

综述与专论

糖基转移酶的研究进展

王克夷

(中国科学院上海生物化学研究所, 上海 200031)

摘要 糖基转移酶参与了聚糖、糖苷和复合糖类中糖部分的生物合成, 具有高度的底物专一性, 已知序列的糖基转移酶没有明显的同源性, 但有相似的域结构. 有些糖基转移酶的基因可因几个碱基的替换而改变酶的专一性; 也因单个碱基的缺失而成为无酶活性的产物. 糖基转移酶和某些疾病密切相关.

关键词 糖基转移酶, 序列研究, 分子遗传

1 糖基转移酶的存在和专一性

糖基转移酶是广泛存在的一大类酶^[1], 它们决定了糖苷键的形成. 自然界中所有寡糖、多糖和各种复合糖类(糖蛋白、糖脂等)中的糖部分, 以及各类糖苷的生物合成均离不开糖基转移酶. 许多糖蛋白中的糖链或多或少地存在着微观不均一性, 即同一蛋白质中同残基上连接的糖链常有不同的组成或不同的结构^[2]. 产生糖链微观不均一有多种原因, 其中非常重要的一点, 是对糖基转移酶的了解甚少.

目前研究较多的是与糖蛋白中糖链合成相关的糖基转移酶, 本文主要介绍这些酶的研究进展. 这些糖基转移酶以及与糖链合成有关的糖苷水解酶, 几乎都是存在于高尔基体上, 但是有些(例如 β -1, 4 半乳糖基转移酶, α -2, 6 唾液酸转移酶, α -1, 3-N-乙酰半乳糖基转移酶)存在于反高尔基体腔(trans-Golgi cisternae)和反高尔基体网(trans-Golgi network)中, 有的酶, 例如 N-乙酰葡萄糖基转移酶(简称 GnT) I 和 α -甘露糖苷水解酶 II, 则在中央高尔基体(medial Golgi)上.

所有的糖基转移酶都有高度的底物专一性: 即对糖基的供体有专一性, 而且对糖基的接受体也有专一性. 大多数糖基的供体是不同

类型的核苷二磷酸激活的单糖, 不同的单糖用不同的核苷二磷酸活化. 但同样是葡萄糖基转移酶, 因合成的产物结构不同或合成的部位不同, 所用的供体上的核苷二磷酸也不同. 例如合成动物糖原时, 葡萄糖的供体是 UDP-Glc; 合成植物淀粉时, 则用 ADP-Glc; 同样是植物多糖, 纤维素的合成用的是 GDP-Glc. 糖基转移酶对接受体的专一性更为复杂. 在糖蛋白的 N-糖苷键连接的糖链的外周有 6 种不同的方式连接的 N-乙酰氨基葡萄糖(GlcNAc), 它们连接在不同甘露糖基(Man)的不同羟基上, 分别有六个不同的糖基转移酶负责它们的转移, 每个酶又对应了不同的接受体(图 1). 更有甚者, 6 个酶作用的先后次序也是非常严格的. 例如: 在 GnT I, GnT III 和 GnT IV 起作用前, 一定要 GnT I 先反应; 同样的, 只有 GnT II 的作用完成后, GnT V 和 GnT VI 才能催化有关的反应. 而在 GnT I 及 GnT II 这两个酶的进行有关催化反应之间还存在一些其它的糖链加工过程, 只有经过加工, 才能得到 GnT II 专一的底物——糖基接受体^[2]. 有的糖基转移酶的接受体还包含了糖蛋白中的某些肽段. 因为作为酶底物的寡糖本身分子较大, 在溶液中能以多种构象存在; 而通过肽链和底物寡糖的相互作用,

可以使某些和酶相适应的构象得以稳定. 对 GnT V 的底物结构的研究支持了这种假设^[3].

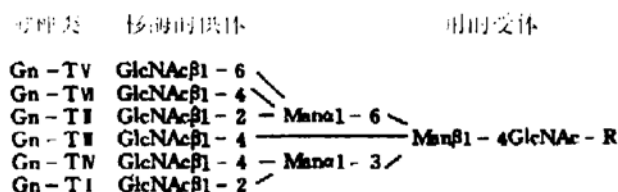


图 1 不同的糖基转移酶所催化的糖基转移反应

糖基转移酶的存在和专一性还因组织而异. 为此不同的组织或细胞株合成的糖蛋白中的糖链结构可以有较大的差别. 有的糖蛋白中, 蛋白部分的结构无明显不同, 但是糖链结构的差异却较大. γ -谷氨酰转肽酶就是一个例子^[4].

