天然产物碳糖基转移酶研究进展

刘 蕊, 陈大伟, 王天宇, 戴均贵*, 卢 多*

(中国医学科学院、北京协和医学院药物研究所, 天然药物活性物质与功能国家重点实验室, 北京 100050)

摘要: *C*-糖基转移酶 (*C*-glycosyltransferases, CGTs) 是生成*C*-糖苷的天然生物催化剂, 能够将活性核苷酸糖供体上的糖基转移至受体底物并形成*C*-*C*糖苷键。*C*-糖苷具有多种药理活性, 如抗氧化、抗炎、抗菌、抗病毒和抗肿瘤活性等, 而且因其糖苷键更耐化学和生物降解, 与*O*-糖苷相比表现出更好的药理性质, 所以CGT在药物研发中有着不可忽视的重要潜力。尽管针对CGT的研究已有很大进展, 但在已鉴定蛋白的数量以及催化反应机制等方面, 与已有广泛研究的*O*-糖基转移酶 (OGT) 相比还存在很大差距。CGT相关晶体结构对糖供体的识别进行了较好阐释, 但有限的受体结合结构限制了对其催化机制的全面理解, 进而影响CGT在生物工程医药领域的应用。本文总结了已被鉴定的天然产物CGT的底物类型、蛋白结构、催化机制以及相关的工程改造和应用, 希望为未来的研究和工作提供一些思路。

关键词: *C*-糖基转移酶; 受体; 结构; 功能; 催化机制

中图分类号: Q946.5 **文献标识码**: A **文章编号**: 0513-4870(2025)07-2186-12

Research progress of *C*-glycosyltransferases for natural products

LIU Rui, CHEN Da-wei, WANG Tian-yu, DAI Jun-gui*, LU Duo*

(*State Key Laboratory of Active Substances and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China*)

Abstract: *C*-Glycosyltransferases (CGTs) are natural biocatalyst responsible for the formation of *C*-glycosides, transferring sugar moieties from activated nucleotide sugar donors to acceptor substrates and establishing *C*-*C* glycosidic bonds. *C*-glycosides show diverse pharmacological activities, such as antioxidant, anti-inflammatory, antibacterial, antiviral, and antitumor activities, and comparing with *O*-glycosides, *C*-glycosides exhibit better pharmacological properties due to their glycosidic bonds being more resistant to chemical and biological degradation. Thus, CGTs have shown considerable potential in drug research and development. Although research on CGTs has made significant progress, with respect to the number of identified CCTs and the understanding of catalytic mechanism, there is still a large distance between CGTs and extensively studied *O*-glycosyltransferases (OGTs). The reported crystal structures of CGTs have provided a good elucidation of the recognition of sugar donors. However, the limited acceptor-bound structure hinders the comprehensive understanding of their catalytic mechanisms, thereby affecting the application in the fields of bioengineering and pharmaceuticals. Here we summarize the information of identified CGTs for natural product regarding substrate types, protein structures, proposed catalytic mechanisms, and CGT-related engineering and applications, aiming to bring out some new ideas for future research and development.

Key words: *C*-glycosyltransferase; acceptor; structure; function; catalytic mechanism

收稿日期: 2025-04-11; 修回日期: 2025-05-16.

基金项目: 中国医学科学院协同创新基金 (2023-I2M-2-006).

*通讯作者 Tel: 86-10-63039979, E-mail: luduo@imm.ac.cn;

Tel: 86-10-63165195, E-mail: jgdai@imm.ac.cn

DOI: 10.16438/j.0513-4870.2025-0411

糖基是许多天然产物的重要组成部分, 一方面作为重要的活性基团, 另一方面可以提高相关分子的稳定性、溶解度和生物利用度等药理特性[1-3]。传统的糖苷生产方法, 如直接从动植物或微生物中提取和纯化, 存在技术要求高、成本高、产量低等问题, 而化学合成则常受到收率低、选择性差以及官能团保护和脱保护步骤繁琐的限制。相比之下, 糖基转移酶 (glycosyltransferase, GT) 在催化糖苷生成过程中效率高、且表现出较好的区域选择性和立体化学选择性, 已成为一种极具前景的生物合成工具[1,4-6]。

目前, 在碳水化合物活性酶数据库 (CAZymes, https://www.cazy.org/) 中根据序列相似性将糖基转移酶分为138个家族。其中, 数量最多的成员是尿苷二磷酸糖基转移酶 (UDP-glycosyltransferase, UGT), 它属于糖基转移酶超家族的1家族 (GT1), 可催化糖基从UDP活化的糖供体 (例如UDP-葡萄糖) 转移到多种受体[7]。糖基转移酶的糖基受体不仅包括糖类、蛋白质、脂质等大分子化合物, 还包含萜类、黄酮类和类固醇等小分子天然产物[8-10]。根据形成糖苷键的不同, 糖基转移酶通常可分为*O*-糖基转移酶 (OGT)、*C*-糖基转移酶 (CGT)、*N*-糖基转移酶 (NGT) 和*S*-糖基转移酶 (SGT), 分别催化形成*O*-、*C*-、*N*-和*S*-糖苷。目前, OGT已得到充分的研究, 其结构和催化机制都较为清晰, NGT和SGT虽然相对较少, 但也提出了自发去质子化的催化机制。糖基转移酶在医药、农业、食品工业等多方面已经进入应用, 破译结构-功能关系和催化机制后, 有助于通过蛋白质工程, 提升应用空间甚至拓展新的应用领域[7,11]。

O-糖苷在已发现的糖苷中占比最大, 其次是*C*-糖苷。然而相比于*O*-糖苷, *C*-糖苷具有对化学或酶催化水解更好的稳定性, 且表现出多种生物活性, 包括: 抗氧化、抗炎、抗菌、抗病毒、抗肿瘤、抗糖尿病活性和心脑血管保护作用等等。例如大豆苷元8-*C*-*β*-*D*-葡萄糖苷 (葛根素) 已经被开发成辅助治疗心血管疾病的临床药物。一些*C*-糖苷作为重要的植物抗毒素, 可以保护植物免受紫外线辐射、干燥环境、昆虫和微生物等外部刺激影响, 可能在农业生产方面提供帮助[4,12,13]。在细菌中, *C*-糖基化修饰也存在于多种抗生素中, 这进一步拓宽了*C*-糖苷成为候选药物的可能性[12,14]。

CGT是生物体中催化*C*-糖基化反应的关键生物合成酶。目前已鉴定并表征的CGT大多来源于植物, 且所有已知的植物CGT都是UGT, 少量CGT来源于细菌、真菌和昆虫[15,16]。随着越来越多的CGT被挖掘, 其底物种类和应用范围也在不断扩展, 从传统的类黄酮类化合物发展到蒽醌类、苯乙酮类、氧杂蒽酮类、香豆素类等多种结构的天然产物[4,17]。通过对CGT进行合理设计, 还有望在已有基础上进一步拓宽其底物谱、改变其催化特性, 开发出更多种类的糖基化修饰产物, 丰富糖苷化合物库, 拓展从中获得候选药物的可能性[18-20]。

尽管CGT在药物研发中展现出巨大潜力, 但其催化机制、底物选择性和结构特征尚未完全阐明。目前, 已有部分CGT的晶体结构被解析, 但只在一个结构中含有受体底物的结合[21]。虽然现有结构揭示了CGT的典型结构类型和某些关键催化残基的作用, 但还缺乏足够的实验证据对其催化机制进行完全揭示。本文将对已报道的CGT的底物类型、蛋白结构和推测的酶催化机制进行总结, 分析CGT的应用前景和目前的局限性, 希望为后续研究提供一些新思路。

1　CGT的底物类型

UDP-葡萄糖 (UDP-Glc) 在CGT中是最常见的糖基供体, 此外根据报道, UDP-阿拉伯糖 (UDP-Ara)、UDP-半乳糖 (UDP-Gal)、UDP-木糖 (Xyl)、UDP-葡萄糖醛酸 (UDP-GlcA) 和UDP-*N*-乙酰葡萄糖胺 (UDP-GlcNAc) 也可以作为一些CGT的底物。与植物CGT相比, 细菌CGT使用多种核苷酸糖基供体, 如UDP-Glc、TDP-橄榄糖 (TDP-olivose)、TDP-碳霉糖 (TDP-mycarose)、NDP-若丹糖 (NDP-rhodinose) 等[15]。

在糖基受体方面, 目前报道的几乎所有CGT都有一个共同的催化特征, 即将糖基转移至受体酚羟基的邻位碳上, 但它们所催化受体的完整化学骨架不尽相同。当前已知的植物CGT的受体主要是类黄酮, 其他结构包括苯乙酮类、二苯甲酮类、香豆素类等[4]。细菌、真菌和昆虫CGT则能催化糖基转移到结构更加多样、更复杂的受体上, 例如蒽醌、苯并异色满醌 (BIQs)、蒽并吡喃、苯并蒽醌、大环内酯、吡喃酮等[16,22-27]。下面将按照化学骨架对CGT作用的受体类型进行总结。

