科学通报 2025 年 第 70 卷 第 15 期: 2302 ~ 2318
复旦大学生物医学研究院建院20周年专辑 评 述
《中国科学》杂志社 SCIENCE CHINA PRESS

位点特异性糖基化蛋白质组学研究回顾与展望

陈乐晗[1], 刘佳林[1], 曹纬倩[1,2*]

1. 复旦大学附属上海市第五人民医院和生物医学研究院, 上海 200032
2. 国家卫生健康委员会糖缀合物重点实验室, 上海 200032
* 联系人, E-mail: wqcao@fudan.edu.cn

2024-12-31 收稿, 2025-04-11 修回, 2025-04-21 接受, 2025-04-21 网络版发表
国家重点研发计划(2024YFA1306300)和国家自然科学基金面上项目(32271490)资助

摘要 糖基化是蛋白质翻译后修饰中最为多样和复杂的类型之一, 具有广泛且重要的生物学功能. 作为研究蛋白质糖基化修饰的核心领域, 糖蛋白质组学, 尤其是位点特异性糖基化蛋白质组学研究, 能够揭示与疾病相关的糖链结构和糖基化位点. 这不仅有助于深入理解特定蛋白质糖基化修饰特征, 还为揭示糖基化在疾病发生和发展中的作用机制提供了重要的信息. 本综述回顾了过去20年基于质谱的糖蛋白质组学的主要进展, 重点聚焦于完整糖肽解析的研究成果. 首先, 概述了糖蛋白质组学的发展历程, 并回顾了从糖基化位点鉴定到完整糖肽解析中的关键研究进展. 随后, 深入探讨了糖肽富集技术、质谱分析方法、数据解析工具及数据库建设等方面的进展, 分析了相关技术的优势与挑战. 最后, 指出了当前糖蛋白质组学研究中亟需解决的问题, 并展望了未来在解析深度、全面性、准确性及生物医学应用等方面的发展方向.

关键词 糖蛋白质组学, 位点特异性糖基化解析, 完整糖肽鉴定和定量, 质谱分析

糖质, 被认为是继核酸和蛋白质之后的第三条生命链, 承载着更为精细的结构信息. 糖类物质的发现可追溯到1857年, 当时法国生理学家克劳德·伯纳德(Claude Bernard)首次发现糖原, 从此揭开糖生物学研究的序幕. 1937年, 格蒂·科里(Gerty Cori)和她的丈夫卡尔·科里(Carl Cori)发现了糖代谢过程中的一种新的化合物——葡萄糖-1-磷酸, 并证明了它是糖转换成糖原的关键步骤. 由于在糖代谢研究中的杰出贡献, 科里夫妇与伯纳德·奥赛(Bernard Houssay)共同获得了1947年诺贝尔生理学或医学奖, 这进一步推动了糖生物学领域的逐渐发展.

随着对糖和蛋白质化学研究的深入, 科学家们逐渐揭示了许多蛋白质在结构上与糖分子的密切联系, 这种共价结合了糖分子的蛋白质被称为糖蛋白[1]. 对生物体内所有糖蛋白的结构和功能进行研究的领域称为糖蛋白质组学[2], 这一领域逐渐引起了研究人员的广泛关注.

质谱分析技术已成为糖蛋白质组研究的重要手段[3]. 20世纪初, 由于工具的缺乏, 无法直接从生物体内大规模筛选具有重要功能的糖蛋白, 研究人员主要通过对单个糖蛋白、特定糖型的研究来探究其生物功能[4]. 糖基化是最多样、最复杂的蛋白质翻译后修饰之一, 它通过序列和结构的动态变化, 反映了在不同生理或病理状态下蛋白质的运作需求, 是一种灵活的生物策略[5]. 糖基化在生物体内具有广泛且重要的生物功能, 比如调节细胞质和核功能, 参与免疫监视和炎症反应, 与自身免疫、激素作用和肿瘤转移等相关[6]. 位点特异性糖蛋白质组解析可获得疾病特异性糖链所修饰的糖蛋白结构和糖基化位点等信息, 有助于理解蛋白质特定位点修饰的糖链特征, 揭示蛋白质糖基化在疾

引用格式: 陈乐晗, 刘佳林, 曹纬倩. 位点特异性糖基化蛋白质组学研究回顾与展望. 科学通报, 2025, 70: 2302–2318
Chen L, Liu J, Cao W. Review and perspectives on site-specific glycoproteomics research (in Chinese). Chin Sci Bull, 2025, 70: 2302–2318, doi: 10.1360/TB-2024-1409

病发生发展中的作用机制[7,8].

位点特异性糖基化蛋白质组学为蛋白质糖基化研究开辟了新视野, 尽管其复杂性导致近二十年的发展道路并非一帆风顺, 但各个阶段均取得了一定的成果. 本文将围绕近20年来基于质谱的糖蛋白质组学进展, 特别是以完整糖肽鉴定为核心的研究展开回顾和讨论.

1 从糖基化位点鉴定到完整糖肽解析

对于完整糖肽的鉴定, 化学家和糖生物学家一直在不懈尝试, 其难点在于: 蛋白质可能含有多个糖修饰位点, 每个位点上的糖链种类及糖链占有率不同, 糖型种类复杂且丰度水平低, 依赖于特异性的富集技术和高灵敏的质谱鉴定; 而糖链的优先碎裂或中性丢失现象为质谱分析提出了更高的要求; 完整糖肽的谱图非常复杂, 需要合适的数据库搜索软件和解析工具[9].

在2012年前, 研究人员主要采用去糖链的方法鉴定糖肽, 即分别鉴定糖组与糖基化修饰位点, 因此成果主要集中在位点鉴定上. 2003年, Ruedi Aebersold团队利用肼化学法开发的N-糖基化位点富集和质谱鉴定技术[10], 是基于糖基化位点水平的大规模糖蛋白质组学研究的标志性成果, 为后续的位点鉴定提供了新思路和高效工具. 2010年, Matthias Mann团队开发了基于超滤膜的快速酶解和凝集素富集方法用于糖基化位点鉴定, 并利用高精度质谱在四种小鼠组织和血浆中鉴定出2352个糖蛋白质上的6367个N-糖基化位点, 这一发表在*Cell*杂志上的工作是位点鉴定领域的里程碑[11], 使位点鉴定的深度达到了前所未有的水平. 2012年, 该团队继续在拟南芥、酵母、线虫、果蝇等6种不同模式生物中进行N-糖基化位点研究, 首次在组学水平证实了真核生物中蛋白质N-糖基化氨基酸序列在识别模式、结构性约束和亚细胞定位中的核心特征[12]. 这一阶段的工作主要集中在高效富集方法的开发和糖基化位点鉴定上, 尽管发展了一系列高效技术, 并在位点鉴定上达到了较高的水平, 但由于丢失了位点特异性糖链信息, 我们称之为位点特异性糖蛋白质组学研究的迂回时期.

2012年后, 糖蛋白质组学领域进入了攻坚时期, 这一时期的特点在于多流程完整糖肽鉴定技术的逐步发展. 由于糖链和肽段在质谱中碎裂行为不同, 我们难以在一次质谱分析中获得含有丰富信息的糖肽碎片, 因此研究者们一般同时采用多实验流程及质谱方法, 如使用多种组合碎裂模式获得足够的糖肽谱图信息[13], 或者利用预先建立位点库的方式缩小检索空间[14,15]. 2016年, Zhang等人[14]提出了一种基于化学酶法的N-糖肽固相萃取和质谱鉴定方法, 采用多实验流程解析不同修饰位点上的糖链组成, 首次实现了基于完整糖肽水平的大规模糖蛋白质组解析. 同年, 复旦大学生物医学研究院杨芃原团队和中国科学院计算技术研究所贺思敏团队开发了pGlyco, 虽仍采用质谱多次碎裂的模式鉴定完整糖肽, 但其首次提出了糖肽水平假阳性质控方法[16], 开启了完整糖肽水平精准鉴定的新篇章.

随着质谱技术的不断发展, 糖蛋白质组学自2017年起迈入了一个全面爆发的新阶段. 国内外多个课题组开始采用高效的富集方法、创新的质谱技术、精准的质谱数据解析软件, 极大简化了糖肽鉴定流程, 显著增加了鉴定数量, 并大幅提升了鉴定的准确性. 2017年, 复旦大学与中国科学院北京计算技术研究所合作, 率先推出pGlyco 2.0[17], 该方法基于HCD阶梯能量碎裂, 在一次实验流程中实现了完整糖肽的高通量鉴定. 随后, 2021年, 开发了pGlyco3[18], 这一工具能够快速、精准地解析位点特异性的N-和O-糖链以及修饰糖单元; 2022年, 开发了pGlycoQuant[19], 可实现高准确性的糖肽定量分析, 从而帮助研究者发现不同样本中蛋白质糖基化的差异; 同一合作团队, 在2024年, 建立了不依赖糖链库的完整糖肽解析策略pGlycoNovo[20], 显著扩展了糖链多样性, 提高了未知糖链的发现能力. 2021年, 复旦大学自主开发了GproDIA[21], 实现了完整糖肽的数据非依赖采集解析; 同年, 西北大学发展了完整N-糖肽解析工具StrucGP, 显著提升了N-糖链结构解析的灵敏度[22]. 2022年, 中国科学院大连化学物理研究所开发了Glyco-Decipher, 这是一种具备谱图拓展功能的完整糖肽检索技术, 提高了谱图鉴定深度[23]. 2024年, 浙江大学和复旦大学相继建立了完整N-糖肽谱图预测技术DeepGlyco[24]和DeepGP[25], 展示了AI在完整糖肽解析中的作用和可能性. 在这个时期, 国外研究团队也取得了相应进展, 比如, 威斯康星大学Lloyd M. Smith团队提出了O-Pair Search, 提供了一种O-糖肽鉴定和O-糖基位点定位的新方法[26]; 密歇根大学Alexey I. Nesvizhskii团队基于MSFragger[27]推出了MSFragger-Glyco[28], 用于完整N-和O-糖肽鉴定. 此外, 糖肽解析的商业化软件也逐渐发展成熟[29]. 2012年, Protein Metrics Inc. 公司推出了早期商业化糖肽分析工具Byonic[30], 由于质谱分析技术的限制, 其常被用于简单样品糖肽解析, 随着技术的发展和软件迭代更新, 目前其在鉴定量

和准确性上均有提升[31]. GlycanFinder[32]是于2023年左右发表并推出的商业化软件, 其结合了深度学习模型, 加强了糖肽鉴定的深度.

