

糖组学研究技术及其进展*

杨 珺^{1,2)**} 蔡绍哲¹⁾ 邹全明²⁾

(¹重庆大学生物力学与组织工程教育部重点实验室, 重庆 400044; ²第三军医大学临床微生物教研室, 重庆 400038)

摘要 多细胞生物机体内, 蛋白质糖基化是一个重要后修饰事件. 蛋白质的糖链不仅仅是区别细胞种类的标志, 且与众多的生物现象有关, 如细胞发育、分化、形态、肿瘤转移、微生物感染等. 糖组学的内容主要涉及单个个体的全部糖蛋白结构分析, 确定编码糖蛋白的基因和蛋白质糖基化的机制. 综述了糖组学的分离和结构鉴定技术及其最新进展.

关键词 糖组 / 糖组学, 糖基化作用, 糖捕获, 前沿亲和色谱, 二维电泳, 多维液相色谱, 质谱

学科分类号 Q513.3, Q25

1 糖组及糖组学的研究意义

所有多细胞生物表面覆盖着反映不同细胞种类和状态的糖链, 它们不仅在细胞与细胞之间, 细胞与基底之间的识别事件起着重要作用, 而且与众多的生物现象有关, 如细胞发育、分化、形态、肿瘤转移、微生物感染等. 聚糖作为第三类生物信息大分子, 对于维持多细胞动物个体十分重要. 因此, 20 世纪末由几个糖生物学家同时提出了“糖组”的概念^[1-3]. 糖组指单个个体的全部聚糖. 类似基因组学和蛋白质组学, “糖组学”指对糖组的全面分析研究, 主要针对糖蛋白, 涉及单个个体的全部糖蛋白结构分析, 确定编码糖蛋白的基因和蛋白质糖基化的机制. 糖组学主要解决 4 个方面问题: a. 什么基因编码糖蛋白, 即基因信息; b. 可能糖基化位点中实际被糖基化的位点, 即糖基化位点信息; c. 聚糖结构, 即结构信息; d. 糖基化功能, 即功能信息^[4,5].

相对于基因组和蛋白质组, 糖组具有以下特点和重要性. a. 所有有机体的所有细胞都被丰富的糖链覆盖, 这种构成反映了细胞不同的种类和状态. b. 自然界糖种类较多, 但聚糖的组成糖种类仅有十几种, 常见的只有葡萄糖(Glu)、乙酰葡萄糖胺(GlcNAc)、甘露糖(Man)、半乳糖(Gal), 乙酰半乳糖胺(GalNAc), 木糖(Xyl), 阿拉伯糖(Ara)等. c. 聚糖如此高的复杂性取决于多变的键连接方式和分支方式(其他生物大分子没有). 聚糖的结构多样性在生物信息上比核酸和蛋白质有更高的影响力, 对于生物情报分子是必需的.

2 糖蛋白的结构与类型

蛋白质和核酸是氨基酸和核苷酸的直链聚合物, 而糖类具有复杂的分支和多样的键连接. 糖蛋白是由多肽和糖通过糖肽键连接而成的. 糖蛋白多肽链的合成与非糖基化蛋白质基本相同, 仅在多肽链合成的同时或之后伴有肽链的糖基化作用. 在真核生物中有 3 种糖基化, 即 N-糖基化, O-糖基化, GPI-糖基磷脂酰肌醇锚区. 而动物糖蛋白的糖基化主要发生在肽链上的天冬酰胺(Asn)、丝氨酸(Ser)、苏氨酸(Thr)、羟赖氨酸(Hyl)、羟脯氨酸(Hyp)残基上.

糖蛋白的糖链可以是直链或支链, 糖基数目一般在 1~15 个, 在生物体内糖链的结构模式还是有一定的规律性. 其中 N-糖基化有三甘露糖核心结构, 根据其他糖与之连接的情况分为高甘露糖型、复杂型和混合型三类. N-糖基化常以 β -N-GlcNAc-Asn 为连接起点, 并且与肽链的氨基酸序列有关, 这种序列称为天冬酰胺顺序子: Asn-X-Ser 或 Asn-X-Thr, X 代表除脯氨酸以外的氨基酸. O-糖基化存在多种形式, 但都由一种或几种单糖与含羟基的氨基酸连接, 没有共同的核心结构, 但在 O-GalNAc 连接糖链中存在 4 种核心结构, 即 Gal β 3(GlcNAc β 6)GalNAc α Ser/Thr(核心 1), Gal β 3GalNAc α Ser/Thr(核心 2), GlcNAc β 3

*国家高技术 863 计划重点资助项目(2001AA21516).

