

N-糖肽的规模化质谱解析方法进展*

曾文锋^{1,2)} 张 扬³⁾ 刘铭琪³⁾ 吴建强^{1,2)} 张晓今^{1,2)} 杨 皓^{1,2)}
 刘 超¹⁾ 迟 浩¹⁾ 张 昆^{1,2)} 孙瑞祥¹⁾ 杨芃原³⁾ 贺思敏^{1,2)**}

¹⁾中国科学院智能信息处理重点实验室, 中国科学院计算技术研究所, 北京 100190;

²⁾中国科学院大学, 北京 100049; ³⁾复旦大学生物医学研究院, 上海 200032)

摘要 蛋白质糖基化修饰的鉴定是蛋白质翻译后修饰分析中最具挑战性的任务之一, 近几年尤其受到关注. 快速发展的质谱技术为规模化的蛋白质糖基化修饰研究提供了有效的手段. 与其他基于质谱技术的翻译后修饰鉴定相比, 糖基化鉴定的难点在于糖链是大分子而且存在微观不均一性, 另外糖链本身可以在串联质谱中碎裂且与肽段的碎裂规律不同, 导致蛋白质组学的质谱解析方法和软件难以完整地鉴定肽段序列和糖链结构. 完整 N-糖肽的鉴定是糖基化分析的热点内容之一, 针对 N-糖肽的鉴定, 近年来, 人们开发了多种多样的质谱解析方法, 其中包括用 N-糖酰胺酶切除糖链后鉴定 N-糖基化位点的方法、基于电子转运裂解的糖肽肽段鉴定、基于高能碰撞裂解与电子转运裂解联用或碰撞诱导裂解与三级谱联用的完整 N-糖肽鉴定等等. 本文对这些质谱解析方法进行了整理和综述, 简要指出了目前完整糖肽鉴定软件存在的一些不足, 展望了未来的发展方向.

关键词 糖蛋白质组学, 糖肽, 糖基化, 质谱

学科分类号 Q51, TP39

DOI: 10.16476/j.pibb.2015.0402

蛋白质的糖基化修饰是最常见、最重要的蛋白质翻译后修饰之一, 它在蛋白质折叠、寄主病原相互作用等生命活动中起到了关键作用, 而且糖蛋白也是主要的药物作用靶点和临床标志物之一^[1-2]. 糖蛋白的质谱分析可以追溯到 1978 年, Morris 等^[3]利用电子碰撞(EI)和化学电离(CI)技术, 并结合埃德曼降解方法, 分析出了一种南极鱼类的某些抗冻糖蛋白. 早期的糖蛋白质谱鉴定方法的历史脉络可以参考前人的综述^[4-9]. 随着蛋白质组学分析技术和计算技术的发展, 尤其是串联质谱技术的快速发展, 人们开始关注糖蛋白的组学分析, 因此出现了蛋白质组学的子领域——糖蛋白质组学^[9]. 糖蛋白质组学更关注的研究点是对蛋白质的糖基化修饰进行规模化分析, 从而可以从全局的角度研究各种糖蛋白在生命活动中所扮演的角色和所起的作用^[10]. 质谱技术已经成为糖蛋白质组学分析的主要手段, 一些综述对近几年的糖蛋白质谱分析方法进行了总

结^[11-15].

作为最主要的蛋白质糖基化修饰之一, N-糖基化是糖基化修饰分析的热点内容. N-糖基化修饰主要发生在氨基酸序列模体“天冬酰胺-X-丝氨酸/苏氨酸(N-X-S/T)”的天冬酰胺上, 极少量发生在“天冬酰胺-X-半胱氨酸(N-X-C)”的天冬酰胺上(其中 X 可以是除脯氨酸以外的任意氨基酸). 与天冬酰胺相连接的糖链以五糖核心为基础, 一般有三种糖型: 高甘露糖、混合型和复杂型. 目前基于质谱技术的糖蛋白质组学是沿袭自鸟枪法蛋白质组

* 国家自然科学基金(21227805, 31300680), 国家重点基础研究发展计划(973)(2013CB911203, 2012CB910602, 2010CB912701, 2011CB910600)和国家高技术研究发展计划(863)(2014AA020901, 2014AA020902, 2012AA020203)资助项目.

** 通讯联系人.

Tel: 010-62600822, E-mail: smhe@ict.ac.cn

收稿日期: 2015-12-26, 接受日期: 2016-04-21

学, 所以要鉴定糖蛋白, 需要先鉴定糖肽. 与常见的简单翻译后修饰不同, 糖链的多样性使得糖肽的鉴定非常复杂, 一条蛋白质上可能会连接有数十到数百条糖链, 常规的肽段鉴定搜索引擎无法搜索如此多的可变修饰. 糖肽上的糖链本身也容易被串联质谱碎裂, 且糖链的碎裂能量与肽段不尽相同, 如果用简单的质谱技术, 我们很难在一张串联质谱图中同时观察到丰富的糖链碎片和肽段碎片, 所以有时需要综合使用多种技术, 才有望达到规模化完整糖肽鉴定的目的^[13, 16]. 近几年, 尤其是 2010 年后, 串联质谱技术有了显著发展, N-糖蛋白质组学技术也有了长足的进步^[17]. 研究人员纷纷开发出了各种质谱方法进行规模化完整 N-糖肽的鉴定, 包括用 N-糖酰胺酶(PNGase)切除糖链与不切糖链结合的方法、用碰撞诱导裂解(CID)或高能碰撞裂解(HCD)技术鉴定糖肽上的糖链, 或者用电子运转裂解(ETD)技术鉴定糖肽上的肽段, 有些方法采用的是 HCD 与 ETD 联用技术, 或 CID 与三级质谱碎裂结合的技术进行糖肽的鉴定等等.

本文就近几年, 特别是 2010 年后发展起来的 N-糖蛋白质组学的质谱解析方法进行综述: 首先简要介绍了 N-糖肽的不同质谱碎裂模式; 然后根据不同的碎裂模式, 对近几年发展起来的多种 N-糖肽鉴定的质谱解析方法进行了介绍; 最后对当前糖蛋白质组学质谱解析方法所存在的问题进行了总结, 并且展望了未来的发展方向.

1 N-糖肽的质谱碎裂

与肽段的串联质谱鉴定一样, N-糖肽样品经过质谱的二级碎裂后得到实验二级谱图, 然后实验二级谱图与糖肽的理论碎片离子进行匹配或者进行从头测序, 从而鉴定出 N-糖肽样品中的肽段序列和糖链结构.

图 1 展示了常见的糖肽的理论碎片离子, 包括糖肽的 b/y、c/z 离子和 Y 离子等, 以及这些离子对应的化学键断裂位置和命名规则. 目前糖肽糖链部分的鉴定主要用的是糖肽的 Y 离子, 其含义可参照表 1, 本文中用到的其他名称也可以参考表 1.