经过近20年的发展, 位点特异性糖基化蛋白质组学已经实现了从单独位点鉴定到完整糖肽解析的巨大飞跃(图1). 这一领域的发展见证了技术的不断演进, 从早期的位点鉴定到现在能够实现完整糖肽水平的糖蛋白质组分析, 展现了该领域技术的成熟和进步, 推动了糖生物学在疾病机制、生物标志物发现等应用中的深度发展. 下面我们将围绕其中的一些关键技术进行回顾和总结, 包括糖肽富集方法、质谱数据分析策略、数据库等. 这些技术的发展不仅推动了位点特异性糖基化蛋白质组学的进步, 也为未来的研究提供了强大的工具和方法.

2 糖肽富集技术进展

糖基化丰度相对较低, 加之糖基化修饰的多样性, 导致在复杂样本中每条糖肽的丰度进一步降低. 此外, 非糖肽会在质谱分析中对糖肽产生离子抑制. 在这些因素的共同影响下, 通过质谱直接检测复杂生物样本中的糖肽变得困难[2]. 为了解决这些问题并提高糖蛋白检测的灵敏度, 对糖肽进行高效富集变得至关重要. 近年来, 开发了多种富集策略, 依据其富集的原理, 主要包括酰肼化学法、凝集素亲和色谱法、体积排阻法、硼酸化学法、协同作用力新材料、亲水作用色谱法等(图2).

酰肼化学法的原理是聚糖的顺式二羟基被氧化生成醛基基团, 与固定有酰肼基团的树脂发生共价偶联, 是一种适用于去糖基化策略的富集方法. 这种方法可与多种糖链类型反应, 因此具有广谱性. 该富集策略最早由Ruedi Aebersold团队的Zhang等人[10]提出, 之后基于此策略发展了多种富集材料和富集方法. 例如, 2014年, Liu等人[33]制备了Fe_3O_4@聚甲基丙烯酸(Fe_3O_4@PMAA), 并通过己二酸二酰肼(Adipic Acid Dihydrazide, ADH)修饰其表面, 得到了Fe_3O_4@聚甲基丙烯酸酰肼(Fe_3O_4@PMAH), 其丰富的酰肼基团可实现糖肽

图 1 近20年来位点特异性糖蛋白质组学研究进展概览

Figure 1 An overview of research advances in site-specific glycoproteomics over the past two decades

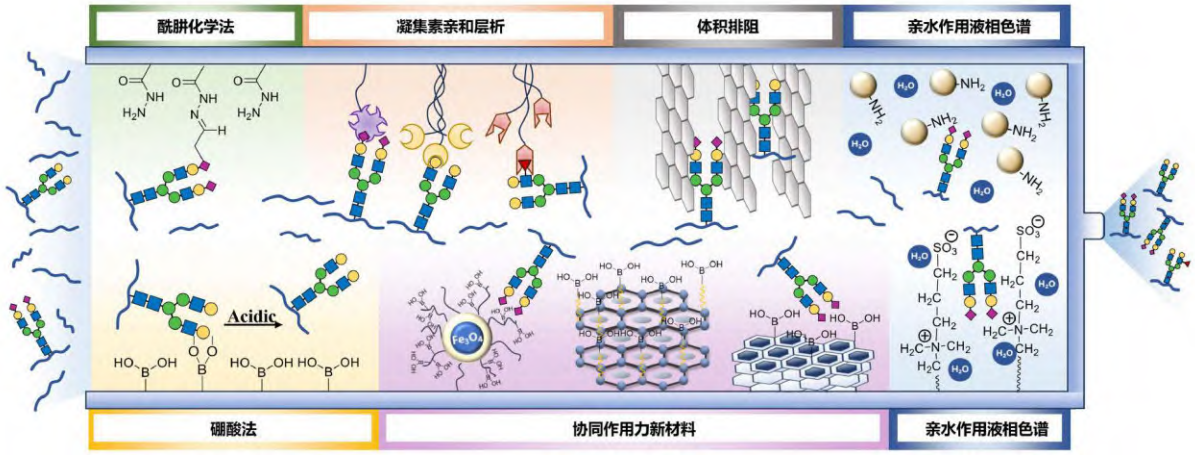

图 2 糖肽富集方法概要
Figure 2 Summary of glycopeptide enrichment methods

的高特异性富集. 2018年, Bai等人[34]开发了一种酰肼官能化的热敏聚合物, 利用这种热敏聚合物从神经胶质瘤患者和健康受试者的血浆外泌体中, 富集并鉴定出180种N-糖蛋白和329个N-糖基化位点. 然而, 由于酰肼化学法具有不可逆性, 通过糖苷酶进行酶切释放去糖基化肽段会破坏糖链结构, 限制了其在完整糖肽分析中的应用[10]. 近年来, 通过改进羟胺裂解方法, 温和断裂脲键, 可以保留较完整的糖链, 推动了酰胺化学法进行完整糖肽解析的进展, 如Li等人[35]结合了酰肼化学标记、点击化学固定的方法, 实现糖肽高效特异性捕获, 而后温和裂解酰脲键, 释放完整糖肽, 保留了糖链和肽段信息, 支持完整糖肽分析.

区别于酰肼化学法的广谱性, 凝集素亲和富集技术通过识别特定的糖链类型对糖肽进行富集[36]. 1888年, Stillmark[37]从蓖麻种子中分离出蓖麻毒素, 揭示了凝集素的凝集特性, 由此开启了对于凝集素结构和富集应用的深入研究. 随后的研究发现了更多具有不同结合特异性的凝集素, 如紫藤花凝素(wisteria floribunda agglutinin, WFA)可识别糖链末端的GalNAc; 小麦胚芽凝集素(wheat germ agglutinin, WGA)可特异性识别糖链末端的GlcNAc等[38], 这些凝集素均可用于糖肽的富集.

在实验过程中, 使用单一凝集素亲和色谱法可以用于富集特定糖链类型的糖肽[39]. 然而, 部分凝集素具有较宽泛的识别特异性, 例如, 伯克霍尔德氏菌凝集素A(Burkholderia Cepacia Lectin A, BC2L-A), 除了已知的对高甘露糖型N-聚糖的结合能力外, 这种凝集素还能保留C-和O-甘露糖修饰肽[40]. 因此, 特异性较低的凝集素在富集过程中可能无法完全排除非目标糖肽的干扰[41]. 近年来发布的一些数据库, 如2015年发布的Glyco3D[42]和2019年发布的UniLectin3D[43], 提供了丰富的凝集素信息, 这些数据库为选择和应用适合的凝集素提供了重要参考. 为了实现广谱的糖蛋白组富集, 通常需要使用多种凝集素. 例如, Gbormittah等人[44]联合使用等量的橙黄网孢盘菌凝集素(aleuria aurantia lectin, AAL)、接骨木凝集素(sambucus nigra lectin, SNA)和菜豆白凝集素(phaseolus vulgaris leucoagglutinin, PHA-L)进行糖肽富集. 2010年, Zielinska等人[12]利用混合凝集素富集糖肽, 开发了一种基于"超滤辅助样品制备"(FASP)的富集方法, 利用高精度质谱在四种小鼠组织和血浆中鉴定到6367个N-糖基化位点, 显著提升了糖基化位点的检测数量. 然而, 即便使用多种凝集素, 仍无法完全解决富集过程中由凝集素的非特异性识别导致的糖肽损失, 因此通常需要结合其他互补富集方法, 以全面覆盖不同类型的糖基化修饰.

硼酸富集法是一种常用的糖肽化学偶联富集策略. 在碱性条件下, 硼酸能够与含1-2和1-3顺式二羟基结构的糖型(例如甘露糖、葡萄糖和半乳糖)发生共价键合, 形成稳定的环酯, 而在酸性条件下, 该反应可逆, 因此构成了一个可逆的化学富集平台[45]. 1970年, Weith等人[46]首次将该方法应用于碳水化合物和核酸的色谱研究. 随后, 硼酸富集法在糖肽富集领域得到了广泛发展[47,48]. 近年来, 基于硼酸法开发了多种协同作用力新材料, 进一步推动了该方法的应用, 如纳米颗粒[49]、聚

合物[50]、石墨烯[51]、介孔二氧化硅[48]等基于衍生化硼酸基团的新型富集材料，在完整糖肽富集方面展现出巨大应用潜力. 纳米颗粒因其高比表面积和可调节的功能化特性，使其能够与衍生化硼酸材料和多种协同作用力材料结合，有效增强对低丰度糖肽的捕获能力，从而提升其富集效率. 2012年，Zou等人[52]首次利用壳聚糖制备纳米球载体材料，结合甲基丙烯酸缩水甘油酯(glycidyl methacrylate, GMA)与3-氨基苯基硼酸(3-aminophenylboronic acid, APB)多功能的化学连接特性，实现了糖肽的选择性富集. 磁性纳米材料，在保留纳米材料高比表面积的优势基础上，能够实现在外加磁场的作用下目标糖肽的快速吸附与分离. 2015年，Zhang等人[49]合成了苯基硼酸和共聚物多功能磁性纳米粒子(nanoparticles, NPs)，该复合物利用硼酸亲和力与亲水单体的协同作用，增强了NPs对糖蛋白的结合强度和选择性，表现出优异的糖肽选择性能. 硼酸官能化聚合物、石墨烯、介孔二氧化硅等复合材料，利用材料本身的特性，同时可有效增加官能团的数量，从而提升其富集效率. 2020年，Mujahid等人[50]报告了一种利用硼酸官能化GMA-MAA-DVB聚合物，其中二乙烯三胺作为与硼酸结合的连接剂，提供多个亲水位点，从而有效增强该材料的结合力和亲水性. 2015年，Wang等人[51]首次采用酚醛树脂作为石墨烯与硼酸基团之间的耦合连接剂，所制备的复合富集材料(magG@P-F@APB)不仅增强了硼酸的亲和性，还结合了磁性石墨烯比表面积大、磁响应性强以及与树脂材料生物相容性好的特点，展现出优异的富集性能. 2009年，Xu等人[48]首次将介孔材料引入糖蛋白质组分析，开发了一种新型的硼酸功能化介孔二氧化硅富集材料，该方法使糖肽的检测量提高了两个数量级. 2021年，Kong等人[53]使用硼酸功能化的介孔石墨烯-二氧化硅复合材料(GO@mSiO$_2$-GLYMO-APB)从复杂的生物样品中分离完整糖肽，材料的高比表面积和介孔材料尺寸排阻功能的协同效应，可高特异性地富集N-糖肽和O-糖肽.

