 634-636.

[8] Lange E, Tautenhahn R, Neumann S, et al. Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. BMC Bioinformatics, 2008, 9: 375.

[9] Röst HL, Sachsenberg T, Aiche S, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods, 2016, 13(9): 741-748.

[10] Li XJ, Yi EC, Kemp CJ, et al. A software suite for the generation and comparison of peptide arrays from sets of data collected by liquid chromatography-mass spectrometry. Mol Cell Proteomics, 2005, 4(9): 1328-1340.

[11] Zhang X, Asara JM, Adamec J, et al. Data pre-processing in liquid chromatography-mass spectrometry-based proteomics. Bioinformatics, 2005, 21(21): 4054-4059.

[12] Kassidas A, MacGregor JF, Taylor PA. Synchronization of batch trajectories using dynamic time warping. AIChE J, 1998, 44(4): 864-875.
[13] Nielsen NPV, Carstensen JM, Smedsgaard J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. J Chromatogr A, 1998, 805(1-2): 17-35.
[14] Tomasi G, van den Berg F, Andersson C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. J Chemometrics, 2004, 18(5): 231-241.
[15] Lommen A. MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. Anal Chem, 2009, 81(8): 3079-3086.
[16] Prince JT, Marcotte EM. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. Anal Chem, 2006, 78(17): 6140-6152.
[17] Prakash A, Mallick P, Whiteaker J, et al. Signal maps for mass spectrometry-based comparative proteomics. Mol Cell Proteomics, 2006, 5(3): 423-432.
[18] Tsou CC, Tsai CF, Tsui YH, et al. IDEAL-Q, an automated tool for label-free quantitation analysis using an efficient peptide alignment approach and spectral data validation. Mol Cell Proteomics, 2010, 9(1): 131-144.
[19] Lange E, Gröpl C, Schulz-Trieglaff O, et al. A geometric approach for the alignment of liquid chromatography-mass spectrometry data. Bioinformatics, 2007, 23(13): i273-i281.
[20] Bloemberg TG, Gerretzen J, Wouters HJP, et al. Improved parametric time warping for proteomics. Chemom Intell Lab Syst, 2010, 104(1): 65-74.
[21] Eilers PHC. Parametric time warping. Anal Chem, 2004, 76(2): 404-411.
[22] Palmblad M, Mills DJ, Bindschedler LV, et al. Chromatographic alignment of LC-MS and LC-MS/MS datasets by genetic algorithm feature extraction. J Am Soc Mass Spectrom, 2007, 18(10): 1835-1843.
[23] Nevedomskaya E, Derks R, Deelder AM, et al. Alignment of capillary electrophoresis-mass spectrometry datasets using accurate mass information. Anal Bioanal Chem, 2009, 395(8): 2527-2533.
[24] Jansen BC, Falck D, de Haan N, et al. LaCyTools: a targeted liquid chromatography-mass spectrometry data processing package for relative quantitation of glycopeptides. J Proteome Res, 2016, 15(7): 2198-2210.
[25] Kirchner M, Saussen B, Steen H, et al. Amsrpm: robust point matching for retention time alignment of LC/MS data with R. J Stat Soft, 2007, 18(4): 1-12.
[26] Tsai TH, Tadesse MG, Di Poto C, et al. Multi-profile Bayesian alignment model for LC-MS data analysis with integration of internal standards. Bioinformatics, 2013, 29(21): 2774-2780.
[27] Ballardini R, Benevento M, Arrigoni G, et al. MassUntangler: a novel alignment tool for label-free liquid chromatography-mass spectrometry proteomic data. J Chromatogr A, 2011, 1218(49): 8859-8868.
[28] Wang J, Lam H. Graph-based peak alignment algorithms for multiple liquid chromatography-mass spectrometry datasets. Bioinformatics, 2013, 29(19): 2469-2476.
[29] Gupta S, Ahadi S, Zhou W, et al. DIAlignR provides precise retention time alignment across distant runs in DIA and targeted proteomics. Mol Cell Proteom, 2019, 18(4): 806-817.
[30] Chambers J, Hastie T, Pregibon D. Statistical models in S. Compstat. Heidelberg: Physica-Verlag HD, 1990: 317-321.
[31] 李龙, 张纪阳, 史秀建, 等. 基于局部回归的色谱保留时间对齐可逆算法. 中南大学学报(自然科学版), 2011, 42(1): 100-105.
Li L, Zhang JY, Shi XJ, et al. Reversible retention time alignment algorithm based on local regression. J Central South Univ (Sci Technol Ed), 2011, 42(1): 100-105 (in Chinese).
[32] Zhang W, Zhang JY, Xu CM, et al. LFQuant: a label-free fast quantitative analysis tool for high-resolution LC-MS/MS proteomics data. Proteomics, 2012, 12(23/24): 3475-3484.
[33] Chang C, Li MS, Guo CP, et al. PANDA: a comprehensive and flexible tool for quantitative proteomics data analysis. Bioinformatics, 2019, 35(5): 898-900.
[34] Bylund D, Danielsson R, Malmquist G, et al. Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography- mass spectrometry data. J Chromatogr A, 2002, 961(2): 237-244.
[35] Wang W, Zhou H, Lin H, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem, 2003, 75(18): 4818-4826.