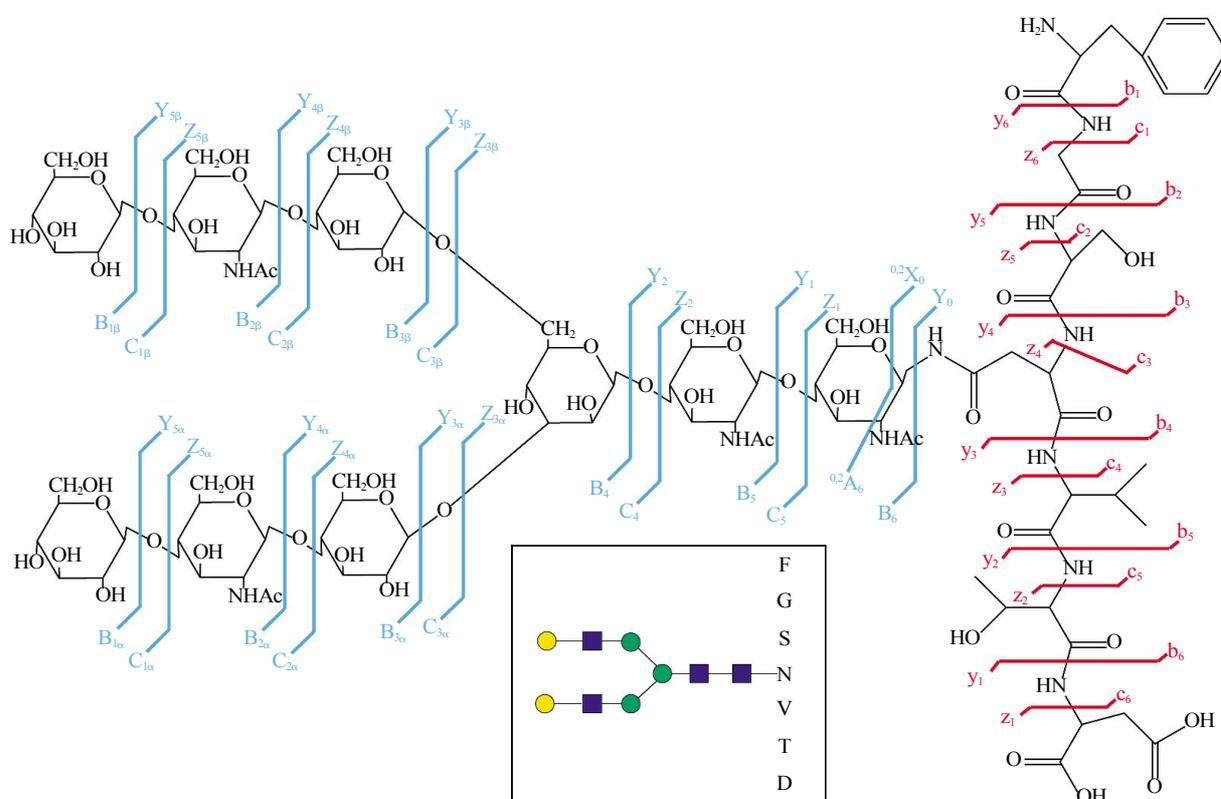


Fig. 1 Theoretical fragments of glycopeptides^[16]

图 1 糖肽的理论碎片离子^[16]

Table 1 Terminologies and their meanings used in this review

表 1 本文用到的名称及其含义

名词	含义
Hex	己糖, 包括葡萄糖、半乳糖和甘露糖等
HexNAc	N-乙酰己糖胺, 包括 N-乙酰葡萄糖胺和 N-乙酰半乳糖胺等
NeuAc	N-乙酰神经氨酸, 唾液酸的一种
NeuGc	N-羟乙酰神经氨酸, 唾液酸的一种
dHex	脱氧己糖, 也称为岩藻糖, 一般也简称为 Fuc
糖肽的 Y 离子	糖链的糖键断裂后产生的连接着完整肽段的碎片离子
糖肽的 $^{2}X_0$ 离子	糖肽的肽段加上 N-糖还原端的 HexNAc 跨环碎片离子
糖肽的裸肽	糖肽的糖链完全碎裂后剩下的肽段部分
糖肽的 b/y 离子	糖肽裸肽的 b/y 离子
糖肽的 c/z 离子	糖肽的完整糖链作为可变修饰附在糖基化位点上(这一点与 b/y 离子不同), 然后肽段的 N-C α 键断裂形成的离子
糖肽的氧鎓离子 (Oxonium Ion)	糖肽的糖链自身碎裂后, 会产生单糖, 或者几个单糖组合, 或者单糖自身碎裂后形成的碎片离子, 详细信息见表 2
一母一谱	质谱中一个母离子被碎裂成一张二级谱图
一母多谱	质谱中一个母离子被碎裂成多张二级谱图

由于单糖 - 单糖之间连接的糖苷键与氨基酸 - 氨基酸之间连接的肽键的键能等理化性质不同, 导致糖链的碎裂与肽段的碎裂有较大差别. 下面我们就最常见的商用质谱仪中的碎裂模式: CID(基于共振激发的离子阱低能量 CID)、HCD 和 ETD, 对糖肽碎裂进行简要介绍.

1.1 N-糖肽的 CID 碎裂

理论上, 糖肽的 Y 离子和 b/y 离子在 CID 碎裂下是可以观测到的. 这里的 CID 特指基于共振激发的离子阱低能量 CID, 它有两个特点: a. 存在 1/3 效应; b. 碎裂能量低, 而且激发碎裂的共振频率只针对母离子质荷比窗口^[18-19]. 所谓 1/3 效应, 是指由于离子阱物理结构和仪器设置的原因, 谱图中小于母离子质荷比 28% 的碎片离子不能由离子阱检测到, 由于糖肽的分子质量一般都比较大, 所以会丢失不少碎片离子. 另外由于 CID 采用的是共振激发的低能量碎裂, 被激发共振的只有母离子质荷比窗口内的离子, 窗口外的离子不会被激发碎裂, 因此当糖肽碎裂以后, 碎片离子就“跳出了”激发窗口, 导致糖肽只发生一次碎裂, 不再发生二次碎裂. 于是低键能的糖苷键优先碎裂, 而肽键就不再碎裂, 所以在 CID 谱图中较容易观察到 Y 离子, 但是很难观察到肽段碎片离子. 如图 2a 所示, 糖肽 VVLHPJYSQVDIGLIK + Hex₅HexNAc₄NeuAc₂(我们为了方便区分, 将糖基化位点 N 改成了字母 J)在 CID 碎裂以后, 低质荷比区域的碎片离子极少(1/3 效应), 而且除了极少量低强度的 b/y 离子(图 2a 中的 b8、b11 离子), 其

他高强度的碎片离子都是糖肽的 Y 离子. 由于 Y 离子较为丰富, 因此 CID 可以鉴定糖链部分, 但是因为缺少肽段碎片离子, 所以 CID 很难独立完成对完整糖肽的鉴定.

1.2 N-糖肽的 HCD 碎裂

糖肽在 HCD 中碎裂的化学键与 CID 是完全相同的, 所以它们的碎片离子类型也完全相同, 然而它们的碎裂环境却完全不同. 因为 HCD 不是由共振激发碎裂, 所以比 CID 更容易发生二次碎裂, 在糖苷键断裂后, 二次碎裂就有可能产生肽段的碎片离子. HCD 的另外一个好处是它利用轨道阱检测离子, 克服了离子阱检测的 1/3 效应, 从而能够检测到低质荷比区域的离子. 因此我们在 HCD 谱图中可以观察到糖肽的氧鎓离子. 图 2b 展示了糖肽 VVLHPJYSQVDIGLIK + Hex₅HexNAc₄NeuAc₂ 的 HCD 碎裂谱图, 其中糖肽的氧鎓离子非常明显, 例如质荷比为 138、204、366 和 657 等碎片离子. 除了氧鎓离子之外, HCD 下糖肽的 Y 离子碎片明显比 CID 下的 Y 离子碎片要少一些, 不过它们依然是 HCD 谱图中的主要离子, 而且更多的 b/y 离子能够在 HCD 谱图中出现. 研究表明如果加大 HCD 的碎裂能量, 则肽段的碎片离子会有所增多^[20]. 随着人们对糖肽 HCD 谱图认识的不断加深, HCD 越来越受到糖肽鉴定流程设计者的青睐^[21-22]. 基于四级杆 - 飞行时间(QTOF)的 CID 碎裂模式与 HCD 基本相同, 理论上所有基于 HCD 的糖肽鉴定方法都可以用在 QTOF-CID 上, 反之亦然.