2 糖基转移酶的结构研究

核酸法测定蛋白质一级结构的迅速发展和广泛使用, 也为糖基转移酶的结构研究提供了有用的方法. 因为大多数糖基转移酶是膜结合蛋白, 量很少, 不足以用经典的蛋白质化学方法测定它们的整个结构. 目前认为仅是参与糖蛋白和糖脂中糖链合成的糖基转移酶就有上百种, 被克隆的哺乳动物来源的糖基转移酶只有十几种. 这些糖基转移酶的 cDNA 的序列已被

测定. 它们大多有相似的域结构^[2], 即: 一段短的 N 末端的肽段尾巴在细胞质中, 接着是 16—20 个残基组成的穿透膜的肽段, 然后是一段长度尚未确定的所谓的“颈”(“neck”)部, 余下的是很长的有催化活性的结构域, 后者在高尔基体的腔内. 肽链中的“颈”部对某些蛋白水解酶敏感, 原来的膜酶经水解后成了可溶性的酶. 后者比前者更为容易分离纯化, 为此多数糖基转移酶的结构研究, 是以水解得到的可溶性酶作起始材料, 得到了它们的 cDNA 后就能进一步测得酶的其它部分. 兔肝 GnT I 就是一个例子^[5]. 在制备此酶时, 仅能得到少量纯化的酶, 大量是不吸附在分离柱上不能进一步纯化的酶. 克隆了编码此酶的基因, 由 2.5kb cDNA 的序列测定得到了此酶的蛋白质一级结构, 由 447 个残基组成. 穿透膜的疏水肽段含有 25 个残基, 可形成 α 螺旋. “颈”部的酶解点在 45 位残基, 在这切点和催化结构域之间的肽段中有异常多的脯氨酸. 其它一些糖基转移酶也有这样高脯氨酸含量的“颈”部, 这提示了糖基转移酶的“颈”部在高尔基体腔内糖基化过程中有特定的作用, 例如有利于糖基化蛋白的移动. 表 1 列举了一些哺乳动物糖基转移酶的总体结构. 已知结构的糖基转移酶的蛋白质一级结构并不具有明显同源性, 但是有些相关的酶之间

表 1 克隆的哺乳动物糖基转移酶的域结构^[2]

酶	接受体	来源	域结构		残基数	
			N 端	跨膜区	C 端	3' 未转译碱基
α -2,6-唾液酰转移酶	Gal-GlcNAc-	大鼠	9	17	377	2.8kb
β -1,4-Gal 转移酶	GlcNAc-	啮齿类	11—24	20	335	2.6
		人	23	20	354	2.9
		牛	11—24	20	358	2.6—2.8
α -1,3-Gal 转移酶	Gal-GlcNAc-	啮齿类	41	19	334	2.1
		牛	6	16	346	2.4
α -1,3-Gal 转移酶 (血型 B 酶)	Gal-R ↑ α 1,2 Fuc	人	16	21	316	
α -1,3-GalNAc 转移酶 (血型 A 酶)	Gal-R ↑ α 1,2 Fuc	人	16	21	316	
GnT I	Man- α 1,3-Man-	人	4	5	418	1.1
		兔	4	25	418	1.1
		小鼠	4	25	418	1.1
α -1,3/4-L-Fuc 转移酶 (Lewis 血型)	GlcNAc- ↑ β 1,4/3 Gal	人	15	19	327	0.9

则有同源性。例如：合成人 AB 血型抗原的 A 酶和 B 酶只差 4 个残基；不同种属的 GnT I 的同源性高达 80%；合成 N-糖苷键连接的糖链中的 α -2, 6-唾液酰转移酶和合成 O-糖苷键连接的糖链时的 α -2, 3-唾液酰转移酶在一段 45 个残基的范围内有 62% 的同源性。

