基于生物质谱的蛋白质O-GalNAc糖基化修饰研究进展

赵新元¹, 沈丙权², 秦伟捷^{2*}, 钱小红^{2*}

1. 北京工业大学生命科学与生物工程学院, 北京 100022;

2. 军事科学院军事医学研究院生命组学研究所, 蛋白质组学国家重点实验室, 北京蛋白质组研究中心, 国家蛋白质科学中心(北京), 北京 102206

* 联系人, E-mail: aunp_dna@126.com; qianxh1@163.com

收稿日期: 2017-06-20; 接受日期: 2017-07-21; 网络版发表日期: 2017-11-29

国家重点研发计划(批准号: 2013CB911204, 2016YFA0501403)和国家自然科学基金(批准号: 21235001, 21405175, 21675172)资助

摘要 糖基化是生物体内蛋白质最常见、最重要的翻译后修饰之一, 普遍存在于细胞膜蛋白及分泌蛋白, 执行重要的生物学功能. 常见的蛋白糖基化修饰有N-糖基化及O-糖基化两种类型, 而O-GalNAc是O-糖基化中重要的存在形式, 在特定生物学进程、癌症发生发展中起着重要作用, 近年来受到广泛重视. 得益于代谢标记、化学衍生化、高分辨和多种碎裂模式质谱技术以及基因编辑技术的高速发展, O-GalNAc的糖基化位点、糖型鉴定和生理功能研究取得一系列重大进展. 本文综述了基于生物质谱技术的蛋白质O-GalNAc糖基化修饰研究进展.

关键词 蛋白质, 糖基化, 质谱, 基因编辑

核酸、蛋白质、脂质与糖类是组成细胞及生物体最基本的化学成分, 而蛋白质糖基化是连接蛋白质和糖复合物这两类重要生物大分子的桥梁. 在已知的蛋白质中, 约有1/2以上存在糖基化修饰^[1]. 糖基化修饰使得糖与蛋白质两种大分子共同协调作用, 调控着糖蛋白在生理病理条件下的生物学功能^[2,3]. 在临床研究中, 多种用于疾病诊断的生物标志物(如AFP, PSA, CA125)都是糖蛋白^[4]. 而在药物研发过程中, 糖基化修饰对蛋白质药物结构和药效的影响也是不可忽略的因素^[5]. 蛋白质的N-糖基化修饰主要发生在特定的氨基酸基序“天冬酰胺-X-丝氨酸/苏氨酸”(X指除脯氨酸之外的氨基酸), 极少量发生在基序“天冬酰胺-X-半胱氨酸”上^[6]. N糖基化以五糖核心为糖链基础, 有3种常见的糖

型: 高甘露糖型、复合型、杂合型. 发生在丝氨酸与苏氨酸上的O-糖基化常指O-GlcNAc及O-GalNAc类糖基化. 其中, O-GlcNAc常发生在细胞核与细胞质中, 是一种重要的细胞内信号调控方式, 常与磷酸化位点重叠, 从而与磷酸化形成cross-talk^[7]. O-GalNAc糖基化修饰最初主要发现存在于mucin类黏蛋白中, 其与蛋白质中丝氨酸或苏氨酸残基直接连接的基本化学结构单元为N-乙酰半乳糖胺, 亦常被成为黏蛋白型O-糖基化. 不同于N-糖基化发生在N-X-S/T/C特定氨基酸基序, O-糖基化发生的受体蛋白序列并无规律化的基序. 在生物体内O-糖基化合成中, 多肽-N-乙酰氨基半乳糖转移酶(polypeptide N-acetyl- α -galactosaminyltransferases, ppGalNAc-Ts)将UDP-GalNAc上的GalNAc单元转移到

引用格式: 赵新元, 沈丙权, 秦伟捷, 等. 基于生物质谱的蛋白质O-GalNAc糖基化修饰研究进展. 中国科学: 生命科学, 2018, 48: 113–123
Zhao X Y, Shen B Q, Qin W J, et al. Advances in O-GalNAc glycosylation research based on biological mass spectrometry (in Chinese). Sci Sin Vitae, 2018, 48: 113–123, doi: [10.1360/N052017-00137](https://doi.org/10.1360/N052017-00137)

丝氨酸(Ser)及苏氨酸(Thr)上^[8]。这一过程是所有黏蛋白O-糖合成的第一步。在人类基因组中有20多个ppGalNAc-Ts基因^[9], 主要定位于细胞内高尔基体中^[10]。ppGalNAc-Ts为二型膜蛋白, 其N端含有穿膜结构域及茎区, 中部含有两个糖基转移酶催化结构域(即GT结构域及Gal/GalNAc糖基转移酶结构域), 而在C端, 含有具有特定底物选择性的凝集素结构域。O-GalNAc糖基化修饰异常改变与多种疾病密切相关, 如Tn综合征^[11]、IgA肾病^[12]等。因此, 对O-GalNAc修饰的研究将有助于更好地了解糖基化的生理功能和在疾病发生发展过程中发挥的作用, 为最终攻克这类疾病提供帮助。