亲水性相互作用色谱(hydrophilic interaction chromatography, HILIC)也是一种重要的富集技术. 自1990年用于分离多肽以来[54]，HILIC因其高效、普适的富集能力和温和的富集条件，受到越来越多的关注[55]. HILIC利用糖链的多羟基亲水性与固定相表面修饰极性基团(如酰胺、二醇、氨基等)互相作用，通过氢键、偶极-偶极作用或静电吸引与糖链结合，实现糖肽与非糖基化肽的分离. 此外，HILIC操作简便，不会引入盐类，且与质谱分析高度兼容. HILIC技术可灵活选择固定相，高效分离传统反相色谱难以处理的极性化合物，其固定相包括胺[56]、酰胺[57]、麦芽糖[58]等酰胺/胺功能化材料，碳水化合物功能化材料[59]，基于金属有机框架(metal-organic framework, MOF)的材料[60]等. 为了提高HILIC的选择性，常与其他富集方法相结合来改进HILIC技术. Ji等制备了交联的环糊精-金属-有机框架材料提升糖肽的富集效率[61]. Dong等人[62]制备了一种具有亲水相互作用和可切换表面电荷的组氨酸键合二氧化硅(histidine-bonded silica, HBS)材料，用于人血清中糖肽富集，富集选择性达到92%. 近年来，HILIC材料的不断改进进一步增强了其糖肽富集能力. Wang等人[63]建立了一种基于HILIC色谱柱的自动富集策略，能够同时富集、分离和表征同一样品中的N-和O-连接的完整糖肽. 两性离子(Zwitterionic, ZIC)表面能保持整体电荷中性，其强水合能力可通过静电诱导水合作用在表面形成致密的水层[64]. 由于具有优异的亲水性和生物相容性，基于ZIC的HILIC色谱柱可以在亲水模式下高效分离亲水性和极性物质，被广泛应用于糖肽的富集，如Zhu等人[65]采用二乙氨基乙醇(diethylaminoethanol, DEAE)Sepharose固相萃取微柱富集N-糖肽，该方法结合了点击麦芽糖法和Zwitterionic HILIC(ZIC-HILIC)法的优点，显著提高了N-糖肽富集的特异性. ZIC-HILIC技术基于两性离子-亲水相互作用，通过静电相互作用增加带电固定相在HILIC中的保留时间，与中性HILIC矩阵相比，具有更高的选择性和分辨能力[66]. 目前，ZIC-HILIC富集方法已成为完整糖肽富集最常用方法之一.

糖肽富集策略的发展为获得高质量糖肽谱图奠定了基础，这些技术通过提高糖肽的选择性富集和降低干扰，显著提高了糖肽在质谱分析中的灵敏度和可检测性.

3 质谱分析与谱图解析技术

由于糖链的分支结构以及糖链和肽段在谱图中的碎裂行为差异，传统蛋白质组质谱分析方法和解析工具无法鉴定完整糖肽. 主要难点在于从质谱中获取丰富的糖肽碎片信息，以及有效解析这些糖肽碎片. 近二十年来，质谱技术和解析软件的开发极大地推动了完整糖肽解析的发展，下面将从糖肽质谱碎裂特点，以及质谱数据解析策略和工具的发展两方面进行总结和回顾.

3.1 糖肽在质谱中的碎裂

质谱策略的发展和相关软件工具的开发极大促进了完整糖肽解析的进展. 无论使用何种策略, 都需要从肽骨架和附着的聚糖中获得足够的糖肽片段. 由于糖苷键和肽键的物理和化学性质不同, 两者之间的碎片化差异显著, 因此难以获得信息丰富的糖肽谱图. 目前, 糖肽鉴定中较为常用的解离方法包括共振激活(离子阱)碰撞诱导解离(collision-induced dissociation, CID)、束型CID、高能碰撞解离(higher-energy collisional dissociation, HCD)和电子转移解离(electron transfer dissociation, ETD), 也常用电子转移/碰撞诱导解离(electron transfer/collision-induced dissociation, ETciD)和电子转移/高能碰撞解离(electron transfer/higher-energy collisional dissociation, EThcD)等组合能量的碎裂方法.

在早期研究中, CID技术主要用于分析简单样品中的糖肽[67]. 由于离子阱CID前体离子的单键裂解、解离能低、低m/z区域的“1/3”截止等问题[68], 其生成的谱图以Y离子为主, 含有少量B离子和有限数量的b/y离子. HCD与CID不同, HCD往往会产生大量的诊断氧鎓离子, 如*m/z*为138.055、204.087和366.140片段. 研究表明, 随着HCD碰撞能量的增加, b/y-离子也会随之增加[69]. 束型CID产生的碎片模式与C-trap中HCD碎裂产生的模式相似. 此外, HCD-MS/MS中不同的碰撞能量会产生互补的糖链和肽段碎片[70], 基于这一特点, 可采用不同能量碎裂糖肽, 如pGlyco 2.0基于HCD阶梯能量在一次质谱分析中获得丰富信息糖肽谱图[17], StrucGP[22]和GlycanFinder[71]使用两种组合能量进行糖肽分析. 许多研究采用了不同的解离方法研究了糖肽的碎裂行为[72]. 目前基于HCD阶梯能量碎裂的方式被广泛应用于完整糖肽的高通量鉴定[73].

ETD主要断裂N-Cα键并产生c/z离子. 理想情况下, 糖肽在ETD中产生大量的c/z离子, 这对糖基化位点和肽序列的鉴定非常有用. 然而, ETD常常存在碎裂不完全的问题, 导致大量前体离子残留[74]. 因此, ETD很少单独使用, 而是经常与HCD或CID方法结合起来用于糖肽的鉴定. 随着MS中离子触发技术的发展, Orbitrap Fusion质谱仪开发的混合解离技术, 如电子转移/碰撞诱导解离[75]、电子转移/高能碰撞解离[76]和活化离子电子转移解离[77]等, 在N-糖肽鉴定中显示出巨大的潜力. 此外, 电子捕获解离(electron capture dissociation, ECD)、电子诱导激发解离(electron-induced excitation dissociation, EED)和紫外光解离(ultraviolet photodissociation, UVPD)等碎裂模式也同样适用于不同分析需求的糖肽碎裂.

混合碎裂法通过结合单个方法的优势, 对糖肽的碎裂和鉴定大有裨益. Swaney等人[78]将ETciD(当时称ETcaD)用于解决双电荷肽段ETD效率低的问题, 其序列覆盖率的中位数为89%, 较ETD(63%)和离子阱CAD(77%)有显著提高. Frese等人[74]使用EThcD方法对ETD形成的所有离子施加补充能量, 提供更全面的糖链特征离子, 生成了信息量更大的谱图. 由于混合碎裂方法的互补性, 混合碎裂在糖蛋白质组学中的应用迅速扩大[9].

3.2 质谱数据解析策略和工具的发展

糖肽解析的发展是随着谱图信息量的增长而进步的. 在过去的研究中, 研究者们无法在一次实验流程中获得足够的谱图信息, 因此, 在2016年之前常采用多流程的方法进行鉴定, 其鉴定通常依赖于预先建立的糖基化位点肽谱库. 例如, Liu等人[79]于2014年开发的GRIP、Cheng等人[80]于2014年开发的ArMone 2.0以及Toghi等人[81]于2015年开发的GPQuest都是基于多流程的鉴定策略. 虽然GRIP、ArMone和GPQuest已被用于复杂样品中完整N-糖肽的质谱分析, 但其工作流程较为繁琐. Yu等人[82]于2016年开发的GlycoSeq旨在仅通过CID谱图鉴定糖肽, 但也需要一个预先鉴定去糖链后的糖基化位点的步骤.

随着糖肽质谱碎裂策略的开发, 越来越多基于特异性解离的串级谱图解析工具被开发, 本文总结了近年来发展的主要检索工具及其特点(表1). 2017年前, 适用于单个解离方法的糖肽解析软件发展迅速, 例如, GlycoPeptide Search[83]、GP Finder[84]和MAGIC[85]可用于Q-TOF-CID MS/MS数据解析. ProteinProspector[86]和GlycoPep Detector[87]可用于ETD-MS/MS分析, GlycoPep Grader[67]用于CID-MS/MS分析. 虽然这些解析工具可对不同类型质谱数据进行解析, 但使用单一解离方法所获得的碎片信息有限, 极大地限制了鉴定的数据量及其可靠性. 为了获得更多的糖肽谱图信息, 研究者们使用组合解离策略. 在N-糖肽解离过程中, HCD、ETD和CID具有互补性, 因此通常联合使用来解离同一前体离子, 其中, HCD/CID与ETD相结合是多种解离策略中应用最广泛的方法之一. 研究者们也同

表 1 用于完整糖肽解析的软件工具特征汇总表[a)]

Table 1 Summary of software tool features for complete glycopeptide resolution

软件名	发表年份	检索策略	糖库依赖	碎裂方式	主要特点
Byonic[30]	2012	肽段优先	是	CID, HCD, ETD	商业软件, 多次更新. 支持多碎片类型搜索
GP Finder[84]	2014	肽段优先	是	CID	自动化匹配N-/O-糖基化位点
Sweet-Heart[93]	2013	糖链优先	部分	CID, HCD, ETD	支持CID-MS^{2}和靶向MS^{3}分析, 优化多模式鉴定流程
GlycoPeptide Search[83]	2013	传统方式	是	CID	适用于Q-TOF-CID分析, 去糖基化分析, 完整糖肽分析
GlycoMaster DB[92]	2014	糖链优先	是	HCD, ETD	去糖基化鉴定, 完整糖肽鉴定, 帮助鉴定游离寡糖
GRIP[79]	2014	传统方式	是	CID	结合去糖基化步骤, 多流程鉴定策略
MAGIC[85]	2015	糖链移除	部分	CID	适用于Q-TOF-CID分析, 基于Y1离子模式匹配算法
GPQuest[81]	2015	肽段优先	是	HCD	多流程鉴定策略, 通过质量差匹配糖链数据库
gFinder[88]	2016	传统方式	是	CID, HCD	合并CID和HCD数据, Mascot自动识别肽段信息
I-GPA[89]	2016	传统方式	是	CID, HCD	集成式分析仪, 自动化N-糖基化位点定量分析
GlycoPAT[91]	2017	糖链优先	部分	CID, ETD, HCD	支持CID/ETD/HCD多模式数据搜索
pGlyco 2.0[17]	2017	糖链优先	是	HCD	阶梯能量HCD-MS/MS, 综合质控(糖链/肽段/糖肽FDR)
MSFragger-Glyco[28]	2020	肽段优先	否	ETD, HCD	基于开放式搜索, 支持高通量糖肽鉴定
MetaMorpheus O-Pair[26]	2020	肽段优先	是	HCD, EThcD	O-糖肽位点特异性解析
StrucGP[22]	2021	糖链移除	部分	HCD	支持糖链结构(拓扑异构体)解析, 部分依赖已知库
Glyco-Decipher[23]	2022	糖链移除	否	HCD	增加二次比对, 对低质量谱图进行深度挖掘
pGlyco 3[18]	2021	糖链优先	是	HCD, ETD, EThcD, ETciD	糖肽和O-糖肽分析, 解析位点特异性糖基化, 增加全面质控
pGlycoNovo[20]	2024	糖链优先	否	HCD	全局Y离子匹配策略, 不依赖糖库, 可用于鉴定未知糖型
GlycanFinder[32]	2023	糖链/肽段双策略	部分	HCD, ETD, EThcD	糖链优先/肽段优先双策略并行, 显著提升低丰度糖肽的解析率