[36] Johnson KJ, Wright BW, Jarman KH, et al. High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. J Chromatogr A, 2003, 996(1/2): 141-155.
[37] Duran AL, Yang J, Wang L, et al. Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioinformatics, 2003, 19(17): 2283-2293.
[38] Torgrip RJO, Åberg M, Karlberg B, et al. Peak alignment using reduced set mapping. J Chemometrics, 2003, 17(11): 573-582.
[39] Radulovic D, Jelveh S, Ryu S, et al. Informatics platform for global proteomic profiling and biomarker discovery using liquid chromatography- tandem mass spectrometry. Mol Cell Proteomics, 2004, 3(10): 984-997.
[40] Listgarten J, Neal RM, Roweis ST, et al. Multiple alignment of continuous time series. Adv Neural Inform Process Sys, 2005, 17: 17-24.
[41] Higgs RE, Knierman MD, Gelfanova V, et al. Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res, 2005, 4(4): 1442-1450.
[42] Walczak B, Wu W. Fuzzy warping of chromatograms. Chemom Intell Lab Syst, 2005, 77(1/2): 173-180.
[43] Skov T, van den Berg F, Tomasi G, et al. Automated alignment of chromatographic data. J Chemometrics, 2006, 20(11/12): 484-497.
[44] Van Nederkassel AM, Daszykowski M, Eilers PH, et al. A comparison of three algorithms for chromatograms alignment. J Chromatogr A, 2006, 1118(2): 199-210.
[45] Jaitly N, Monroe ME, Petyuk VA, et al. Robust algorithm for alignment of liquid chromatography-mass spectrometry analyses in an accurate mass and time tag data analysis pipeline. Anal Chem, 2006, 78(21): 7397-7409.
[46] Jaffe JD, Mani DR, Leptos KC, et al. PEPPeR, a platform for experimental proteomic pattern recognition. Mol Cell Proteomics, 2006, 5(10): 1927-1941.
[47] Fischer B, Grossmann J, Roth V, et al. Semi-supervised LC/MS alignment for differential proteomics. Bioinformatics, 2006, 22(14): e132-e140.
[48] Sadygov RG, Maroto FM, Hühmer AF. ChromAlign: a two-step algorithmic procedure for time alignment of three-dimensional LC-MS chromatographic surfaces. Anal Chem, 2006, 78(24): 8207-8217.
[49] De Souza DP, Saunders EC, McConville MJ, et al. Progressive peak clustering in GC-MS metabolomic experiments applied to *Leishmania* parasites. Bioinformatics, 2006, 22(11): 1391-1396.
[50] Mueller LN, Rinner O, Schmidt A, et al. SuperHirn- a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics, 2007, 7(19): 3470-3480.
[51] Wang P, Tang H, Fitzgibbon MP, et al. A statistical method for chromatographic alignment of LC-MS data. Biostatistics, 2007, 8(2): 357-367.
[52] Yao WF, Yin XY, Hu YZ. A new algorithm of piecewise automated beam search for peak alignment of chromatographic fingerprints. J Chromatogr A, 2007, 1160(1-2): 254-262.
[53] Fischer B, Roth V, Buhmann JM. Time-series alignment by non-negative multiple generalized canonical correlation analysis. BMC Bioinformatics, 2007, 8(S10): S4.
[54] Christin C, Smilde AK, Hoefsloot HC, et al. Optimized time alignment algorithm for LC-MS data: correlation optimized warping using component detection algorithm-selected mass chromatograms. Anal Chem, 2008, 80(18): 7012-7021.
[55] Suits F, Lepre J, Du PC, et al. Two-dimensional method for time aligning liquid chromatography- mass spectrometry data. Anal Chem, 2008, 80(9): 3095-3104.
[56] Podwojski K, Fritsch A, Chamrad DC, et al. Retention time alignment algorithms for LC/MS data must consider non-linear shifts. Bioinformatics, 2009, 25(6): 758-764.
[57] Valkenborg D, Thomas G, Krols L, et al. A strategy for the prior processing of high-resolution mass spectral data obtained from high-dimensional combined fractional diagonal chromatography. J Mass Spectrom, 2009, 44(4): 516-529.
[58] Tang Z, Zhang L, Amrita KC, et al. A new method for alignment of LC-MALDI-TOF data, 2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)[EB/OL], [2021-05-30]. http://ieeexplore.ieee. org/document/5706589/.
[59] Voss B, Hanselmann M, Renard BY, et al. SIMA: simultaneous multiple alignment of LC/MS peak lists. Bioinformatics, 2011, 27(7): 987-993.
[60] Zhang Z. Retention time alignment of LC/MS data by a divide-and-conquer algorithm. J Am Soc Mass Spectrom, 2012, 23(4): 764-772.
[61] Struck W, Wiczling P, Waszczuk-Jankowska M, et al. New supervised alignment method as a preprocessing tool for chromatographic data in metabolomic studies. J Chromatogr A, 2012, 1256: 150-159.