1 蛋白质O-GalNAc糖基化概述

当前, 蛋白质组学研究常用的策略是bottom-up模式, 即首先利用特异性蛋白酶如胰蛋白酶(trypsin)将蛋白质酶解成肽段, 继而利用质谱对肽段进行分析, 最后对质谱鉴定结果进行数据库检索来获得相应的蛋白或肽段信息。因此, 对糖蛋白的分析往往会归结到对糖肽的分析上。传统的糖肽解析中, 主要是通过特定的糖苷酶如PNGase F对N-糖链进行切除^[1], 或者利用 β 消除反应对O-糖链进行切除^[13], 并结合O¹⁸或DTT对糖基化位点进行标记, 进而进行质谱分析, 得到糖基化位点和肽段序列的信息。为了得到糖链的单糖组成及排列方式, 一方面, 可利用基质辅助激光解吸飞行时间质谱(matrix-assisted laser desorption ionization-time of flight-mass spectrometry, MALDI-TOF-MS)对切除下的糖链进行一级质量数鉴定并匹配特定糖型; 或采用多级质谱如MALDI-QIT-TOF-MSⁿ (基质辅助激光解吸-离子阱-飞行时间结合多级质谱)进行精确糖结构的解析^[14]。然而, 糖链与肽段的分别解析割裂了蛋白质与糖链之间的桥梁关系, 无法准确获得位点特异的糖基化信息, 因此, 2011年*Nature Methods*发文呼吁研究人员对完整糖肽的解析发起挑战^[15]。然而, 由于糖类化合物的高度复杂性, 完整糖肽的质谱解析仍面临诸多技术瓶颈^[16]。不同于基因组以及蛋白质组存在生物合成模板, 生物体内糖基化修饰的形成由一系列糖基转移酶调控。据统计, 生物体内共有约200种糖基转移酶^[17]。不同于其他翻译后修饰, 在糖基化修饰中, 组成糖链的糖单元种类多样, 连接方式各异。因此, 在同一修饰位点存在大量不同的糖链组成形式, 被称为糖

基化修饰的微不均一性, 对完整糖肽的鉴定技术提出了极高要求。然而对特定位点糖型的精确解析对于理解其生理功能和病理意义具有关键作用。

就O-GalNAc类型糖基化而言, 其研究主要面临四大技术难点^[18]: (i) 修饰位点无特定氨基酸模体序列; (ii) 糖链构成更为复杂, 缺乏固定的核心结构; (iii) 缺乏可特异性切除O-糖链的糖苷内切酶; (iv) O-GalNAc糖基化位点在蛋白/肽链上往往连续高密度存在, 准确解析极为困难。近年来, 迅速发展的基因编辑技术及以ECD, ETD为代表的质谱碎裂技术的出现, 为位点特异的O-GalNAc类型糖基化修饰研究提供了强大的技术支持。本文将围绕O-GalNAc类型糖基化在生物体合成的途径以及当前一系列先进解析手段进行详细的描述。

2 O-GalNAc糖基化修饰的发现

O-GalNAc类型糖基化是从最初的Tn抗原结构开始合成的。Tn抗原的发现是糖研究中具有里程碑意义的事件。Tn syndrome or Tn polyagglutination syndrome在1957年由Moreau等人^[19]首先发现。其在一个溶血性贫血病人的红细胞表面发现一种不寻常的抗原, 可以导致血液中红细胞的聚集。进一步研究发现, 这一现象产生的原因是T抗原(Gal β 1-3GalNAc α 1-O-Ser/Thr)在糖苷酶作用下产生的。Tn抗原也被命名为CD175, 其唾液酸化形式, sialyl Tn也被称为sTn或者CD175s。T抗原最早由Hübener和Thomsen在1925年左右研究血液凝集时发现, 并由Friedenreich在1930年证实T抗原的存在^[20], 其结构最早由Dahr等人^[21]在1975年通过气相色谱测定。T及Tn抗原的化学合成研究始于1977年^[22], 而对应的检测方法有化学检测法、凝集素检测法及抗体检测法^[23]。化学检测方法中最直接的方法是 β 消除反应, 反应机理是在NaOH或NH₄OH等碱性条件下, 丝氨酸或苏氨酸上的羰基碳带负电, 因此在 α 碳与此羰基碳之间易形成双键, 进而双键迁移至 α 碳与 β 碳之间, 使得连接在 β 碳上的氧原子及连接的糖链断裂, 糖链组分被切除, 同时可以利用肽段骨架结构上存在的双键对丝氨酸或苏氨酸进行修饰, 从而实现检测。大量从植物及动物中提取的GalNAc结合凝集素(如VVA-B4及HPA)亦被用于Tn抗原的检测。Tn抗原是大量O-GalNAc糖基化修饰的前体组成部分, 如core1,

core2, core3, core4型O糖(表1). 在之前文献报道的对20余种ppGalNAcTs的基因功能研究发现, 其在众多生理通路及疾病进程中发挥作用, 如对GALNT2进行基因敲除后发现, ApoC-III被证实为GALNT2的特定O-GalNAc修饰底物^[24]. 已有的研究也证实, GALNT2在异常血脂症中起着重要作用^[25]. 2013年, Khokha研究组^[26]证实GALNT11对Notch通路具有调控作用. 不同于以往认为O-GalNAc修饰仅发生在S或T上, 酪氨酸(Tyr)的O-GalNAc类型修饰近来在脑脊液中的淀粉

样前体蛋白上被发现^[27], 补充了科学界对这一修饰的认识. 在生物体内, 有一类富含O-GalNAc修饰的蛋白, 即黏蛋白(mucin), 其通常为上皮黏液的组成部分^[28]. 黏蛋白往往富含丝氨酸与苏氨酸重复序列的结构域, 在这些结构域上具有高度糖基化的特点, 使得糖基化质量分数占比在50%以上^[29]. 由于黏蛋白含有大量糖链的特性, 使之主要用途是为暴露在环境中的生物体表面保存水分, 并大量出现在消化道、呼吸系统表面.

