

黏蛋白型 O-糖基化蛋白质组学分析进展

刘兆亮^{1,2}, 于永亮¹, 叶明亮²

(1. 东北大学理学院化学系, 辽宁 沈阳 110819; 2. 中国科学院大连化学物理研究所, 辽宁 大连 116023)

摘要: 黏蛋白型糖基化作为蛋白质重要的翻译后修饰, 具有高度异质性和多位点修饰的特点, 参与细胞识别、免疫应答及信号转导等关键生物学过程, 并广泛存在于多种肿瘤中。然而, 由于黏蛋白型糖基化具有缺乏保守序列、复杂的糖链结构且在生物样本中丰度较低的特点, 使其精确分析面临极大挑战。发展高效的富集技术对于从复杂生物样品中捕获低丰度 O-糖肽具有重要意义, 先进的质谱数据库检索策略是从复杂碎片信息中准确解读糖肽结构的关键。本文综述了近 5 年来用于富集 O-糖肽的方法及其质谱数据库检索策略的最新进展, 并对其发展前景进行展望。

关键词: 黏蛋白型糖基化; 完整糖肽富集; 质谱; 亲和色谱; O-糖蛋白酶; 数据库检索

中图分类号: O657.63

文献标志码: A

文章编号: 1004-2997(2026)03-0303-10

DOI: 10.7538/zpxb.2025.0156

CSTR: 32365.14.zpxb.2025.0156

Recent Advances in Mucin-Type O-Glycosylation Proteomic Analysis

LIU Zhao-liang^{1,2}, YU Yong-liang¹, YE Ming-liang²

(1. Department of Chemistry, College of Sciences, Northeastern University, Shenyang 110819, China;

2. Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian 116023, China)

Abstract: Mucin-type O-GalNAc glycosylation is a pivotal post-translational modification of proteins, characterized by high glycan structural heterogeneity and multisite modifications. It regulates core biological processes including cellular recognition, immune response, and signal transduction—aberrations in this modification are linked to autoimmune diseases, diabetes, cardiovascular disorders, inflammation, viral infections, neurodegeneration, and cancers. Elucidating disease-specific O-glycosylation patterns will advance disease detection, decode molecular pathogenesis, and inform the discovery of novel therapeutic targets. The bottom-up strategy is the gold standard for large-scale glycoproteomic analysis, yet comprehensive O-glycosylation profiling remains challenging due to non-conserved flanking sequences around glycosylation sites, complex glycan structures, and the low abundance of O-glycopeptides. High-efficiency enrichment techniques are critical for capturing low-abundance O-glycopeptides from complex matrices, while advanced database search strategies enable accurate interpretation of tandem mass spectrometry (MS/MS) data. This review summarized key progress in O-glycopeptide enrichment and MS-based database search methods over the past five years. Enrichment methods have seen significant innovations: hybrid materials integrating hydrophilic interaction liquid chromatography (HILIC) with complementary affinity techniques (e.g., immobilized metal ion affinity chromatography (IMAC), boronic acid

chemistry) greatly enhance enrichment efficiency. For example, Ti-IMAC materials capture sialylated *O*-glycopeptides from 0.1 μL human serum and enable the identification of ~ 200 *O*-glycopeptides, while boronic acid-functionalized mesoporous composites analyzed using 1 μL of serum yield 724 *N*-glycopeptides and 152 *O*-glycopeptides. Automated high performance liquid chromatography (HPLC) workflows enable simultaneous *N/O*-glycopeptide separation, identifying 181 *N*- and 17 *O*-glycopeptides with significant changes in gastric cancer serum. *O*-Glycoprotease-based methods (OgpA/IMPa) combined with solid-phase chemoenzymatic approaches allow specific enrichment: MOTAI distinguishes Tn/sTn from other *O*-glycopeptides in colon cancer tissues, identifying 32 upregulated Tn/sTn glycoproteins. Bioorthogonal strategies (GalNAz metabolic labeling-click chemistry) such as Click-iG identify 262 *O*-glycosylation sites in mouse tissues. Database search tools have overcome traditional limitations: O-Search-Pattern uses Y-ion pattern matching to boost *O*-glycopeptide identifications by 15.4%-199.0% compared with other tools; MSFragger-Glyco leverages open search and ion indexing to increase identifications by 4-6-fold and reduce analysis time to minutes; pGlyco3's glycan-first strategy enables fast, precise intact glycopeptide analysis. Machine learning approaches show promise: CandyCrunch predicts glycan structures from LC-MS/MS data with 90.3% accuracy in seconds; DeepGlyco uses tree-LSTM and graph neural networks to distinguish glycan isomers; GlyPep-Quant integrates random forests and DBSCAN to improve quantitative performance; MarkerPredict identifies cancer biomarkers through disordered protein and signal network features. These tools address bottlenecks such as low-abundance glycopeptide detection and isomer differentiation. Overall, this review provides a systematic overview of *O*-glycopeptide analysis methods, guiding technical innovation and deepening understanding of disease mechanisms-ultimately accelerating the translation of glycoproteomic insights into clinical applications.

Key words: mucin-type *O*-glycosylation; intact glycopeptide enrichment; mass spectrometry; affinity chromatography; *O*-glycoprotease; database search

黏蛋白型糖基化是真核生物中最重要的蛋白质翻译后修饰之一,在丝氨酸或苏氨酸残基上以 *O*-糖苷键连接 *N*-乙酰半乳糖胺 (GalNAc) 并延伸形成高度异质化的糖链,又被称为 *O*-GalNAc 糖基化。黏蛋白型糖基化构建的亲水性聚糖屏障是黏膜组织的核心结构,深度参与细胞识别、免疫调节、信号转导等关键生理过程,维系机体多系统稳态^[1-3]。异常黏蛋白型糖基化与多种癌症^[4-6]、神经系统综合征^[7]、代谢性疾病^[8]和胃肠道炎症^[9-10]等疾病密切相关。因此,系统解析黏蛋白型 *O*-糖基化蛋白质组学,可为疾病机制的诠释提供分子锚点,为靶向糖基化通路的新颖诊疗策略提供理论依据和技术支持。