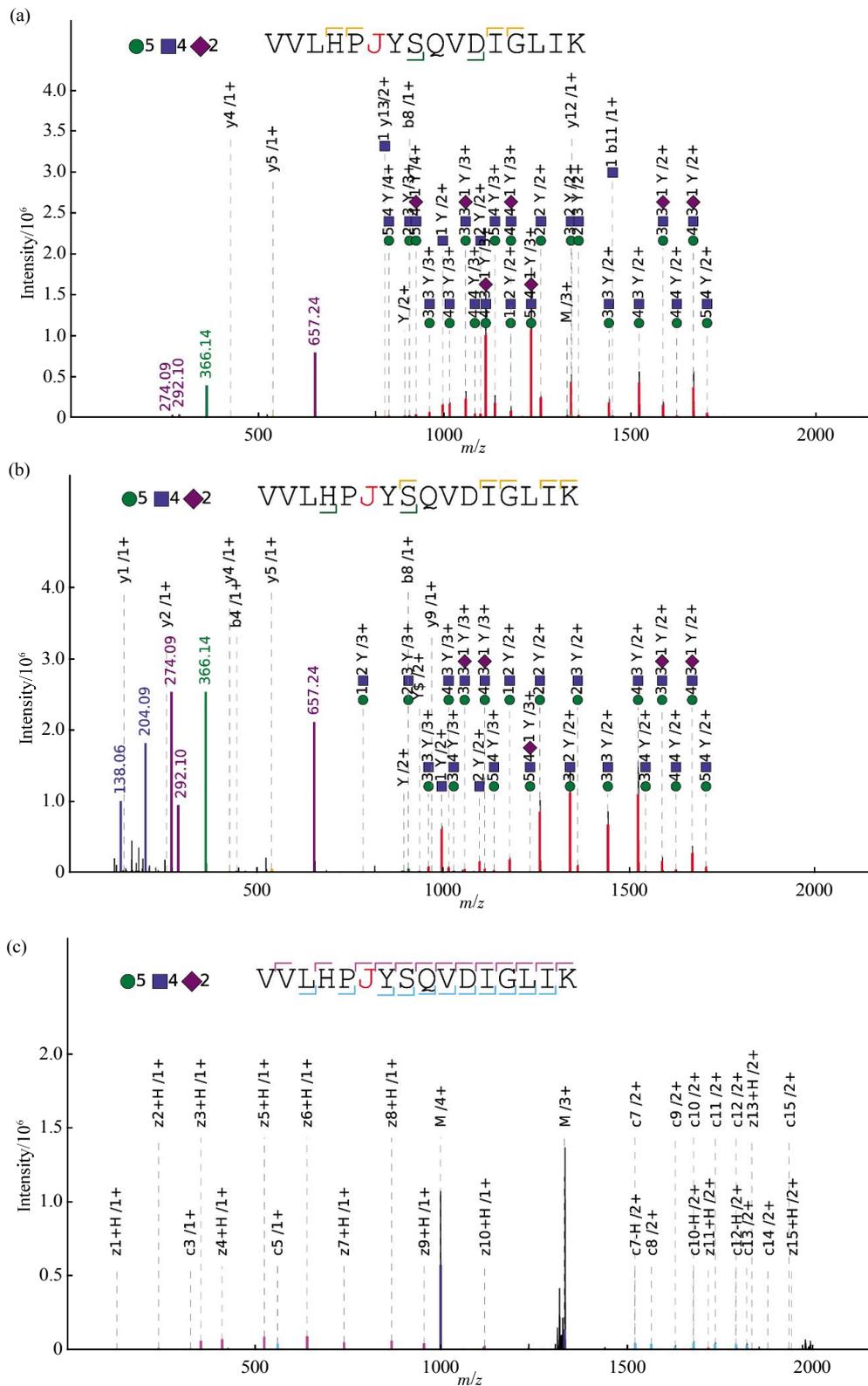


Fig. 2 CID, HCD and ETD MS/MS of an N-glycopeptide

图 2 N-糖肽的 CID、HCD 和 ETD 谱图

(a) CID. (b) HCD. (c) ETD.

1.3 N-糖肽的 ETD 碎裂

ETD 的碎裂与 CID/HCD 完全不同，它主要断裂的化学键为 N-C α 键^[23]，所以一般认为 ETD 主要碎裂的是糖肽肽段部分的 N-C α 键，产生糖肽的 c/z 离子。理想情况下，ETD 能够产生丰富的糖肽 c/z 离子，此时糖链并不碎裂，如同一个完整修饰一样连接在糖基化位点上。图 2c 展示了糖肽 VVLHPJYSQVDIGLIK + Hex₅HexNAc₄NeuAc₂ 的 ETD 谱图，谱图中能观察到高质荷比的 c11、z13、c14 等离子，它们就是由没有碎裂的完整糖链连接在带有糖基化位点的肽段碎片上得到的。另外由于氢重排现象，图 2c 中还可以观察到大量的 z+H 和 c-H 离子，这一点与纯肽段的 ETD 碎裂一致^[23-24]。ETD 的一个缺点是其碎裂效率受到母离子电荷与母离子质荷比等因素的制约^[25-26]，例如图 2c 中的母离子(即图 2c 中的 M 离子)及其同位素峰簇的谱峰

强度很高，说明大部分糖肽母离子都没有充分碎裂。人们也试图采用提高糖肽带电能力的方法以提高 ETD 的碎裂效率^[27-28]。另外 ETD 需要的反应时间较长，导致单位时间内的谱图产出率比 CID 或 HCD 低。虽然 ETD 比之 CID 和 HCD 有自身的缺点，但是 ETD 能够提供较好的糖肽肽段部分的碎片离子信息，所以目前依然是糖肽鉴定的重要手段之一。

2 N-糖肽质谱鉴定流程

不同的碎裂模式提供了不同的碎片离子信息，人们根据这些信息进行了质谱流程设计，以便能够同时鉴定糖肽的糖链部分和肽段部分。图 3 展示了主流的 N-糖肽鉴定的质谱解析流程，主要包括切糖链和不切糖链两个方面。

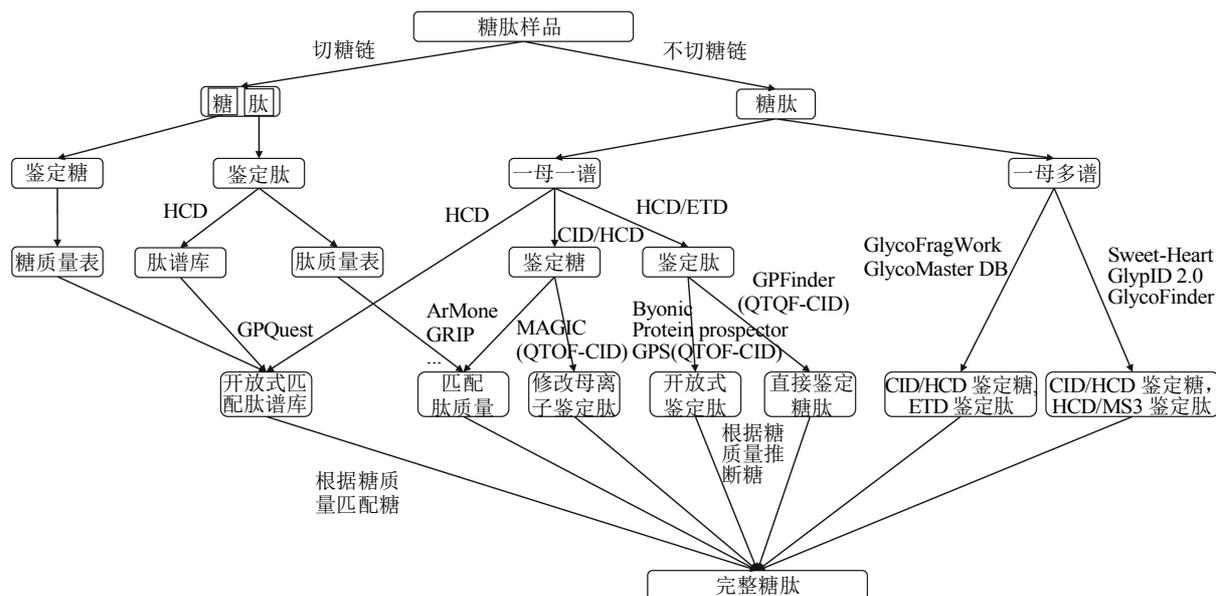


Fig. 3 The overview of N-glycopeptide identification pipelines and software tools

图 3 N-糖肽鉴定流程及其软件概览

在介绍质谱解析方法之前，我们先介绍如何判定一张谱图是否为糖肽谱图。

2.1 N-糖肽二级谱图的筛选

为了提高 N-糖肽鉴定的准确性，需要从所有谱图中筛选出糖肽谱图，使得糖肽鉴定能够避免肽段谱图的干扰，从而提高鉴定的可靠性。糖肽谱图的筛选有通过母离子判定和通过二级谱图判定两种方式。