1.1　类黄酮 (C6-C3-C6结构)　类黄酮是一类以2-苯基色原酮为基本母核, 具有C6-C3-C6基本骨架的化合物。已报道的类黄酮底物覆盖其大多数亚型。

许多CGT能够以2-羟基黄烷酮为受体底物, 属于I型CGT[4]。尽管2-羟基黄烷酮可以在开环和闭环两种结构之间快速转换, 但多数CGT只能作用于开环形式, 在6位和/或8位生成单碳糖苷或者双碳糖苷 (**1**, 图1A)。只有来自玉米 (*Zea mays*) 的UGT708A6以2-羟基黄烷酮的闭环形式为底物 (**2**, 图1A), 能够将2-羟基柚皮素转化为异牡荆素 (芹菜素6-*C*-葡萄糖苷) 和牡荆素 (芹菜素8-*C*-葡萄糖苷)[15,28-30]。鉴于闭环形式下, 6位与8位的化学环境存在一定差异, 该CGT如何在两个位点之间进行选择尚不清楚。此外, 一些具有底

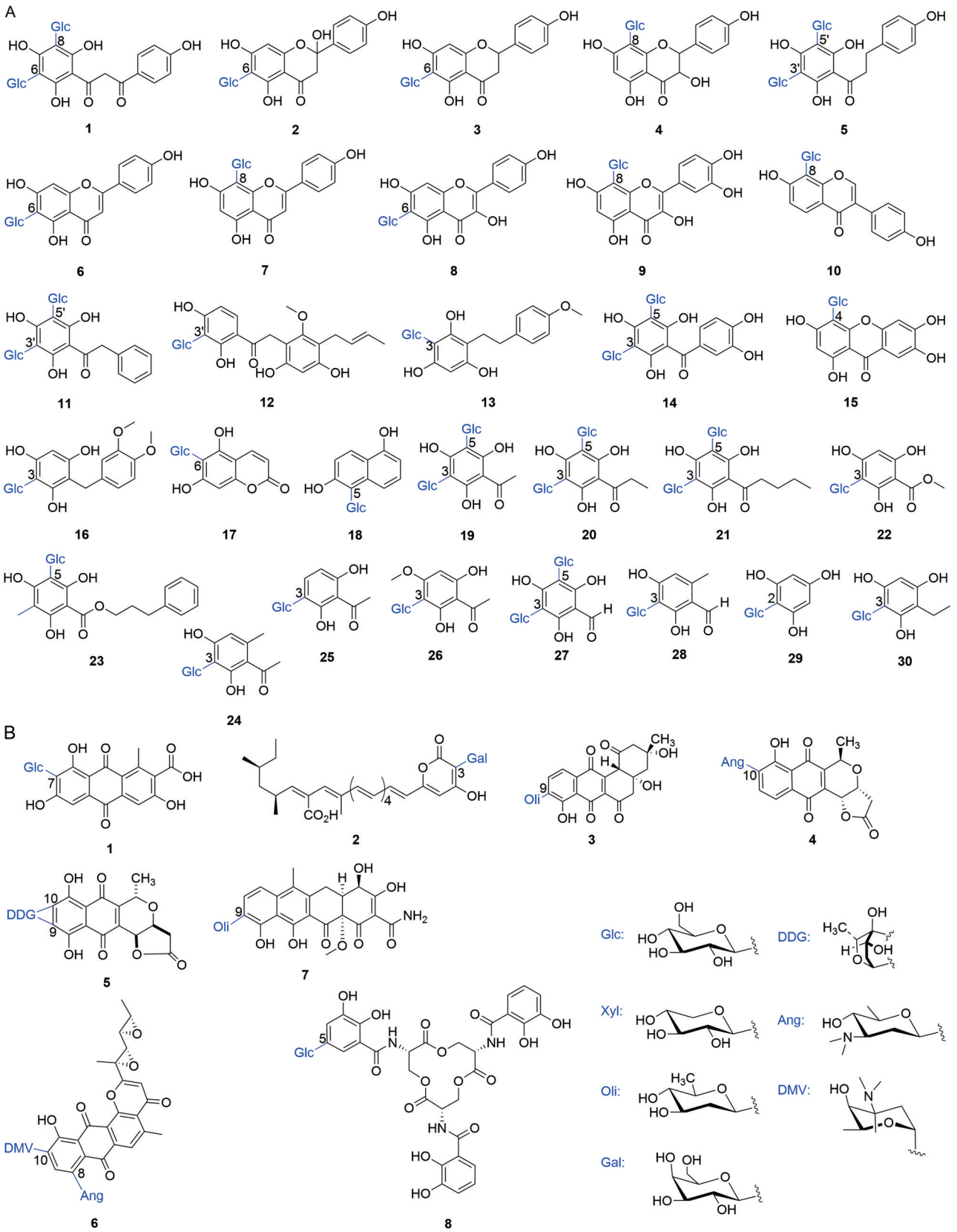


Figure 1 Structural diversity of sugar-acceptors for CGTs. A: Representatives of different type of acceptors for plant CGTs; B: Representatives of different type of acceptors for CGTs from other organisms. The structures depicted with two glucosyl groups indicate that in addition to general mono-*C*-glycoside products, di-*C*-glycosides can be produced by certain CGTs. CGTs: *C*-Glycosyltransferases; Glc: Glucose; Xyl: Xylose; Oli: Olivose; Gal: Galactose; Ang: Angolosamine; DDG: 4-Keto-2,6-dideoxyglucose; DMV: *N*,*N*-Dimethylvancosamine

物混杂性的CGT也能够作用于黄烷酮 (**3**, 图1A) 或黄烷酮醇 (**4**, 图1A), 例如来自金莲花 (*Trollius chinensis*) 的TcCGT1可以生成6位或8位单糖苷[5]。

鉴于二氢查尔酮与2-羟基黄烷酮的开环形式非常相似 (**5**, 图1A), 大多数2-羟基黄烷酮CGT也能够作用于二氢查尔酮结构 (如根皮素)。当以根皮素为受体时, CGT可以将糖基转移到根皮素的3′位 (相当于黄酮骨架的6位) 生成nothofagin, 或者以nothofagin为底物继续生成nothofagin-5′-糖苷[21,28,30-34]。

黄酮与异黄酮可以被少数II型CGT修饰[4]。区别于催化2-羟基黄烷酮的I型CGT, 作用于黄酮或异黄酮的CGT通常具有高度的区域特异性, 例如来自三花龙胆 (*Gentiana triflora*) 的GtUF6CGT1[35]和来自芥末 (*Eutrema japonicum*) 的WjGTs[36]只能将葡萄糖基转移到黄酮骨架的6位 (**6**, 图1A), 而TcCGT1[5]则特异性地催化黄酮的8-*C*-糖基化 (**7**, 图1A)。来自葛根 (*Pueraria lobata*) 的PlUGT43[37]和PlCGT[38]可以特异性地将糖基转移到异黄酮的8位 (**10**, 图1A)。WjGTs和TcCGT1还对黄酮醇表现出CGT活性, 分别催化生成黄酮醇6-*C*-糖苷 (**8**, 图1A) 和黄酮醇8-*C*-糖苷 (**9**, 图1A)。

可以看到, 不同的CGT表现出对不同位置的偏好, 这种差异可能源于不同CGT反应腔结构的差异, 也可能源于化合物自身结构与化学特性的不同导致其在CGT反应腔中空间取向的变化。

1.2　C6-C2-C6结构　研究发现当二氢查尔酮两个六元环之间的距离缩短到两个原子时, 也可以得到有效*C*-糖基修饰。不同CGT对该类底物的催化类型不同, 例如: PlCGT可以催化3′位单糖基修饰[38], 而GgCGT可以完成3′,5′-双糖基修饰 (**11**, 图1A)[21]。此外, 1′位无酰基取代的化合物也可以作为一些CGT的催化底物 (**13**, 图1A), 如来自芦荟 (*Aloe barbadensis*) 的AbCGT[39]。