a) 传统方法: 从肽段库和聚糖库中枚举所有可能的候选糖肽, 与谱图进行匹配, 找出与每个谱图最匹配的候选糖肽. 糖链优先: 先匹配谱图中的糖链Y离子确定糖链组成, 后通过前体离子和糖链之间的质量差推断肽段组成; 肽段优先: 先匹配谱图中的肽段b和y离子, 后通过前体离子和肽段之间的质量差推断糖链组成; 糖链移除: 先基于独特的Y0离子确定肽段部分, 后基于糖库对糖链部分进行匹配

步开发出适用于多种碎片解析的工具. Byonic是最常用的糖肽解析商业化软件[30], 它支持搜索多种碎片类型的MS/MS数据. Yoo等开发了可以用于解析CID/HCD-MS/MS谱图的gFinder[88]和I-GPA[89]软件. Tang等对GlypID工具进行了改进, 升级后的GlypID 2.0[90]可以支持HCD和CID碎片离子的搜索. GlycoPAT[91]能够支持CID/ETD-MS/MS或CID/HCD/ETD-MS/MS数据的搜索. GlycoMaster DB[92]设计用于解析HCD/ETD-MS/MS数据. 2013年, Sweet-Heart作为一套集成的计算工具, 采用HCD触发CID和ETD的模式, 基于HCD产生的糖肽碎片筛选糖肽谱图, 进而触发后续CID和ETD分析, 极大地优化了多种碎裂模式下完整糖肽的鉴定流程[93].

在2014年之前, 几乎所有的鉴定策略和软件工具都只适用于简单或标准糖蛋白样品中的糖肽分析, 且分析通量较低. 例如, 2012年, Carrie等人开发的Glyco-Pep Grader仅鉴定了45个N-糖肽[67], Pompach等人[94]使用GlycoPeptide Search在血红蛋白中鉴定出57个N-糖基化位点. 2016年, Sun等人[14]发表了一项最具代表性的研究, 开发了一种固相萃取N-聚糖和N-糖肽鉴定新方法. 虽然此策略依然采取多流程实验方法, 但其鉴定量大大提升, 使用其团队在2015年开发的GPQuest软件[81], 在不同细胞系中成功鉴定到了1562个N-糖肽, 这在当时是一大亮点工作.

高效的质谱分析流程和数据解析工具极大地提高了完整糖肽鉴定的深度和精度(图3). 2017年, pGlyco 2.0建立了一次实验流程完整糖肽解析新方法, 突破了完整糖肽需要预先建库和多流程的瓶颈, 同时建立了完整糖肽水平的假阳性控制和评估方法, 实现了完整糖肽水平糖蛋白质组的高通量精准鉴定, 在五种小鼠组织中共鉴定出10009个完整的N-糖肽[17]. 随后, 完整糖肽解析策略和解析工具的不断发展, 促进了糖蛋白质组鉴定覆盖度, 完整糖肽鉴定规模从2010年至今增加了3个数量级[95]. 目前, 完整糖肽解析策略, 根据糖链和肽段部分检索的先后, 主要分为肽段优先、糖链移除(也被视为三步糖链优先策略, 其中首先使用糖肽离子来缩小肽段搜索空间, 先鉴定肽段, 然后再全面表

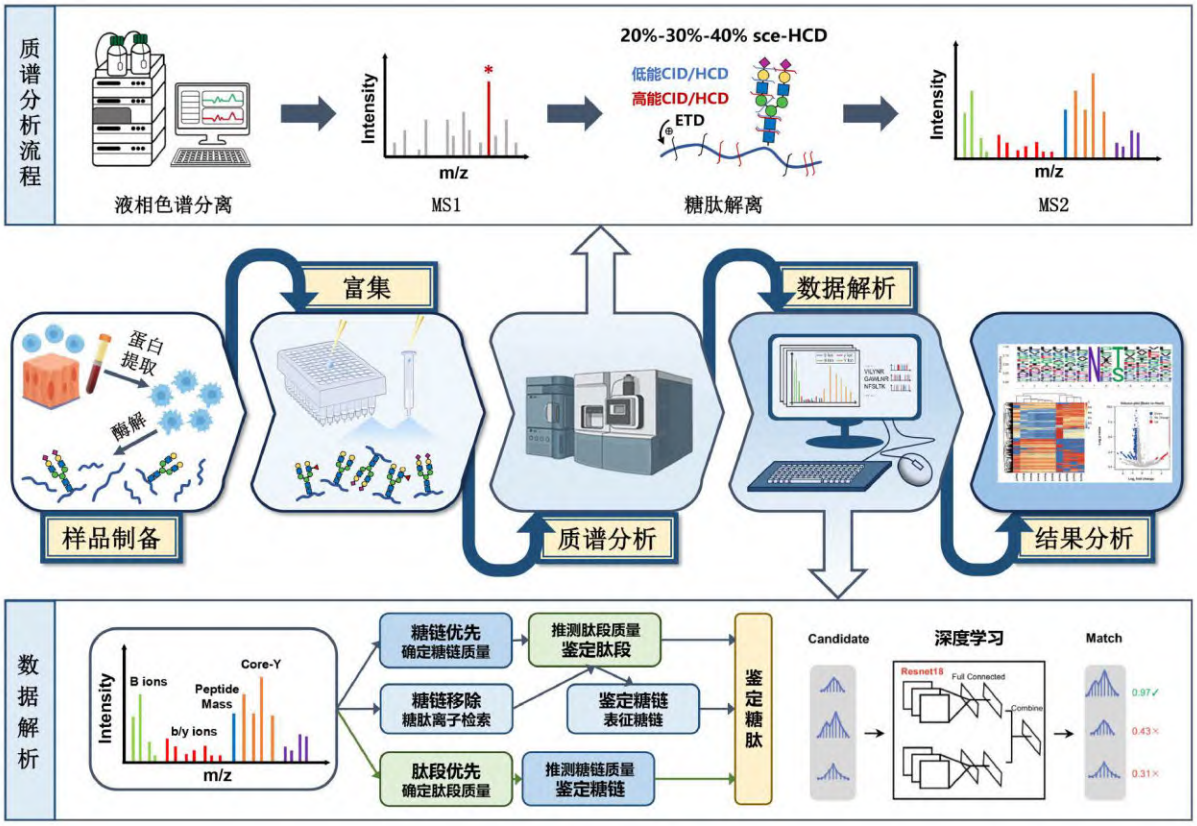

图 3 高效质谱分析流程和谱图解析工具用于完整糖肽水平糖蛋白质组解析. 该图中的样品制备和富集中的部分图元素，包括细胞、试管等，引自GDP素材库[100]; 质谱分析流程和数据解析中的彩色质谱图修改自参考文献[19], 数据解析中的深度学习模块图元素摘自参考文献[20], 采用知识共享署名4.0国际公共许可协议(https://creativecommons.org/licenses/by/4.0/); 结果分析中的图元素由GlycoAP平台(https://project.omicsolution.com/GlycAP/)生成.

Figure 3 Efficient mass spectrometry analytical workflows and data interpretation tools for glycoproteomic analysis at intact glycopeptide-level. Some of the graph elements in sample preparation and enrichment in this figure, including cells and tubes, are quoted from the GDP material library[100]; the color mass spectrometry diagrams in the mass spectrometry analysis process and data parsing are modified from ref. [19], and the graph elements of the Deep Learning module in data parsing are extracted from ref. [20], under Creative Common Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/); graph elements in the result analysis were generated by the GlycoAP platform (https://project.omicsolution.com/GlycAP/).

征糖链)和糖链优先检索策略. 肽段优先搜索策略被多个工具广泛采用[14,28,30,81,84], 其基于蛋白质组学鉴定方法的进步, 直接匹配谱图中的肽段b和y离子, 然后通过前体离子和肽段之间的质量差推断糖链组成, 其中MSFragger-Glyco[28]和MetaMoroheus O-Pair[26]通过加入肽段片段离子索引来加速搜索过程, 提高了鉴定的通量和检索速度. MAGIC[85]、StrucGP[22]和Glyco-Decipher[23]等工具采用的是糖链移除策略, 与肽段优先搜索不同, 它首先基于独特的Y0离子确定肽段部分, 然后基于糖库对糖链部分进行匹配. 糖链优先检索策略, 广泛应用于pGlyco系列工具[17,18,20], 以及Sweet-Heart[93]和GlycoMaster DB[92]中, 这种策略通过优先检索糖链组成来排除不可靠的糖链, 从而增强了糖链鉴定的可靠性. 基于这种方法, pGlyco 2.0[17]首次在糖链、肽段和糖肽水平上实现全面质控. pGlyco3进一步加入糖链离子索引方法, 优化Y和B离子的使用, 从而提高准确性并加快糖肽匹配速度[18]. pGlycoNovo基于糖链优先检索的优势, 建立了基于全局糖肽Y离子检索的策略, 实现了不依赖糖库的完整糖肽解析, 用于发现含有未知糖链的糖肽[20]. GlycanFinder集成了基于肽段和基于糖基的搜索策略, 通过两种策略同时进行来增加谱图的解析率[32].