用母离子判定糖肽的方法就是直接判定一级谱

的母离子谱峰信号是否为糖肽。IsoTag 方法在培养细胞时加入含叠氮的单糖作为糖链自然合成的原料，而糖蛋白上的叠氮会与二溴(Br)化的硅烷探针(dibrominated biotin probe)发生反应，从而在糖链上引入了成对的溴元素。由于自然状态下，溴的同位素的丰度接近相等(⁷⁹Br : ⁸¹Br = 1 : 1)，所以发生二溴化修饰的母离子会存在一个三元组的质量标签，即 M : (M + 2) : (M + 4) = 1 : 2 : 1，通过判断母离子信号的三元组质量标签，就可以判断该母离

子是否为糖肽母离子^[29]。另外一种通过母离子判断糖肽的方法利用的是糖肽与非糖肽母离子质量的差异信息。由于糖链的氧元素比较多, 所以糖肽母离子的质量与其取整质量之间的差和非糖肽是不同的(氧原子的分子质量为 15.99 u, 比它取整后的分子质量 16 略小, 而氢原子和氮原子都比它们取整后的分子质量略大, 碳原子的分子质量与取整后的质量相等)。利用这一点, 文献[30]通过模拟实验, 得到了一个可以区分糖肽母离子与肽段母离子的简单分类器, 它在实际数据上可以达到 89% 以上的灵敏度和 93% 以上的特异性。不过此方法要求糖肽的糖链比较大, 如此糖肽分子中氧原子才会比较多, 可区分性较强。

目前最常用的方法是通过 HCD(或 QTOF-CID) 二级谱图中的氧鎓离子判断谱图是否为糖肽谱图。本文对常见的氧鎓离子进行了收集和总结, 如表 2 所示。在 N-糖肽谱图的氧鎓离子中, 一般强度最高的为 204 和 366 两种离子, 它们分别是 HexNAc 和 Hex + HexNAc 的单电荷形式, 有时 138 谱峰也是强度最高的氧鎓离子之一, 它是 HexNAc 某种跨环碎裂的碎片离子的单电荷形式。现在所有的基于 HCD 和 QTOF-CID 的鉴定软件在设计时, 或多或少都会根据 204 和 366(或 138)谱峰判断谱图是否为糖肽谱图。

Table 2 Oxonium ions of glycopeptides, their types and formulas

表 2 糖肽的氧鎓离子及其对应的类型和化学式

氧鎓离子(<i>m/z</i>)	类型	化学式
109.028	Hex 的碎片	C ₆ H ₆ O ₂
115.039	Hex 的碎片	C ₅ H ₆ O ₃
126.055	HexNAc 的碎片	C ₈ H ₁₀ O ₂ N ₁
127.039	Hex - 2H ₂ O	C ₆ H ₆ O ₃
138.055	HexNAc 的碎片	C ₈ H ₁₀ O ₂ N ₁
144.066	HexNAc 的碎片	C ₈ H ₉ O ₃ N ₁
163.060	Hex	C ₆ H ₁₀ O ₅
168.066	HexNAc - 2H ₂ O	C ₈ H ₉ O ₃ N ₁
186.076	HexNAc - H ₂ O	C ₈ H ₁₁ O ₄ N ₁
204.087	HexNAc	C ₈ H ₁₅ O ₅ N ₁
274.092	NeuAc - H ₂ O	C ₁₁ H ₁₅ O ₇ N ₁
290.087	NeuGc - H ₂ O	C ₁₁ H ₁₅ O ₈ N ₁
292.103	NeuAc	C ₁₁ H ₁₇ O ₈ N ₁
308.098	NeuGc	C ₁₁ H ₁₇ O ₉ N ₁
366.140	Hex + HexNAc	C ₁₄ H ₂₅ O ₁₀ N ₁
657.140	Hex + HexNAc + NeuAc	C ₂₅ H ₄₀ O ₁₈ N ₂
673.230	Hex + HexNAc + NeuGc	C ₂₅ H ₄₀ O ₁₉ N ₂

2.2 切糖链的 N-糖肽鉴定流程

N-糖肽的糖链可以用 N-糖酰胺酶进行酶切, 使得糖链从天冬酰胺 N 上切除, 留下的 N 脱酰胺后变成天冬氨酸 D(D 的分子质量 = N 的分子质量 + 0.984 u), 如果切糖链是在 H₂¹⁸O 中完成的, 那么该分子质量差就变成 2.988 u。我们对切糖链后的肽段样品进行常规的质谱鉴定, 如果鉴定结果中合理的糖基化位点 N 上发生了脱酰胺修饰, 则认为这个 N 是切糖链后得到的糖基化位点, 从而完成糖肽肽段部分的鉴定^[17]。接着需要进行糖肽糖链部分的鉴定, 其鉴定是不需要切除糖链的, 可以直接将(富集后的)完整糖肽样品进行 CID 或者 HCD 碎裂, 而后用从头测序或者糖库搜索的方法鉴定糖链。复旦大学杨芄原 / 张扬团队开发的 GRIP 使用的是糖链从头测序方法鉴定糖链^[31]。他们先利用切糖链后鉴定到的去糖基化糖肽构建了肽段库(本文中去除糖基化糖肽或去糖基化肽段指的是糖肽经切除糖链后剩下的肽段部分, 而糖肽指的是完整糖肽), 然后基于此肽段库进行糖链组成的从头测序鉴定, 并且利用 HCD 的氧鎓离子筛选糖肽谱图, 最后用 Y₁ 和 ⁰²X₀ 离子验证糖链鉴定的正确性。中国科学院大连化学物理研究所邹汉法 / 叶明亮团队开发的 ArMone 2.0 也是基于切糖链流程的 N-糖肽鉴定软件^[32]。他们首先根据 HCD 谱图的氧鎓离子筛选糖肽谱图, 然后利用糖链数据库 Glycome-DB^[33]和糖肽的 HCD 谱图进行匹配, 利用糖肽的五糖核心 Y 离子确定糖链, 并且推断出糖肽的肽段质量。他们发现相同的肽段序列, 加糖链和不加糖链的保留时间相差不大, 所以他们结合保留时间信息, 将推断出的肽段质量与去糖基化的糖肽进行了匹配。GPQuest 则采用了不同的切糖链鉴定策略^[34], 根据切糖链鉴定得到的肽段结果建立一个肽段谱图库。在糖肽经过 HCD 碎裂以后, 他们利用糖肽的 Y₀、Y₁、⁰²X₀ 和 Y₂ 等离子与谱图库中的肽段母离子匹配确定肽段质量, 其他糖肽肽段的理论碎片离子与谱库中的碎片离子匹配以鉴定糖肽的肽段, 从而可以推断出糖链的质量。GPQuest 将切糖链后得到的纯糖链进行质谱鉴定, 得到可能的糖链组成表。最后将推断出的糖肽糖链质量与糖链组成表进行匹配, 从而完成完整糖肽的鉴定。

N-糖肽的切糖链鉴定流程能够达到较高的鉴定通量。但是去糖基化的糖肽是通过脱酰胺修饰鉴定的, 用它来百分之百确定糖肽的肽段部分可能存在风险^[35], 我们不能排除生物体内或者处理样品时

天冬酰胺自身引入的脱酰胺修饰对切糖链鉴定造成的影响。为了避免这个问题，人们开发了针对 N-糖肽的固相萃取方法，该方法在切糖链之前加入苯胺与天冬氨酸进行反应，使天然的天冬氨酸加入了苯胺修饰，则切糖链后的没有苯胺修饰的天冬氨酸可以认为是 N-糖基化位点^[36]。

然而即使能够解决切糖链后引入的位点误判问题，根据糖链质量推断糖链的方法依然存在一个难以避免的风险，即假如两个糖链质量相等，仅靠质量是难以区分二者的。例如只靠质量是不能区分 Hex₁NeuAc₁ 与 dHex₁NeuGc₁ 的，因为 Hex₁NeuAc₁ 与 dHex₁NeuGc₁ 的化学式完全一样，它们质量也完全相等。同理，根据糖肽的肽段质量去匹配肽段库的方法也可能存在同样的缺陷。因此如果可行，对于糖肽的鉴定，不切糖链是更为可靠的方法。

2.3 不切糖链的 N-糖肽鉴定流程

不切糖链的 N-糖肽鉴定流程指的是直接将完整 N-糖肽样品进行质谱分析鉴定的流程。根据同一个母离子是否产生多张二级谱图，该流程可以分为一母一谱和一母多谱两种模式，下面我们进行详细讨论。