基于质谱的“自下而上”策略是目前应用最广泛的解析蛋白质组学技术。该策略将生物蛋白样品酶解为肽段,通过质谱技术鉴定肽段的氨基酸序列和翻译后修饰。由于糖肽在酶解物中

的相对丰度较低且其质谱信号易受非糖肽干扰,因此在质谱分析前需要对糖肽进行富集。此外,糖链结构的多样性、修饰位点的不确定性以及现有数据库缺乏全面的 *O*-糖基化修饰信息,导致数据库检索时匹配难度较大、假阳性率偏高,进一步制约了黏蛋白型糖肽的高效鉴定。

近年来, *O*-糖肽富集方法的不断开发^[11-13] (列于表 1) 和新数据库检索策略的提出^[14-15], 极大推动了黏蛋白型糖基化蛋白质组学的解析。本文将综述新兴的糖肽富集方法和数据库检索策略,并对黏蛋白型糖基化蛋白质组学分析方法的发展前景进行展望。

1 *O*-GalNAc 糖肽的富集方法

1.1 新型色谱富集方法

亲水作用色谱 (hydrophilic interaction liquid chromatography, HILIC) 法利用糖肽与亲水性固

表1 O-糖肽富集方法

Table 1 Enrichment methods of O-glycopeptide

方法 Method	原理 Theory	优势 Advantage	局限 Disadvantage
亲水作用色谱 (HILIC)法	糖肽与亲水性固相之间的相互作用	保留完整糖肽结构,可富集几乎所有类型的糖肽,质谱兼容性好	存在其他亲水性肽段的共洗脱
固定金属离子亲和色谱(IMAC)法	金属阳离子与唾液酸型糖肽发生静电相互作用	非破坏性,对唾液酸型糖肽具有高亲和力和高选择性	对中性或带正电的糖肽无富集效果,易受磷酸肽的干扰
硼酸化学法	糖链上的邻二羟基与硼酸形成可逆的硼酸酯键	反应可逆,可保留完整糖链结构	结合力相对较弱,对环境pH敏感
O-糖蛋白酶法	通过不同的O-糖蛋白酶特异性酶切产生对应的糖肽	特异性高	成本高,通量较低
生物正交标记富集法	通过代谢标记、点击化学等方法使糖链带有亲和和标签,利用亲和材料特异性捕获糖肽	特异性高,适用于活细胞动态研究	依赖代谢标记,操作步骤复杂

相之间的氢键相互作用实现对糖肽的富集。然而,O-糖链相对较短的特点使其与固相间的氢键作用较弱。除通过 PNGase F 去除 N-糖干扰外,研究者还采用“HILIC+”策略^[13,16-17],即在亲水性材料中设计引入其他富集机制,以提升对 O-糖肽的富集效率。

固定金属离子亲和色谱(immobilized metal ion affinity chromatography, IMAC)法基于固定金属阳离子与磷酸基团之间的螯合相互作用,主要用于磷酸肽的富集。由于唾液酸基团的负电性,IMAC被广泛用于唾液酸修饰的糖肽富集。Yue等^[16]开发了一种 Ti-IMAC(IV)材料,对唾液酸修饰的 O-GalNAc 糖肽及非唾液酸修饰的 O-GalNAc 糖肽进行富集。该材料兼具亲水性与螯合亲和力,含有大量的 Ti(IV)离子,能够与唾液酸型 O-GalNAc 糖肽发生螯合作用。该方法从 0.1 μL 人血清样本中鉴定到近 200 条 O-GalNAc 糖肽,其富集能力在肝癌患者血清中得到进一步验证。Cui等^[18]合成了一种超亲水四通道的 TiO₂ 静电纺纤维,用于 O-GalNAc 糖肽与磷酸肽的同时富集。该材料表面丰富的 Ti-OH 使其具有超亲水性。此外,相比于球形固相,多通道结构使得液体在通道内快速扩散,有助于形成稳定的水化层,促进糖肽的吸附和扩散。该材料在 200 μg 结肠癌细胞裂解物中鉴定到 183 条 O-GalNAc 糖肽与 3 198 条磷酸肽,并进一步应用于抗癌药物曲妥珠单抗耐药性机制的研究。

硼酸化学法基于硼酸与顺式邻二羟基发生的可逆共价反应实现糖肽的富集。在碱性条件下,固定在固相上的硼酸可以与顺式-1,2-二醇的聚糖形成五元或六元环酯,从而将糖肽保留在固

相上;在酸性条件下,硼酸酯键断裂,糖肽被释放。Kong等^[17]构建了一种硼酸功能化的介孔石墨烯-二氧化硅复合材料(GO@mSiO₂-GLYMO-APB),用于人血清中完整 N-糖肽和 O-糖肽的富集。得益于介孔的排阻效应、二氧化硅的亲水性和硼酸基团的可逆共价结合,该材料在 1 μL 人血清样本中鉴定到 724 条 N-糖肽及 152 条 O-糖肽,其富集性能显著优于单一富集机制的商品化材料。

由于大多数 O-糖肽富集策略会去除 N-糖链,以避免高丰度 N-糖肽的干扰,使得 N-糖肽与 O-糖肽的富集需分别进行,因此,结合高效液相色谱(high performance liquid chromatography, HPLC)策略为高效分离 N-糖肽与 O-糖肽提供了选择。相比于手动富集操作,将富集材料装载于色谱柱中并搭载液相系统的分离方法,可以精确控制流动相梯度、洗脱温度等参数,从而实现 N-糖肽与 O-糖肽精准、稳定地分离富集。此外,在临床生物标志物挖掘中,临床发现与验证需要处理大规模样本以减弱个体差异的干扰。自动化液相分离策略为此提供了标准化操作流程,可实现糖基化分析的高通量、高重现性。Wang等^[19]开发了一种可同时富集鉴定 N-糖肽与 O-糖肽的集成平台,该平台搭载了自动化亲水色谱系统,利用 N-糖肽(通常糖链更长、更复杂,亲水性更强)和 O-糖肽(糖链较短、亲水性较弱)在 HILIC 上保留能力的差异实现分离。在乙腈梯度洗脱下,通过分析不同时间洗脱出的馏分发现,当乙腈浓度大于 76.4%时,85.1%的 O-糖肽可被洗脱;当乙腈浓度在 76.4%~30.0%时,93.4%的 N-糖肽可被洗脱。最后,该团队将该平台应