表 1 O-GalNAc修饰核心结构汇总

O-GalNAc基本 结构单元	化学式	结构式
Tn	GalNAcαSer/Thr	
sTn	Siaα2-6GalNAcαSer/Thr	
Core 1 (T抗原)	Galβ1-3GalNAcαSer/Thr	
Core 2	GlcNAcβ1-6(Galβ1-3)GalNAcαSer/Thr	
Core 3	GlcNAcβ1-3GalNAcαSer/Thr	
Core 4	GlcNAcβ1-6(GlcNAcβ1-3)GalNAcαSer/Thr	

3 基于生物质谱技术的 *O*-GalNAc修饰位点的富集鉴定及完整糖肽解析

O-GalNAc糖基化修饰发挥生物学功能是由其糖蛋白骨架及其糖链结构共同协调作用的, 因此, 以完整糖肽作为研究对象, 可同时得到蛋白及糖链信息, 能够将糖基化位点及特定位点上糖型信息完整解析. 通过保留其发挥作用的桥梁, 即糖基化连接形式, 将有助于更深刻地理解糖基化修饰对于生理及病理下如何发挥生物学功能, 将为糖基化研究领域带来更全面、深刻的分析能力. 质谱技术作为近年来生命科学领域快速发展的高灵敏度分析手段, 在蛋白质翻译后解析领域发挥着重要作用. 当前, 基于生物质谱技术的 *O*-GalNAc糖基化的修饰位点鉴定及完整糖肽解析主要包括以下5个方面: (i) *O*-GalNAc糖基化蛋白或肽段的富集; (ii) *O*-GalNAc糖肽的质谱鉴定; (iii) 自动化分析软件对质谱数据的解析; (iv) 酶促反应及化学方法用于 *O*-GalNAc糖链的糖型解析; (v) 基因编辑技术在 *O*-GalNAc研究领域的应用.

3.1 *O*-GalNAc糖基化富集研究进展

由于 *O*-糖肽的化学组成与 *N*-糖肽近似, 其理化性质接近, 众多 *N*-糖肽研究中使用的富集方法同样可实现 *O*-糖肽的富集. 亲水色谱法是最常见的糖肽富集方法, 利用 *O*-糖肽的亲水性质实现 *O*-糖肽的富集^[30]. 凝集素弱亲和色谱法是利用凝集素与特定糖单元的亲和作用来实现特定糖型的富集. 其中, Jacalin凝集素是最常见的用于GalNAc类型糖单元富集的凝集素^[31-33]. 唾液酸是 *N*糖及 *O*糖链末端常见的糖结构单元. 2007年, Larsen等人^[34]和Palmisano等人^[35]利用二氧化钛成功实现了唾液酸化糖肽的富集. 继而有一系列利用二氧化钛对唾液酸化 *N*糖肽进行富集的研究出现^[36,37]. 这一类富集方法可以保留完整的糖肽信息, 将来可作为末端带有唾液酸的 *O*-GalNAc糖肽的富集方法, 但目前这一领域还未见研究成果报道^[38]. 另外, 硼酸分子可与糖结构单元中的邻二醇结构在碱性条件下反应, 生成环状二酯结构; 而在酸性条件下发生解离, 亦被应用于糖肽的富集. 以上方法的目的是实现低丰度 *O*-糖肽的富集, 并去除非糖肽的干扰, 使其在质谱分析时具有更高的信号响应和检测灵敏度.

3.2 质谱技术应用于 *O*-GalNAc糖基化的鉴定

质谱是当前糖蛋白质组学研究中最重要分析手段, 其分析能力的飞速发展为近十几年来糖蛋白质学研究的高速发展提供了强有力的技术支持^[39,40]. 当前生物质谱仪器按质量分析器的不同主要分为以下5种: 四级杆质谱仪(quadrupole)、飞行时间质谱仪(TOF)、离子阱质谱仪(ion trap)、傅里叶变换离子回旋共振质谱仪(Fourier transform ion cyclotron resonance, FTICR)、静电场轨道阱质谱(orbitrap). 而从电离技术上, 主要区分为基质辅助激光解吸电离(matrix assisted laser desorption/ionization, MALDI)^[41,42]与电喷雾电离(electrospray ionization, ESI)^[26]两种. 各品牌商业化质谱往往由2~3种质量分析器组合而成, 可实现多级质谱碎裂模式^[14,43], 极大方便了糖基化鉴定. 糖肽由于其自身特点, 如其由糖链部分与肽段部分组成, 分子量远大于普通肽段, 因此, 在数据依赖性质谱扫描模式(data dependent analysis, DDA)中, 一级质量分析器需要有足够宽的质量扫描范围(m/z 需可以检测到2000 Da). 另外, 由于糖肽的二级质谱碎片含有大量的糖链碎片特征离子, 因此需要二级质谱能够涵盖尽可能低的质量范围, 能检测到质荷比(m/z)在100 Da左右的碎片. 而线性离子阱质量分析器存在1/3 cut off效应, 会丢失低质量端的糖链特征碎片, 在糖基化研究中存在一定缺陷^[44]. 因此, 可避免1/3效应的高精度轨道离子阱质量分析器, 可对糖碎片离子进行有效的检测, 具有得天独厚的优势. 另外, 由于唾液酸的强负电性, 在MALDI-TOF质谱仪器中无法得到很好的质谱信号响应, 更适合在ESI质谱中进行检测^[45].

CID (collision-induced dissociation)与HCD (higher energy collision dissociation)碎裂方式是近年来蛋白质组学研究中最常用的质谱碎裂方式, 其基于碰撞能量再分配的基本原理, 碎裂肽段中的肽键, 产生b, y类型的碎片离子, 在肽段序列鉴定中具有很大优势. 然而, 由于糖苷键键能小于肽键, 糖基化修饰在CID及HCD碎裂中极易断裂, 糖链丢失的中性丢失效应比较常见, 易形成一系列糖链离子特征碎片峰(图1). 在CID碎裂中, 由于能量低, 糖肽仅发生单次碎裂, 因此在二级谱图中往往能只检测到糖链碎片离子, 而缺失肽段碎片离子, 难以获得肽段序列信息. HCD与CID形成碎片的离子形式相同, 主要产生b, y离子, 然而由于其采用

图 1 糖基化碎片在CID及HCD碎裂中的糖碎片谱图(网络版彩图)