2.3.1 一母一谱的 N-糖肽鉴定流程

一母一谱的糖肽鉴定就是同一个母离子只产生一张二级谱图(CID、HCD 或者 ETD 谱图)，然后直接通过对该谱图的解析完成一条糖肽的鉴定。该流程对前期实验的要求较低，质谱流程也较为简单，但是对软件的要求比较高。Protein Prospector^[37] 在 HCD 谱图中筛选出糖肽谱图以后，它先开放地鉴定糖肽的 ETD 谱图，而后将开放式搜索得到的意外修饰质量认为是候选的糖链质量，通过查糖链质量表得到糖链的组成。Byonic 将糖链组成库作为可变修饰搜索糖肽的 ETD 或者 HCD 谱图以鉴定肽段，也具备类似功能。GlycoPep Detector^[38] 也支持糖肽 ETD 谱图鉴定，它预先生成了糖肽组合然后计算出所有糖肽的理论 c/z 离子，接着去匹配 ETD 谱图。GlycoPeptideSearch(GPS)^[39-40] 在对糖肽的 QTOF-CID 谱图进行搜索时，除了考虑肽段的 b、y 离子外，还考虑了糖链的 Y₁、Y₂ 和 Y₃ 离子，以确定肽段质量和糖链质量，最后糖链质量与 Glycome-DB 糖库匹配得到糖链组成。

利用 QTOF-CID 谱图同时鉴定糖链与肽段的软件主要有两款：GlycoPeptide Finder(GPFinder)^[41] 和 MAGIC^[42]。GPFinder 是针对非特异酶切糖肽的

QTOF-CID 谱图所设计的软件，它对糖链的鉴定并不是根据糖链质量推断的，它同时考虑了 QTOF-CID 谱图中的糖链碎片和肽段碎片，根据 Y 离子糖组成的合法性过滤掉不合理的 Y 离子，然后设计了一种同时考虑糖链碎片与肽段碎片的打分方法以同时鉴定糖肽。MAGIC 采取了较为不同的鉴定策略。它首先根据(Y₀-NH₃, Y₀, Y₂)或(Y₀, Y₁, Y₂)谱峰三元组确定糖链质量和肽段质量，然后用 0-1 背包模型对糖链碎片离子的匹配进行建模以鉴定糖链组成。完成糖链组成的鉴定后，MAGIC 删除了二级谱图中所有糖链碎片离子，并且将二级谱图母离子质量改成肽段质量，最后用常规的蛋白质搜索引擎鉴定肽段。

目前，一母一谱存在的最大问题是有时单张谱图提供的碎片离子信息有限，不足以同时鉴定糖链和肽段。结合图 2 我们可以看到，CID 或者 HCD 可以提供丰富的糖肽糖链碎片离子，但是肽段的碎片离子还不够丰富，虽然 ETD 可以碎裂肽段，但是它没有提供糖链碎片离子的信号。所以在谱图碎片离子信息不够的情况下，利用一母一谱鉴定糖肽的流程还具有较大的改进空间。

2.3.2 一母多谱的 N-糖肽鉴定流程

在当前常规质谱条件下，单种碎裂提供的谱图信息有时不足以同时鉴定糖肽的糖链和肽段，所以需要综合多种碎裂的信息以达到完整糖肽鉴定的目的。当前有一些质谱仪提供了一个母离子顺序完成两种或者两种以上碎裂的功能，使得基于一母多谱的 N-糖肽鉴定成为可能。

比较常用的一母多谱的碎裂模式为 HCD + ETD，其中 HCD 提供了氧鎓离子以判断谱图是否糖肽，而后还可以进行糖链的鉴定，而 ETD 则可以提供肽段的碎片离子以鉴定肽段。近几年，由于仪器提供了碎片离子触发(product-dependent)的功能，使得一母多谱的技术更加方便了用户。母离子进行 HCD 碎裂以后，仪器能够根据用户设定的氧鎓离子是否出现来判断母离子是否为糖肽，进而触发该糖肽母离子的 ETD 碎裂，此模式称为 HCD-pd-ETD^[43-44]。由于 ETD 反应时间较长，离子触发技术能够避免 ETD 的反应时间浪费在非糖肽样品上。GlycoMasterDB 可以支持 HCD + ETD 碎裂模式的糖肽鉴定，如果只有 HCD 谱图，则它只给出糖链组成的鉴定结果，而肽段需要人工解析^[45]。由于 HCD 提供的糖链碎片离子信息可能不如 CID

丰富, 所以人们也开发了一母三谱的糖肽鉴定模式, 即 CID + HCD + ETD. 目前支持此三组合碎裂的软件为 GlycoFragwork^[46], 它用从头测序的方法在 CID 谱图上推测糖链结构, 用 ETD 鉴定肽段, 而 HCD 则提供了氧鎓离子以判断糖肽谱图以及糖链类型. 如果没有 ETD 谱图, GlycoFragwork 或者软件 GlypID 2.0^[47]也可以只利用 CID + HCD 完成糖肽的鉴定, 但是此时肽段就只能通过肽段质量匹配得到. GlycoFinder^[48]则使用低能量的 HCD 代替 CID 以鉴定糖链组成, 并且使用高能量的 HCD 鉴定糖肽的肽段.

另一种一母多谱的 N-糖肽鉴定流程为 Sweet-Heart^[49-50], 它首先建立了一批基于低精度 CID 的糖肽标注数据集, 接着考虑糖链碎片的序列标签、N-糖肽的五糖核心离子和匹配谱峰强度等特征, 训练得到一个随机森林模型. 而后 Sweet-Heart 用该模型对糖肽低精度 CID 谱图的糖链组成进行预测, 从而可以推断出肽段的可能质量, 再用靶向三级谱(targeted-MS3)的方式碎裂并且鉴定肽段, 进一步确认糖链预测结果, 最后糖链与肽段都完成了鉴定. 这种方式并不是传统的一母多谱, 因为目标三级谱实验必须在糖链预测之后进行, 需要额外进行一次质谱实验, 所以这种方式的自动化程度不够高. 后来 Sweet-Heart 又发展了另外一种流程, 即 Sweet-Heart for HCD^[51], 它利用 HCD 谱图中的 Y_0/Y_1 或 Y_1/Y_2 离子二元组推测出 Y_1 离子的质量, 然后将谱图的母离子质量改成 Y_1 离子的质量, 再用常规的蛋白质搜索引擎鉴定修改后的谱图得到肽段, 而糖链依然可以依靠 CID 谱图上的随机森林模型进行预测, 最终结合 CID 和 HCD 完成完整糖肽的鉴定. 当然该流程也可以额外配合 HCD/ETD 谱图的开放式肽段搜索以鉴定肽段.

文献[51]用 Sweet-Heart for HCD 流程的鉴定结果与 Byonic 的糖肽鉴定结果进行比较并且手工验证后, 发现虽然 Byonic 报告的鉴定结果的错误率为 0%, 但是它实际错误率却高达 37%, 他们认为造成高假阳鉴定的主要原因是 Byonic 缺少对糖链鉴定的错误率控制, 说明了对于完整糖肽的鉴定, 只针对肽段部分进行错误率控制是不够的.

我们通过切糖链和不切糖链两种流程的对比可以看出, 除了利用质量匹配可能降低切糖链流程鉴定结果的可靠性外, 这两种流程的另一个区别主要体现在肽段部分的鉴定上. 在一般的切糖链流程

中, 去糖基化的糖肽会与非糖肽样品混合在一起进入质谱进行二级碎裂, 由于缺少像氧鎓离子这样的谱图特征去区分去糖基化糖肽与非糖肽谱图, 因此所有的谱图都需要搜索同时包含糖肽肽段和非糖肽肽段的蛋白质序列数据库, 这会引起两种错误鉴定, 导致错误鉴定的可能性增加: a. 去糖基化的糖肽谱图被鉴定成了非糖肽的肽段; b. 非糖肽谱图被鉴定成了去糖基化糖肽. 所以总体上讲, 不切糖链的流程天然保留了糖链连接位点的信息, 它的鉴定结果更容易被研究人员所接受.