1.3　C6-C1-C6结构　将两个六元环之间的距离进一步缩短得到C6-C1-C6化合物, 如二苯甲酮和氧杂蒽酮 (xanthone)。从芒果 (*Mangifera indica*) 中鉴定出的CGT MiCGT[40]和MiCGTb[41]可以催化二苯甲酮桑橙素 (maclurin) 生成桑橙素3-*C*-葡萄糖苷, AbCGT[39]还可以继续生成桑橙素3,5-di-*C*-葡萄糖苷 (**14**, 图1A)。将二苯甲酮非反应端的苯环替换成吲哚、喹啉等杂环仍可以反应, 但转化率有所降低。从贯叶连翘 (*Hypericum perforatum*) 中鉴定出两种催化闭环氧杂蒽酮糖基化反应的CGT: N4CGT1和N4CGT2[42]。它们能够将糖基转移至norathyriol的C4位, 生成异芒果苷 (isomangiferin, **15**, 图1A)。除二苯甲酮和氧杂蒽酮外, AbCGT还可以作用于C1原子无酰基取代的结构 (**16**, 图1A)[39]。

1.4　双环结构　香豆素受体核心结构是一个苯并*α*-吡喃酮环, 通常在5,7位有二羟基取代。来自桑树 (*Morus alba*) 的MaCGT可以作用于香豆素受体, 能够高效和区域特异性地合成结构多样化的香豆素6-*C*-葡萄糖苷 (**17**, 图1A)[43]。另外, 具有底物混杂性的TcCGT1还可以作用于羟基萘结构, 将糖基转移至羟基萘结构的5位 (**18**, 图1A)[5]。

1.5　单环结构　许多CGT能够作用的最小受体是只有一个芳香环的三羟基苯乙酮类结构 (**19**, 图1A), 以三羟基苯乙酮为基础在末端具有更长脂肪链的化合物也可以成为CGT的底物 (**20**、**21**, 图1A), 包括一些苯甲酸酯结构 (**22**、**23**, 图1A)。此外, 没有酰基的底物依然可以得到*C*-糖基化修饰 (**29**、**30**, 图1A)[21,39,40]。

利用芒果和芦荟来源的CGT (MiCGT和AbCGT), 研究人员考察了三羟基苯乙酮上羟基数目与取代位置对反应的影响。综合两项工作的实验结果发现, 苯环上只有一个羟基时, 只能发生*O*-糖基化反应; 3,5位二羟基取代化合物也只能发生*O*-糖基修饰, 2,4位与2,6位二羟基取代则既可以发生单-*C*-糖基修饰也可以发生*O*-糖基修饰 (**24**、**25**, 图1A); 2,4,6位三羟基取代则可能生成双-*C*-糖基修饰 (**19**, 图1A), 而4位换成甲氧基时只能生成单-*C*-糖苷产物 (**26**, 图1A)[39,40]。由此可以明显地看出, 苯环上羟基的数量、位置, 以及取代基团种类对*C*-糖基化修饰都能够产生重大影响。

在GgCGT的研究中发现, 生成双-*C*-糖苷产物似乎需要三羟基苯丙酮基本结构单元, 三羟基苯乙酮虽然能生成双-*C*-糖苷, 但转化率显著降低, 而三羟基苯甲酸酯则只能得到单-*C*-糖基化修饰[21]。

1.6　蒽醌类　来自美洲胭脂虫 (*Dactylopius coccus*) 的DcUGT2可以作用于蒽醌。蒽醌是由三个苯环线性稠合并在中央环的9,10位引入两个羰基形成的醌式结构。DcUGT2将葡萄糖基转移至具有蒽醌骨架的前体黄胭脂酸 (flavokermesic acid, FK) 和胭脂虫酸 (kermesic acid, KA) 的7位 (与黄酮骨架的6位相似), 生成胭脂红酸 (carminic acid, CA, **1**, 图1B)[17]。

1.7　多烯吡喃酮　Epipyrone A是从真菌黑附球菌 (*Epicoccum nigrum*) 中分离出来的水溶性抗真菌分子, 是一种独特的*C*-半乳糖基化4-羟基-2-吡喃酮多烯 (**2**, 图1B)。近期从此菌中鉴定出的EpiB能够将半乳糖转移至Epipyrone A前体吡喃酮环的3位, 且研究显示多烯链上的羧基对于糖基化是必不可少的[16]。

1.8　芳香族聚酮类　一些细菌中鉴定的CGT作用于芳香族聚酮类。来自链霉菌 *Streptomyces fradiae* Tü2717的UrdGT2催化*D*-olivose (2,6-二脱氧-*D*-葡萄糖) 从供体dTDP-*D*-olivose转移到苯并蒽醌衍生的聚酮受体的C9原子上, 生成urdamycin前体 (**3**, 图1B)[26,44]。另外两

个已知化合物 medermycin (**4**, 图 1B)[25,45]和 granaticin (**5**, 图 1B)[27]都是芳香族聚酮苯并异色满醌类的 *C*-糖苷, 研究人员在它们的生物合成簇里鉴定出了可能的 CGT 基因。同样, 由链霉菌 *Streptomyces griseoruber* 产生的 hedamycin[24](**6**, 图 1B) 也是一种芳香族聚酮类化合物, 含有一个平面的 4-*H*-蒽并(1,2-*b*)吡喃生色团, 且具有双-*C*-糖苷结构, 研究人员推测 HedJ 和 HedL 是合成 hedamycin 的两个候选 CGT。来自链霉菌的 SsfS6[22] 可以催化四环素样聚酮 C-9 位的 *D*-olivose 糖基化, 生成具有抗肿瘤活性的 SF2575 (**7**, 图 1B)。

1.9 大环内酯类 肠杆菌素 (enterobactin, Ent) 是肠道细菌 (例如大肠杆菌和肠道沙门氏菌) 的典型铁载体, 是由 *N*-(2,3-二羟基苯甲酰基)丝氨酸 (DHB-Ser) 组成的三聚体大环内酯。IroB 能够以其为受体在三个 2,3-二羟基苯单元的 5 位生成单糖基-*C*-肠杆菌素 (MGE, **8**, 图 1B)、双糖基-*C*-肠杆菌素 (DGE) 和三糖基-*C*-肠杆菌素 (TGE)[23]。

综合各种类型的结构可看到, 底物结构上化学性质的变化对反应有明显影响, 而不同 CGT 对反应区域及类型的确定至关重要, 比如上述 WjGTs 和 TcCGT1 对黄酮修饰位点的差异, 又比如 GgCGT 不能催化三羟基苯甲醛生成双-*C*-糖苷, 而 AbCGT 可以 (**27**, 图 1A)。

2 CGT 的结构

2.1 总体结构 已报道的糖基转移酶表现出两种主要拓扑结构: GT-A 和 GT-B 折叠, 此外也有少量 GT-C 等变体折叠的报道[8]。迄今为止, 所有已表征的天然产物 CGT 按序列同源性分类均属于 GT1 家族, 而 GT1 家族的糖基转移酶结构均采用 GT-B 折叠[46]。该折叠由交替的 β 链和 α 螺旋构成两个在空间上对向排列的 Rossmann 样结构域, 分别称为 C 端结构域 (CTD) 和 N 端结构域 (NTD)。两者之间形成一个深裂隙, 催化活性位点即位于此裂隙中。CGT 的 CTD 包含一个相对保守的供体结合基序, 负责核苷糖供体的识别和结合, 而 NTD 则负责与受体互作, 表现出更大的拓扑可塑性, 以适应结构多样的受体类型。两个结构域通过连接环疏松连接, 可能方便二者相对空间位置的移动, 而蛋白 C 末端序列形成一个长断续 α 螺旋, 跨接在两个结构域外侧, 给整个蛋白结构提供一定的稳定性 (图 2A, E)[5,7,8,47-49]。

当前 12 个天然产物 CGT 的结构得到解析: 10 个植物 CGT (包括 TcCGT1、GgCGT、UGT708C1、SbCGTa、SbCGTb、LpCGTa、LpCGTb、ZmCGTa、MiCGT 和 GcCGT) 和 2 个细菌 CGT (UrdGT2 和 SsfS6)。另外, 还有一个由细菌 OGT 改造成的变体 CGT (LanGT2S8Ac) 的结构[50]。在所有结构中, 11 个包含核苷酸结合的信息,