高能碎裂的方式,使得糖肽易发生二次碎裂,相比CID碎裂可产生更多的肽段碎片信息。CID及HCD碎裂谱图中含有大量的糖链碎片离子特征峰,常用于糖肽离子的谱图筛选^[46]。但肽段碎片离子中往往存在糖链部分的脱落缺失,使得O-GalNAc糖基化精确位点确认变得困难,尤其是在目标肽段含有多个S或T时,位点指认难度较大。糖肽质谱解析的另一大难点是糖链存在同分异构形式,如GlcNAc与GalNAc,在一级质谱中分子量完全相同,无法进行有效区分。然而,近年针对这两种糖结构单元在HCD碎裂中碎片离子分布特征的研究发现,质量数为144的碎片离子信号在GalNAc中明显强于GlcNAc^[47]。通过这一发现,可以在糖基化鉴定中通过建立糖碎片离子比值模型进行GalNAc与GlcNAc的区分^[48]。

ECD (electron capture dissociation)与ETD (electron transfer dissociation)在传统的蛋白质组学研究中受到肽段碎裂效率低、分析灵敏度差等弊端限制,应用范围有限。而随着ECD及ETD技术在近些年的逐步发展,为糖肽二级质谱碎裂进行结构解析提供了全面而强大的工具^[49-52]。ECD主要内置于傅里叶变换离子回旋质谱仪,而ETD常内置于静电场轨道阱质谱仪中。在ECD及ETD碎裂中,肽段骨架中N-C α 键断裂,理论上可以保留完整的不稳定修饰,实现修饰位点的精确定位^[53,54]。得益于这一特点,大量依赖ECD及ETD技术的完整糖肽研究在近年不断涌现^[55]。然而,ECD/ETD由

于自身碎裂机理的特殊性,相比CID/HCD碎裂,其碎片离子更复杂,其中最具特点的是z离子氢原子重排效应的产生,使得z离子质量数产生1 Da的质量偏移,增大了谱图解析的难度。

Medzihradskyaff等人^[56]早在1996年即开始用CID裂解方式对O-糖基化进行鉴定。Mirgorodskaya等人^[57]于1999年系统性研究了ECD条件下O-GalNAc的碎裂,发现在ECD碎裂模式下,糖肽骨架c, z离子的形成快于糖碎片的断裂,从而在c, z离子中保留完整的糖链信息,有利于糖基化位点的解析。ETD技术是近年来出产的一系列高端orbitrap组合型质谱中集成的碎裂技术,已在多种蛋白质分析研究中发挥重要作用^[58-61]。离子在经过ETD活化后,进一步进行HCD碎裂,从而将糖肽碎裂成包含b, y, c, z在内的共4种碎片离子。既得到了肽段鉴定需要的包含肽段序列信息的b, y离子;又含有保留了修饰位点糖链的c, z离子,是当前最有利于大规模糖肽分析的质谱技术之一。

多反应质谱检测技术MRM,是当前蛋白质组学研究中一项靶向肽段定量技术,也被用于糖肽定量^[62-64]。当前,对糖基化肽段的分析往往仅涉及位点鉴定,但是更深层次的分析其含量的动态变化将会是未来精准糖基化研究的热门领域^[65]。得益于MRM方法自身得天独厚的灵敏度及定量准确度,已有采用该技术对糖肽进行精确定量的研究,即通过采用前体离子扫描方法,设定特定糖基化碎片离子,继而进行母离子

扫描, 从而得到含有糖基化的母离子肽段. Ritchie等人^[63]在2002年即利用MRM方法靶向定量了质荷比204的HexNAc+H离子及含有一个HexNAc的Y1离子. 由于*O*-GalNAc糖肽也会产生大量的糖链碎片, 因此, 理论上此策略同样可应用于*O*-GalNAc的糖基化母离子筛选, 但尚未见文献报道.

3.3 自动化分析软件对质谱数据的解析

在对糖肽的质谱谱图解析研究中, 最基本的研究方案是在一级质量数水平上对糖链结构及组成进行预测, 众多算法均采用这一策略, 包括Coooper等人^[66]开发的Glycomod软件等. 然而, 由于一级质谱仅包含质量数这一最基本信息, 无法利用糖链在二级质谱中的碎裂信息, 因此难以提供准确的糖链组成, 大大限制了该软件的应用. 另外, 当前高通量蛋白质组学鉴定所采用的HCD和ETC/D碎裂模式, 能产生大量的高质量二级碎裂质谱图, 因此, 基于二级谱图的高准确性完整糖肽鉴定是当前的研究热点.

当前, 对完整糖肽的鉴定主要集中在*N*-糖肽鉴定中, 例如, 中国科学院计算技术研究所贺思敏及复旦大学杨芑原研究组^[67]开发的pGlyco软件, 可利用HCD及CID碎片信息进行*N*-糖肽的鉴定. 而在*O*-GalNAc类型糖基化修饰领域, 亦有一些专门针对糖肽二级碎片进行算法优化的软件出现, 极大方便了*O*-GalNAc修饰的完整糖肽鉴定^[68]. 其中Byonic软件是将糖链按照组成设为可变修饰, 并在出厂数据库中添加了大量人源及哺乳动物中常见的*N*糖及*O*糖型, 用户无需设定特定的糖型修饰, 即可实现完整糖肽的检索^[69]. Medzihradszky研究组^[70-72]在近年来利用Protein Prospector软件做了大量关于*N*糖及*O*糖基化完整糖肽的鉴定工作. Protein Prospector软件可开放性地在谱图中筛选出糖基化肽段, 可在200~2000 Da范围内检索发生在Asn, Ser及Thr上可能的糖基化修饰. 完成筛选后, 进一步通过相应的二级质谱碎片信息与精确糖链组成列表确定糖肽的糖链组成. GlycoSuiteDB是由Cooper团队建立的包括大量*N*-糖链及*O*-糖链的结构及文献溯源的数据库^[73], 可为研究人员对完整*O*糖的研究提供极大便利.