** 通讯联系人.

Tel: 023-68753591, Fax: 023-68752315, E-mail: W8301991@263.net

收稿日期: 2004-09-01, 接受日期: 2004-09-28

GalNAc α Ser/Thr(核心 3), GlcNAc β 6 (GlcNAc β 3) GalNAc α Ser/Thr(核心 4). GPI-糖基磷脂酰肌醇锚区为糖链通过磷脂酰肌醇与蛋白质连接, 主要分布在细胞膜.

3 糖组学分离技术

糖组学研究技术的关键在于糖蛋白的分离和富集. 目前, 糖蛋白和聚糖分离策略主要有两条途径: 一是经典的凝集素亲和色谱“糖捕获”方法, 另一种是二维电泳结合特殊染色的分离方法. 近年来又发展了快捷易于自动化的多维液相色谱分离多糖技术.

3.1 凝集素亲和色谱“糖捕获”方法

3.1.1 植物凝集素探针. 植物凝集素是一类非免疫来源的, 无酶活性的, 能与聚糖特异性结合的蛋白质, 它象探针一样可以捕获到混合物中的聚糖, 为目标细胞、组织或机体的糖组学研究提供第一手资料. 在凝集素亲和色谱中最常用的植物凝集素有刀豆素 A (concanavalin A, ConA)、半乳糖凝集素 (galectin, LEC-6 或 Gal6)、花生凝集素 (peanut agglutinin, PNA)、橙黄网胞盘菌凝集素 (*Aleuria aurantia* lectin, AAL) 等. 不同的植物凝集素用于分离不同类型的糖蛋白和糖肽, 如 ConA 和 Gal6 特异吸附高甘露糖型和复杂型 N-聚糖; 在 N-聚糖和 O-聚糖中 Gal6 专门吸附含有 LacNAc 的聚糖; 在 O-聚糖中 AAL 专门吸附 T 抗原, AAL 对含有 L-Fuc-聚糖有较宽的特异性^[9]. 显然, 在糖捕获操作中植物凝集素的应用还是有一定限制的.

3.1.2 糖基捕获的过程. 利用植物凝集素亲和柱捕获糖蛋白的过程: a. 植物凝集素亲和柱 1 捕获一组糖蛋白; b. 糖蛋白被蛋白酶彻底水解; c. 水解产物再经植物凝集素亲和柱 1 捕获到糖肽; d. 用 N-寡聚糖糖肽酶消化糖肽释放出肽链和糖链; e. 肽链和糖链分别经 HPLC/MS 分离鉴定; f. 获得肽序列和糖链分子质量; g. 分析蛋白质序列并查询数据库获得相关遗传信息; h. 分析聚糖的结构获得糖基化信息. 使用不同的植物凝集素柱进行第二和第三次循环, 捕获其他类型的糖肽, 可以对某个细胞和机体进行较全面的糖组学研究. 例如, 在第一循环中利用 ConA 亲和色谱, 第二循环用 Gal6 亲和糖蛋白, 第三循环用 PNA 亲和糖蛋白. 使用串联的植物凝集素亲和层析有效地覆盖不同类型的糖蛋白.

3.1.3 前沿亲和色谱获得生物亲和常数. 测定亲和常数的其他方法包括经典平衡渗析和最近发展的生

物传感器技术, 但大多数方法精度差, 操作复杂. 前沿亲和色谱 (frontal affinity chromatography, FAC) 是一个定量方法^[7], 可以精确地测定不同生物分子之间的亲和常数, 如酶-底物, 植物凝集素-聚糖. FAC 分析提供可靠的亲和信息, 即每种聚糖和一种植物凝集素之间的结合常数. 事实上, 植物凝集素对聚糖结构的细微差异具有高分辨能力, 植物凝集素被假定为体内复杂聚糖的翻译器. 使用串联的植物凝集素亲和色谱获得对不同植物凝集素的结合常数 K_a 值, 为聚糖结构分析提供第一手信息.