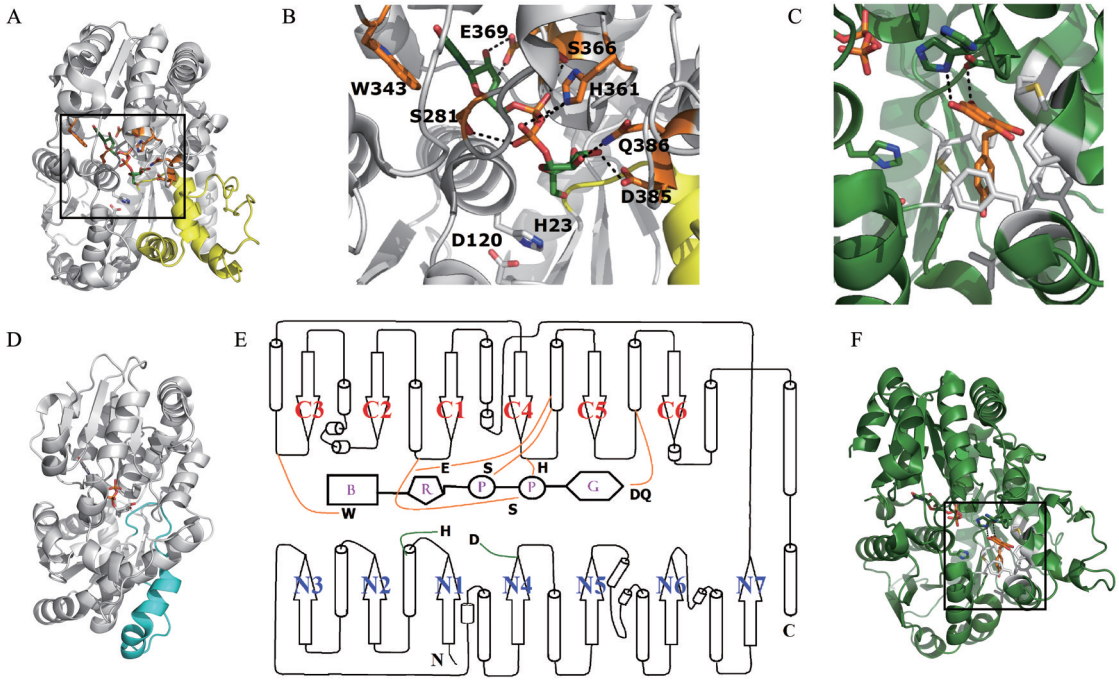


Figure 2 Key structural features of CGTs. A: Crystal structure of MiCGT as a representative of plant CGTs. The section between Nβ5 (the fifth β-strand of the N-terminal domain) and Nβ6 is colored in yellow, which is the most different part between plant and bacterial CGTs; B: Enlarged region of MiCGT showing key residues for sugar-donor UDP-Glc recognition; C: Enlarged region of GgCGT showing key residues for sugar-acceptor binding; D: Crystal structure of LanGT2S8Ac as a representative of bacterial CGTs. The section between Nβ5 and Nβ6 is colored in cyan to make comparison with plant CGTs; E: Schematic view of MiCGT structure with key residues for sugar-donor recognition and catalytic activity depicted; F: Overview of GgCGT structure; UDP-Glc: Uridine diphosphate glucose

6个有核苷糖基供体结合, 1个有受体底物结合, 1个有修饰后产物分子结合 (表1[5,18,21,22,50-55])。从结构的角度看CGT可以分为两类, 且可以按其生物来源区分: 细菌来源和植物来源。它们之间CTD相似度较高, 而NTD差异非常明显。植物来源CGT的NTD在第5、6β链之间有更复杂的结构 (图2A, D)。

Table 1　Existing CGT structures

CGT	Apo/complex	PDB ID	Reference
UrdGT2	Apo	2P6P	[51]
SsfS6	Apo	4FZR	[22]
	SsfS6/TDP	4G2T	
LanGT2S8Ac	Apo	4RIG	[50]
	LanGT2S8Ac/TDP	4RII	
	LanGT2S8Ac/TDP-carba-*D*-olivose	4RIH	
TcCGT1	TcCGT1/UDP	6JTD	[5]
GgCGT	GgCGT/UDP-Glc	6L5P	[21]
	GgCGT/UDP-Gal	6L5Q	
	GgCGT/UDP/phloretin	6L5S/6L5R	
	GgCGT/UDP/nothofagin	6L7H	
UGT708C1	Apo	6LLG	[52]
	UGT708C1/UDP	6LLW	
	UGT708C1/UDP-Glc	6LLZ	
SbCGTa	SbCGTa/UDP	6LG0	[53]
SbCGTb	SbCGTb/UDP-Glc	6LFZ	
LpCGTa	LpCGTa/UDP	6LG1	
LpCGTb	Apo	6LFN	
ZmCGTa	ZmCGTa/UDP	6LF6	
MiCGT	WT/UDP-Glc	7VA8	[54, 55]
	VFAH/UDP	7VAA	
	MiCGT-QDP/UDP	8XFH	
GcCGT	Apo	8IE4	[18]
	GcCGT/UDP	8W7Y	

2.2　与糖基供体的结合　可能由于所识别的糖基供体化学结构相似, 所有植物UGT (包括OGT和CGT)的CTD通常高度保守。最保守的区域被称为植物次生产物糖基转移酶盒 (PSPG盒), 由44个氨基酸组成, 从CTD中C端Rossmann样结构域第3β链 (Cβ3) 的C端连接环延伸到第5β链 (Cβ5) C端α螺旋的N端, 主要决定了对UDP糖供体的识别与结合 (图2B, E)。其中, 第3β链C端连接环上的色氨酸与核苷的碱基形成π-π堆叠; 第4与第5β链之间的α螺旋中部贡献出一个谷氨酸 (或天冬氨酸) 残基与核糖的2′-及3′-羟基结合; 同一个α螺旋N端的一个丝氨酸或苏氨酸识别紧邻核糖的第一个磷酸基团; 第4β链C端连接环上一个高度保守的组氨酸与第二个磷酸基团结合; 在组氨酸的对侧, 第1β链C端连接环上也有一个相对保守的氨基酸与该磷酸基团结合 (经常是丝氨酸/苏氨酸/天冬酰胺), 与组氨酸一起夹住该磷酸基团。这样糖供体的UDP部分各个基团都通过保守氨基酸的结合而得到稳定, 特别是与糖基直接连接的磷酸基团。PSPG盒的最后两个氨基酸, 对糖基的识别和特异性至关重要。在UGT708C1晶体结构中, 第5β链C端α螺旋的N端酸性残基Asp382与葡萄糖部分的O3′和O4′原子形成氢键, D382A突变导致催化活性丧失。PSPG基序的最后一个残基被认为是决定糖供体特异性的关键, 例如谷氨酰胺通常决定葡萄糖特异性 (如UGT708C1中的Gln383), 而组氨酸用于鼠李糖特异性的识别 (如UGT89C1中的His357)。在CGT和OGT中糖供体的结合位置相对保守且表现出相似偏好, 即: UDP-Glc > UDP-Xyl > UDP-Gal > UDP-Ara[5,7,21,48,52,56]。

相比之下, 细菌CGT由于利用的核苷酸糖供体更加多样, 似乎没有类似于植物CGT中PSPG盒的保守基序。在现有结构中, 对核苷二磷酸 (NDP) 部分的识别与植物CGT相比存在一定保守性, 特别是对两个磷酸基团的识别, 但是也有一些因为核苷的不同而特异的识别模式。例如SsfS6对dTDP的识别, 其特异性部分源于NTD和CTD间连接环上的N196残基与脱氧核糖的3′-羟基形成的氢键。在植物CGT中结合3′-羟基的谷氨酸位置由一个苏氨酸 (T300) 取代, 失去了对3′-羟基的结合能力[22]。对于糖基的识别, 根据LanGT2S8Ac结构, 糖基的3′-羟基可以与CTD第4β链C端α螺旋的N端主链互作。这与植物CGT类似, 但是没有了与CTD上特定氨基酸侧链的结合。4′-羟基则与NTD第5β链C端连接环上的一个天冬氨酸结合, 而且这个氨基酸在细菌CGT中并不保守[50]。鉴于细菌CGT与核苷酸糖供体结合的复合物结构只有一个, 其对糖基识别的特异性还需要进一步的研究。

2.3　与糖基受体的结合　糖基转移酶NTD中受体结合口袋的大小、形状和关键氨基酸残基决定着受体结合特异性以及糖基化类型。结合有受体底物的CGT结构当前只有GgCGT与根皮素的复合物。其NTD的一个疏水腔结合根皮素非修饰六元环一端, 反应端六元环指向CTD (图2C, F)。疏水腔由第4β链两侧的两个α螺旋N端氨基酸、第5β链的C端氨基酸及一系列第5和第6β链之间复杂结构中的氨基酸构成。根皮素的2′-羟基与CTD上PSPG盒末端用于结合糖基的天冬氨酸之间形成氢键, 4′-羟基与在天冬氨酸N端的一个组氨酸结合。然而该组氨酸在结构中显示两种构象, 只有一个构象允许与4′-羟基结合, 说明该作用可能不是很稳定 (图2C)[21]。

其他受体结合模式都是通过分子建模得到的。UrdGT2与受体的建模分析显示, UrdGT2因为有足够大的口袋才能容纳UWM6和茜素[51]; TcCGT1宽敞的结合口袋赋予其广泛的底物混杂性, 而*C*-或*O*-糖基化