3.4 酶促反应及化学方法用于*O*-GalNAc糖基化的解析

目前, 针对*O*-GalNAc的糖苷内切酶很少, 尚无通

用的将*O*-糖从肽段骨架上释放的内切酶, 只有一些特异性识别core1或core3结构的*O*-糖基化内切酶^[74]. 利用化学方法切除糖链的 β 消除反应可同时进行*O*-GalNAc糖基化位点的标记. 这种方法的原理是在碱性环境中, 丝氨酸及苏氨酸上的*O*糖链通过 β 消除反应被去除, 并产生一个不饱和双键, 随后受到亲核试剂的攻击, 通过加成反应在原来的*O*-糖基化位点产生额外的质量数标记. 通过这一标记, 可在质谱鉴定及数据库检索中, 准确确认*O*-糖基化位点信息^[13,75,76]. 然而需要特别注意的是, β 消除反应不仅与*O*-GalNAc位点发生反应, 对于其他发生在Ser及Thr上的磷酸化及*O*-GlcNAc修饰, 同样会发生反应^[77,78]. 也有研究表明, 当温度超过25℃的强碱环境下, 即便是自由Ser及Thr残基亦有强烈倾向发生 β 消除反应^[79]. 这些潜在问题是利用 β 消除反应进行*O*-GalNAc位点鉴定时需要注意的. 另外, 通过糖苷酶及化学方法简化*O*-GalNAc糖基化修饰中糖单元种类的策略已得到成功应用^[31]. 半乳糖氧化酶可对*N*糖及*O*糖末端Gal或GalNAc进行氧化, 将半乳糖C-6位羟基氧化为醛基, 从而可以实现肼化学富集^[80], 是一种潜在的*O*糖基化鉴定策略. 糖基化代谢标记策略是糖蛋白质组研究领域中的重要方法. 2003年, Hang等人^[81]将叠氮标记的GalNAz加入到细胞培养基中, 从而在细胞代谢过程中, 将GalNAz标记到特定糖基化位点上, 继而利用叠氮基团参与的点击化学反应, 实现*O*-GalNAc糖基化的体内标记或体外富集. 2012年, Wang等人^[82]利用合成具有二硫键及炔基官能团的硅球材料, 利用点击化学反应, 特异性富集标记有叠氮基团的GalNAz-肽段, 继而采用二硫苏糖醇对二硫键进行还原, 实现GalNAz-肽段的释放, 从而实现GalNAz-肽段的富集与质谱检测. 2017年, Xu等人^[83]利用ppGalNAc-T将UDP-GalNAz转移至蛋白的丝氨酸/苏氨酸残基上, 利用基于芯片及叠氮炔基点击化学反应的方法, 进行3种不同的多肽-*N*-乙酰氨基半乳糖转移酶的底物筛选(ppGalNAc-T1, T2和T3), 共鉴定到570个潜在的底物蛋白, 并发现仅有128个底物蛋白是3种酶的共同底物. 2015年, Woo等人^[84]发明了以代谢标记为基础的IsoTag方法, 该方法利用代谢标记方法, 标记Ac₄GalNAz进入到细胞内*O*-GalNAc修饰中, 进而利用点击化学反应及链霉亲和素-生物素亲和纯化实现糖基化肽段的富集, 继而结合基于溴元素同位素分布谱图筛选技术, 实现了糖基化糖肽的质谱鉴定.

在这一策略里,可同时进行糖肽的全方位分析,即(i)肽段信息,(ii)糖基化位点,(iii)完整糖链解析,可实现O-GalNAc糖基化修饰的微不均一性分析.

3.5 基因编辑技术的发展对O-GalNAc质谱鉴定的突破性贡献

2011年, Steentoft等人^[85-87]利用锌指核酸酶基因编辑方法建立了SimpleCell技术,使得O-GalNAc修饰位点的大规模鉴定成为可能. COSMC基因编码蛋白是 β 3半乳糖转移酶的分子伴侣蛋白,这项技术通过对COSMC基因的敲除,使得Core1类型O-GalNAc糖型合成受阻(图2). 另一方面, core3合成途径主要发生在肠胃组织中,而在癌细胞系中几乎不存在^[88]. 因此, COSMC基因的敲除将使得细胞特异性表达Tn抗原及sTn抗原,即GalNAc α 1-O-Ser/Thr及NeuAc α 2-6GalNAc α 1-O-Ser/Thr结构,大大简化了O-GalNAc糖基化的糖链组成,消除了糖链微不均一性对鉴定造成的困难. 辅助以唾液酸酶的使用,产物中只存在Tn抗原这一种糖型,更有利用质谱对O-GalNAc糖基化位点的鉴定. 结合LWAC串联凝集素富集方法,对简化的O-GalNAc糖肽进行富集^[89]. 利用这一策略最多可在12个改造的Simple-cells癌细胞系中鉴定到3000个O-GalNAc糖基化位点,对应600多个糖蛋白^[87]. 在未来O-GalNAc糖基化研究中,得益于ZFN,TALEN及CRISPR/Cas9基因编辑系统的迅速发展,基于基因编辑方法的O-糖基化研究策略将会拥有强大的潜力^[90].

图2 O-glycosylation体内合成路径^[85]

4 总结与展望

蛋白质O-GalNAc糖基化修饰由于其结构的复杂性,位点修饰的不均一性,以及缺乏特定氨基酸结构域等特点,使得O-GalNAc糖基化修饰研究面临众多技术挑战. 得益于生化技术、质谱技术及基因编辑技术的飞速发展,基于ETD的修饰保留碎裂技术、代谢标记方法和基因编辑技术成为O-GalNAc修饰研究的三大核心手段^[38]. 与此同时,用于完整O-GalNAc糖肽的质谱数据自动化解析软件也取得了众多突破. 综上所述,日新月异新技术的出现使得O-GalNAc糖基化修饰研究由点及面,从最初的单一蛋白、单一糖型的研究拓展到整个细胞、组织层面的研究. 蛋白质O-GalNAc糖基化修饰的解析将在未来糖蛋白结构和功能研究、生物工程药物糖基化分析、分子生物学水平上的关键蛋白质修饰调控研究以及重大疾病的诊断标志物和治疗靶标筛选中发挥重要作用.