2.4 N-糖肽中核心岩藻糖的鉴定流程

核心岩藻糖化的 N-糖基化是非常重要的 N-糖基化修饰^[52-54], 设计专门针对核心岩藻糖鉴定的流程是非常有意义的. 带有核心岩藻糖的糖肽会被内切糖苷酶(endoglycosidase)F3 切成“肽段 + HexNAc₁dHex₁”形式的糖肽. 基于此, 军事医学科学院-北京蛋白质组研究中心的钱小红/应万涛团队开发了一系列核心岩藻糖的鉴定流程. 在早先的带线性离子阱 CID 的质谱仪上, 由于共振激发与低能量碎裂的原因, 核心岩藻糖很容易断裂, 形成高强度的“肽段 + HexNAc₁”谱峰, 但是肽段不容易碎裂, 所以在“肽段 + HexNAc₁”谱峰与谱图母离子之间存在一个固定的 146 u 质量差(dHex 质量为 146 u), 根据此质量差触发三级碎裂, 接着可以用 pFind(<http://pfind.ict.ac.cn>)对该三级谱图进行搜索鉴定, 得到“肽段 + HexNAc₁”, 从而推断出“肽段 + HexNAc₁dHex₁”的形式^[55]. 后来该团队通过解析谱图发现, 低能量的 HCD 与线性离子阱一样只能碎裂得到“肽段 + HexNAc₁”谱峰, 而中等能量的 HCD 碎裂则可以得到肽段的碎片, 高能量的 HCD 碎裂则可以产生丰富的氧鎓离子. 于是他们利用仪器提供的阶梯能量碎裂功能, 将低、中、高能量的 HCD 碎裂到一张谱图上, 从而可以完成高通量的核心岩藻糖化修饰肽段的鉴定, 并且由常规肽段鉴定的质控方法进行质控^[56]. 在用技术重复保证可靠性的前提下, 他们在 78 min 的梯度上能够完成鼠肝样品的 1 364 条核心岩藻糖化肽段的鉴定. 该流程的缺点就是它只能鉴定核心岩藻糖化修饰的肽段, 但是无法鉴定核心岩藻糖化的完整糖链.

上述不少流程的分析软件都可以免费下载, 有些甚至是源代码开放的. 我们在表 3 中列出了相关软件的下链接.

Table 3 Download link of N-glycopeptide identification software tools

表 3 N-糖肽鉴定软件下载链接

软件	描述	下载链接
ArMone 2.0	支持切糖链流程, 切糖链的肽段需要由其他肽段鉴定软件完成鉴定.	http://www.bioanalysis.dicp.ac.cn/proteomics/software/ArMone2.html
MAGIC	支持 QTOF-CID 的一母一谱流程. 肽段需要由其他肽段鉴定软件完成鉴定.	http://ms.iis.sinica.edu.tw/COMics/Software_MAGIC.html
Sweet-Heart 或 Sweet-Heart for HCD	支持 CID/MS3 或 CID/HCD 一母双谱的流程. 肽段需要由其他肽段鉴定软件完成鉴定.	http://sweet-heart.glycoproteomics.proteome.bc.sinica.edu.tw/download
GlycoFragwork	支持 CID/HCD(ETD)一母多谱的流程.	http://darwin.informatics.indiana.edu/col/GlycoFragwork/download.html
GlypID 2.0	支持 CID/HCD 一母双谱的流程.	http://mendel.informatics.indiana.edu/~chuyu/glypID/software.html
ProteinProspector	开放式搜索 ETD 的一母一谱流程. 网页版软件.	http://prospector.ucsf.edu/prospector/mshome.htm
GPFinder	支持 QTOF-CID 的一母一谱流程. 支持糖肽错误率估计. 源代码开放.	https://github.com/gpfinder/gpfinder
GlycoPeptideSearch	支持 QTOF-CID 的一母一谱流程. 支持糖肽错误率估计.	https://edwardslab.bmcb.georgetown.edu/trac/GlycoPeptideSearch/wiki/GPS/Download
GlycoMaster DB	支持 HCD 一母一谱或者 HCD/ETD 一母双谱流程. 网页版软件.	http://www-novo.cs.uwaterloo.ca:8080/GlycoMasterDB/

3 总结与展望

虽然目前存在诸多糖肽质谱解析方法, 但是不同方法得到的鉴定结果其可靠性是不同的, 我们将之分为 4 个层次, 分别为:

a. 完整糖肽的二级谱图用于鉴定糖链, 肽段靠质量进行匹配. 单通过质量不能确定糖肽谱图的肽段, 例如肽段 ACDEFK 和 FEDCAK 的质量是完全相等的. 蛋白质组学研究结果表明, 如果仅靠肽段质量而没有二级谱碎片离子的匹配打分信息还不足以达到“鉴定”的目标. 此问题主要出现在切糖链流程上, 该流程只根据谱图推测得到的肽段质量与切糖链鉴定的肽段进行匹配得到肽段结果.

b. 完整糖肽的二级谱图用于鉴定肽段, 糖链靠质量进行匹配. 与层次 a 类似, 仅靠质量有时也无法区分一些糖链组成, 例如 $\text{Hex}_3\text{HexNAc}_4\text{NeuAc}_1\text{dHex}_1$ 和 $\text{Hex}_6\text{HexNAc}_4\text{NeuGc}_1$ 是同时存在于自然界的糖组成, 但是二者的质量是完全相等的, 所以也需要糖链的二级谱碎片离子信息来区分. 许多软件都只通过糖肽谱图推测出糖链质量, 然后匹配糖组成库, 比如 GPQuest、Protein Prospector、Byonic 和 GPS 等等.

c. 完整糖肽的二级谱图同时用于糖链和肽段的匹配打分, 但是缺少糖链鉴定的错误率估计. 大

部分基于 QTOF-CID(或 HCD)和一母多谱的流程都能在糖肽谱图中同时鉴定糖链和肽段, 而且可以用基于肽段的目标诱饵库模型进行肽段鉴定的错误率估计, 例如 GlycoFragwork、GlycoMaster DB、MAGIC 和 GlycoPep Evaluator^[57]等, 然而它们依然没有专门考虑糖链鉴定的错误率.

d. 完整糖肽的二级谱图同时用于糖链和肽段的匹配打分, 而且同时具有糖链和肽段鉴定错误率控制的机制. 目前只有基于 QTOF-CID 的 GPFinder 单独考虑了糖链鉴定的错误率, 它利用糖链加 11 u 建立诱饵糖链, 从而使用目标诱饵库模型以估计糖链鉴定错误率^[41]. 如果加上肽段的目标诱饵库策略, 则可以同时估计肽段和糖链的错误率. 当然该方法目前只使用在简单样品中, 如果要用在复杂体系中, 则需要优化模型以得到更可靠的错误率估计结果.

对于多数流程, 其鉴定可靠性的不足主要来自糖肽解析软件的不足, 其中一个主要因素是解析软件缺少同时针对糖链和肽段的错误率估计机制. 对于未来的糖肽鉴定软件, 我们还是希望它的可靠性能够达到层次 d, 只有肽段和糖链的鉴定可靠性同时得到了保证, 后期的生物学分析才更加有意义.

除了可靠性, 糖肽鉴定还存在一些问题需要解决: a. 通量问题. 虽然该流程不能鉴定完整糖链,

但是钱小红 / 应万涛团队的核心岩藻糖鉴定流程能够在 78 min 的梯度上完成鼠肝样品上 1 364 条核心岩藻糖化肽段的鉴定。而在同时控制糖肽的肽段和糖链错误率的前提下, 目前的完整糖肽鉴定流程很难达到相同的鉴定通量。b. 糖结构库的完整性问题。由于我们对糖基转移酶的研究还处于探索阶段, 所以糖链在生物体内的合成还未发现模板, 现在使用的大部分糖结构都是由生物化学研究人员从实验中得出的, 例如 Glycome-DB。然而在实际应用中我们也许希望鉴定到新的糖链结构, 以发现新的功能。从头测序也许是解决方案之一^[58], 但是用从头测序方法推测糖链结构是极为困难的^[59]。所以如何生成合理、完整的理论糖结构库是需要解决的问题之一。c. 精确的糖结构鉴定。即使有了糖链结构库, 也并不意味着我们就能够精确地鉴定糖链的结构, 例如单糖与单糖的六圆环之间存在连接位点, 只利用 Y 离子信息是无法得到六圆环连接位点信息的。此时鉴定软件需要挖掘更加精细的糖肽碎片离子信息, 以辅助推断单糖与单糖之间的连接位点, 然后推测出更加精确的糖链结构。d. 软件的易用性、软件的速度和糖肽谱图的自动标注等都是糖肽解析软件需要考虑的问题, 文献[60]对这方面的需求进行了很好的综述。