用于临床胃癌血清的定量分析,发现了胃癌血清中181条N-糖肽及17条O-糖肽存在显著性变化。该方法具有操作简单、重现性高的特点,适用于大规模临床样本的糖蛋白质组学分析。Daramola等^[20]基于介孔石墨化碳柱开发了一种液相色谱-串联质谱(LC-MS/MS)法,用于N-糖肽与O-糖肽的分离,示于图1。当糖肽样品进入石

墨化碳柱时,受疏水相互作用、静电相互作用、诱导偶极作用等协同作用,N-糖肽与O-糖肽因产生了不同的色谱保留行为而分离。研究者基于牛胎球蛋白,验证了该材料分离 $\alpha 2,3/\alpha 2,6$ 唾液酸型糖肽的能力优于多孔石墨化碳和HILIC材料。该方法在跨批次、跨设备、跨实验室的验证中展现出高重现性特点。

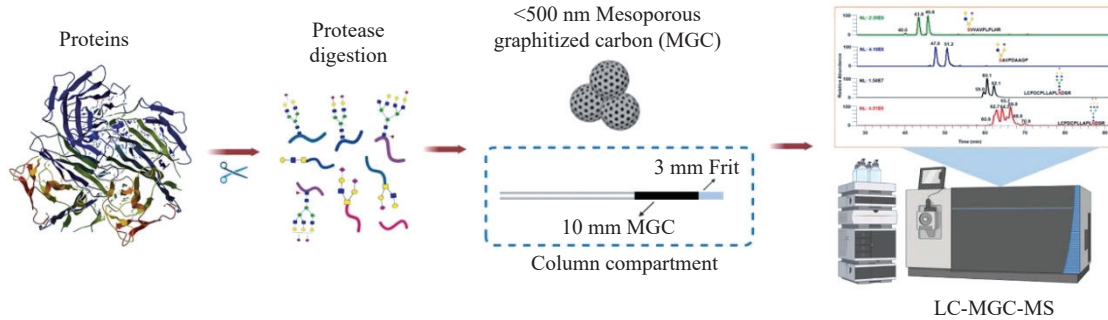


图1 介孔石墨化碳液相色谱柱用于N/O-糖肽的分离^[20]

Fig. 1 Mesoporous graphitized carbon liquid chromatography columns for separation of N/O-glycopeptides^[20]

1.2 基于O-糖蛋白酶的分析方法

随着OgpA、IMPα等O-糖蛋白酶的发现及商品化,许多基于O-糖蛋白酶的分析方法被开发。OpeRATOR是OgpA酶的商品化产品,其酶切位点在O-糖连接的丝氨酸或苏氨酸的N端,但其对唾液酸修饰的O-糖肽活性较低。免疫调节蛋白酶IMPα具有广泛的特异性,可消化大多数O-糖基化位点丝氨酸或苏氨酸的N端。研究者利用O-糖蛋白酶对O-糖基化位点的特异性识别,对其所在位点附近的肽链进行酶切,从而简化O-糖肽结构,便于质谱检测及数据库检索。Liu等^[21]开发了一种O-糖肽截短方法,用于生物样本中蛋白质O-半乳糖酰化的检测。该方法首先对蛋白质样品进行胰酶消化生成肽段, PNGase F酶处理去除N-聚糖, HILIC富集酶解液中的O-糖肽,再使用Glu-C酶或唾液酸酶与OpeRATOR的O-糖蛋白酶组合,以截短O-糖肽。其中,利用OpeRATOR方案将O-糖肽的鉴别量提升了50%。在大规模分析中,2种方案从75 μL人血清中共鉴定出对应于391种糖蛋白的2 000条O-糖肽。Kang等^[22]设计了一种揭示未知O-糖基化的流程,将蛋白质样品去N糖后,采用IMPα处理后使用蛋白酶消化,然后通过阴离子交换柱富集O-糖肽。该团队利用该方法检测并验证了胎球蛋白-A、纤维蛋白原、载脂蛋白

E等标准蛋白及人血浆蛋白中的未知O-糖基化。此外, Finn等^[23]报道了一种“一锅法”快速便捷获取O-糖肽的通用型方法GlycoFASP,首先利用分子截留滤器离心除去样本中的电解质、脂质、蛋白质碎片等干扰物,再使用OpeRATOR及IMPα等O-糖蛋白酶解截留下来的蛋白质样品,离心获得O-糖肽。通过对人血清样本及精浆样本的验证发现,10 ku分子截留滤器对糖基化密度高的黏蛋白序列覆盖率较高,且离心得到的O-糖肽除盐后可直接进行质谱检测;而30 ku分子截留滤器无论对高度糖基化还是糖基化密度相对较低的序列均有较好的覆盖率,但通过30 ku滤器的O-糖肽需再使用胰蛋白酶进一步酶解。

Tn(GalNAcα1-O-S/T)和sTn(Neu5Acα2-6GalNAcα1-O-S/T)抗原在胰腺癌、肠癌、乳腺癌样本中高表达^[24-26],与癌症侵袭性、预后不良、转移和免疫抑制密切相关^[27-29]。Tn/sTn抗原通常高密度地存在于富含S/T残基但缺乏R/K残基的蛋白质区域,胰蛋白酶对该类O-糖蛋白酶解时会产生带有多种O-聚糖的较大糖肽,这对Tn/sTn糖基化位点的检测提出了挑战。Wang等^[30]提出了一种用于复杂蛋白质样品中Tn/sTn糖基化位点的无标检测方法,该方法将胰酶消化物用PNGase F酶处理去除N-聚糖,通过自动化HILIC平台富集其中的O-糖肽,再用OpeRATOR/

IMPα处理得到被截短的O-糖肽,然后根据氧鎓离子的信号强度比值判定Tn/sTn糖基化位点。该方法在HeLa细胞和血清样本中得到了验证,并且在胰腺癌细胞系和血清样本中检测到来自于136个糖蛋白的191个Tn/sTn糖基化位点。

固相化学酶法是将样品固定化,然后利用化学法对其衍生化,通过酶的特异性识别将聚糖/糖肽酶解,实现对复杂样品中聚糖/糖肽的高纯度富集,简化了后续质谱分析的难度。AminoLink树脂是一种商品化的醛基树脂材料,表面的醛基可与蛋白质或肽段的氨基发生还原胺化反应,从而使其固定化。Yang等^[31]报道了一种EXoO-Tn方法用于大规模Tn糖基化蛋白和糖基化位点的鉴定,将样品蛋白酶酶解后用凝集素VVA富集,利用AminoLink树脂固定肽段,通过糖基转移酶C1GalT1将带有同位素标记的半乳糖连接到Tn上,使Tn转化为Gal(¹³C₆)-Tn,再用OpeRATOR释放Gal(¹³C₆)-Tn糖肽。该团队利用该方法在人T细胞中鉴定到来自480个糖蛋白的947个