参考文献

- Bunkenborg J, Pilch B J, Podtelejnikov A V, et al. Screening for N-glycosylated proteins by liquid chromatography mass spectrometry. *Proteomics*, 2004, 4: 454-465
- Ohtsubo K, Marth J D. Glycosylation in cellular mechanisms of health and disease. *Cell*, 2006, 126: 855-867
- Rudd P M, Elliott T, Cresswell P, et al. Glycosylation and the immune system. *Science*, 2001, 291: 2370-2376
- Kim E H, Misek D E. Glycoproteomics-based identification of cancer biomarkers. *Int J Proteomics*, 2011, in press, doi: 10.1155/2011/601937
- Solá R J, Griebenow K. Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharmaceut Sci*, 2009, 98: 1223-1245
- 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
- Chalkley R J, Thalhammer A, Schoepfer R, et al. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. *Proc Natl Acad Sci USA*, 2009, 106: 8894-8899
- Hang H C, Bertozzi C R. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg Med Chem*, 2005, 13: 5021-5034
- Ten Hagen K G, Fritz T A, Tabak L A. All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*, 2003, 13: 1R-16
- Rottger S, White J, Wandall H H, et al. Localization of three human polypeptide galnac-transferases in hela cells suggests initiation of O-linked

- glycosylation throughout the golgi apparatus. *J Cell Sci*, 1998, 111: 45–60
- 11 Berger E G. Tn-syndrome. *Biochim Biophys Acta*, 1999, 1455: 255–268
- 12 Novak J, Rizk D, Takahashi K, et al. New insights into the pathogenesis of IgA nephropathy. *Kidney Dis*, 2015, 1: 8–18
- 13 Mirgorodskaya E, Hassan H, Clausen H, et al. Mass spectrometric determination of *O*-glycosylation sites using β -elimination and partial acid hydrolysis. *Anal Chem*, 2001, 73: 1263–1269
- 14 Amano J, Sugahara D, Osumi K, et al. Negative-ion MALDI-QIT-TOFMSn for structural determination of fucosylated and sialylated oligosaccharides labeled with a pyrene derivative. *Glycobiology*, 2009, 19: 592–600
- 15 Doerr A. Glycoproteomics. *Nat Meth*, 2012, 9: 36–36
- 16 Krasnova L, Wong C H. Understanding the chemistry and biology of glycosylation with glycan synthesis. *Annu Rev Biochem*, 2016, 85: 599–630
- 17 Moremen K W, Tiemeyer M, Nairn A V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol*, 2012, 13: 448–462
- 18 Jensen P H, Kolarich D, Packer N H. Mucin-type *O*-glycosylation-putting the pieces together. *FEBS J*, 2010, 277: 81–94
- 19 Moreau R, Dausset J, Bernard J, et al. Acquired hemolytic anemia with polyagglutinability of erythrocytes by a new factor present in normal blood. *Bull Mem Soc Med Hop Paris*, 1957, 73: 569–587
- 20 Reepmaker J. The relation between polyagglutinability of erythrocytes *in vivo* and the hbener-thomsen-friedenreich phenomenon. *J Clin Pathol*, 1952, 5: 266–270
- 21 Dahr W, Uhlenbruck G, Gunson H H, et al. Molecular basis of Tn-polyagglutinability. *Vox Sanguinis*, 1975, 29: 36–50
- 22 Kaifu R, Osawa T. Synthesis of *O*-(2-acetamido-2-deoxy- α -D-galactopyranosyl)-*N*-tosyl-L-serine. *Carbohydr Res*, 1977, 58: 235–239
- 23 Ju T, Otto V I, Cummings R D. The Tn antigen-structural simplicity and biological complexity. *Angew Chem Int Ed*, 2011, 50: 1770–1791
- 24 Schjoldager K T B G, Vakhrushev S Y, Kong Y, et al. Probing isoform-specific functions of polypeptide GalNAc-transferases using zinc finger nuclease glycoengineered SimpleCells. *Proc Natl Acad Sci USA*, 2012, 109: 9893–9898
- 25 Willer C J, Sanna S, Jackson A U, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet*, 2008, 40: 161–169
- 26 Boskovski M T, Yuan S, Pedersen N B, et al. The heterotaxy gene GALNT11 glycosylates Notch to orchestrate cilia type and laterality. *Nature*, 2013, 504: 456–459
- 27 Halim A, Brinkmalm G, Rtschi U, et al. Site-specific characterization of threonine, serine, and tyrosine glycosylations of amyloid precursor protein/amyloid beta-peptides in human cerebrospinal fluid. *Proc Natl Acad Sci USA*, 2011, 108: 11848–11853
- 28 王胜, 邹霞, 张延. 基于质谱的蛋白质 *O*-糖基化分析. *化学进展*, 2010, 22: 2428–2435
- 29 Rose M C, Voynow J A. Respiratory tract mucin genes and mucin glycoproteins in health and disease. *Physiol Rev*, 2006, 86: 245–278
- 30 Hoffmann M, Marx K, Reichl U, et al. Site-specific *O*-glycosylation analysis of human blood plasma proteins. *Mol Cell Proteomics*, 2016, 15: 624–641
- 31 Darula Z, Medzihradsky K F. Affinity enrichment and characterization of mucin core-1 type glycopeptides from bovine serum. *Mol Cell Proteomics*, 2009, 8: 2515–2526
- 32 Durham M, Regnier F E. Targeted glycoproteomics: serial lectin affinity chromatography in the selection of *O*-glycosylation sites on proteins from the human blood proteome. *J Chromatogr A*, 2006, 1132: 165–173
- 33 Hortin G L. Isolation of glycopeptides containing *O*-linked oligosaccharides by lectin affinity chromatography on jacalin-agarose. *Anal Biochem*, 1990, 191: 262–267
- 34 Larsen M R, Jensen S S, Jakobsen L A, et al. Exploring the sialome using titanium dioxide chromatography and mass spectrometry. *Mol Cell Proteomics*, 2007, 6: 1778–1787
- 35 Palmisano G, Lendal S E, Engholm-Keller K, et al. Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry. *Nat Protoc*, 2010, 5: 1974–1982
- 36 Cao W, Cao J, Huang J, et al. Enhanced *N*-glycosylation site exploitation of sialoglycopeptides by peptide IPG-IEF assisted TiO₂ chromatography. *Glycoconj J*, 2012, 29: 433–443
- 37 Zhao X, Ma C, Han H, et al. Comparison and optimization of strategies for a more profound profiling of the sialylated *N*-glycoproteomics in human plasma using metal oxide enrichment. *Anal Bioanal Chem*, 2013, 405: 5519–5529
- 38 Lavery S B, Steentoft C, Halim A, et al. Advances in mass spectrometry driven *O*-glycoproteomics. *Biochim Biophys Acta*, 2015, 1850: 33–42
- 39 Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*, 2003, 422: 198–207
- 40 Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*, 2016, 537: 347–355
- 41 Tanaka K, Waki H, Ido Y, et al. Protein and polymer analyses up to *m/z* 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid*