半乳糖基转移酶和唾液酰转移酶的 DNA 研究结果表明^[6,7]，它们的基因都很大，50kb 或更大，至少含有 6 个外显子，而且第一个外显子和其它几个外显子相距较远。在 GnT I 的基因中至少有两个启动子，一个在上游外显子的前面，另一个在 3' 末段的外显子的前面。其它一些糖基转移酶的基因中也有两个启动子^[6,7]。虽然这些研究还是初步的，但是这些酶基因中同时存在着多个启动子可能对酶的专一性以及它们酶活力的短暂的改变起了重要作用。

3 糖基转移酶的分子遗传学

糖链的结构没有特定的模版，糖链的合成是通过糖基转移酶将糖基由糖基的供体转移到接受体上。为此，糖链可认为是基因的次级产物，一个基因编码了一个糖基转移酶，一个糖基转移酶专一地催化一个糖苷键的合成；这样一条糖链的合成就需要一个多酶系统，也就对应了一个基因组^[1]。不同的个体，因遗传基因的不同，致使不同个体的有些糖链结构表现出差异。有的差异极微，但具有明显的生理意义。很多糖链的靠近还原端处常具有相同的内核结构，所不同的是糖链的非还原端的 1—2 个糖基。典型的例子是 ABH 血型-组织抗原以及有关的抗原^[8]：A 抗原和 B 抗原的不同仅在于非还原端的一个糖基，前者是 N-乙酰氨基半乳糖，后者是半乳糖。转移这两个糖基的酶分别称为 A 酶和 B 酶。这两个酶在遗传学上是由一个等位基因，A 基因和 B 基因编码的。另一个等位基因是 H 基因，它所编码的是一个无酶活性的产物，结果是 H 抗原上就比 A 抗原或 B 抗原少了一个糖基。这些等位基因及其表达产物在结构上有没有关系？这是一个很有趣的问题。

近几年中，对糖基转移酶的结构研究取得了一些结果，特别是对 ABH 血型物质的合成中的关键酶的研究，使人们能从分子生物学的角度更深刻地了解血型的分子遗传学基础。在用核酸法测定了 A 和 B 这两个酶的一级结构^[9]后，发现了它们的一级结构基本相同，所不同的只有 4 个氨基酸残基，176 位的氨基酸残基分别为精氨酸和甘氨酸，235 位分别为甘氨酸和丝氨酸，266 位分别为赖氨酸和甲硫氨酸，268 位分别为甘氨酸和丙氨酸；编码这两个酶的 DNA 中仅有 4 个碱基的更换：C→G，G→A，C→A，G→C。这就是说，血型为 A 型的个体和血型为 B 型的个体，它们的编码合成血型物质的关键酶的等位基因中仅有 4 个碱基的不同。如果 AB 酶的等位基因中 258 位核苷酸 G 的缺失，就使这一基因的表达产物前一部分的结构和 AB 酶相同，而从缺失的核苷酸以后的结构则不同于 AB 酶，成为一个没有酶活性的产物，结果是 H 基因的携带者，H 抗原不再被转化为其它的糖链结构。更有甚者，在 AB 基因中 4 个不同的碱基人为地改变其中某几个，就可能得到 16 种不同的 A 酶和 B 酶的嵌合体^[10]，所得的产物结果表现出不同的酶活性，有的仅有 A 酶活性，有的仅有 B 酶活性，另一些兼有两酶的活性，但两酶的活性强弱不一（表 2）。更多的糖基转移酶的结构测定必将揭

