

• 药材与资源 •

何首乌糖基转移酶基因的克隆、生物信息学及表达分析

蔡启忠, 周良云, 刘 露, 刘长征, 王 浩, 钟 磊, 杨 全*

广东药科大学中药学院 国家中医药管理局岭南药材生产与开发重点实验室 国家中药材产业技术体系广州综合试验站
广东省南药规范化种植与综合开发工程技术研究中心, 广东 广州 510006

摘要: 目的 克隆何首乌 *Polygonum multiflorum* 糖基转移酶基因 *PmUGTs*, 并对其生物信息学以及差异表达分析。方法 基于何首乌转录组数据, 利用 PCR 扩增 *PmUGTs* 全长 cDNA 并进行生物信息学分析; 使用 autodock4 软件进行分子模拟对接; 运用 qRT-PCR 检测基因在不同器官中的表达差异。结果 克隆得到 4 条 *PmUGTs*, 分别命名为 *PmUGT1*、*PmUGT2*、*PmUGT3*、*PmUGT4* 基因, 开放阅读框 (ORF) 长度分别为 1509、1506、1464 和 1476 bp; 编码 503、502、488 与 492 个氨基酸; 蛋白相对分子质量分别为 56 940、54 780、54 350 和 54 620。*PmUGTs* 氨基酸序列中含有 UGT 高度保守的 PSPG 盒结构域, 其二级结构主要由无规则卷曲与 α -螺旋组成。进化树分析表明, *PmUGT1*、*PmUGT2*、*PmUGT3* 与拟南芥的 UGT 蛋白距离较近, *PmUGT4* 则与蒺藜苜蓿 UGT 亲缘关系最近。根据 autodock4 结果可知 *PmUGT1*、*PmUGT3*、*PmUGT2* 与 *PmUGT4* 蛋白上的残基分别可以与鹰嘴豆芽素 A、罗汉果醇、白藜芦醇通过氢键连接。qRT-PCR 结果表明, *PmUGTs* 表达量均在叶中最高。结论 为进一步研究何首乌糖基转移酶基因的功能及何首乌中糖苷类化合物的生物合成途径奠定基础。

关键词: 何首乌; 糖基转移酶; 基因克隆; 原核表达; qRT-PCR

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2021)19-6013-10

DOI: 10.7501/j.issn.0253-2670.2021.19.024

Cloning, bioinformatic, and expression analysis of glycosyltransferase gene from *Polygonum multiflorum* Thunb.

CAI Qi-zhong, ZHOU Liang-yun, LIU Lu, LIU Chang-zheng, WANG Hao, ZHONG Lei, YANG Quan

School of Traditional Chinese Medicine (TCM), Key Laboratory for Production & Development of Cantonese Medicinal Materials Under State Administration of TCM, Guangzhou Comprehensive Experimental Station Chinese Materia Medica Industry Technology System, Guangdong Provincial Research Center on Good Agricultural Practice & Comprehensive Agricultural Development Engineering Technology of Cantonese Medicinal Materials, Guangdong Pharmaceutical University, Guangdong Guangzhou 510006, China

Abstract: Objective To clone the *Polygonum multiflorum* Thunb. glycosyltransferase genes (*PmUGTs*) and analyze their bioinformatics and relative expression levels. **Methods** Based on the transcriptomic data of *Polygonum multiflorum* Thunb., the *PmUGTs* full length cDNA were amplified by PCR and their bioinformatics were analysis. Autodock4 software was used for molecular docking, and qRT-PCR was used to detect *PmUGTs* organ-specific expression levels. **Results** The length of *PmUGT1*, *PmUGT2*, *PmUGT3* and *PmUGT4* ORF were from 1509, 1506, 1464 and 1476 bp, encoding 503, 502, 488 and 492 amino acids with the molecular mass of 56 940, 54 780, 54 350 and 54 620, respectively. *PmUGTs* amino acid sequences contain a highly conserved PSPG box domain, and their secondary structures are mainly composed of Random Coil and α -Helix. Phylogenetic analysis showed that *PmUGT1*, *PmUGT2* and *PmUGT3* were closed to the *Arabidopsis thaliana* UGTs, and *PmUGT4* was closed to which from *Medicago truncatula*. According to the autodock4 results, *PmUGT1*, *PmUGT3*, *PmUGT2* and *PmUGT4* were able to be hydrogen bonded with biochanin A, mogrol and resveratrol, respectively. Measurement of organ-specific expressions showed that the levels of *PmUGTs* leaves expression were the highest. **Conclusion** This research will laid a foundation for further study on the function of *PmUGTs* genes and biosynthesis pathway of *P. multiflorum* glycosides.