- Commun Mass Spectrom*, 1988, 2: 151–153
- 42 Fenn J B, Mann M, Meng C K, et al. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 1989, 246: 64–71
  - 43 Zhang S, Williamson B L. Characterization of protein glycosylation using chip-based nanoelectrospray with precursor ion scanning quadrupole linear ion trap mass spectrometry. *J Biomol Tech*, 2005, 16: 209–219
  - 44 Yang Y H, Lee K, Jang K S, et al. Low mass cutoff evasion with qz value optimization in ion trap. *Anal Biochem*, 2009, 387: 133–135
  - 45 Ito H, Takegawa Y, Deguchi K, et al. Direct structural assignment of neutral and sialylated *N*-glycans of glycopeptides using collision-induced dissociation MSn spectral matching. *Rapid Commun Mass Spectrom*, 2006, 20: 3557–3565
  - 46 Hahne H, Gholami A M, Kuster B. Discovery of *O*-GlcNAc-modified proteins in published large-scale proteome data. *Mol Cell Proteomics*, 2012, 11: 843–850
  - 47 Halim A, Westerlind U, Pett C, et al. Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC-MS/MS of glycopeptides. *J Proteome Res*, 2014, 13: 6024–6032
  - 48 Wang X, Yuan Z F, Fan J, et al. A novel quantitative mass spectrometry platform for determining protein *O*-GlcNAcylation dynamics. *Mol Cell Proteomics*, 2016, 15: 2462–2475
  - 49 Zubarev R A, Horn D M, Fridriksson E K, et al. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*, 2000, 72: 563–573
  - 50 Syka J E P, Coon J J, Schroeder M J, et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA*, 2004, 101: 9528–9533
  - 51 Sarbu M, Ghiulai R M, Zamfir A D. Recent developments and applications of electron transfer dissociation mass spectrometry in proteomics. *Amino Acids*, 2014, 46: 1625–1634
  - 52 Zubarev R A, Kelleher N L, McLafferty F W. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc*, 1998, 120: 3265–3266
  - 53 Zhurov K O, Fornelli L, Wodrich M D, et al. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem Soc Rev*, 2013, 42: 5014–5030
  - 54 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, (1): 94–102
  - 55 Halim A, Nilsson J, Ruetschi U, et al. Human urinary glycoproteomics attachment site specific analysis of *N*- and *O*-linked glycosylations by CID and ECD. *Mol Cell Proteomics*, 2012, 11: M111 013649
  - 56 Medzihradszky K F, Gillece-Castro B L, Townsend R R, et al. Structural elucidation of *O*-linked glycopeptides by high energy collision-induced dissociation. *J Am Soc Mass Spectrom*, 1996, 7: 319–328
  - 57 Mirgorodskaya E, Roepstorff P, Zubarev R A. Localization of *O*-glycosylation sites in peptides by electron capture dissociation in a fourier transform mass spectrometer. *Anal Chem*, 1999, 71: 4431–4436
  - 58 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
  - 59 Frese C K, Zhou H, Taus T, et al. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res*, 2013, 12: 1520–1525
  - 60 Liu F, van Breukelen B, Heck A J R. Facilitating protein disulfide mapping by a combination of pepsin digestion, electron transfer higher energy dissociation (EThcD), and a dedicated search algorithm SlinkS. *Mol Cell Proteomics*, 2014, 13: 2776–2786
  - 61 Mommen G P M, Frese C K, Meiring H D, et al. Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci USA*, 2014, 111: 4507–4512
  - 62 Song E, Pyreddy S, Mechref Y. Quantification of glycopeptides by multiple reaction monitoring liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom*, 2012, 26: 1941–1954
  - 63 Ritchie M A, Gill A C, Deery M J, et al. Precursor ion scanning for detection and structural characterization of heterogeneous glycopeptide mixtures. *J Am Soc Mass Spectrom*, 2002, 13: 1065–1077
  - 64 Sandra K, Devreese B, Van Beeumen J, et al. The Q-trap mass spectrometer, a novel tool in the study of protein glycosylation. *J Am Soc Mass Spectrom*, 2004, 15: 413–423
  - 65 王继峰, 江静, 彭博, 等. 完整糖肽质谱解析的研究进展. *质谱学报*, 2014, 35: 108–117
  - 66 Cooper C A, Gasteiger E, Packer N H. GlycoMod—A software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*, 2001, 1: 340–349
  - 67 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
- 68 曾文锋, 张扬, 刘铭琪, 等. *N*-糖肽的规模化质谱解析方法进展. *生物化学与生物物理进展*, 2016, 43: 550–562
- 69 Bern M, Kil Y J, Becker C. Byonic: advanced peptide and protein identification software. *Curr Protoc Bioinformatics*, 2012, Chapter 13: Unit13 20
- 70 Darula Z, Chalkley R J, Baker P, et al. Mass spectrometric analysis, automated identification and complete annotation of *O*-linked glycopeptides. [Eur J Mass Spectrom](#), 2010, 16: 421–428
- 71 Medzihradsky K F, Kaasik K, Chalkley R J. Tissue-specific glycosylation at the glycopeptide level. [Mol Cell Proteomics](#), 2015, 14: 2103–2110
- 72 Trinidad J C, Schoepfer R, Burlingame A L, et al. *N*- and *O*-glycosylation in the Murine synaptosome. [Mol Cell Proteomics](#), 2013, 12: 3474–3488