3.2 二维电泳结合特殊染色的分离方法

PAS 法 (periodic acid Schiff's mechanism, 高碘酸 Schiff's 法) 最初是作为寡糖的荧光标记技术发展的. 后来 PAS 运用在糖蛋白的二维电泳 (two-dimensional gel electrophoresis, 2DE) 中, 后修饰染色和鉴别染色技术对糖蛋白与非糖蛋白的差异染色, 染色蛋白质在胶内水解用质谱测定准确的相对分子质量. 该方法灵敏度较高, 检测限可达到 300 pg, 而且 MS 检测 PA 衍生糖蛋白的灵敏度远高于非衍生糖蛋白^[8]. Steinberg 等^[9] 使用高碘酸将糖蛋白的聚糖氧化成醛, 然后荧光染料 Pro-Q Emerald 300 共价结合到糖醛上产生高荧光共轭化合物. 通过 2DE 分离蛋白质, 2DE 凝胶用 SYPRO Ruby 染料染色, 所有蛋白质被染成红色, 而糖蛋白与 Pro-Q Emerald 300 的共轭化合物在 530 nm 呈绿色, 将非糖蛋白和糖蛋白区别开. 切下 2DE 凝胶上糖蛋白点, 可以进行后续的鉴定.

3.3 液相色谱分离多糖方法

多维液相色谱法 (multidimensional liquid chromatography) 是一种精确的糖组分离系统. Takahashi 等^[10] 建立三维糖谱图技术分离并鉴定人血清糖蛋白. 用 N-寡聚糖糖肽酶消化糖肽释放出聚糖, 经 PA 衍生在 HPLC 上, 先后经过二乙氨基乙基离子交换柱、C18 疏水柱和硅酰胺亲水柱分离, 洗脱数据分别列在 x 、 y 和 z 轴, 获得人血清糖蛋白的三维液相糖谱图, 借助连续的外切糖苷酶消化目标聚糖点获得结构信息.

利用多维液相色谱法分离蛋白聚糖方法具有以下优势: a. 与质谱直接连接自动化程度高、重复性好; b. 避免了 2DE 显色、切胶的损失; c. 放大效应, 分离得到较多的蛋白质能满足后续的鉴定; d. 可以进行多维分离, 分离更完全彻底. 基于这些优点, 随着多维液相色谱-串联质谱联用技术的发展和改进, 已成为糖组学研究的核心技术之一.

用毛细管液相色谱 / 串联质谱联机 (CapLC/MS/MS) 技术可以胜任细胞、组织内低丰度糖蛋白的结构分析. 毛细管液相色谱的优势在于对痕量糖蛋白 (ng 水平) 的分离, 与串联质谱连用结合外切糖苷酶获得每个聚糖的定位和结构. Kawasaki 等^[11]用 CapLC/MS/MS 鉴定了鼠骨髓瘤细胞产生的肝细胞生长因子为混合型 N- 连接聚糖. CapLC/MS/MS 得到的糖谱对于检测一个低丰度糖蛋白的糖基化模式和糖类的结构分析很有用, 在生物学、制药学和临床研究中将成为有力的工具.

4 糖组学的鉴定技术

如果从事某个有机体的糖组研究, 确定所有聚糖的共价结构是一个主要的内容. 而聚糖是以非常复杂的方式结合在一起, 鉴定一个有机体中成千上万聚糖的共价结构是一个庞大的工程. 传统方法定义共价聚糖不够灵敏也太费事, 目前还没有自动化聚糖序列方法, 因此急切需要发展一种鉴定聚糖结构的新方法. 高精度质谱可以获得准确的质量数, 为共价聚糖的剖析提供准确的信息, 成为糖组学研究的生力军.

4.1 基质辅助激光解吸离子化飞行质谱

当前, 基质辅助激光解吸离子化飞行质谱 (matrix-assisted laser desorption ionization time-of-flight, MALDI-TOF) 在蛋白质组研究中扮演着主要的角色. MALDI-TOF-MS 具有从非衍生分子中获得离子的能力, 正确反映糖蛋白的构成, 同时方便与其他技术, 如 HPLC、GC 或外切糖苷酶结合, 提供更多的结构信息. 但是 MALDI-TOF 的分辨率和质量精度较低会导致蛋白质鉴定的不确定性, 同时会引起分子中较弱的键断裂, 引起糖链部分或全部丢失, 糖链的定位信息和结构信息就会丢失, 所以 MALDI-TOF 在糖组学研究中还存在遗憾^[12].