表 2 AB 酶的不同嵌合体

表现的酶活性	A	B	AB
各种嵌合体	AAAA	AABB	AABA
	AAAB	ABBB	ABAB (B 较弱)
	ABAA	BABB	ABBA
	BAAA	BBBB	BABA
	BAAB		BBAB (B 较弱)
	BBAA		BBBA

示其它糖类血型抗原（及其不同类别）的分子遗传基础。由此还得到了另一个启示，通过蛋白质工程的方法，完全能将对一种糖类专一的糖结合蛋白、糖苷水解酶的基因中的某些碱基

作为数不多的替代, 就能改变成为对另一种糖类专一的结合蛋白或酶. 有一例子是^[11]用基因工程方法将羊蹄甲凝集素中一段活性九肽换成兵豆中对应的八肽后, 所得表达产物的糖结合专一性, 就由原来的半乳糖变成了甘露糖, 但仍有微弱的半乳糖结合能力. 人、猿、非美洲猴 (旧世纪猴) 和其它的哺乳类动物, 包括美洲猴 (新世纪猴), 在糖链结构上有一个不同是 Gal ($\alpha 1-3$) Gal ($\beta 1-4$) GlcNAc-R 异质体的表达^[12]. 在后一类动物的细胞表面和分泌糖蛋白中 α Gal 异质体是高度表达的; 而人类和另一些高等动物中则不表达, 反而产生大量的对其专一的抗体 (占循环 IgG 的 1%). 产生这一差别的原因是 $\alpha 1-3$ 半乳糖基转移酶 ($\alpha 1-3$ GT) 活性的有无. 以牛和小鼠的 $\alpha 1-3$ GT 的 cDNA 为基础, 从人的基因库中也找到了相似的区域, 由此得到了人的 $\alpha 1-3$ GT 的基因^[13]. 比较了人和牛、小鼠的有关基因, 显示了 88% 和 81% 的相同. 进而以三者相同的保守的序列为基础设计引物, 得到了 13 种人和动物的 $\alpha 1-3$ GT 基因片段, 并比较了它们的 DNA 序列. 结果是在 820—824 和 902—906 这两段碱基中有差别: 有 $\alpha 1-3$ GT 酶活性的动物基因中, 这两段序列均相同; 而一些无此酶活性的动物基因中存在着不同. 随着不同的进化, 出现了两种情况: 一是新世纪的猴成为老世纪的猴, 与此同时, 仅出现了碱基的突变; 另一种进化是变成猿和人, 在这过程中, 除了有碱基的突变外, 还有碱基的缺失, 在人和黑猩猩的 $\alpha 1-3$ GT 基因中, 则丢失了 882 位碱基 C 和 904 位碱基 G. 如果在人体内这一酶的活性重新产生, 就会引起疾病, 例如自身免疫疾病.

4 糖基转移酶和疾病

$\alpha 1-3$ GT 的活性在人体内的重现会引发自身免疫反应. 相反, 如果人体内 $\beta 1-4$ 半乳糖基转移酶的活性下降, 致使正常的免疫球蛋白 G (IgG) 分子中的糖链半乳糖含量降低. IgG 是一种糖蛋白, 它的糖含量不高, 约 5%, 最主要的糖链是在 CH-2 结构域中的 Asn-297 上的二分

支复杂型的糖链, 即使是这样比较简单的糖链也至少有 30 种结构类似仅有极小差异的不同形式, 每种结构被称为一种糖型 (glycoform). 这种半乳糖缺少的异质体对正常机体而言也是异物, 因此, 它能诱导抗体的产生, 从而也会出现自身免疫, 其症状是类风湿^[14].

糖基转移酶的表达和细胞的周期密切相关. 在细胞分化阶段, 许多糖基转移酶的基因是表达的, 为此出现了一系列糖类异质体作为分化抗原; 一旦发育成熟, 在细胞表面出现了另一些糖链异质体. 如果糖基转移酶在成熟细胞中活性很高, 就会产生癌变, 同时出现了早期的分化抗原. 在 N-糖苷键连接的糖链中多聚 N-乙酰乳糖胺链的出现被看成是肿瘤的重要标志, 这类糖链可以降低细胞和基质间的粘着, 有利于癌细胞的进一步的入侵. 嬰幼田鼠肾细胞 (BHK) 被 Rous 肉瘤病毒感染后, GnT V 的活性升高 2.5 倍; 一些有转移倾向的肿瘤细胞和非转移的细胞相比, GnT V 的活性高出 3—10 倍^[15]. 岩藻糖基转移酶的活性在肿瘤细胞中也明显升高.

糖基转移酶的研究还处于起步阶段, 仅知道一些酶的结构还是远远不够的. 更为重要的是了解这些糖基转移酶在多酶系统中是如何协调的, 以及这些酶的表达是受那些因素调控的. 这是糖基转移酶今后研究的方向.