活性之间的选择性取决于底物在口袋中的结合姿态[5]; GgCGT中宽敞的底物结合隧道对其di-*C*-糖基化活性和广泛的底物混杂至关重要[21]; UGT708C1的受体结合口袋比VvGT1的更浅、更小, 导致能够结合在VvGT1口袋入口附近的刚性山柰酚不能在UGT708C1口袋中稳定结合[52]; MiCGT的变体VFAH通过微调结合口袋便获得了新的催化活性, 从CGT变成一个OGT[55]。这些研究结果显示受体结合口袋的大小可以影响受体结合, 进而影响反应类型的改变。

植物UGT (包括OGT和CGT) 的NTD上通常存在一对保守的氨基酸在空间上靠近, 即: 第1β链C端α螺旋上的一个组氨酸和第4β链C端的一个酸性氨基酸 (多为天冬氨酸)。在OGT中该对氨基酸负责将受体分子上参与反应的羟基去质子化[7]。在多个CGT模型中研究人员认为它们存在类似的底物结合模式[21,43,52,57], 然而在GgCGT结构中, 虽然该组氨酸确实与受体底物羟基接近, 但是并不在氢键距离之内。

3 CGT的催化机制

糖基转移酶可以根据它们催化反应的整体立体化学过程分为翻转型和保留型[58], 而GT1家族的糖基转移酶均属于翻转型酶, 即: 催化产物的端基构型与底物相比发生翻转[59]。它们通常利用类似S_N2的直接置换来催化糖基转移反应, 并翻转UDP糖供体异头碳处的构型, 从而产生糖苷产物[7]。

经过充分表征的植物UGT都具有一定的同源性和相似的性质, 不论是CGT还是OGT。如上文所述, 它们绝大多数在NTD上存在组氨酸−天冬氨酸催化二元体, 两者侧链之间互作导致组氨酸咪唑环上电子云偏转而成为一个广义碱性氨基酸, 对糖基化修饰非常重要。在已经明确的*O*-糖基化机制中, 酚羟基经与组氨酸作用而去质子化, 以S_N2方式亲核攻击供体葡萄糖基团的1-位碳原子, 形成*O*-糖基化产物[60,61]。这个机制得到了结构生物学数据的强力支持, 各种结构特征都符合S_N2取代的立体化学要求。大多数CGT的晶体结构与OGT非常相似, 且许多蛋白同时具备OGT与CGT的功能, 如TcCGT1、GgCGT1和MiCGT等, 表明CGT和OGT可能具有相似的催化机制。

3.1 CGT的推测催化机制

目前对CGT催化机制的研究仍较为有限。由于CGT和OGT在结构上具有高度相似性, 基于OGT的研究成果, 已提出了几种可能的CGT催化机制 (图3)。

已鉴定的几乎所有CGT都将糖基转移到受体酚

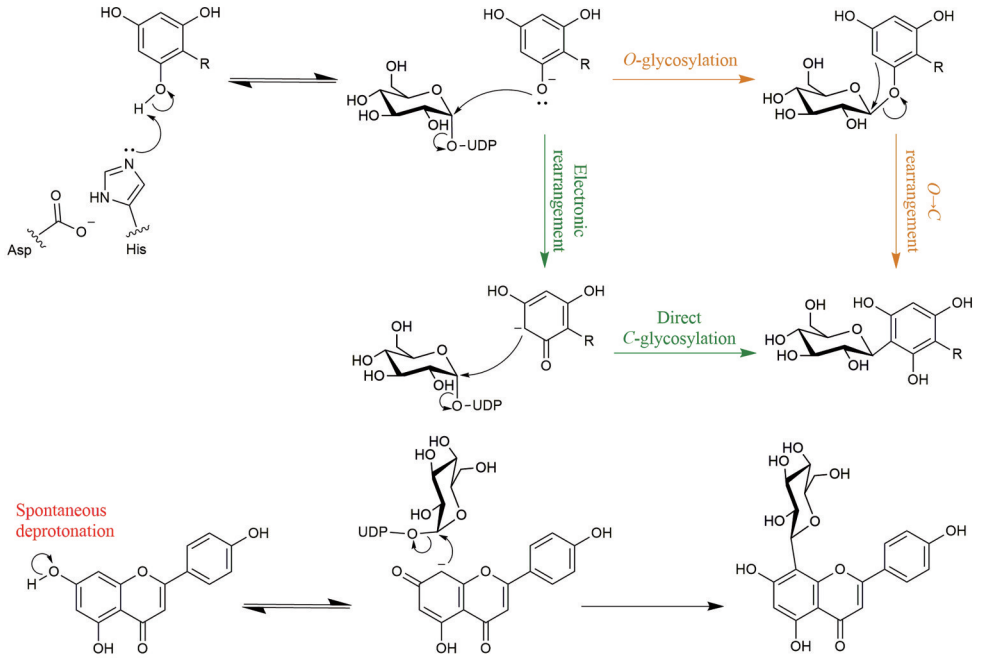


Figure 3 The proposed possible catalytic mechanisms of CGTs. The upper route indicates that the His-Asp catalytic dyad deprotonates the phenolic hydroxyl group of the acceptor molecule. The mechanism denoted by the orange arrows is that the deprotonated oxygen attacks the anomeric carbon of UDP-Glc, first forming an *O*-glycoside, followed by non-enzymatic *O*→*C* glycosyl transfer to form a *C*-glycoside. The mechanism described by the green arrow shows that after phenolic hydroxyl deprotonation, resonance (electronic rearrangement) renders the adjacent aromatic carbon electron-rich, enabling direct nucleophilic attack on the anomeric carbon of the sugar donor. The lower route shows the spontaneous deprotonation mechanism proposed in TcCGT1, where the phenolic hydroxyl group of the substrate spontaneously deprotonates and through resonance (electronic rearrangement) enables the adjacent aromatic carbon to directly nucleophilically attack the anomeric carbon of the sugar donor

羟基的邻位C上, 因此一种设想认为, 糖基首先连接到受体分子去质子化的酚羟基上形成*O*-糖苷, 再经非酶催化的O→C糖基转移形成*C*-糖苷[62]。另一种被普遍接受的机制与上述机制类似, 也需要组氨酸–天冬氨酸催化二元组启动糖基化反应。具体而言, 受体酚羟基去质子化, 同时通过共振 (电子重排) 使邻位芳香碳处于富电子状态, 能够直接对糖供体的端基碳进行亲核攻击, 从而生成*C*-糖苷产物[7]。这一机制在ScCGT1[57]、UGT708C1[52]和双-*C*-糖基化活性的GgCGT[21]中得到了认可。

然而, C/O-糖基化双功能酶TcCGT1是一个例外。在TcCGT1/UDP-Glc/受体的对接模型中, 本该作为催化碱基的H24远离糖基受体的任何羟基 (包括7-OH), 而且H24向丙氨酸的突变虽然减弱了糖基化活性, 但并没有完全消除。使用Schördinger® GLIDE软件包进行配体分析的结果表明: 在生理条件下, 受体的7-OH会发生自发去质子化。研究人员因此提出了*C*/*O*-糖基化的受体底物自发去质子化的模型。在该模型中, 残基H24和E396发挥的作用是稳定和定向底物, 而不是启动糖基化反应所必需[5]。

少数CGT没有上述组氨酸–天冬氨酸催化二元体, 比如细菌来源的UrdGT2和SsfS6、植物来源的GcCGT[18]。在细菌CGT的结构里, 结合异头碳端磷酸基团的丝氨酸/苏氨酸/天冬酰胺的C端紧邻一个精氨酸, 指向反应中心。这个碱性氨基酸或许能帮助受体去质子化, 可惜相关研究没有对其功能进行检测。在荞麦来源的UGT708C1中存在同样序列, 将该精氨酸点突变成丙氨酸确实会彻底破坏催化反应。然而UGT708C1结构中该精氨酸与NTD中的一个天冬氨酸形成盐桥, 可能影响精氨酸活化受体分子。有意思的是, 突变该天冬氨酸成为丙氨酸同样会彻底破坏蛋白活性[52]。相似的结果在黄豆来源的UGT708D1中得到验证[63]。这两个氨基酸在使用2-羟基黄烷酮作为受体的CGT中高度保守, 但在其他植物UGT中并不保守[52]。研究人员一般认为它们对催化反应非常重要。然而, 在没有这对氨基酸的TcCGT1中, 经过定点突变在对应位置引入一个精氨酸或者一对氨基酸, CGT活性不仅没有得到提升反而遭到破坏[5]。目前, 形成盐桥的这对氨基酸具体在催化反应中起什么作用还不清楚。在GcCGT中既没有组氨酸–天冬氨酸催化二元体, 也没有这个精氨酸。事实上, 整个反应腔里都没有碱性氨基酸出现。该现象在某种程度上支持底物受体可以自发去质子化的观点。