在不同样品中蛋白质糖基化差异研究过程中, 不仅需要精准的谱图解析, 还需要高准确性的定量策略. 传统的蛋白质组定量方法, 如非标定量、同位素化学标记和代谢标记方法, 已被部分应用于完整糖肽水平

糖蛋白质组的定量分析. 然而, 针对糖肽定量数据处理的专用工具仍然至关重要. 2022年, SugarQuant工作中, 开发了基于pGlyco 2.0的完整糖肽定量工具GlycoBinder[96], 专门针对N-糖肽的鉴定和基于MS3的TMT定量. IonQuant[97]可用于糖肽非标记定量分析, 具有FDR控制和Match-between-runs(MBR)功能, 是MSFragger-Glyco平台[28]中强有力的定量工具. pGlycoQuant能够进行基于MS1/MS2的完整糖肽定量, 并利用深度学习技术有效解决缺失值问题, 相比其他软件工具, 其缺失值降低了19%到89%. 该工具还开发了Match-in-run (MIR)算法, 增加了糖肽的覆盖度, 并支持与多个搜索引擎兼容, 扩展了其应用[19]. 数据非依赖采集(DIA)作为一种能够实现全面覆盖和无标记定量的技术, 在糖蛋白质组学中得到了越来越广泛的应用. 最初, DIA用于去糖链的糖肽定量分析, 现在已拓展至完整糖肽的定量分析, 支持通过靶向提取丰富的Y离子并通过谱图库检索进行分析. 基于DIA的Glyco-DIA解析策略, 依赖于其先前开发的SimpleCell O-glycoproteomics平台构建的O-糖肽谱图库[98], 专门用于复杂生物样本中O-GalNAc型糖蛋白的高通量定量分析[99]. GproDIA[21]应用了肽段为中心DIA分析的概念, 通过二维FDR控制方法和糖型推断算法, 实现了在较宽的隔离窗口中完整糖肽的识别与定量. 这些工具和技术的发展, 推动了完整糖肽水平糖蛋白质学定量分析的进步, 尤其在高通量、准确性和特定位点糖基化的深入分析中, 展现出巨大的潜力.

4 数据库及数据挖掘工具

随着富集技术和质谱分析技术的不断进步, 糖肽的定性和定量数据实现了显著增长. 这一增长对糖蛋白质组学的两个关键方面产生了重要影响. 首先, 完整糖肽的鉴定主要依赖于肽段库和糖库. 肽段库可以直接从UniProt数据库获取, 而全面的糖库对于深入解析糖肽结构和功能至关重要. 其次, 如何对获得的糖蛋白质组数据进行深入挖掘, 以揭示糖基化在生物过程中的作用, 成为了研究的重点. 因此, 以下将围绕糖链数据库的构建, 包含蛋白质信息的糖链数据库的开发, 以及下游数据挖掘工具几方面进行回顾总结.

4.1 糖链数据库

糖链信息储存可追溯至1997年. 当时首个糖生物学数据库CarbBank成立, 收录了超过40000个聚糖结构条目[101]. 如今, 这些数据被整合存储在多个数据库中, 包括CFG糖数据库[102]、KEGG糖数据库[103]和glycosciences.de[104]. KEGG糖数据库自2005年起, 开始收集实验确定的聚糖结构及其生物合成和代谢途径[103]. CFG糖结构数据库建立于2006年, 它不仅包含聚糖结构, 还涵盖了糖基转移酶、糖结合蛋白(glycan binding protein, GBP)和糖基因敲除小鼠的数据, 并提供了糖基因微阵列和聚糖阵列等信息[102]. glycosciences.de是一个综合性门户网站, 自2006年起提供糖分析相关的数据库和工具, 特别关注多糖的三维结构[104]. GlycoSuiteDB自2001年起收集了所有已发表的聚糖结构和糖基化位点信息, 包括多糖结构、蛋白质糖基化位点、生物来源、文献参考以及疾病相关信息[93,105]. JCGGDB由Kameyama等人[106]在2005年建立, 是一个包含15个原始数据库的元数据库, 涉及质谱数据、凝集素阵列数据、糖蛋白数据、糖基因数据和糖表位数据等, 并提供了MS^n谱图采集协议, 便于用户通过谱图匹配进行比较和分析, 从而方便地检索多糖的质谱数据.

随着聚糖结构数据的积累和质谱技术的发展, 聚糖结构鉴定变得更加快速、准确和自动化. UniCarbKB于2012年推出, 由GlycoSuiteDB、GlycoBase和EUROCarbDB组成, 旨在构建一个信息存储和搜索平台[107]. 该平台包括从GlycoSuiteDB获得的数据以及理论N-聚糖结构数据库UniCorn[108], 并且其数据模型在不断改进, 以构建一个提供糖蛋白质组位点特异性糖基化信息的公共平台, 解决信息差距问题. UniCarbKB还包括了从GlycoBase[109]获得的数百个糖链的高效液相色谱(high performance liquid chromatography, HPLC)分析数据, 以及EUROCarbDB[110]中的糖链MS、核磁共振(nuclear magnetic resonance, NMR)和HPLC实验数据.

2013年后, 糖链数据库在数据多样性和全面性方面得到了进一步的完善. UniCarb-DB[111]是一个2014年推出的公共数据库, 提供从糖蛋白释放的游离寡糖的LC-MS/MS详细信息, 包括游离寡糖类型、键、结构以及相应的MS数据. 2016年, Mariethoz等人[112]创建了Glycomics@ExPASy, 集中了基于网络的糖信息学资源, 包括糖组学的数据库和工具, 以推广生物信息学在糖生物学中的应用. GlyTouCan[113]是一个聚糖结构存储库, 便于研究人员引用聚糖结构. Glycosciences.DB[114]是一个链接糖组学和蛋白质组学数据的注释数据库, 包含25000个聚糖结构条目和12500个3D结构模型.

2018年, Zhao等人[115]开发了GlycoStore, 一个基于GlycoBase公开实验数据集的注释数据库, 整合了N-、O-聚糖和鞘糖脂(GSL)聚糖以及相关糖蛋白、糖脂和生物治疗药物的色谱、电泳和质谱数据.

近年来, 为了促进糖组学研究, 还开发了其他类型的数据库, 如ProGlycProt V2.0[116], 一个经实验验证的原核生物糖蛋白和蛋白质糖基转移酶的储存库; UniLectin3D[43], 一个碳水化合物结合蛋白数据库, 提供3D结构和相互作用配体的信息; MatrixDB[117], 关于细胞外基质蛋白、蛋白多糖和多糖相互作用的数据库; 以及SugarBindDB[118], 涵盖人病原体凝集素和黏附素的聚糖结合信息的数据库. 这些数据库为糖基化研究提供了宝贵的资源, 促进了该领域的深入发展.

4.2 含修饰位点的糖链库和数据挖掘工具

随着完整糖肽鉴定技术的快速发展, 建立包含修饰位点信息的糖链库对于储存肽段、糖基化位点和糖链信息变得尤为重要.

GlycoProtDB[119](https://acgg.asia/gpdb2)是一个包含实验数据的N-糖蛋白质数据库, 自2006年建立以来, 通过凝集素亲和柱结合同位素编码的糖基化位点特异性标记(IGOT)方法鉴定N-糖基化位点, 用户可以通过基因编号、基因名称或蛋白质名称进行搜索. UniPep[120](http://www.unipep.org/)则是人类N-糖蛋白质数据库, 提供了详细的N-糖蛋白质信息, 支持通过多种方式进行搜索. N-GlycositeAtlas[121](http://nglycositeatlas.biomarkercenter.org)是一个大型数据库, 于2019年发布, 包含了7200多种N-糖蛋白, 30000多种糖肽和超过14000个N-糖基化位点, 这些数据来源于100多篇出版物和未发表的数据集, 并可映射到UniProt数据库.

糖蛋白质数据库逐渐向数据深度挖掘和多元化的资源集合等方面发展. 2019年, Ye等[99]在哥本哈根糖组学中心建立的GlycoDomain Viewer(https://glycodomain.glycomics.ku.dk/)用于存储和共享通过SimpleCell技术鉴定的O-GalNAc糖蛋白质组, 包括实验鉴定的来自人和动物细胞系的629种O-GalNAc糖蛋白和2942个O-糖基化位点. 2021年, Huang等人[122]建立了OGP(https://www.oglyp.org/), 这是一个O-糖蛋白质专家知识库, 收集了来自不同来源的O-糖蛋白数据, 包含2133个O-糖蛋白质, 9354个O-糖基化位点和11633个特定位点的O-糖链, 是迄今为止最大的O-糖蛋白质资源库. 该数据库网站还包括统计分析、数据库搜索、位点预测和数据提交等模块, 有效支持O-糖基化数据挖掘.

在糖蛋白质组数据下游分析方面, 2022年, Wu等人[123]开发了一个综合的糖蛋白质组学分析平台GlycAP(https://project.omicsolution.com/GlycAP/), 该平台嵌入了不同的分析模块, 包括定性分析、定量分析、功能分析和临床分析等, 促进糖蛋白质组学在位点特异性糖基化中的探索. 最近, Ives等人[124]建立了GlycoShape(https://glycoshape.org), 一个开放存取的聚糖结构数据库. 迄今为止, GlycoShape数据库已收录了500多个独特的聚糖, 这些结构可以通过一种名为ReGlyco的强大算法链接到蛋白质上, 直接与蛋白质数据库(RCSB PDB)和AlphaFold蛋白结构数据库等开放存取库中的结构数据兼容. 此外, 数据库也在往AI与机器学习的集成网站、跨数据源的整合分析、一站式数据分析和实时数据更新等方向提升.

5 总结与展望

糖蛋白质组学已在多个领域展现了其研究潜力, 逐渐成为生命科学中的一个重要领域. 高分辨率质谱技术、糖蛋白富集技术、数据解析软件以及分析工具等方面不断地创新, 进一步促进了糖蛋白质组学在基础研究和临床应用中的发展[125]. 目前, 糖蛋白质组学研究已经从单一的糖基化位点鉴定发展到对糖基化宏观和微观异质性的研究, 这种深入的研究有助于我们更好地理解糖基化在生物过程中的作用, 以及其在疾病发生和发展中的影响[126]. 近年来, 糖蛋白质组学的研究已不再单纯关注糖肽鉴定数量的增加, 而是转向提升分析的灵敏度、准确性和全面性, 特别是在大规模的完整糖肽水平糖蛋白质组分析中.