[62] Kaya H, Gündüz-Öğüdücü Ş. SAGA: a novel signal alignment method based on genetic algorithm. Inf Sci, 2013, 228: 113-130.
[63] Ghanat Bari M, Ma X, Zhang J. PeakLink: a new peptide peak linking method in LC-MS/MS using wavelet and SVM. Bioinformatics, 2014, 30(17): 2464-2470.
[64] Li L, Ren W, Kong H, et al. An alignment algorithm for LC-MS-based metabolomics dataset assisted by MS/MS information. Anal Chim Acta, 2017, 990: 96-102.
[65] Wang Y, Ma L, Zhang M, et al. A simple method for peak alignment using relative retention time related to an inherent peak in liquid chromatography-mass spectrometry-based metabolomics. J Chromatogr Sci, 2019, 57(1): 9-16.
[66] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol, 2008, 26(12): 1367-1372.
[67] Cox J, Hein MY, Luber CA, et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics, 2014, 13(9): 2513-2526.
[68] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry- based shotgun proteomics. Nat Protoc, 2016, 11(12): 2301-2319.
[69] Lim MY, Paulo JA, Gygi SP. Evaluating false transfer rates from the match-between-runs algorithm with a two-proteome model. J Proteome Res, 2019, 18(11): 4020-4026.
[70] Shen XM, Shen SC, Li J, et al. IonStar enables high-precision, low-missing-data proteomics quantification in large biological cohorts. Proc Natl Acad Sci USA, 2018, 115(21): 4767-4776.
[71] The M, Käll L. Focus on the spectra that matter by clustering of quantification data in shotgun proteomics. Nat Commun, 2020, 11: 32-34.
[72] Röst HL, Liu Y, D'Agostino G, et al. TRIC: an automated alignment strategy for reproducible protein quantification in targeted proteomics. Nat Methods, 2016, 13(9): 777-783.
[73] Tsou CC, Avtonomov D, Larsen B, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods, 2015, 12(3): 258-264.
[74] Li Y, Zhong CQ, Xu X, et al. Group-DIA: analyzing multiple data-independent acquisition mass spectrometry data files. Nat Methods, 2015, 12(12): 1105-1106.
[75] Mitra V, Smilde AK, Bischoff R, et al. Tutorial: correction of shifts in single-stage LC-MS(/MS) data. Anal Chim Acta, 2018, 999: 37-53.
[76] Li M, Wang XR. Peak alignment of gas chromatography-mass spectrometry data with deep learning. J Chromatogr A, 2019, 1604: 460-476.
[77] Jiang Y, Sun A, Zhao Y, et al. Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature, 2019, 567(7747): 257-261.

(本文责编　陈宏宇)