Tn糖基化位点,并进一步应用于胰腺癌血清中Tn糖基化的研究。Suttapitugsakul等^[32]提出利用IMPα可以无需同位素标记,直接酶切Tn型糖肽,由此改进了EXoO-Tn方法,将蛋白质样品酶解,固化凝集素VVA富集肽段后,用IMPα酶切释放Tn型糖肽。此外,该团队提出将凝集素VVA富集替换为RAX强阴离子交换亲水富集,可实现全O-糖基化的分析。该策略在小鼠大脑样品中的应用揭示了其中多种唾液酸化O-聚糖的高频出现以及Tn抗原的存在。Yue等^[33]利用OpeRATOR对Tn和唾液酸化型O-糖位点酶切活性低的特点,开发了一种可将Tn/sTn型糖肽与其他O-糖肽区分的方法MOTAI,示于图2。该方法首先将蛋白酶解物固定到AminoLink树脂表面,用PNGase F酶处理去除N-聚糖,再依次用OpeRATOR和IMPα酶解,分别释放复杂O-聚糖型糖肽和Tn/sTn型糖肽。该方法在结肠癌组织中鉴定到32个Tn/sTn糖蛋白和43个T/sT糖蛋白上调。

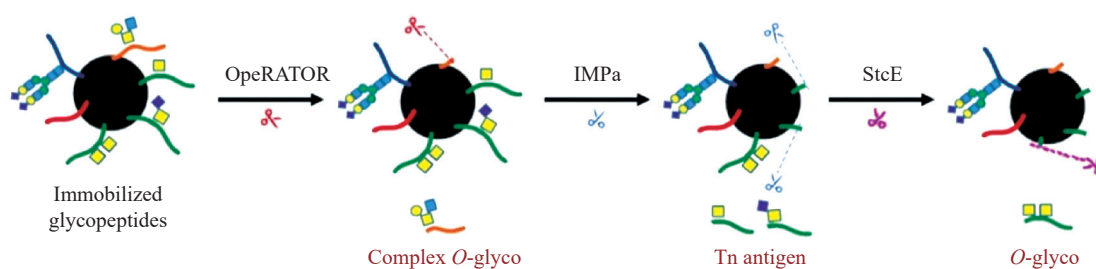


图2 MOTAI策略用于富集多种O-GalNAc型完整糖肽^[33]

Fig. 2 MOTAI strategy allows for enrichment of multiple intact O-GalNAc glycopeptides^[33]

1.3 生物正交标记富集法

生物正交标记法是一类在活细胞内进行、不干扰细胞原本生物化学过程的方法。该方法首先将带有叠氮基或炔基的单糖在细胞内酶促反应的作用下合成到目标聚糖上,实现对目标聚糖的标记,然后通过点击化学反应与生物素标签结合或者直接与固相结合,带有生物素标签的目标聚糖可被链霉素固相捕获。*N*-叠氮基乙酰半乳糖胺(GalNAz)是*N*-乙酰半乳糖氨(GalNAc)的叠氮类似物,可以通过GalNAc补救通路在细胞中高效代谢转化为UDP-GalNAz。Liu等^[34]基于生物正交化学开发了一种富集完整*N/O*-糖肽的方法Click-iG。GalNAz的1,6-二丙酰化利于细胞摄取。该团队利用1,6-二丙酰化的GalNAz标记

细胞内的O-GalNAc型聚糖,并通过点击化学对其添加生物素标签,将样品酶解后,利用链霉素固相捕获O-GalNAc型糖肽,照射365 nm紫外光可使标签断裂,从而释放糖肽。利用该方法在小鼠的肺、心和脾中共鉴定到262个O-GalNAc糖基化位点。Tian等^[35]开发了一种不依赖生物素-链霉素亲和作用的一步标记策略,用于多类型糖基化的鉴定,示于图3。该策略通过2种底物特异性唾液酸转移酶,使*N*-糖肽和O-糖肽被温度敏感性型探针标记。该探针在40℃时沉淀,低温时溶解。因此,40℃时使糖肽沉淀,去除未标记的非糖肽,低温时使糖肽溶解;再依次使用Endo F3和PNGase F处理,40℃时沉淀,分别释放核心岩藻糖修饰的*N*-糖肽与非核心岩藻

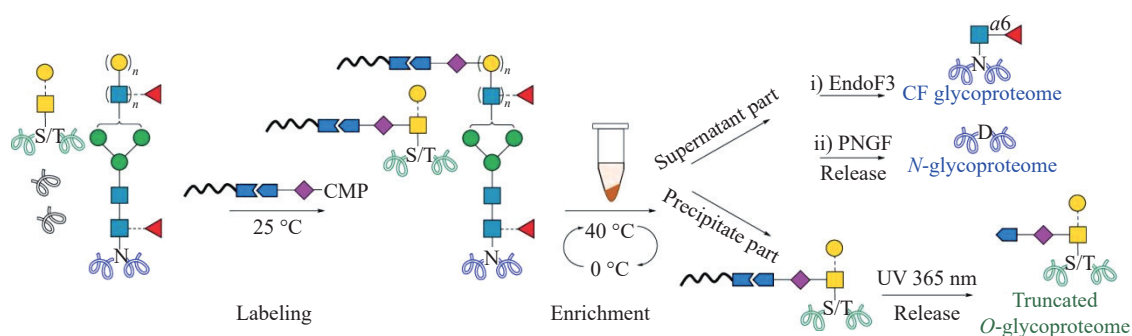


图3 一步标记策略用于分析多种类型的蛋白质糖基化^[35]

Fig. 3 One-step labeling strategy for profiling of multiple types of protein glycosylation^[35]

糖修饰的 *N*-糖肽, 低温时剩余的 *O*-糖肽溶解; 将溶液照射 365 nm 紫外光, 破坏探针, 释放 *O*-糖肽。相比于生物素-链霉素亲和方法, 该策略更经济、高效。