首先, 糖基化的复杂性使得对低丰度糖肽的高灵敏度检测和定量变得尤为重要, 而深度学习技术的应用为这一挑战提供了新的解决方案. 深度学习通过构建神经网络和训练算法, 能够在糖肽谱图预测、结构推断以及定量分析中实现前所未有的准确度. 例如, 神经网络模型在糖肽谱图的预测中已展示出优势, 能够提高分析的灵敏度[24,25]. 目前, 较高的糖肽解析率也仅为10%左右[23], 这一数据的局限性主要受到现有质谱技术和分析方法的制约. 通过引入AI计算科学, 尤其是深度学习在谱图预测中的应用, 有望突破传统技术的局限, 提升糖肽的定性和定量分析精度, 增强糖蛋白质组学在复杂生物样本中的应用潜力. 其技术应用有以下切入点: AI深度学习和注意力机制鉴定低丰度糖肽;

知识图谱嵌入的方式构建迁移学习模型提升糖肽识别率; 多模态数据融合提升异构体区分能力. 目前, 在谱图预测方面, 通过神经网络等模型对谱图进行深度学习, 预测糖肽的二级谱图, 实现高精度匹配[24,25]; 在特征提取方面, 利用卷积神经网络(CNN)等模型学习质谱中的各个糖肽碎片离子的碎裂模式及对应谱图, 对糖链拓扑结构进行预测[25]. 利用AI技术可系统性解决糖蛋白质的复杂性、动态性、低丰度等问题, 开发AI驱动的糖基化标志物发现平台, 结合分子动力学模拟预测糖链构象对质谱行为的影响等, 推动其在精准医学和生物制药中的转化应用. 这不仅能提升糖肽分析的解析深度, 还将促进整个领域向更高通量、精准度和灵敏度方向发展.

其次, 糖肽的精细结构解析仍然是糖蛋白质组学研究中的一大挑战. 糖肽的异构体、多糖链连接等复杂结构的精确鉴定, 一直是困扰研究者的难题. 糖基化修饰的异质性包括糖链的多样性、糖链长度、分支方式及其空间结构, 尤其是在复杂生物样本中, 这些细节对于理解糖基化对细胞功能和疾病的影响至关重要[127]. 虽然新兴技术如HCD触发紫外光解离(UVPD)在糖链连接的分析中已取得一定进展[128], 但仍然面临着若干挑战, 例如需要不同软件组合来处理不同数据、限速问题以及准确性问题等. 因此, 如何利用新技术更高效地解析糖肽的精细结构、提高数据的通量和准确性, 将成为未来糖蛋白质组学发展的关键方向. 结合创新的质谱分析技术、改善的数据处理算法和自动化分析平台, 能够为糖蛋白质组学的深度解析提供更强大的支持.

除了解析深度, 完整糖肽的精度和准确性也是糖蛋白质组学中一个亟待解决的重要问题. 在糖肽定量分析中, 糖基化位点上糖链的准确性直接影响分析结果的可靠性和生物学意义. 为提高糖肽的定性和定量精度, 结合化学合成技术进行标准糖肽的合成[129], 并通过构建标准谱图库来优化质谱分析流程, 已经成为一种行之有效的策略. 标准糖肽的合成可以确保谱图库的高质量, 从而减少由于实验误差或样本复杂性引起的偏差. 通过不断更新和扩展标准谱库, 糖蛋白质组学分析的精度和准确性将得到显著提升. 这一策略的核心在于通过高质量的参考数据来优化分析流程, 帮助研究人员在面对不同生物样本时进行更加精确的糖肽鉴定和定量分析. 标准谱库的建设与不断完善, 将推动糖蛋白质组学研究朝着更高精度、更广泛应用的方向迈进.

最后, 糖蛋白质组学作为研究糖基化修饰的关键领域, 与生物功能调控和疾病发生关系密切, 尤其在精准医学和疾病生物标志物的研究中展现出巨大的潜力. 然而, 糖蛋白质组学研究中产生的数据庞大且复杂, 如何高效整合、存储并利用这些数据, 已成为推动该领域发展的重要课题. 为了有效整合不同来源的数据, 必须依赖于新的数据筛选方法和生物信息技术的发展. 例如, 通过多中心交互平台, 能够将来自不同研究团队和实验室的数据汇集在一起, 形成具有广泛代表性的数据集. 这些平台不仅需要支持大数据的存储与管理, 还要能够提供数据分析、挖掘与共享的功能, 以促进跨学科、跨领域的合作. 此外, 随着研究的深入, 糖蛋白质组学领域需要建立更加智能化的数据库和专家知识库, 以实现数据的高效存取与智能分析. 这些数据库不仅应包含基本的糖基化信息, 还应整合蛋白质结构信息、生物学功能、疾病机制和临床相关数据, 为研究人员提供全面的参考框架, 推动糖蛋白质组学与生物功能调控的深度融合.

随着新技术和新方法的不断涌现, 糖蛋白质组学有望在精准医学、疾病诊断与治疗以及生物学研究等多个领域发挥越来越重要的作用. 尤其是通过深度学习和人工智能技术的广泛应用, 糖蛋白质组学的分析能力将得到前所未有的提升, 推动糖基化研究进入更加精细和高效的阶段. 未来, 糖蛋白质组学将不断拓展其在基础科学和临床研究中的广泛应用, 深刻改变我们对生物系统、疾病机制及其治疗的理解, 并为个性化医疗和早期疾病诊断提供更加精准的工具和策略.

致谢 谨以最崇高的敬意与最热烈的祝贺, 献礼复旦大学生物医学研究院建院20周年. 感谢学院糖蛋白质组学团队全体成员的精诚合作与不懈拼搏, 使得糖蛋白组学的研究不断取得突破. 感谢研究院初创院长贺福初院士的战略引领、持续关怀与大力支持. 感谢并深切缅怀生物医学研究院原常务副院长杨芃原教授(1949年6月12日~2021年5月31日), 先生作为我国糖蛋白质组学领域的先行者, 为该领域的发展奠定了坚实基础, 对学院人才培养、学科交叉、科技创新作出了卓越贡献. 最后, 衷心感谢国家自然科学基金委员会、科学技术部、国家卫生健康委员会对我们科研工作的长期关注与大力支持.

参考文献

1 Ruhaak L R, Xu G, Li Q, et al. Mass spectrometry approaches to glycomic and glycoproteomic analyses. Chem Rev, 2018, 118: 7886–7930
2 Thomas D R, Scott N E. Glycoproteomics: growing up fast. Curr Opin Struct Biol, 2021, 68: 18–25
3 Riley N M, Bertozzi C R, Pitteri S J. A pragmatic guide to enrichment strategies for mass spectrometry–based glycoproteomics. Mol Cell Proteomics, 2021, 20: 100029
4 Taniguchi N. Human disease glycomics/proteome initiative (HGPI). Mol Cell Proteomics, 2008, 7: 626–627
5 Hart G W, Copeland R J. Glycomics hits the big time. Cell, 2010, 143: 672–676
6 Christiansen M N, Chik J, Lee L, et al. Cell surface protein glycosylation in cancer. PROTEOMICS, 2014, 14: 525–546
7 Liu X, Fu B, Chen J, et al. High-throughput intact glycopeptide quantification strategy with targeted-MS (HTiGQs-target) reveals site-specific IgG N-glycopeptides as biomarkers for hepatic disorder diagnosis and staging. Carbohydrate Polyms, 2024, 325: 121499
8 Sun Z, Fu B, Wang G, et al. High-throughput site-specific *N*-glycoproteomics reveals glyco-signatures for liver disease diagnosis. Natl Sci Rev, 2023, 10: nwac059
9 Cao W, Liu M, Kong S, et al. Recent advances in software tools for more generic and precise intact glycopeptide analysis. Mol Cell Proteomics, 2021, 20: 100060
10 Zhang H, Li X, Martin D B, et al. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. Nat Biotechnol, 2003, 21: 660–666
11 Zielinska D F, Gnad F, Wiśniewski J R, et al. Precision mapping of an *In Vivo* N-glycoproteome reveals rigid topological and sequence constraints. Cell, 2010, 141: 897–907
12 Zielinska D F, Gnad F, Schropp K, et al. Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. Mol Cell, 2012, 46: 542–548
13 Mayampurath A, Yu C Y, Song E, et al. Computational framework for identification of intact glycopeptides in complex samples. Anal Chem, 2014, 86: 453–463
14 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, 2016, 34: 84–88
15 Kaji H, Saito H, Yamauchi Y, et al. Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins. Nat Biotechnol, 2003, 21: 667–672
16 Zeng W F, Liu M Q, Zhang Y, et al. pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. Sci Rep, 2016, 6: 25102
17 Liu M Q, Zeng W F, Fang P, et al. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat Commun, 2017, 8: 438
18 Zeng W F, Cao W Q, Liu M Q, et al. Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. Nat Methods, 2021, 18: 1515–1523
19 Kong S, Gong P, Zeng W F, et al. pGlycoQuant with a deep residual network for quantitative glycoproteomics at intact glycopeptide level. Nat Commun, 2022, 13: 7539
20 Zeng W F, Yan G, Zhao H, et al. Uncovering missing glycans and unexpected fragments with pGlycoNovo for site-specific glycosylation analysis across species. Nat Commun, 2024, 15: 8055
21 Yang Y, Yan G, Kong S, et al. GproDIA enables data-independent acquisition glycoproteomics with comprehensive statistical control. Nat Commun, 2021, 12: 6073
22 Shen J, Jia L, Dang L, et al. StrucGP: *de novo* structural sequencing of site-specific N-glycan on glycoproteins using a modularization strategy. Nat Methods, 2021, 18: 921–929
23 Fang Z, Qin H, Mao J, et al. Glyco-Decipher enables glycan database-independent peptide matching and in-depth characterization of site-specific N-glycosylation. Nat Commun, 2022, 13: 1900
24 Yang Y, Fang Q. Prediction of glycopeptide fragment mass spectra by deep learning. Nat Commun, 2024, 15: 2448
25 Zong Y, Wang Y, Qiu X, et al. Deep learning prediction of glycopeptide tandem mass spectra powers glycoproteomics. Nat Mach Intell, 2024, 6: 950–961
26 Lu L, Riley N M, Shortreed M R, et al. O-pair search with metaMorpheus for O-glycopeptide characterization. Nat Methods, 2020, 17: 1133–1138
27 Kong A T, Leprevost F V, Avtonomov D M, et al. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. Nat Methods, 2017, 14: 513–520

28 Polasky D A, Yu F, Teo G C, et al. Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-glyco. Nat Methods, 2020, 17: 1125–1132

29 Banazadeh A, Veillon L, Wooding K M, et al. Recent advances in mass spectrometric analysis of glycoproteins. ELECTROPHORESIS, 2017, 38: 162–189