对于双-*C*-糖基化的机制研究主要关注于反应腔容量的大小, 但是存在看似相反的结论。在MiCGTb中Nβ5链C端α螺旋上的I152在双-*C*-糖基化中起关键作用。将其按照MiCGT对应位置序列突变为谷氨酸, 可将酶转化为单-*C*-糖基转移酶; 反过来, MiCGT在对应位置突变成异亮氨酸后, 则能够有效催化双-*C*-葡萄糖基化。Chen及其同事提出, 该残基对活性位点空间有影响, 而较宽的活性位点有利于双-*C*-葡萄糖基化[19]。在GgCGT的研究中也报告了类似的观察结果。同样是对反应空间的分析, 另一项研究中的建模结构比较显示, 能够进行双-*C*-糖基化的UGT708D9和UGT708B8与不能进行的UGT708F2和UGT708A60相比, 具有更小的结合位点。在这四种酶中, 活性位点越小, 酶在双-*C*-糖基化方面的能力越强[31]。或许对于不同的CGT来说, 实现双-*C*-糖基化的关键残基和结构特性可能有所不同。

3.2　*O*/*C*-糖基化选择性　在前文对催化机制的介绍中可以看出, 无论是底物自发去质子化, 还是利用保守氨基酸残基进行酸碱催化, OGT和CGT表现出很大的相似性[5]。那么到底是什么决定了不同的糖基化类型呢？近期一个关于MiCGT的研究为这一问题提供了一些新的见解。研究者认为UGT中受体底物的不同结合模式决定了*C*-糖苷或*O*-糖苷的产生, 而底物的结合模式可能是由底物结构及蛋白活性位点空间设置共同决定。分子对接辅以分子动力学模拟计算的结果显示, 在MiCGT野生型和QDP突变体 (仅具有*C*-糖基化活性的三联变体) 中, 由于Nβ5链C端α螺旋上的M148残基空间占位较大, 底物分子在反应腔中以“水平”姿态靠近活性位点; 而在ATD突变体 (仅具有*O*-糖基化活性的三联变体) 中, M148A突变导致受体分子远端羟基苯可以在该方向延伸, 并与残基E152形成氢键, 由此表现出“垂直”姿态。研究人员由此提出, “水平”与“垂直”姿态之间的变化决定了*C*-糖基还是*O*-糖基修饰反应的发生[54]。TcCGT1的*C*/*O*-糖基化活性研究对关键残基空间占位也有类似观点, 但是所关注的残基位点不同。相较于MiCGT处于受体底物远端的M148, TcCGT1在异头碳磷酸端引入具有大体积侧链的赖氨酸乃至盐桥。分子建模分析认为底物的结合模式发生改变, 底物由原本的8-C原子靠近糖供体的C-1′位切换成7-O原子, 导致TcCGT1由*C*-糖基化活性转化为*O*-糖基化[5]。除了受体底物结合方式以外, 糖供体相对于亲核试剂的特定取向也可能是区分*O*-糖基化和*C*-糖基化的关键特征。在经过杂交突变由OGT转变为CGT的LanGT2S8Ac中, 其底物对接模型显示, 虽然突变不在反应腔内, 但却使反应腔空间被挤压, 导致苷元相对于反应腔的空间位置重新取向, 从而更有利于C-糖基化的进行。值得注意的是, 突变前后

底物也发生了在“水平”与“垂直”姿态之间的变化[50]。

4 CGT的工程和应用

对比OGT在工业生产上的广泛应用, CGT用于生物催化工程的实例则相对较少。一方面是由于目前表征的CGT数量有限, 另一方面, 已发现的天然CGT通常不同时具备用于生物合成的全部理想特性, 包括: ① 在简单表达系统 (如大肠杆菌) 中的高表达水平; ② 高催化效率; ③ 最小的不良活性 (例如供体底物水解); ④ 广泛的底物混杂性; ⑤ 高区域与立体选择性; ⑥ 存储与应用时的稳定性[64]。另外, 昂贵的核酸糖供体也是限制其工业化应用的一个重要因素。通过酶工程技术对CGT进行改造, 可以改善这些特性, 从而获得更适应工业生产的生物催化工具。

酶工程是通过改变氨基酸序列使得生物催化剂具有改进特性的过程。目前, 有两种广泛使用的酶工程策略: 理性设计和定向进化。理性设计基于酶的结构分析和深入计算建模, 通过考虑氨基酸的物理化学性质并模拟它们与环境的相互作用, 有针对性地设计突变。定向进化则模拟自然进化过程, 通过诱变迭代产生突变文库, 然后筛选具有所需特性的酶变体[65]。目前, 研究人员已利用上述方法对多种CGT进行了多方面改造。

由于OGT的数量较多且与CGT在结构和特性上高度相似, 研究者们乐于将天然OGT或多功能酶改造成具备反应专一性的CGT, 且已经获得了几个成功案例。LanGT2S8Ac是在细菌天然OGT (LanGT2) 基础上, 参考CGT (UrdGT2) 结构的对应区域, 对其51−62位残基 (Nβ3链到Nβ3链C端α螺旋中部) 进行置换杂交而获得的CGT。如上文所述, 进一步的结构研究发现杂交链的使用引起反应腔空间的缩小和苷元的重新取向, 可能因此改变了糖基化模式[50]。PcOGT的I117D-D118I双突变体显示出很好的*C*-糖基化活性, 能够将葡萄糖基转移到根皮素且只产生唯一*C*-糖苷产物nothofagin[66], 而D118正是上述对催化功能起重要作用的组氨酸−天冬氨酸催化二元体的成员。

Li等[54]近期则是将MiCGT的反应腔利用定向进化进行了改造, 获得了高度特异且活性增强的C-糖基化突变体MiCGT-QDP和*O*-糖基化突变体MiCGT-ATD, 由此建立了一个具有较好化学选择性、区域选择性和反应活性的多功能糖苷合成平台。此外, 也有研究利用蛋白质工程试图拓宽或改变酶的底物特异性。例如通过对原本作用于二苯甲酮的MiCGTb进行改造, 获得了可以作用于香豆素的变体MiCGTb-GAGM, 它对结构多样的香豆素衍生物表现出强大的*C*-葡萄糖基化能力, 已有两种*C*-糖苷产物表现出了对人钠−葡萄糖协同转运蛋白2 (SGLT2) 的抑制活性[20]。

利用OGT的可逆性催化特征, 通过偶联反应可以实现从*O*-糖苷到*C*-糖苷的一锅法转化。Gutmann等[67]将PcOGT的*O*-糖基化活性和OsCGT的*C*-糖基化活性偶联, 仅需催化量的UDP即可在一锅反应中实现根皮苷向nothofagin的转化。Chen等[40]利用*C*/*O*双功能酶MiCGT也成功实现了*O*-糖苷到*C*-糖苷的转化; 随后的一项研究中他们又利用MiCGTb独特的去*O*-葡萄糖基化活性, 开发了一种利用2-*O*-葡萄糖苷高效合成酰基间苯三酚3-*C*-葡萄糖苷的方法[41]。蔗糖合酶 (能够催化生成NDP-葡萄糖) 通过与CGT偶联, 可以实现UDP-Glc的原位再生循环。该方法成功应用于Gt6CGT和GmSUS偶联, 催化合成异荭草素和异牡荆素[68]。上述这些方法避免了昂贵的UDP-Glc的使用, 为生物活性*C*-糖苷的多样化合成提供了一种经济、绿色的途径。

5 结语与展望

CGT作为一类关键的生物催化剂, 在药物研究与开发方面具有巨大的应用潜力。然而, 尽管近年来对CGT的研究取得了显著进展, 但与OGT相比, 已鉴定和表征的CGT数量仍然较少, 其催化机制和结构特征尚未完全阐明。此外, 仍有不少天然*C*-糖苷 (如具有喹啉查耳酮、蒽酮、香豆酮、生物碱和萜类等结构的*C*-糖苷)[4]没有鉴定出负责合成它们的CGT, 更多新颖CGT有待进一步挖掘。

当前已知的CGT在结构上采用模块化构造方式: NTD进行受体的识别, 而CTD负责捕捉核苷酸糖供体。植物与细菌来源的两类结构在CTD中的相似度较高, 而NTD差异较大。这个特征也体现了它们识别受体分子结构的不同。多种不同骨架的受体可以作为CGT的催化底物, 同时它们之间也呈现相对统一的局部结构特征, 即: 修饰位点都在酚羟基的邻位或者对位。这或许与其具备相对较高的电子密度相关, 有利于形成碳负离子从而完成对糖基端头碳的攻击。各种受体的体积和结构可以有较大差异, 从简单的间苯三酚, 到复杂的大环内酯, 或许对应于CGT反应腔的大小及其内部化学性质。多个CGT展现了底物受体的混杂性, 可以催化多种骨架受体的糖基化修饰, 甚至可以同时催化*C*-糖基化和*O*-糖基化修饰, 表现出较高的可塑性, 有利于进一步的生物工程改造。