另外在蛋白质糖基化修饰中, O 糖也是人们关注的一类糖基化修饰, 但是其高通量的鉴定流程与软件却不如 N-糖基化的丰富, 而且我们在实际研究中发现, 一条肽段上经常有两个甚至两个以上位点都发生了 O 糖基化修饰, 文献[61]也报道了羟基赖氨酸 O 糖基化的多修饰位点问题, 这给鉴定带来更大的挑战。肽段也可能发生多位点上的 N-糖基化修饰, 但是实际研究中我们发现该现象还没有那么普遍。

近几年随着糖科学越来越受到人们关注, 作为糖科学分支的糖蛋白质组学也有了不小的进展。2014 年美国国家卫生研究院专门设立了基金以支持糖生物学、糖化学方法的研究和糖软件工具的开发, 而国内的一些研究机构也正不断地推进糖蛋白质组学的研究。相信在国内外糖生物学家、糖化学家和生物信息学家的共同努力下, 不仅质谱流程会更加高效, 软件工具也会更加简单易用, 最终上述问题将会得到有效解决。

参 考 文 献

- [1] Kolarich D, Lepenies B, Seeberger P H. Glycomics, glycoproteomics and the immune system. *Curr Opin Chem Biol*, 2012, **16**(1-2): 214-220
- [2] Moremen K W, Tiemeyer M, Nairn A V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*, 2012, **13**(7): 448-462
- [3] Morris H R, Thompson M R, Osuga D T, *et al.* Antifreeze glycoproteins from the blood of an antarctic fish. The structure of the proline-containing glycopeptides. *J Biol Chem*, 1978, **253**(14): 5155-5162
- [4] Dell A, Morris H R. Glycoprotein structure determination by mass spectrometry. *Science*, 2001, **291**(5512): 2351-2356
- [5] Geyer H, Geyer R. Strategies for analysis of glycoprotein glycosylation. *Biochim Biophys Acta*, 2006, **1764**(12): 1853-1869
- [6] Haslam S M, North S J, Dell A. Mass spectrometric analysis of N- and O-glycosylation of tissues and cells. *Curr Opin Struct Biol*, 2006, **16**(5): 584-591
- [7] Wührer M, Catalina M I, Deelder A M, *et al.* Glycoproteomics based on tandem mass spectrometry of glycopeptides. *J Chromatogr B Analyt Technol Biomed Life Sci*, 2007, **849**(1-2): 115-128
- [8] Zaia J. Mass spectrometry and the emerging field of glycomics. *Chem Biol*, 2008, **15**(9): 881-892
- [9] Tissot B, North S J, Ceroni A, *et al.* Glycoproteomics: past, present and future. *FEBS Lett*, 2009, **583**(11): 1728-1735
- [10] Pan S, Chen R, Aebersold R, *et al.* Mass spectrometry based glycoproteomics--from a proteomics perspective. *Mol Cell Proteomics*, 2011, **10**(1): R110 003251
- [11] Leymarie N, Zaia J. Effective use of mass spectrometry for glycan and glycopeptide structural analysis. *Anal Chem*, 2012, **84**(7): 3040-3048
- [12] Woodin C L, Maxon M, Desaire H. Software for automated interpretation of mass spectrometry data from glycans and glycopeptides. *Analyst*, 2013, **138**(10): 2793-2803
- [13] Alley W R, Jr., Mann B F, Novotny M V. High-sensitivity analytical approaches for the structural characterization of glycoproteins. *Chem Rev*, 2013, **113**(4): 2668-2732
- [14] Zhang Y, Jiao J, Yang P, *et al.* Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clin Proteomics*, 2014, **11**(1): 18
- [15] Zhu Z, Desaire H. Carbohydrates on proteins: site-specific glycosylation analysis by mass spectrometry. *Annu Rev Anal Chem (Palo Alto Calif)*, 2015, **8**: 463-483
- [16] Dodds E D. Gas-phase dissociation of glycosylated peptide ions. *Mass Spectrom Rev*, 2012, **31**(6): 666-682
- [17] Zielinska D F, Gnad F, Wisniewski J R, *et al.* Precision mapping of an *in vivo* N-glycoproteome reveals rigid topological and sequence constraints. *Cell*, 2010, **141**(5): 897-907
- [18] de Hoffmann E, Stroobant V. *Mass Spectrometry: Principles and Applications*. 3rd ed, Wiltshire: Wiley. 2012: 100-117
- [19] Hu H, Khatri K, Klein J, *et al.* A review of methods for interpretation of glycopeptide tandem mass spectral data. *Glycoconj J*, 2015(published online)
- [20] Khatri K, Staples G O, Leymarie N, *et al.* Confident assignment of

- site-specific glycosylation in complex glycoproteins in a single step. *J Proteome Res*, 2014, **13**(10): 4347–4355
- [21] An Y, Rininger J A, Jarvis D L, *et al.* Comparative glycomics analysis of influenza Hemagglutinin (H5N1) produced in vaccine relevant cell platforms. *J Proteome Res*, 2013, **12**(8): 3707–3720
- [22] An Y, Cipollo J F. An unbiased approach for analysis of protein glycosylation and application to influenza vaccine hemagglutinin. *Anal Biochem*, 2011, **415**(1): 67–80
- [23] 孙瑞祥, 董梦秋, 迟浩, 等. 基于电子捕获裂解/电子转运裂解串联质谱技术的蛋白质组学研究. *生物化学与生物物理进展*, 2010, **37**(1): 94–102
- Sun R X, Dong M Q, Chi H, *et al.* *Prog Biochem Biophys*, 2010, **37**(1): 94–102
- [24] Sun R X, Dong M Q, Song C Q, *et al.* Improved peptide identification for proteomic analysis based on comprehensive characterization of electron transfer dissociation spectra. *J Proteome Res*, 2010, **9**(12): 6354–6367
- [25] Frese C K, Altalear A F, Hennrich M L, *et al.* Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J Proteome Res*, 2011, **10**(5): 2377–2388
- [26] Blackburn K, Mbeunkui F, Mitra S K, *et al.* Improving protein and proteome coverage through data-independent multiplexed peptide fragmentation. *J Proteome Res*, 2010, **9**(7): 3621–3637
- [27] Yin X, Bern M, Xing Q, *et al.* Glycoproteomic analysis of the secretome of human endothelial cells. *Mol Cell Proteomics*, 2013, **12**(4): 956–978
- [28] Ye H, Boyne M T, 2nd, Buhse L F, *et al.* Direct approach for qualitative and quantitative characterization of glycoproteins using tandem mass tags and an LTQ Orbitrap XL electron transfer dissociation hybrid mass spectrometer. *Anal Chem*, 2013, **85**(3): 1531–1539
- [29] Woo C M, Iavarone A T, Spicirich D R, *et al.* Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods*, 2015, **12**(6): 561–567
- [30] Froehlich J W, Dodds E D, Wilhelm M, *et al.* A classifier based on accurate mass measurements to aid large scale, unbiased glycoproteomics. *Mol Cell Proteomics*, 2013, **12**(4): 1017–1025
- [31] Liu M, Zhang Y, Chen Y, *et al.* Efficient and accurate glycopeptide identification pipeline for high-throughput site-specific N-glycosylation analysis. *J Proteome Res*, 2014, **13**(6): 3121–3129
- [32] Cheng K, Chen R, Seebun D, *et al.* Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. *J Proteomics*, 2014, **110**(14): 145–154
- [33] Ranzinger R, Herget S, von der Lieth C W, *et al.* GlycomeDB—a unified database for carbohydrate structures. *Nucleic Acids Res*, 2011, **39**(Issue suppl 1): D373–376
- [34] Toghi Eshghi S, Shah P, Yang W, *et al.* GPQuest: a spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact N-glycopeptides. *Anal Chem*, 2015, **87**(10): 5181–5188
- [35] Palmisano G, Melo-Braga M N, Engholm-Keller K, *et al.* Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res*, 2012, **11**(3): 1949–1957
- [36] Sun S, Shah P, Eshghi S T, *et al.* Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat Biotechnol*, 2015 (published online)
- [37] Medzhradszky K F, Kaasik K, Chalkley R J. Tissue-specific glycosylation at the glycopeptide level. *Mol Cell Proteomics*, 2015, **14**(8): 2103–2110
- [38] Zhu Z, Hua D, Clark D F, *et al.* GlycoPep Detector: a tool for assigning mass spectrometry data of N-linked glycopeptides on the basis of their electron transfer dissociation spectra. *Anal Chem*, 2013, **85**(10): 5023–5032
- [39] Pompach P, Chandler K B, Lan R, *et al.* Semi-automated identification of N-Glycopeptides by hydrophilic interaction chromatography, nano-reverse-phase LC-MS/MS, and glycan database search. *J Proteome Res*, 2012, **11**(3): 1728–1740
- [40] Chandler K B, Pompach P, Goldman R, *et al.* Exploring site-specific N-glycosylation microheterogeneity of haptoglobin using glycopeptide CID tandem mass spectra and glycan database search. *J Proteome Res*, 2013, **12**(8): 3652–3666
- [41] Strum J S, Nwosu C C, Hua S, *et al.* Automated assignments of N- and O-site specific glycosylation with extensive glycan heterogeneity of glycoprotein mixtures. *Anal Chem*, 2013, **85**(12): 5666–5675
- [42] Lynn K S, Chen C C, Lih T M, *et al.* MAGIC: an automated N-linked glycoprotein identification tool using a Y1-ion pattern matching algorithm and in silico MS (2) approach. *Anal Chem*, 2015, **87**(4): 2466–2473
- [43] Saba J, Dutta S, Hemenway E, *et al.* Increasing the productivity of glycopeptides analysis by using higher-energy collision dissociation-accurate mass-product-dependent electron transfer dissociation. *Int J Proteomics*, 2012, **2012**: 560391
- [44] Singh C, Zampronio C G, Creese A J, *et al.* Higher energy collision dissociation (HCD) product ion-triggered electron transfer dissociation (ETD) mass spectrometry for the analysis of N-linked glycoproteins. *J Proteome Res*, 2012, **11**(9): 4517–4525
- [45] He L, Xin L, Shan B, *et al.* GlycoMaster DB: software to assist the automated identification of N-linked glycopeptides by tandem mass spectrometry. *J Proteome Res*, 2014, **13**(9): 3881–3895
- [46] Mayampurath A, Yu C Y, Song E, *et al.* Computational framework for identification of intact glycopeptides in complex samples. *Anal Chem*, 2014, **86**(1): 453–463
- [47] Mayampurath A M, Wu Y, Segu Z M, *et al.* Improving confidence in detection and characterization of protein N-glycosylation sites and microheterogeneity. *Rapid Commun Mass Spectrom*, 2011, **25**(14): 2007–2019
- [48] Cao L, Tolic N, Qu Y, *et al.* Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. *Anal Biochem*, 2014, **452**: 96–102