30 Bern M, Kil Y J, Becker C. Byonic: advanced peptide and protein identification software. CP BioInf, 2012, 40

31 Kawahara R, Chernykh A, Alagesan K, et al. Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. Nat Methods, 2021, 18: 1304–1316

32 Sun W, Zhang Q, Zhang X, et al. Glycopeptide database search and de novo sequencing with PEAKS GlycanFinder enable highly sensitive glycoproteomics. Nat Commun, 2023, 14: 4046

33 Liu L, Yu M, Zhang Y, et al. Hydrazide functionalized core–shell magnetic nanocomposites for highly specific enrichment of *N*-glycopeptides. ACS Appl Mater Interfaces, 2014, 6: 7823–7832

34 Bai H, Pan Y, Qi L, et al. Development a hydrazide-functionalized thermosensitive polymer based homogeneous system for highly efficient N-glycoprotein/glycopeptide enrichment from human plasma exosome. Talanta, 2018, 186: 513–520

35 Li M, Huang J, Ma M, et al. Selective enrichment of sialylglycopeptides enabled by click chemistry and dynamic covalent exchange. Anal Chem, 2022, 94: 6681–6688

36 Goumenou A, Delaunay N, Pichon V. Recent advances in lectin-based affinity sorbents for protein glycosylation studies. Front Mol Biosci, 2021, 8: 746822

37 Stillmark H. Ueber ricin, ein giftiges ferment aus den samen von ricinus comm. L. und einigen anderen euphorbiaceen: inaugural-dissertation. Dorpat: Schnakenburg, 1888

38 Chao X, Zhang B, Yang S, et al. Enrichment methods of N-linked glycopeptides from human serum or plasma: a mini-review. Carbohydr Res, 2024, 538: 109094

39 Yin H, Tan Z, Wu J, et al. Mass-selected site-specific core-fucosylation of serum proteins in hepatocellular carcinoma. J Proteome Res, 2015, 14: 4876–4884

40 Hütte H J, Tiemann B, Shcherbakova A, et al. A bacterial mannose binding lectin as a tool for the enrichment of c- and o-mannosylated peptides. Anal Chem, 2022, 94: 7329–7338

41 Zhang C, Ye Z, Xue P, et al. Evaluation of different N-glycopeptide enrichment methods for N-glycosylation sites mapping in mouse brain. J Proteome Res, 2016, 15: 2960–2968

42 Pérez S, Sarkar A, Rivet A, et al. Glyco3D: a portal for structural glycosciences. Methods Mol Biol, 2015, 1273: 241–258

43 Bonnardel F, Mariethoz J, Salentin S, et al. UniLectin3D, a database of carbohydrate binding proteins with curated information on 3D structures and interacting ligands. Nucleic Acids Res, 2019, 47: D1236–D1244

44 Gbormittah F O, Lee L Y, Taylor K O, et al. Comparative studies of the proteome, glycoproteome, and N-glycome of clear cell renal cell carcinoma plasma before and after curative nephrectomy. J Proteome Res, 2014, 13: 4889–4900

45 Wang X, Xia N, Liu L. Boronic acid-based approach for separation and immobilization of glycoproteins and its application in sensing. Int J Mol Sci, 2013, 14: 20890–20912

46 Weith H L, Wiebers J L, Gilham P T. Synthesis of cellulose derivatives containing the dihydroxyboryl group and a study of their capacity to form specific complexes with sugars and nucleic acid components. Biochemistry, 1970, 9: 4396–4401

47 Chen W, Smeekens J M, Wu R. A universal chemical enrichment method for mapping the yeast n-glycoproteome by mass spectrometry (MS). Mol Cell Proteomics, 2014, 13: 1563–1572

48 Xu Y, Wu Z, Zhang L, et al. Highly specific enrichment of glycopeptides using boronic acid-functionalized mesoporous silica. Anal Chem, 2009, 81: 503–508

49 Zhang X, Wang J, He X, et al. Tailor-made boronic acid functionalized magnetic nanoparticles with a tunable polymer shell-assisted for the selective enrichment of glycoproteins/glycopeptides. ACS Appl Mater Interfaces, 2015, 7: 24576–24584

50 Mujahid Ali M, Hussain D, Xu B, et al. Diethylenetriamine assisted functionalization of boronic acid on poly GMA-MAA-DVB for selective enrichment of glycoproteins and glycopeptides. Talanta, 2020, 219: 121178

51 Wang J, Wang Y, Gao M, et al. Multilayer hydrophilic poly(phenol-formaldehyde resin)-coated magnetic graphene for boronic acid immobilization as a novel matrix for glycoproteome analysis. ACS Appl Mater Interfaces, 2015, 7: 16011–16017

52 Zou X, Liu D, Zhong L, et al. Synthesis and characterization of a novel boronic acid-functionalized chitosan polymeric nanosphere for highly specific enrichment of glycopeptides. Carbohydrate Polyms, 2012, 90: 799–804

53 Kong S, Zhang Q, Yang L, et al. Effective enrichment strategy using boronic acid-functionalized mesoporous graphene–silica composites for intact N- and O-linked glycopeptide analysis in human serum. Anal Chem, 2021, 93: 6682–6691

54 Alpert A J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr A,

1990, 499: 177–196
55 Sun N, Wu H, Chen H, et al. Advances in hydrophilic nanomaterials for glycoproteomics. Chem Commun, 2019, 55: 10359–10375
56 Yu M, Di Y, Zhang Y, et al. Fabrication of alkoxyamine-functionalized magnetic core-shell microspheres via reflux precipitation polymerization for glycopeptide enrichment. Polymers, 2016, 8: 74
57 Wang Y, Wang J, Gao M, et al. An ultra hydrophilic dendrimer-modified magnetic graphene with a polydopamine coating for the selective enrichment of glycopeptides. J Mater Chem B, 2015, 3: 8711–8716
58 Bi C, Zhao Y, Shen L, et al. Click synthesis of hydrophilic maltose-functionalized iron oxide magnetic nanoparticles based on dopamine anchors for highly selective enrichment of glycopeptides. ACS Appl Mater Interfaces, 2015, 7: 24670–24678
59 Snovida S I, Bodnar E D, Viner R, et al. A simple cellulose column procedure for selective enrichment of glycopeptides and characterization by nano LC coupled with electron-transfer and high-energy collisional-dissociation tandem mass spectrometry. Carbohydr Res, 2010, 345: 792–801
60 Wu Y, Liu Q, Xie Y, et al. Core-shell structured magnetic metal-organic framework composites for highly selective enrichment of endogenous N-linked glycopeptides and phosphopeptides. Talanta, 2018, 190: 298–312
61 Ji Y, Xiong Z, Huang G, et al. Efficient enrichment of glycopeptides using metal–organic frameworks by hydrophilic interaction chromatography. Analyst, 2014, 139: 4987–4993
62 Dong X, Qin H, Mao J, et al. In-depth analysis of glycoprotein sialylation in serum using a dual-functional material with superior hydrophilicity and switchable surface charge. Anal Chem, 2017, 89: 3966–3972
63 Wang Z, Fang Z, Liu L, et al. Development of an integrated platform for the simultaneous enrichment and characterization of N- and O-linked intact glycopeptides. Anal Chem, 2023, 95: 7448–7457
64 Qing G, Yan J, He X, et al, Recent advances in hydrophilic interaction liquid interaction chromatography materials for glycopeptide enrichment and glycan separation, et al. TrAC Trends in Anal Chem, 2020, 124: 115570
65 Zhu H, Li X, Qu J, et al. Diethylaminoethyl sepharose (DEAE-Sepharose) microcolumn for enrichment of glycopeptides. Anal Bioanal Chem, 2017, 409: 511–518
66 Jiang B, Liang Y, Wu Q, et al. New GO–PEI–Au–L-Cys ZIC-HILIC composites: synthesis and selective enrichment of glycopeptides. Nanoscale, 2014, 6: 5616–5619
67 Woodin C L, Hua D, Maxon M, et al. GlycoPep grader: a web-based utility for assigning the composition of *N*-linked glycopeptides. Anal Chem, 2012, 84: 4821–4829
68 Hu H, Khatri K, Klein J, et al. A review of methods for interpretation of glycopeptide tandem mass spectral data. Glycoconj J, 2016, 33: 285–296
69 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: 4347–4355
70 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: 6804–6811
71 Cao L, Tolić N, Qu Y, et al. Characterization of intact N- and O-linked glycopeptides using higher energy collisional dissociation. Anal Biochem, 2014, 452: 96–102
72 Kuo C W, Guu S Y, Khoo K H. Distinctive and complementary MS^2 fragmentation characteristics for identification of sulfated sialylated *N*-glycopeptides by nanoLC-MS/MS workflow. J Am Soc Mass Spectrom, 2018, 29: 1166–1178
73 Fang P, Xie J J, Sang S, et al. Multilayered N-glycoproteome profiling reveals highly heterogeneous and dysregulated protein n-glycosylation related to Alzheimer's disease. Anal Chem, 2020, 92: 867–874
74 Frese C K, Altelaar A F M, van den Toorn H, et al. Toward Full peptide sequence coverage by Dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. Anal Chem, 2012, 84: 9668–9673
75 Kolbowski L, Mendes M L, Rappsilber J. Optimizing the parameters governing the fragmentation of cross-linked peptides in a tribrid mass spectrometer. Anal Chem, 2017, 89: 5311–5318
76 Chen G, Zhang Y, Trinidad J C, et al. Distinguishing sulfotyrosine containing peptides from their phosphotyrosine counterparts using mass spectrometry. J Am Soc Mass Spectrom, 2018, 29: 455–462
77 Riley N M, Hebert A S, Westphall M S, et al. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. Nat Commun, 2019, 10: 1311
78 Swaney D L, McAlister G C, Wirtala M, et al. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. Anal Chem, 2007, 79: 477–485
79 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: 3121–3129
80 Cheng K, Chen R, Seebun D, et al. Large-scale characterization of intact N-glycopeptides using an automated glycoproteomic method. J Proteomics, 2014, 110: 145–154

81 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: 5181–5188

82 Yu C Y, Mayampurath A, Zhu R, et al. Automated glycan sequencing from tandem mass spectra of N-linked glycopeptides. Anal Chem, 2016, 88: 5725–5732

83 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: 3652–3666

84 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: 5666–5675

85 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: 2466–2473

86 Medzihradszky K F, Kaasik K, Chalkley R J. Tissue-specific glycosylation at the glycopeptide level. Mol Cell Proteomics, 2015, 14: 2103–2110