对CGT应用造成限制的主要因素在于对催化机制缺乏深入理解。结构生物学研究结果提供了对糖供体分子识别的详细信息, 但是对受体识别的分子模式了解有限。当前只有一个受体-蛋白复合物结构的报道, 且根据结构信息难以对现有几种催化机制的假设进行分辨。究其本质是受体分子上的碳负离子如何活

化这个问题没有得到很好的回答。更多底物受体结合的蛋白结构应该能够在这方面提供较好的帮助, 特别是受体–糖供体–蛋白三元复合物结构。

对天然糖基转移酶的改造或许能在多个方面取得不错的收获。除了提高催化效率和特异性进而向工程化应用发展, OGT 向 CGT 的改造很大程度上可以推进对催化机制的理解。在方法学方面, 人工智能机器学习 (ML) 方法近年来成为一种新兴的酶工程方法, 以数据驱动的方式, 通过识别现有数据中的模式, 预测并生成具有潜在应用价值的新型酶变体[65]。在糖基转移酶改造方面引入人工智能应该是一个合理的选择。

CGT 在天然产物合成和工业应用中的重要性日益凸显。随着结构生物学、酶工程和合成生物学技术的不断发展, CGT 的研究和应用将迎来新的突破。通过深入理解其催化机制、优化酶的性能以及拓展其应用领域, CGT 有望在药物开发、食品工业和农业中发挥更大的作用, 为人类健康和环境保护做出重要贡献。

作者贡献: 刘蕊负责文献资料收集和综述撰写; 陈大伟、王天宇、戴均贵、卢多负责指导与修改。

利益冲突: 本文所有作者声明不存在利益冲突关系。

References

[1] Thuan NH, Huong QTT, Lam BD, et al. Advances in glycosyltransferase-mediated glycodiversification of small molecules [J]. 3 Biotech, 2024, 14: 209.

[2] Thibodeaux CJ, Melancon CE, Liu HW. Natural-product sugar biosynthesis and enzymatic glycodiversification [J]. Angew Chem Int Ed Engl, 2008, 47: 9814-9859.

[3] Méndez C, Salas JA. Altering the glycosylation pattern of bioactive compounds [J]. Trends Biotechnol, 2001, 19: 449-456.

[4] Zhang YQ, Zhang M, Wang ZL, et al. Advances in plant-derived *C*-glycosides: phytochemistry, bioactivities, and biotechnological production [J]. Biotechnol Adv, 2022, 60: 108030.

[5] He JB, Zhao P, Hu ZM, et al. Molecular and structural characterization of a promiscuous *C*-glycosyltransferase from *Trollius chinensis* [J]. Angew Chem Int Ed Engl, 2019, 58: 11513-11520.

[6] Lim EK. Plant glycosyltransferases: their potential as novel biocatalysts [J]. Chemistry, 2005, 11: 5486-5494.

[7] Wang M, Ji Q, Lai B, et al. Structure-function and engineering of plant UDP-glycosyltransferase [J]. Comput Struct Biotechnol J, 2023, 21: 5358-5371.

[8] Zhang P, Zhang Z, Zhang L, et al. Glycosyltransferase GT1 family: phylogenetic distribution, substrates coverage, and representative structural features [J]. Comput Struct Biotechnol J, 2020, 18: 1383-1390.

[9] Bowles D, Isayenkova J, Lim EK, et al. Glycosyltransferases: managers of small molecules [J]. Curr Opin Plant Biol, 2005, 8: 254-263.

[10] Gachon CMM, Langlois-Meurinne M, Saindrenan P. Plant secondary metabolism glycosyltransferases: the emerging functional analysis [J]. Trends Plant Sci, 2005, 10: 542-549.

[11] Srivastava P, Ghosh S. Insights into functional divergence, catalytic versatility and specificity of small molecule glycosyltransferases [J]. Int J Biol Macromol, 2025, 292: 138821.

[12] Bililign T, Griffith BR, Thorson JS. Structure, activity, synthesis and biosynthesis of aryl-*C*-glycosides [J]. Nat Prod Rep, 2005, 22: 742-760.

[13] Wan HX, Shen JK. Research progress of sodium-glucose co-transporter 2 inhibitors for treatment of type 2 diabetes [J]. Acta Pharm Sin (药学学报), 2012, 47: 716-724.

[14] Li L, Wang P, Tang Y. *C*-glycosylation of anhydrotetracycline scaffold with SsfS6 from the SF2575 biosynthetic pathway [J]. J Antibiot (Tokyo), 2013, 67: 65-70.

[15] Putkaradze N, Teze D, Fredslund F, et al. Natural product *C*-glycosyltransferases - a scarcely characterised enzymatic activity with biotechnological potential [J]. Nat Prod Rep, 2021, 38: 432-443.

[16] Fan A, Zhong B, Liu D, et al. Biosynthesis of epipyrone A reveals a highly specific membrane-bound fungal *C*-glycosyltransferase for pyrone galactosylation [J]. Org Lett, 2024, 26: 1160-1165.

[17] Kannangara R, Siukstaite L, Borch-Jensen J, et al. Characterization of a membrane-bound *C*-glucosyltransferase responsible for carminic acid biosynthesis in *Dactylopius coccus* Costa [J]. Nat Commun, 2017, 8: 1987.

[18] Bao Y-o, Zhang M, Li H, et al. Functional characterization and protein engineering of a glycosyltransferase GcCGT to produce flavone 6,8-Di-*C*- and 6-*C*-4′-*O*-glycosides [J]. ACS Catal, 2024, 14: 1075-1082.

[19] Chen D, Fan S, Chen R, et al. Probing and engineering key residues for Bis-*C*-glycosylation and promiscuity of a *C*-glycosyltransferase [J]. ACS Catal, 2018, 8: 4917-4927.

[20] Chen DW, Chen RD, Xie KB, et al. Biocatalytic *C*-glucosylation of coumarins using an engineered *C*-glycosyltransferase [J]. Org Lett, 2018, 20: 1634-1637.

[21] Zhang M, Li F, Li K, et al. Functional characterization and structural basis of an efficient Di-*C*-glycosyltransferase from *Glycyrrhiza glabra* [J]. J Am Chem Soc, 2020, 142: 3506-3512.

[22] Wang F, Zhou M, Singh S, et al. Crystal structure of SsfS6, the putative *C*-glycosyltransferase involved in SF2575 biosynthesis [J]. Proteins, 2013, 81: 1277-1282.

[23] Fischbach MA, Lin HN, Liu DR, et al. *In vitro* characterization of IroB, a pathogen-associated *C*-glycosyltransferase [J]. Proc Natl Acad Sci U S A, 2005, 102: 571-576.

[24] Bililign T, Hyun CG, Williams JS, et al. The Hedamycin locus implicates a novel aromatic PKS priming mechanism [J]. Chem Biol, 2004, 11: 959-969.

[25] Ichinose K, Ozawa M, Itou K, et al. Cloning, sequencing and heterologous expression of the medermycin biosynthetic gene cluster of *Streptomyces* sp. AM-7161: towards comparative analysis of the benzoisochromanequinone gene clusters [J]. Microbiology (Reading), 2003, 149: 1633-1645.

[26] Künzel E, Faust B, Oelkers C, et al. Inactivation of the *urdGT2* gene, which encodes a glycosyltransferase responsible for the *C*-glycosyltransfer of activated *D*-olivose, leads to formation of the novel urdamycins I, J, and K [J]. J Am Chem Soc, 1999, 121: 11058-11062.

[27] Ichinose K, Bedford DJ, Tornus D, et al. The granaticin biosynthetic gene cluster of *Streptomyces violaceoruber* Tü22: sequence analysis and expression in a heterologous host [J]. Chem Biol, 1998, 5: 647-659.

[28] Ito T, Fujimoto S, Suito F, et al. *C*-glycosyltransferases catalyzing the formation of di-*C*-glucosyl flavonoids in citrus plants [J]. Plant J, 2017, 91: 187-198.

[29] Falcone Ferreyra ML, Rodriguez E, Casas MI, et al. Identification of a bifunctional maize *C*- and *O*-glucosyltransferase [J]. J Biol Chem, 2013, 288: 31678-31688.