- [49] Wu S W, Liang S Y, Pu T H, *et al.* Sweet-Heart - an integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides. *J Proteomics*, 2013, **84**(12): 1-16
- [50] Liang S Y, Wu S W, Pu T H, *et al.* An adaptive workflow coupled with Random Forest algorithm to identify intact N-glycopeptides detected from mass spectrometry. *Bioinformatics*, 2014, **30**(13): 1908-1916
- [51] Wu S W, Pu T H, Viner R, *et al.* Novel LC-MS (2) product dependent parallel data acquisition function and data analysis workflow for sequencing and identification of intact glycopeptides. *Anal Chem*, 2014, **86**(11): 5478-5486
- [52] Ferrara C, Grau S, Jager C, *et al.* Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcγRIIIb and antibodies lacking core fucose. *Proc Natl Acad Sci USA*, 2011, **108**(31): 12669-12674
- [53] Wang X, Gu J, Ihara H, *et al.* Core fucosylation regulates epidermal growth factor receptor-mediated intracellular signaling. *J Biol Chem*, 2006, **281**(5): 2572-2577
- [54] Akasaka-Manyo K, Manyo H, Sakurai Y, *et al.* Increased bisecting and core-fucosylated N-glycans on mutant human amyloid precursor proteins. *Glycoconj J*, 2008, **25**(8): 775-786
- [55] Jia W, Lu Z, Fu Y, *et al.* A strategy for precise and large scale identification of core fucosylated glycoproteins. *Mol Cell Proteomics*, 2009, **8**(5): 913-923
- [56] Cao Q, Zhao X, Zhao Q, *et al.* Strategy integrating stepped fragmentation and glycan diagnostic ion-based spectrum refinement for the identification of core fucosylated glycoproteome using mass spectrometry. *Anal Chem*, 2014, **86**(14): 6804-6811
- [57] Zhu Z, Su X, Go E P, *et al.* New glycoproteomics software, GlycoPep Evaluator, generates decoy glycopeptides *de novo* and enables accurate false discovery rate analysis for small data sets. *Anal Chem*, 2014, **86**(18): 9212-9219
- [58] Serang O, Froehlich J W, Muntel J, *et al.* SweetSEQer, simple *de novo* filtering and annotation of glycoconjugate mass spectra. *Mol Cell Proteomics*, 2013, **12**(6): 1735-1740
- [59] Shan B Z, Ma B, Mang K Z, *et al.* Complexities and algorithms for glycan structure sequencing using tandem mass spectrometry. *Proc of the 5th APBC*, 2007, **5**: 297-306
- [60] Dallas D C, Martin W F, Hua S, *et al.* Automated glycopeptide analysis—review of current state and future directions. *Brief Bioinform*, 2013, **14**(3): 361-374
- [61] Zhang Y, Yu C Y, Song E, *et al.* Identification of glycopeptides with multiple hydroxylysine O-glycosylation sites by tandem mass spectrometry. *J Proteome Res*, 2015, **14**(12): 5099-5108

Trends in Mass Spectrometry-Based Large-scale N-Glycopeptides Analysis*

ZENG Wen-Feng^{1,2)}, ZHANG Yang³⁾, LIU Ming-Qi³⁾, WU Jian-Qiang^{1,2)}, ZHANG Xiao-Jin^{1,2)}, YANG Hao^{1,2)},
LIU Chao¹⁾, CHI Hao¹⁾, ZHANG Kun^{1,2)}, SUN Rui-Xiang¹⁾, YANG Peng-Yuan³⁾, HE Si-Min^{1,2)**}

¹⁾ Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences, Institute of Computing Technology,
Chinese Academy of Sciences, Beijing 100190, China;

²⁾ University of Chinese Academy of Sciences, Beijing 100049, China;

³⁾ Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China)

Abstract Identification of post-translational modifications is one of the most challenging tasks in proteomics, and the analysis of glycosylation is a very important yet difficult one among all post-translational modifications, which has attracted more and more attention in recent years. Mass spectrometry provides an effective way for the high-throughput analysis of glycosylation. Comparing with most of the other post-translational modifications, glycans are large and heterogeneous, and glycans themselves could be fragmented in tandem mass spectrometry, in particular, the fragmentation patterns of glycans are quite different from those of peptides, resulting in difficulties in simultaneously identifying glycans and peptides of intact glycopeptides using proteomic analytical methods and software tools. The identification of intact N-glycopeptides is a hot spot in glycosylation research, for which various mass spectrometry-based analytical methods have been developed in recent years, including deglycosylation for the identification of N-glycosylated sites, electron transfer dissociation for the identification of peptide backbones, the combination of higher energy collisional dissociation and electron transfer dissociation or the combination of collision-induced dissociation and MS3 for complete identification of intact N-glycopeptides. In this article, we reviewed these analytical methods, and briefly pointed out the deficiencies of existing software tools, and suggested some future work.

Key words glycoproteomics, glycopeptide, glycosylation, mass spectrometry

DOI: 10.16476/j.pibb.2015.0402

* This work was supported by grants from The National Natural Science Foundation of China (21227805, 31300680), The National Basic Research Program of China (2013CB911203, 2012CB910602, 2010CB912701, 2011CB910600), The Hi-Tech Research and Development Program of China (2014AA020901, 2014AA020902, 2012AA020203).

**Corresponding author.

Tel: 86-10-62600822, E-mail: smhe@ict.ac.cn

Received: December 26, 2015 Accepted: April 21, 2016