- 73 Cooper C A, Harrison M J, Wilkins M R, et al. GlycoSuiteDB: a New Curated Relational. Oxford: Oxford University Press, 2001, 29: 332–335
- 74 O'Neill R A. Enzymatic release of oligosaccharides from glycoproteins for chromatographic and electrophoretic analysis. *J Chromatogr A*, 1996, 720: 201–215
- 75 Hanisch F G, Jovanovic M, Peter-Katalinic J. Glycoprotein identification and localization of *O*-glycosylation sites by mass spectrometric analysis of deglycosylated/alkylaminylated peptide fragments. [Anal Biochem](#), 2001, 290: 47–59
- 76 Czeszak X, Ricart G, Tetaert D, et al. Identification of substituted sites on MUC5AC mucin motif peptides after enzymatic *O*-glycosylation combining beta-elimination and fixed-charge derivatization. [Rapid Commun Mass Spectrom](#), 2002, 16: 27–34
- 77 McLachlin D T, Chait B T. Improved β -elimination-based affinity purification strategy for enrichment of phosphopeptides. [Anal Chem](#), 2003, 75: 6826–6836
- 78 Wells L, Vosseller K, Cole R N, et al. Mapping sites of *O*-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. [Mol Cell Proteomics](#), 2002, 1: 791–804
- 79 Li W, Backlund P S, Boykins R A, et al. Susceptibility of the hydroxyl groups in serine and threonine to β -elimination/Michael addition under commonly used moderately high-temperature conditions. [Anal Biochem](#), 2003, 323: 94–102
- 80 Ramya T N C, Weerapana E, Cravatt B F, et al. Glycoproteomics enabled by tagging sialic acid- or galactose-terminated glycans. [Glycobiology](#), 2013, 23: 211–221
- 81 Hang H C, Yu C, Kato D L, et al. A metabolic labeling approach toward proteomic analysis of mucin-type *O*-linked glycosylation. [Proc Natl Acad Sci USA](#), 2003, 100: 14846–14851
- 82 Wang S, Xie W, Zhang X, et al. Disulfide- and terminal alkyne-functionalized magnetic silica particles for enrichment of azido glycopeptides. [Chem Commun](#), 2012, 48: 5907–5909
- 83 Xu Z, Li X, Zhou S, et al. Systematic identification of the protein substrates of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase-T1/T2/T3 using a human proteome microarray. *Proteomics*, 2017, 17: 1600485
- 84 Woo C M, Iavarone A T, Spiciari D R, et al. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact *N*- and *O*-glycopeptide discovery and analysis. [Nat Meth](#), 2015, 12: 561–567
- 85 Steentoft C, Vakhrushev S Y, Vester-Christensen M B, et al. Mining the *O*-glycoproteome using zinc-finger nuclease-glycoengineered SimpleCell lines. [Nat Meth](#), 2011, 8: 977–982
- 86 Steentoft C, Bennett E P, Clausen H. Glycoengineering of human cell lines using zinc finger nuclease gene targeting: simplecells with homogeneous galnac *O*-glycosylation allow isolation of the *O*-glycoproteome by one-step lectin affinity chromatography. *Methods Mol Biol*, 2013, 1022: 387–402
- 87 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
- 88 Iwai T, Kudo T, Kawamoto R, et al. Core 3 synthase is down-regulated in colon carcinoma and profoundly suppresses the metastatic potential of carcinoma cells. [Proc Natl Acad Sci USA](#), 2005, 102: 4572–4577
- 89 Vakhrushev S Y, Steentoft C, Vester-Christensen M B, et al. Enhanced mass spectrometric mapping of the human GalNAc-type *O*-glycoproteome with SimpleCells. [Mol Cell Proteomics](#), 2013, 12: 932–944
- 90 Steentoft C, Bennett E P, Schjoldager K T B G, et al. Precision genome editing: a small revolution for glycobiology. [Glycobiology](#), 2014, 24: 663–680

Advances in *O*-GalNAc glycosylation research based on biological mass spectrometry

ZHAO XinYuan, SHEN BingQuan, QIN WeiJie & QIAN XiaoHong

1 School of Life Science and Bioengineering, Beijing University of Technology, Beijing 100022, China;

2 State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

Glycosylation is one of the most common and important post-translational modifications of proteins in eukaryotic cells. It is prevalent in cell membrane proteins and secreted proteins, performing important biological functions. Major types of protein glycosylation include *N*-glycosylation and *O*-glycosylation. *O*-GalNAc glycosylation is an important form of *O*-glycosylation and plays vital roles in regulating many key biological processes. It has close correlation with the occurrence and development of cancer. Thanks to powerful metabolic labeling methods, chemical derivatization approaches, mass spectrometry techniques and gene editing technology, the identification and elucidation of *O*-GalNAc glycosylation sites and sugar chains are developing rapidly. This paper reviews the advances in research of *O*-GalNAc glycosylation in recent years.

protein, glycosylation, mass spectrometry, gene editing

doi: [10.1360/N052017-00137](https://doi.org/10.1360/N052017-00137)