[30] Brazier-Hicks M, Evans KM, Gershater MC, et al. The *C*-glycosylation of flavonoids in cereals [J]. J Biol Chem, 2009, 284: 17926-17934.

[31] Putkaradze N, Gala VD, Vaitkus D, et al. Sequence mining yields 18 phloretin *C*-glycosyltransferases from plants for the efficient biocatalytic synthesis of nothofagin and phloretin-di-*C*-glycoside [J]. Biotechnol J, 2023, 18: e2200609.

[32] Huang J, She Y, Yue J, et al. Exploring the catalytic function and active sites of a novel *C*-glycosyltransferase from *Anemarrhena asphodeloides* [J]. Synth Syst Biotechnol, 2022, 7: 621-630.

[33] Ren ZY, Ji XY, Jiao ZB, et al. Functional analysis of a novel *C*-glycosyltransferase in the orchid *Dendrobium catenatum* [J]. Hortic Res, 2020, 7: 111.

[34] Nagatomo Y, Usui S, Ito T, et al. Purification, molecular cloning and functional characterization of flavonoid *C*-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon [J]. Plant J, 2014, 80: 437-448.

[35] Sasaki N, Nishizaki Y, Yamada E, et al. Identification of the glucosyltransferase that mediates direct flavone *C*-glucosylation in *Gentiana triflora* [J]. FEBS Lett, 2014, 589: 182-187.

[36] Mashima K, Hatano M, Suzuki H, et al. Identification and characterization of apigenin 6-*C*-glucosyltransferase involved in biosynthesis of isosaponarin in Wasabi (*Eutrema japonicum*) [J]. Plant Cell Physiol, 2019, 60: 2733-2743.

[37] Wang X, Li C, Zhou C, et al. Molecular characterization of the *C*-glucosylation for puerarin biosynthesis in *Pueraria lobata* [J]. Plant J, 2017, 90: 535-546.

[38] Bao YO, Zhang M, Qiao X, et al. Functional characterization of a *C*-glycosyltransferase from *Pueraria lobata* with dual-substrate selectivity [J]. Chem Commun (Camb), 2022, 58: 12337-12340.

[39] Xie K, Zhang X, Sui S, et al. Exploring and applying the substrate promiscuity of a *C*-glycosyltransferase in the chemoenzymatic synthesis of bioactive *C*-glycosides [J]. Nat Commun, 2020, 11: 5162.

[40] Chen D, Chen R, Wang R, et al. Probing the catalytic promiscuity of a regio- and stereospecific *C*-glycosyltransferase from *Mangifera indica* [J]. Angew Chem Int Ed Engl, 2015, 127: 12869-12873.

[41] Chen D, Sun L, Chen R, et al. Enzymatic synthesis of acylphloroglucinol 3-*C*-glucosides from 2-*O*-glucosides using a *C*-glycosyltransferase from *Mangifera indica* [J]. Chemistry, 2016, 22: 5873-5877.

[42] Uchida K, Akashi T, Hirai MY. Identification and characterization of glycosyltransferases catalyzing direct xanthone 4-*C*-glycosylation in *Hypericum perforatum* [J]. FEBS Lett, 2021, 595: 2608-2615.

[43] Chen DW, Fan S, Yang ZY, et al. Biocatalytic application of a membrane-bound coumarin *C*-glucosyltransferase in the synthesis of coumarin and benzofuran *C*-glucosides [J]. Adv Synth Catal, 2021, 363: 5072-5078.

[44] Trefzer A, Hoffmeister D, Künzel E, et al. Function of glycosyltransferase genes involved in urdamycin A biosynthesis [J]. Chem Biol, 2000, 7: 133-142.

[45] Cai XF, Taguchi T, Wang HL, et al. Identification of a *C*-glycosyltransferase involved in medermycin biosynthesis [J]. ACS Chem Biol, 2021, 16: 1059-1069.

[46] Tegl G, Nidetzky B. Leloir glycosyltransferases of natural product *C*-glycosylation: structure, mechanism and specificity [J]. Biochem Soc Trans, 2020, 48: 1583-1598.

[47] Lairson LL, Henrissat B, Davies GJ, et al. Glycosyltransferases: structures, functions, and mechanisms [J]. Annu Rev Biochem, 2008, 77: 521-55.

[48] McIntosh CA, Owens DK. Advances in flavonoid glycosyltransferase research: integrating recent findings with long-term citrus studies [J]. Phytochem Rev, 2016, 15: 1075-1091.

[49] Breton C, Snajdrová L, Jeanneau C, et al. Structures and mechanisms of glycosyltransferases [J]. Glycobiology, 2006, 16: 29r-37r.

[50] Tam HK, Härle J, Gerhardt S, et al. Structural characterization of *O*- and *C*-glycosylating variants of the landomycin glycosyltransferase LanGT2 [J]. Angew Chem Int Ed Engl, 2015, 54: 2811-2815.

[51] Mittler M, Bechthold A, Schulz GE. Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin [J]. J Mol Biol, 2007, 372: 67-76.

[52] Liu M, Wang D, Li Y, et al. Crystal structures of the *C*-glycosyltransferase UGT708C1 from buckwheat provide insights into the mechanism of *C*-glycosylation [J]. Plant Cell, 2020, 32: 2917-2931.

[53] Wang ZL, Gao HM, Wang S, et al. Dissection of the general two-

step di-*C*-glycosylation pathway for the biosynthesis of (iso) schaftosides in higher plants [J]. Proc Natl Acad Sci U S A, 2020, 117: 30816-30823.

[54] Li M, Zhou Y, Wen Z, et al. An efficient *C*-glycoside production platform enabled by rationally tuning the chemoselectivity of glycosyltransferases [J]. Nat Commun, 2024, 15: 8893.

[55] Wen Z, Zhang ZM, Zhong L, et al. Directed evolution of a plant glycosyltransferase for chemo- and regioselective glycosylation of pharmaceutically significant flavonoids [J]. ACS Catal, 2021, 11: 14781-14790.

[56] Akere A, Chen SH, Liu X, et al. Structure-based enzyme engineering improves donor-substrate recognition of *Arabidopsis thaliana* glycosyltransferases [J]. Biochem J, 2020, 477: 2791-2805.

[57] Ni R, Liu XY, Zhang JZ, et al. Identification of a flavonoid *C*-glycosyltransferase from fern species *Stenoloma chusanum* and the application in synthesizing flavonoid *C*-glycosides in *Escherichia coli* [J]. Microb Cell Fact, 2022, 21: 210.

[58] Thibodeaux CJ, Melançon CE, Liu HW. Unusual sugar biosynthesis and natural product glycodiversification [J]. Nature, 2007, 446: 1008-1016.

[59] Louveau T, Leitao C, Green S, et al. Predicting the substrate specificity of a glycosyltransferase implicated in the production of phenolic volatiles in tomato fruit [J]. FEBS J, 2011, 278: 390-400.

[60] Offen W, Martinez-Fleites C, Yang M, et al. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification [J]. EMBO J, 2006, 25: 1396-1405.

[61] Brazier-Hicks M, Offen WA, Gershater MC, et al. Characterization and engineering of the bifunctional *N*- and *O*-glucosyltransferase involved in xenobiotic metabolism in plants [J]. Proc Natl Acad Sci U S A, 2007, 104: 20238-20243.

[62] Dürr C, Hoffmeister D, Wohlert SE, et al. The glycosyltransferase UrdGT2 catalyzes both *C*- and *O*-glycosidic sugar transfers [J]. Angew Chem Int Ed Engl, 2004, 43: 2962-2965.

[63] Hirade Y, Kotoku N, Terasaka K, et al. Identification and functional analysis of 2-hydroxyflavanone *C*-glucosyltransferase in soybean (*Glycine max*) [J]. FEBS Lett, 2015, 589: 1778-1786.

[64] McArthur JB,Chen X. Glycosyltransferase engineering for carbohydrate synthesis [J]. Biochem Soc Trans, 2016, 44: 129-142.

[65] Mazurenko S, Prokop Z, Damborsky J. Machine learning in enzyme engineering [J]. ACS Catal, 2020, 10: 1210-1223.

[66] Gutmann A, Nidetzky B. Switching between *O*- and *C*-glycosyltransferase through exchange of active-site motifs [J]. Angew Chem Int Ed Engl, 2012, 51: 12879-12883.

[67] Gutmann A, Krump C, Bungaruang L, et al. A two-step *O*- to *C*-glycosidic bond rearrangement using complementary glycosyltransferase activities [J]. Chem Commun, 2014, 50: 5465-5468.

[68] Pei J, Sun Q, Gu N, et al. Production of isoorientin and isovitexin from luteolin and apigenin using coupled catalysis of glycosyltransferase and sucrose synthase [J]. Appl Biochem Biotechnol, 2019, 190: 601-615.