4.2 傅立叶变换离子回旋共振质谱

高分辨率、高质量精度的傅立叶变换离子回旋共振质谱 (fourier transform ion cyclotron resonance mass spectrometry, FT-ICR-MS) 吸引了研究者的注意. 1974 年 Comisarow 等^[13]发明了 FT-ICR-MS, 几经改进, 直到 Henry 等^[14]将电喷雾离子化技术 (electrospray ionization, ESI) 引入 FT-ICR-MS 才真正开启了高精度生物学分析的大门. ESI 高灵敏度改进, 如 microelectrospray 和 nanoelectrospray, FT-ICR-MS 可以分析真正生物水平 (nmol) 的生物

样品^[15]. FT-ICR-MS 也受到糖组学研究的关注. FT-ICR-MS 多重的独特的分裂技术碰撞活化的分裂 (collisional activated dissociation, CAD), 红外线多光子分裂 (infrared multiphoton dissociation, IRMPD) 以及电子捕获分裂 (electron-capture dissociation, ECD), 可以确定后修饰的位置和修饰的结构^[16]. ECD 劈开氨基酸主链而不影响后修饰部分, 可以进行后修饰定位. IRMPD 劈开后修饰部分, 分析后修饰的结构信息^[17]. Mark^[18]将小胶质细胞全蛋白质经 PA 衍生后 2DE 分离, 糖蛋白被 SYPRO-Ruby 染色, 在胶内用胰蛋白酶水解, 先用 MALDI-TOF-MS 分析肽链组成, 余下的样品用 micro-ES -FT-ICR-MS 测定糖链的结构信息, 很轻松地获得几个糖蛋白的全部结构.

5 展 望

目前糖组研究方法较少, 糖蛋白为主的糖组学技术问题和发展趋势是: a. 糖蛋白分离的问题. 使用植物凝集素柱捕获不同的糖蛋白是目前许多研究者的首选技术, 但是一方面没有一种能吸附所有类型糖蛋白的植物凝集素, 另一方面可能会丢失一些能通过任何植物凝集素柱的糖肽, 而研究不被任何植物凝集素识别可能具有生物学上的重大意义的聚糖. 具有更广泛的吸附范围内源植物凝集素越来越多地用于捕获糖蛋白, 而全能糖吸附剂的发展是必然趋势. b. 使用当前技术, 糖蛋白的基因识别效率较低. 非常短的糖肽 (小于 6 个氨基酸) 在基因组数据库中是不可能鉴定目标基因的. 非常长的糖肽 (大于 20 个氨基酸) 即使可以成功鉴定目标基因, 关于糖基化位点的信息却有限. c. 糖链的结构分析和共价键的确定仍然不是一件容易的工作, 虽然高精度的质谱可以完成一些鉴定, 但是该技术要求纯的样品和熟练的操作人员, 且仪器昂贵. d. 目前这些糖组研究方法几乎是静态的, 而生物体的糖基化过程是动态的, 研究方法的动态化也是一种必然.

糖组学是后基因组学的一个重要领域. 许多生物信息都体现在丰富多样的复合糖形式, 应用基因组 - 蛋白质组 - 糖组的概念来全面解释复杂的生命系统是完全必要的, 随着样品分离技术和相关仪器的发展, 快速准确的糖组鉴定将能够实现.

参 考 文 献

- Hirabayashi J, Arata Y, Kasai K. Glycome project: concept, strategy and preliminary application to *Caenorhabditis elegans*. *Proteomics*, 2001, 1 (2): 295~303