2 数据库检索方法

完整糖肽的富集和质谱技术的突破产生了大量高质量的数据集。然而, 由于黏蛋白型糖基化具有缺少固定的基序、高度微观不均一性等特点, 诠释大规模的 *O*-GalNAc 型糖肽质谱数据仍面临挑战。传统的检索方法是在丝氨酸和苏氨酸残基上每种可能的 *O*-糖链设置为“可变修饰”, 该方法对于均质、单一的翻译后修饰(如磷酸化、乙酰化)非常有效, 但对于高度异质、多位点的 *O*-糖基化而言, 由于搜索空间巨大、碎片模型不匹配, 导致其鉴定效率低下, 难以应用于复杂的 *O*-糖蛋白质组学研究。目前, 开展临床相关 *O*-糖基化研究亟需快速、准确的新型检索引擎。

本文总结了近年来用于 *O*-糖基化分析的检索引擎, 列于表 2。Mao 等^[36]将糖基化视为肽段

水平的质量标签, 提出一种用于鉴定完整 *O*-糖肽的检索策略 *O*-Search, 利用氧鎓离子筛选出糖肽谱图, 再在母离子质量中减去可能的 *O*-糖质量并生成理论去糖基化谱图, 据此检索肽段序列。该策略大幅缩小了 *O*-糖肽的检索空间, 与传统方法相比, 显著提升了糖肽的鉴定数量。MS-Decipher 是一款支持传统肽段搜索并集成的 *O*-Search 模块数据库检索软件, 支持 rank score、hyperscore 评分算法以及 E-value、Percolator 验证方法, 拥有用户友好的图形界面且可跨平台使用^[37]。Zhu 等^[14]在 *O*-Search 基础上引入 Y 离子模式匹配, 开发了 *O*-Search-Pattern, 进一步缩小了 *O*-糖肽的搜索空间, 示于图 4。*O*-Search-Pattern 根据已知的 *O*-糖链结构, 理论推导出 Y0、Y1、Y2、Y3 等离子的可能组合及其质量差, 构建理论 Y 离子。利用氧鎓离子筛选出糖肽谱图, 以理论 Y 离子与实验谱图进行匹配。如果理论 Y 离子模式与实验谱图中的峰匹配, 则可获得肽的质量, 并且可通过比较 *O*-糖质量以及前体离子和 Y0 离子之间的质量差来确定

表 2 *O*-糖肽数据库检索方法对比

Table 2 Comparison of database search methods for *O*-glycopeptide

检索方法 Search method	检索策略 Search strategy	碎裂模式 Fragmentation pattern	适配软件 Compatible software
Bionic	肽段优先	CID/HCD/ETD	Bionic, Protein Metrics, Proteome Discoverer
<i>O</i> -Search	糖链移除	HCD	MS-Decipher
<i>O</i> -Search-Pattern	Y离子模式匹配+糖链移除	sceHCD	MS-Decipher
MSFragger-Glyco	肽段优先+离子索引	HCD/AI-ETD/ETcD	MSFragger, FragPipe
pGlyco3	糖链优先+糖链离子索引	HCD/ETxxD	pGlyco3
<i>O</i> -Pair Search	离子索引开放搜索+图论位点定位	HCD+ETxxD	MetaMorpheus, FragPipe

注: CID. 碰撞诱导解离; HCD. 高能碰撞解离; ETD. 电子转移解离; ETcD. 电子转移/碰撞诱导解离; sceHCD. 阶梯式高能碰撞解离; AI-ETD. 活化离子电子转移解离; ETxxD. ETD/ETcD/ETciD

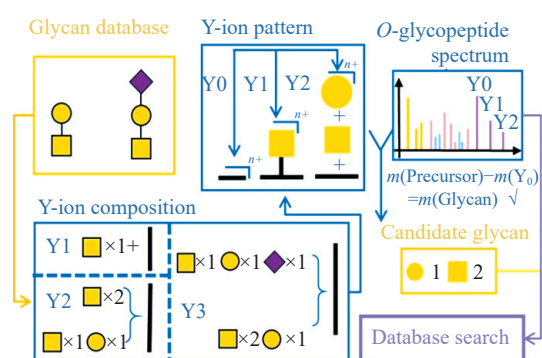


图4 O-Search-Pattern 检索策略用于完整 O-GalNAc 糖肽的谱图解析^[14]

Fig. 4 O-Search-Pattern searching strategy for the interpretation of intact O-GalNAc glycopeptide spectra^[14]

候选 O-糖组成, 缩小了聚糖检索的范围。比较 O-糖质量以及前体和匹配的 Y0 离子之间的质量差来确定候选 O-糖组成, 减小了聚糖数据库的大小。为避免遗漏, 将未匹配成功的谱图通过 O-Search 方法进行检索。相比于其他软件工具, O-Search-Pattern 在对人血清数据库检索中鉴定到的糖肽谱图匹配数提升了 15.4%~199.0%, O-糖肽鉴定数提升了 19.6%~107.1%。此外, Polasky 等^[38]基于开放检索与质量偏移策略, 在 MSFragger 软件中开发了一种用于糖蛋白质组学的模块 MSFragger-Glyco。该模块利用碎片离子索引检索其肽段序列, 报告所有的质量偏移, 根据用户提供的聚糖质量列表与糖肽特异性碎片离子(Y 离子和 b/y+HexNAc 离子)检索糖链信息, 该模块对 O-糖肽谱图匹配数相比传统的“可变修饰”策略提升了 4~6 倍; 同时, 结合碎片离子索引技术可将分析时间缩短至数分钟。Zeng 等^[39]报道了基于聚糖优先策略开发的糖肽检索引擎 pGlyco3, 用于分析完整的 N/O-糖肽。pGlyco3 首先利用聚糖离子索引快速筛选聚糖信号, 并对其匹配的候选聚糖进行评分, 然后根据聚糖质量推导肽段质量, 在蛋白质数据库中检索进而获得肽段序列, 最后利用 pGlycoSite 算法定位糖基化位点。pGlyco3 基于聚糖优先策略缩小了糖肽的检索空间, 实现了对完整糖肽的快速、精准分析。