参 考 文 献

- 1 Beyer T A, Sadler J E, Rearick J I *et al.* *Advances Enzymol*, 1981; **52**: 23
- 2 Schachter H. *Trends Glycosci Glycotechnol*, 1992; **4**: 241
- 3 Lindh I, Hindsgaul O. *J Am Chem Soc*, 1991; **113**: 216
- 4 Kobata A. In: Ginsberg V, Robbins P W eds, *Biology of Carbohydrate*, New York: John Wiley & Sons, 1984; **2**: 87
- 5 Sarkar M, Hull E, Nishikawa Y *et al.* *Proc Natl Acad Sci USA*, 1991; **88**: 234
- 6 Hollis G F, Douglas J G, Shaper N L *et al.* *Biochem Biophys Res Commun*, 1989; **162**: 1069
- 7 Wang X C, O' Hanlon T P, Young R F *et al.* *Glycobiology*, 1990; **1**: 25
- 8 Clausen H, Hakomori S. *Vox Sang*, 1989; **56**: 1

- 9 Yamamoto F, Clausen H, White T *et al.* Nature, 1990; **345**: 229
- 10 Yamamoto H, Hakomori S. J Biol Chem, 1990; **265**: 19257
- 11 Yamamoto K, Konami Y, Osawa T *et al.* J Biochem, 1992; **111**: 87
- 12 Galili U, Thall A, Macher B A. Trends Glycosci Glycotechnol, 1990; **2**: 304
- 13 Galili U, Swanson K. Proc Natl Acad Sci USA, 1991; **88**: 7401
- 14 Furukawa K. Trends Glycosci Glycotechnol, 1990; **2**: 243
- 15 Yamashita K, Tachibana Y, Ohkura T *et al.* J Biol Chem, 1985; **260**: 3963

密码结构域及其功能表现

劳为德

(中国科学院生物物理研究所, 北京 100101)

摘要 能组织成超级结构的各种序列组块,因其具有特定一级序列、或者具有某种卷曲的螺旋构象和某种非 β 螺旋结构,可以作为有特异功能的结构域、被特异的结合蛋白质识别和结合,因而可称为密码结构域。密码结构域作为特异的分子相互作用和过程的遗传指令,参与细胞周期的各种事件乃至发育和分化过程中基因有区别的表达。

关键词 密码结构域, 染色质超级结构, 串列重复序列, 着丝粒, 端粒, 核基质, 同源异型基因, 发育与分化

从事分子生物学研究的人大都在习惯上把编码特性与 DNA 序列联系起来,看看该序列是否有某种读出框架可转译出某种氨基酸序列。DNA 这种编码能力约在 30 年前(1966 年)已得到认识。基本的遗传密码广泛存在于迄今研究过的古今生物种类而且高度保守。但现在越来越明显,DNA 具有比此要大得多的编码潜力来完成其它重要的任务。除了编码特异的 RNA(如 rRNA, snRNA 和 tRNA)分子之外,DNA 特异结构型式和序列型式本身也呈现出调节密码和表达密码的遗传活性。比如,染色质密码规定着核小体中 8 体的组蛋白复合物的相位调整;转译框架密码决定蛋白质合成时三联子的正确辨认;剪接过程中套环密码则组织 RNA 新生链的单链与蛋白质的相互作用;剪接密码使相继的 5' 和 3' 拼接点的正确剪接。这些 DNA 密码,大多数毫无例外地是基于 DNA 的一级序列本身;但看来也包涵于相应的超级结构的特异性特征,因为各种 DNA 密码

都是在间期和细胞分裂时在细胞核内起重要作用的特定分子相互作用(或过程)的遗传指令。

1 串列重复 DNA 结构的编码能力

大多数真核生物基因组都存在具有不同总长度的串列重复序列结构,其特征是两个以上的序列单位以头对尾的形式重复排列,其总长度在两个与上千个重复之间。这种串列的组块式结构或是位于特异的染色体位点,如着丝粒和端粒的染色体区域;或是弥散于基因组各处。各种组块结构几乎被组织成某种重复的座位特异的超级结构^[1]。

1.1 串列重复 DNA 序列之座位特异的超级结构

虽然串列重复的 DNA 序列结构都可在每个染色体中找到,但它们的总体序列组成(超级结构)都是座位特异的。这种特异性或是基于各种较小的重复序列组块在特异的座位上的扩增,或是基于基因组其它部分的非同源序列