- 2 Feizi T. Progress in deciphering the information content of the 'glycome' — a crescendo in the closing years of the millennium. *J Glycoconj*, 2000, **17** (7-9): 553~565
- 3 Taniguchi N, Ekuni A, Ko J H, *et al.* A glycomic approach to the identification and characterization of glycoprotein function in cells transfected with glycosyltransferase genes. *Proteomics*, 2001, **1** (2): 239~247
- 4 Demetriou M, Granovsky M, Quaggin S, *et al.* Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. *Nature*, 2001, **409** (2): 733~739
- 5 Griffiths J S, Whitacre J L, Stevens D, *et al.* The toxin resistance from loss of a putative carbohydrate-modifying enzyme. *Science*, 2001, **293** (5531): 860~864
- 6 Seetharaman J, Kanigsberg A, Slaaby R, *et al.* X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution. *J Biol Chem*, 1998, **273** (21): 13047~13052
- 7 Malmqvist M. An affinity biosensor system for characterization of biomolecular interactions. *Biochemical Society Transactions*, 1999, **27** (2): 335~340
- 8 Arata Y, Hirabayashi J, Kasai K. Sugar binding properties of the two lectin domains of the tandem repeat-type galectin LEC-1 (N32) of *Caenorhabditis elegans*. Detailed analysis by an improved frontal affinity chromatography method. *J Biol Chem*, 2001, **276** (5): 3068~3077
- 9 Steinberg T H, Top K P O, Berggren K N, *et al.* Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots. *Proteomics*, 2001, **1** (7): 841~855
- 10 Takahashi N. Three-dimensional mapping of N-linked oligosaccharides using anion-exchange, hydrophobic and hydrophilic interaction modes of high-performance liquid chromatography. *J Chromatogr A*, 1996, **720** (1~2): 217~255
- 11 Kawasaki N, Itoh S, Ohta M, *et al.* Microanalysis of N-linked oligosaccharides in a glycoprotein by capillary liquid chromatography/mass spectrometry and liquid chromatography/tandem mass spectrometry. *Anal Biochem*, 2003, **316** (1): 15~22
- 12 David J. Matrix-assisted laser desorption/ionization mass spectrometry of carbohydrates and glycoconjugates. *International J Mass Spectrometry*, 2003, **226** (1): 1~35
- 13 Comisarow M B, Marshall A G. Fourier transform ion cyclotron resonance spectroscopy. *Chem Phys Lett*, 1974, **25** (2): 282~283
- 14 Henry K D, Williams E R, Wang B H, *et al.* Fourier-transform mass spectrometry of large molecules by electrospray ionization. *Proc Natl Acad Sci*, 1989, **86** (23): 9075~9078
- 15 Emmett M R, White F M, Hendrickson C L, *et al.* Application of micro-electrospray liquid chromatography techniques to FT-ICR MS to enable high-sensitivity biological analysis. *J American Society for Mass Spectrometry*, 1998, **9** (4): 333~340
- 16 Zubarev R A, Haselmann K F, Budnik B, *et al.* Towards an understanding of the mechanism of electron-capture dissociation: a historical perspective and modern ideas. *Eur J Mass Spectrom*, 2002, **8**: 337~349
- 17 Bruce E W, Christopher L, Hendrickson C L, *et al.* Improved ion extraction from a linear octopole ion trap: SIMION analysis and experimental demonstration. *J American Society for Mass Spectrometry*, 2002, **13** (11): 1304~1312
- 18 Mark R. Determination of post-translational modifications of proteins by high-sensitivity, high-resolution Fourier transform ion cyclotron resonance mass spectrometry. *J Chromatography A*, 2003, **1013** (1~2): 203~213

Advances in Analysis Techniques of Glycomics*

YANG Jun^{1,2)**}, CAI Shao-Xi¹⁾, ZOU Quan-Ming²⁾

¹⁾Key Laboratory of Biomechanics & Tissue Engineering, Chongqing University, Chongqing 400044, China;

²⁾Department of Clinical Microbiology, Third Military Medical University, Chongqing 400038, China)

Abstract In multicellular organisms protein glycosylation is a key post-translational modifications' event. Glycans of glycoproteins are not merely markers to characterize each cell type but are more aggressively involved in numerous biological phenomena, such as cell development, differentiation, implantation, morphogenesis, tumor metastasis and microbe infection. Glycomics is defined to analyse mostly the whole set of glycans of glycoproteins produced in a single organism. Several analysis techniques for separation and identification of glycoproteins and glycans are outlined, and the progress in these techniques is discussed.

Key words glycome/glycomics, glycosylation, glyco-catch, frontal affinity chromatography, two-dimensional gel electrophoresis, multidimensional liquid chromatography, mass spectrometry

*This work was supported by a grant from The state 863 High Technology R&D Project of China (2001AA21516).

**Corresponding author . Tel: 86-23-68753591, Fax: 86-23-68752315, E-mail: W8301991@263.net

Received: September 1, 2004 Accepted: September 28, 2004