3 总结与展望

近年来, 黏蛋白型糖基化蛋白质组学分析技术在富集方法及谱图解析方面取得了显著进

展。HILIC 与其他亲和富集方法相结合的策略被广泛用于富集 O-糖肽, 以提高对 O-糖肽的富集效率。自动化液相系统的引入为大规模临床样本中 O-糖基化的解析提供了高效、稳定的技术平台。基于 O-糖蛋白酶的分析方法和生物正交化学的发展, 实现了高特异性富集和精确位点定位。此外, 聚丝氨酸精氨酸材料特异性富集 N/O-糖肽^[40]、酸性低温条件下纳米孔富集 O-GalNAc 糖肽^[41]等策略为新型 O-糖肽富集方法的开发提供了思路。在 O-糖基化数据库检索方面, 以 O-Search-Pattern、MSFragger-Glyco、pGlyco3 为代表的检索策略突破了传统“在氨基酸残基上设置可变修饰”的思维定式, 极大缩小了 O-糖肽的检索空间, 提升了鉴定的灵敏度和准确性。此外, 机器学习已成功应用于临床标志物的发现^[42-43]、糖基化位点预测^[44]及非标定量^[45]等领域。生物标志物分类算法 MarkerPredict 整合了内在无序蛋白及其信号网络拓扑特征, 用于癌症相关生物标志物的发现^[46]。残差神经网络模型 CandyCrunch 可以在几秒钟内从 LC-MS/MS raw 文件中预测聚糖结构, 准确率达 90.3%^[47]。深度学习模型 DeepGlyco 通过树状 LSTM 与图神经网络有效建模糖基结构, 实现了完整 N-糖肽的高精度预测及异构体区分^[48]。非标定量软件 GlyPep-Quant 借助随机森林模型及无监督聚类算法 DBSCAN 显著提高了可量化的糖肽数量^[49]。然而, 现阶段机器学习在糖基化分析领域的应用仍以 N-糖基化研究为核心。依托机器学习从复杂多维数据中挖掘深层内在规律的独特优势, 未来有望构建 O-糖基化分析模型, 解决 O-糖基化解析中低丰度糖肽的灵敏鉴定、多位点糖型的精准解析等问题。黏蛋白型糖基化蛋白质组学分析方法的不断革新, 深化了对其生物学功能的解析, 为基础研究发现向临床诊疗策略的转化提供了动力。

参考文献:

- [1] SCHJOLDAGER K T, NARIMATSU Y, JOSHI H J, CLAUSEN H. Global view of human protein glycosylation pathways and functions[J]. *Nature Reviews Molecular Cell Biology*, 2020, 21(12): 729-749.
- [2] WAGNER C E, WHEELER K M, RIBBECK K. Mucins and their role in shaping the functions of mucus barriers[J]. *Annual Review of Cell and Developmental*

- Biology*, 2018, 34: 189-215.
- [3] SHON D J, MALAKER S A, PEDRAM K, YANG E, KRISHNAN V, DORIGO O, BERTOZZI C R. An enzymatic toolkit for selective proteolysis, detection, and visualization of mucin-domain glycoproteins[J]. *PNAS*, 2020, 117(35): 21 299-21 307.
- [4] REILY C, STEWART T J, RENFROW M B, NOVAK J. Glycosylation in health and disease[J]. *Nature Reviews Nephrology*, 2019, 15(6): 346-366.
- [5] HUSSAIN M R M, HOESSLI D C, FANG M. *N*-Acetylgalactosaminyltransferases in cancer[J]. *Oncotarget*, 2016, 7(33): 54 067-54 081.
- [6] ARAI J, HAYAKAWA Y, TATENO H, MURAKAMI K, HAYASHI T, HATA M, MATSUSHITA Y, KINOSHITA H, ABE S, KUOKAWA K, OYA Y, TSUBOI M, IHARA S, NIIKURA R, SUZUKI N, IWATA Y, SHIOKAWA T, SHIOMI C, UEKURA C, YAMAMOTO K, FUJIWARA H, KAWAMURA S, NAKAGAWA H, MIZUNO S, KUDO T, TAKAHASHI S, USHIKU T, HIRATA Y, FUJII C, NAKAYAMA J, SHIBATA S, WOODS S, WORTHLEY D L, HATAKEYAMA M, WANG T C, FUJISHIRO M. Impaired glycosylation of gastric mucins drives gastric tumorigenesis and serves as a novel therapeutic target[J]. *Gastroenterology*, 2024, 167(3): 505-521. e19.
- [7] BÜLL C, NASON R, SUN L, van COILLIE J, MADRIZ SØRENSEN D, MOONS S J, YANG Z, ARBITMAN S, FERNANDES S M, FURUKAWA S, McBRIDE R, NYCHOLAT C M, ADEMA G J, PAULSON J C, SCHNAAR R L, BOLTJE T J, CLAUSEN H, NARI-MATSU Y. Probing the binding specificities of human Siglecs by cell-based glycan arrays[J]. *PNAS*, 2021, 118(17): e2026102118.
- [8] ZHANG J, HUANG Y, LI H, XU P, LIU Q, SUN Y, ZHANG Z, WU T, TANG Q, JIA Q, XIA Y, XU Y, JING X, LI J, MO L, XIE W, QU A, HE J, LI Y. B3galt5 functions as a PXR target gene and regulates obesity and insulin resistance by maintaining intestinal integrity[J]. *Nature Communications*, 2024, 15: 5 919.
- [9] BERGSTROM K, SHAN X, CASERO D, BATUSHANSKY A, LAGISHETTY V, JACOBS J P, HOOVER C, KONDO Y, SHAO B, GAO L, ZANDBERG W, NOY-OVITZ B, McDANIEL J M, GIBSON D L, PAKPOUR S, KAZEMIAN N, McGEE S, HOUCHE C W, RAO C V, GRIFFIN T M, SONNENBURG J L, McEVER R P, BRAUN J, XIA L. Proximal colon-derived *O*-glycosylated mucus encapsulates and modulates the microbiota[J]. *Science*, 2020, 370(6 515): 467-472.
- [10] SINGH A, BEAUPRE M, VILLEGAS-NOVOA C, SHIOMITSU K, GAUDINO S J, TAWCH S, DAMLE R, KEMPEN C, CHOUDHURY B, McALEER J P, SHERIDAN B S, DENOYA P, BLUMBERG R S, HEARING P, ALLBRITTON N L, KUMAR P. IL-22 promotes mucin-type *O*-glycosylation and MATH1+ cell-mediated amelioration of intestinal inflammation[J]. *Cell Reports*, 2024, 43(5): 114 206.
- [11] XIONG Y, LU Z, SHAO Y, MENG P, WANG G, ZHOU X, YAO J, BAO H, LU H. Rapid and large-scale glycopeptide enrichment strategy based on chemical ligation[J]. *National Science Review*, 2024, 11(11): nwae341.
- [12] MAHONEY K E, CHANG V, LUCAS T M, MARUSZKO K, MALAKER S A. Mass spectrometry-compatible elution technique enables an improved mucin-selective enrichment strategy to probe the mucinome[J]. *Analytical Chemistry*, 2024, 96(13): 5 242-5 250.
- [13] HU Z, GAO W, LIU R, YANG J, HAN R, LI J, YU J, MA D, TANG K. An efficient strategy with a synergistic effect of hydrophilic and electrostatic interactions for simultaneous enrichment of *N*- and *O*-glycopeptides[J]. *Analyt*, 2024, 149(4): 1 090-1 101.
- [14] ZHU H, LIU L, FANG Z, QIN H, DONG M, YE M. O-Search-pattern: a searching tool utilizing the Y-ion pattern to enhance *O*-glycopeptide identification for the analysis of *O*-GalNAc glycoproteome[J]. *Analytical Chemistry*, 2023, 95(26): 9 761-9 768.
- [15] LU L, RILEY N M, SHORTREED M R, BERTOZZI C R, SMITH L M. O-Pair search with MetaMorpheus for *O*-glycopeptide characterization[J]. *Nature Methods*, 2020, 17(11): 1 133-1 138.
- [16] YUE X, QIN H, CHEN Y, FANG Z, LIU L, ZHU H, LIU X, ZHOU J, TIAN K, QIAO X, YE M. Highly efficient enrichment of *O*-GalNAc glycopeptides by using immobilized metal ion affinity chromatography[J]. *Analytical Chemistry*, 2021, 93(21): 7 579-7 587.
- [17] KONG S, ZHANG Q, YANG L, HUANG Y, LIU M, YAN G, ZHAO H, WU M, ZHANG X, YANG P, CAO W. Effective enrichment strategy using boronic acid-functionalized mesoporous graphene-silica composites for intact *N*- and *O*-linked glycopeptide analysis in human serum[J]. *Analytical Chemistry*, 2021, 93(17): 6 682-6 691.
- [18] CUI Y, LI S, WAN H, DONG X, CHEN C, YUE G, ZHAO Y, LI X, LIANG X. Superhydrophilic multichannel TiO₂ fibers for simultaneous enrichment of *O*-