87 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: 5023–5032

88 Kim J W, Hwang H, Lim J S, et al. gFinder: a web-based bioinformatics tool for the analysis of *N*-glycopeptides. J Proteome Res, 2016, 15: 4116–4125

89 Park G W, Kim J Y, Hwang H, et al. Integrated glycoproteome analyzer (I-GPA) for automated identification and quantitation of site-specific N-glycosylation. Sci Rep, 2016, 6: 21175

90 Wu Y, Mechref Y, Klouckova I, et al. Mapping site-specific protein N-glycosylations through liquid chromatography/mass spectrometry and targeted tandem mass spectrometry. Rapid Comm Mass Spectrom, 2010, 24: 965–972

91 Liu G, Cheng K, Lo C Y, et al. A comprehensive, open-source platform for mass spectrometry-based glycoproteomics data analysis. Mol Cell Proteomics, 2017, 16: 2032–2047

92 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: 3881–3895

93 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: 1–16

94 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: 1728–1740

95 Chau T H, Chernykh A, Kawahara R, et al. Critical considerations in *N*-glycoproteomics. Curr Opin Chem Biol, 2023, 73: 102272

96 Fang P, Ji Y, Silbern I, et al. A streamlined pipeline for multiplexed quantitative site-specific N-glycoproteomics. Nat Commun, 2020, 11: 5268

97 Yu F, Haynes S E, Teo G C, et al. Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. Mol Cell Proteomics, 2020, 19: 1575–1585

98 Steentoft C, Vakhrushev S Y, Joshi H J, et al. Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. EMBO J, 2013, 32: 1478–1488

99 Ye Z, Mao Y, Clausen H, et al. Glyco-DIA: a method for quantitative O-glycoproteomics with *in silico*-boosted glycopeptide libraries. Nat Methods, 2019, 16: 902–910

100 Jiang S, Li H, Zhang L, et al. Generic diagramming platform (GDP): a comprehensive database of high-quality biomedical graphics. Nucleic Acids Res, 2025, 53: D1670–D1676

101 Li X, Xu Z, Hong X, et al. Databases and bioinformatic tools for glycobiology and glycoproteomics. Int J Mol Sci, 2020, 21: 6727

102 Raman R, Venkataraman M, Ramakrishnan S, et al. Advancing glycomics: implementation strategies at the consortium for functional glycomics. Glycobiology, 2006, 16: 82R–90R

103 Hashimoto K, Goto S, Kawano S, et al. KEGG as a glycome informatics resource. Glycobiology, 2006, 16: 63R–70R

104 Lütteke T, Bohne-Lang A, Loss A, et al. GLYCOSCIENCES.de: an internet portal to support glycomics and glycobiology research. Glycobiology, 2006, 16: 71R–81R

105 Cooper C A. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. Nucleic Acids Res, 2001, 29: 332–335

106 Kameyama A, Kikuchi N, Nakaya S, et al. A strategy for identification of oligosaccharide structures using observational multistage mass spectral library. Anal Chem, 2005, 77: 4719–4725

107 Campbell M P, Packer N H. UniCarbKB: new database features for integrating glycan structure abundance, compositional glycoproteomics data, and disease associations. Biochim Biophys Acta Gen Subj, 2016, 1860: 1669–1675

108 Akune Y, Lin C H, Abrahams J L, et al. Comprehensive analysis of the N-glycan biosynthetic pathway using bioinformatics to generate UniCorn: a theoretical N-glycan structure database. Carbohydr Res, 2016, 431: 56–63

109 Royle L, Radcliffe C M, Dwek R A, et al. Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. Methods Mol Biol, 2006, 347: 125–143
110 von der Lieth C W, Freire A A, Blank D, et al. EUROCarbDB: an open-access platform for glycoinformatics. Glycobiology, 2011, 21: 493–502
111 Campbell M P, Nguyen-Khuong T, Hayes C A, et al. Validation of the curation pipeline of UniCarb-DB: building a global glycan reference MS/MS repository. Biochim Biophys Acta, 2014, 1844: 108–116
112 Mariethoz J, Alocci D, Gastaldello A, et al. Glycomics@ExPASy: bridging the gap. Mol Cell Proteomics, 2018, 17: 2164–2176
113 Tiemeyer M, Aoki K, Paulson J, et al. GlyTouCan: an accessible glycan structure repository. Glycobiology, 2017, 27: 915–919
114 Böhm M, Bohne-Lang A, Frank M, et al. Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update). Nucleic Acids Res, 2019, 47: D1195–D1201
115 Zhao S, Walsh I, Abrahams J L, et al. GlycoStore: a database of retention properties for glycan analysis. Bioinformatics, 2018, 34: 3231–3232
116 Choudhary P, Nagar R, Singh V, et al. ProGlycProt V2.0, a repository of experimentally validated glycoproteins and protein glycosyltransferases of prokaryotes. Glycobiology, 2019, 29: 461–468
117 Clerc O, Deniaud M, Vallet S D, et al. MatrixDB: integration of new data with a focus on glycosaminoglycan interactions. Nucleic Acids Res, 2019, 47: D376–D381
118 Mariethoz J, Khatib K, Alocci D, et al. SugarBindDB, a resource of glycan-mediated host–pathogen interactions. Nucleic Acids Res, 2016, 44: D1243–D1250
119 Kaji H, Yamauchi Y, Takahashi N, et al. Mass spectrometric identification of N-linked glycopeptides using lectin-mediated affinity capture and glycosylation site–specific stable isotope tagging. Nat Protoc, 2006, 1: 3019–3027
120 Zhang H, Loriaux P, Eng J, et al. UniPep - a database for human N-linked glycosites: a resource for biomarker discovery. Genome Biol, 2006, 7: R73
121 Sun S, Hu Y, Ao M, et al. N-GlycositeAtlas: a database resource for mass spectrometry-based human N-linked glycoprotein and glycosylation site mapping. Clin Proteom, 2019, 16: 35
122 Huang J, Wu M, Zhang Y, et al. OGP: a repository of experimentally characterized *O*-glycoproteins to facilitate studies on *O*-glycosylation. Genomics Proteomics BioInf, 2021, 19: 611–618
123 Wu M, Liu H, Wang X, et al. GlycAP, a glycoproteomic analysis platform for site-specific N-glycosylation research. Int J Mass Spectrom, 2022, 482: 116947.
124 Ives C M, Singh O, D'Andrea S, et al. Restoring protein glycosylation with GlycoShape. Nat Methods, 2024, 21: 2117–2127
125 Xiong Y, Zheng Y, Yan Y, et al. Circulating proteomic panels for risk stratification of intracranial aneurysm and its rupture. EMBO Mol Med, 2022, 14: e14713
126 Liu L, Liu L, Wang Y, et al. Robust glycoproteomics platform reveals a tetra-antennary site-specific glycan capping with sialyl-lewis antigen for early detection of gastric cancer. Adv Sci, 2024, 11: 2306955
127 Reily C, Stewart T J, Renfrow M B, et al. Glycosylation in health and disease. Nat Rev Nephrol, 2019, 15: 346–366
128 Brodbelt J S, Morrison L J, Santos I. Ultraviolet photodissociation mass spectrometry for analysis of biological molecules. Chem Rev, 2020, 120: 3328–3380
129 Xu Z, Liu Y, Liu J, et al. Integrated chemoenzymatic synthesis of a comprehensive sulfated ganglioside glycan library to decipher functional sulfoglycomics and sialoglycomics. Nat Chem, 2024, 16: 881–892

Summary for “位点特异性糖基化蛋白质组学研究回顾与展望”

Review and perspectives on site-specific glycoproteomics research

Lehan Chen[1], Jialin Liu[1] & Weiqian Cao[1,2*]

[1] *Shanghai Fifth People's Hospital and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China*
[2] *NHC Key Laboratory of Glycoconjugates Research, Fudan University, Shanghai 200032, China*
* Corresponding author, E-mail: wqcao@fudan.edu.cn

Glycosylation is one of the most intricate and variable post-translational modifications (PTMs), playing a pivotal role in various biological processes such as cell signaling, immune response, protein folding, and molecular recognition. The dynamic and diverse nature of glycosylation makes it an essential modification in cellular function and organismal development, as well as in disease progression. Glycoproteomics has emerged as a core field that enables the in-depth analysis of glycoproteins, providing valuable insights into the glycosylation sites and the underlying glycan structures. This is particularly important for understanding how changes in glycosylation patterns are linked to disease states, such as cancer, autoimmune disorders, and neurodegenerative diseases.

Over the past two decades, significant advancements in mass spectrometry (MS)-based glycoproteomics have revolutionized the study of protein glycosylation. Initially, the identification of glycosylation sites was primarily based on peptide mapping techniques. However, the field has evolved to focus on the analysis of intact glycopeptides, which preserve both the peptide backbone and the attached glycans, providing a more complete view of glycosylation. The development of cutting-edge MS technologies has enabled researchers to detect and characterize a wide variety of glycopeptides with high sensitivity and specificity. These advancements have allowed for the identification of glycosylation sites at unprecedented levels of detail, unveiling complex glycan structures and modifications that were previously difficult to study.

One of the key developments in glycoproteomics has been the enhancement of glycopeptide enrichment methods, which facilitate the isolation of glycosylated peptides from complex biological samples. These techniques, such as lectin affinity chromatography and hydrophilic interaction liquid chromatography (HILIC), have significantly improved the sensitivity and throughput of glycoproteomics analyses. In parallel, advances in mass spectrometry analytical strategies, including high-resolution MS and tandem MS (MS/MS), have provided deeper insights into glycopeptide fragmentation patterns, allowing for the determination of site-specific glycosylation of glycoproteins.

Despite these advancements, several challenges remain in glycoproteomics, such as the need for improved analytical depth and reproducibility, the complexity of glycan structures, and the incomplete characterization of low-abundance glycosylated proteins. Furthermore, data interpretation tools and glycan databases are still evolving to keep pace with the growing complexity of glycosylation data. Looking ahead, the future of glycoproteomics will likely involve the development of more accurate, high-throughput methods for glycoprotein analysis, the integration of multi-omics approaches, and the application of glycoproteomics in personalized medicine, particularly for the discovery of biomarkers and therapeutic targets in disease.

glycoproteomics, site-specific glycosylation analysis, identification and quantification of intact glycopeptides, mass spectrometry analysis

doi: 10.1360/TB-2024-1409