- GalNAc glycopeptides and phosphopeptides[J]. *Journal of Proteome Research*, 2025, 24(7): 3 188-3 201.
- [19] WANG Z, FANG Z, LIU L, ZHU H, WANG Y, ZHAO C, GUO Z, QIN H, NIE Y, LIANG X, DONG M, YE M. Development of an integrated platform for the simultaneous enrichment and characterization of *N*- and *O*-linked intact glycopeptides[J]. *Analytical Chemistry*, 2023, 95(19): 7 448-7 457.
- [20] DARAMOLA O, GAUTAM S, GUTIERREZ REYES C D, FOWOWE M, ONIGBINDE S, NWAIWU J, MECHREF Y. LC-MS/MS of isomeric *N*- and *O*-glycopeptides on mesoporous graphitized carbon column[J]. *Analytica Chimica Acta*, 2024, 1 317: 342 907.
- [21] LIU L, ZHU H, LIU L, YOU X, MAO J, WANG Y, LIU X, QIN H, DONG M, YE M. *O*-Glycopeptide truncation strategy for heterogeneous *O*-GalNAc glycoproteomics characterization[J]. *Analytical Chemistry*, 2023, 95(26): 10 017-10 024.
- [22] KANG T, BUDHRAJA R, KIM J, JOSHI N, GARAPATI K, PANDEY A. Global *O*-glycoproteome enrichment and analysis enabled by a combinatorial enzymatic workflow[J]. *Cell Reports Methods*, 2024, 4(4): 100 744.
- [23] FINN S M, MAHONEY K E, LUCAS T M, RANGEL-ANGARITA V, CHEN R J, MALAKER S A. GlycoFASP: a universal method to prepare complex mixtures for *O*-glycoproteomic analysis[J]. *Analytical Chemistry*, 2025, 97(43): 23 751-23 756.
- [24] GUILLEN-POZA P A, SÁNCHEZ-FERNÁNDEZ E M, ARTIGAS G, GARCÍA FERNÁNDEZ J M, HINOUE H, ORTIZ MELLET C, NISHIMURA S I, GARCIA-MARTIN F. Amplified detection of breast cancer autoantibodies using MUC1-based tn antigen mimics[J]. *Journal of Medicinal Chemistry*, 2020, 63(15): 8 524-8 533.
- [25] LIU F, FU J, BERGSTROM K, SHAN X, McDANIEL J M, McGEE S, BAI X, CHEN W, XIA L. Core 1-derived mucin-type *O*-glycosylation protects against spontaneous gastritis and gastric cancer[J]. *The Journal of Experimental Medicine*, 2020, 217(1): e20182325.
- [26] KUDELKA M R, JU T, HEIMBURG-MOLINARO J, CUMMINGS R D. Simple sugars to complex disease—mucin-type *O*-glycans in cancer[J]. *Advances in Cancer Research*, 2015, 126: 53-135.
- [27] NIVEAU C, CETTOUR-CAVE M, MOURET S, SOSA CUEVAS E, PEZET M, ROUBINET B, GIL H, de FRAIPONT F, LANDEMARRE L, CHARLES J, SAAS P, ASPORD C. MCT1 lactate transporter blockade reinvigorates anti-tumor immunity through metabolic rewiring of dendritic cells in melanoma[J]. *Nature Communications*, 2025, 16: 1 083.
- [28] PARK S, HO C E, DARNELL E P, WOLFF A N, TAKEI H, BIROCCHI F, BOUFFARD A A, SALAS-BENITO D, ESCOBAR G, LEICK M B, MUCCI A, BERGER T R, MAUS M V. Tuning CAR-T cells by targeting cancer-associated glycan in pancreatic cancer[J]. *Nature Communications*, 2025, 16: 11 246.
- [29] Da COSTA V, van VLIET S J, CARASI P, FRIGERIO S, GARCÍA P A, CROCI D O, FESTARI M F, COSTA M, LANDEIRA M, RODRÍGUEZ-ZRAQUIA S A, CAGNONI A J, CUTINE A M, RABINOVICH G A, OSINAGA E, MARIÑO K V, FREIRE T. The Tn antigen promotes lung tumor growth by fostering immunosuppression and angiogenesis *via* interaction with Macrophage Galactose-type lectin 2 (MGL2)[J]. *Cancer Letters*, 2021, 518: 72-81.
- [30] WANG Z, WANG Y, MA T, ZHU H, SUN D, GUO Z, LIANG X, YE M, DONG M. A label-free method combining chromatography enrichment with *N*-terminal specific proteolysis enables sensitive identification of Tn/sTn glycosites[J]. *Analytical Chemistry*, 2025, 97(28): 15 038-15 048.
- [31] YANG W, AO M, SONG A, XU Y, SOKOLL L, ZHANG H. Mass spectrometric mapping of glycoproteins modified by Tn-antigen using solid-phase capture and enzymatic release[J]. *Analytical Chemistry*, 2020, 92(13): 9 230-9 238.
- [32] SUTTAPITUGSAKUL S, MATSUMOTO Y, ARYAL R P, CUMMINGS R D. Large-scale and site-specific mapping of the murine brain *O*-glycoproteome with IMPa[J]. *Analytical Chemistry*, 2023, 95(36): 13 423-13 430.
- [33] YUE S, WANG X, WANG L, LI J, ZHOU Y, CHEN Y, ZHOU Z, YANG X, SHI X, GAO S, WEN Z, ZHU X, WANG Y, YANG S. MOTAI: a novel method for the study of *O*-GalNAcylation and complex *O*-glycosylation in cancer[J]. *Analytical Chemistry*, 2024, 96(28): 11 137-11 145.
- [34] LIU J, CHENG B, FAN X, ZHOU X, WANG J, ZHOU W, LI H, ZENG W, YANG P, CHEN X. Click-iG: simultaneous enrichment and profiling of intact *N*-linked, *O*-GalNAc, and *O*-GlcNAcylated glycopeptides[J]. *Angewandte Chemie (International Ed in English)*, 2023, 62(36): e202303410.
- [35] TIAN Y, WANG Y, ZHANG Y, GUO J, ZHANG P, LI X, ZHOU H, WEN L. One-step labeling strategy for the profiling of multiple types of protein glycosylation[J].

- Analytical Chemistry*, 2025, 97(14): 7 833-7 841.
- [36] MAO J, YOU X, QIN H, WANG C, WANG L, YE M. A new searching strategy for the identification of *O*-linked glycopeptides[J]. *Analytical Chemistry*, 2019, 91(6): 3 852-3 859.
- [37] MAO J, ZHU H, LIU L, FANG Z, DONG M, QIN H, YE M. MS-Decipher: a user-friendly proteome database search software with an emphasis on deciphering the spectra of *O*-linked glycopeptides[J]. *Bioinformatics*, 2022, 38(7): 1 911-1 919.
- [38] POLASKY D A, YU F, TEO G C, NESVIZHSHKII A I. Fast and comprehensive *N*- and *O*-glycoproteomics analysis with MSFragger-glyco[J]. *Nature Methods*, 2020, 17(11): 1 125-1 132.
- [39] ZENG W, CAO W, LIU M, HE S, YANG P. Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3[J]. *Nature Methods*, 2021, 18(12): 1 515-1 523.
- [40] DU G, CHEN C, CUI Y, HUANG B, ZHENG Y, YAN J, SHANG D, DONG X, JIN M, LI X. Glycosylation profiling of monkeypox virus structural proteins with poly Ser-Arg materials[J]. *Analyst*, 2025, 150(4): 680-691.
- [41] WANG J, MA W, HU Z, GAO Z, LONG Y, LI T, YING Y. Direct identification of *O*-glycopeptides by low-temperature assisted nanopore technique[J]. *Research*, 2025, 8: 0850.
- [42] WANG Z, GU Y, HUANG L, LIU S, CHEN Q, YANG Y, HONG G, NING W. Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data[J]. *Cardiovascular Diabetology*, 2024, 23(1): 351.
- [43] WANG S, WU C, HE M, YONG J, CHEN Y, QIAN L, ZHANG J, ZENG Z, XU R, WANG F, ZHAO Q. Machine learning-based extrachromosomal DNA identification in large-scale cohorts reveals its clinical implications in cancer[J]. *Nature Communications*, 2024, 15: 1 515.
- [44] WEN B, WANG C, LI K, HAN P, HOLT M V, SAVAGE S R, LEI J T, DOU Y, SHI Z, LI Y, ZHANG B. DeepMVP: deep learning models trained on high-quality data accurately predict PTM sites and variant-induced alterations[J]. *Nature Methods*, 2025, 22(9): 1 857-1 867.
- [45] KONG S, GONG P, ZENG W F, JIANG B, HOU X, ZHANG Y, ZHAO H, LIU M, YAN G, ZHOU X, QIAO X, WU M, YANG P, LIU C, CAO W. pGlycoQuant with a deep residual network for quantitative glycoproteomics at intact glycopeptide level[J]. *Nature Communications*, 2022, 13: 7 539.
- [46] VERES D V, CSERMELY P, SCHULC K. MarkerPredict: predicting clinically relevant predictive biomarkers with machine learning[J]. *npj Systems Biology and Applications*, 2025, 11: 132.
- [47] URBAN J, JIN C, THOMSSON K A, KARLSSON N G, IVES C M, FADDA E, BOJAR D. Predicting glycan structure from tandem mass spectrometry *via* deep learning[J]. *Nature Methods*, 2024, 21(7): 1 206-1 215.
- [48] YANG Y, FANG Q. Prediction of glycopeptide fragment mass spectra by deep learning[J]. *Nature Communications*, 2024, 15: 2 448.
- [49] ZHU H, FANG Z, LIU L, WANG Y, QIN H, NIE Y, DONG M, YE M. Library-based virtual match-between-runs quantification in GlyPep-Quant improves site-specific glycan identification[J]. *Nature Communications*, 2025, 16: 6 483.

(收稿日期: 2025-11-21; 修回日期: 2025-12-26)