Key words: *Polygonum multiflorum* Thunb.; glycosyltransferase; gene cloning; prokaryotic expression; qRT-PCR

收稿日期: 2021-03-02

基金项目: 国家重点研发计划项目 (2017YFC1700704); 2017 年广东省岭南中药材保护资金专项 (粤财社[2017]60 号); 广东省普通高校青年创新人才类项目 (2018KQNCX133)

作者简介: 蔡启忠 (1996—), 男, 硕士研究生, 主要从事中药资源开发与品质评价研究。Tel: (020)39252353 E-mail: 574382026@qq.com

*通信作者: 杨 全, 男, 教授, 主要从事道地药材道地性机制研究。Tel: (020)39252353 E-mail: yangquan7208@vip.163.com

糖基化是自然界中广泛存在的重要修饰反应之一。糖基转移酶(glycosyltransferase, GT, EC 2.4.x.y)是催化糖基化反应的一类酶^[1],在生物体内广泛存在的超基因家族^[2],在酶催化反应中具有高效性和立体选择性^[3],对植物的各种生命活动及次生代谢产物的形成上都具有重要影响^[4]。随着植物糖基转移酶及催化底物谱的不断探索与发现,糖基化工具酶库不断扩充,反应类型不断丰富,糖基化反应将变得更绿色与经济,甚至使某些化学法难以糖基化的反应得以实现^[5]。在糖基转移酶的研究与开发利用中,针对不同糖基酶的催化特点,有针对性的对其底物识别的特异性进行探索,从而发现特异性催化的糖基转移酶,系统性的总结其催化特点,进而建立糖基化的新方法,与化学合成进行优势互补,取长补短,以低成本、更绿色的手段来合成糖苷类化合物。

何首乌 *Polygonum multiflorum* Thunb. 为蓼科多年生草本植物,主要以块根入药,是我国著名的传统中药。何首乌中含二苯乙烯类、蒽醌类、黄酮类、磷脂类、多糖类主要活性化学成分^[6]。目前何首乌的单组分药理研究主要集中在对二苯乙烯苷的研究上。研究表明二苯乙烯苷具有抗衰老、抗动脉粥样硬化、抗高血脂、抗肿瘤、抗炎、清除自由基、保肝等方面的生物活性^[7-8]。

目前,从何首乌中分离得到了多种以二苯乙烯类结构为母核的糖苷,在何首乌中主要以 2,3,5,4'-四羟基二苯乙烯-2-O- β -D-葡萄糖苷的形式存在,是二苯乙烯苷元侧链结合一个单糖基,这意味着何首乌中存在相应的糖基转移酶参与何首乌二苯乙烯苷的糖基化修饰,且主要在 2 位上形成 O-糖苷。研究催化特定位点糖苷化的酶,不仅能阐述何首乌中二苯乙烯苷的 O-糖基引入机制,也有助于阐明何首乌中活性成分的生物合成途径。迄今为止尚未有将何首乌中获得的糖基转移酶作为工具酶用于结构多样的糖苷化合物生物合成的报道。因此,挖掘并研究何首乌糖基转移酶并应用于多种结构类型的天然产物特定位点糖苷的酶法合成,具有重要的理论研究与实际应用价值。

1 材料与试剂

1.1 材料

何首乌植物采自广东省肇庆市德庆县大桥村,经广东药科大学杨全教授鉴定为蓼科植物何首乌 *P. multiflorum* Thunb.。取其根、茎、叶组织样品,液

氮速冻后置 -80 °C 保存备用。

1.2 菌株、载体及试剂

原核表达载体 pET-28a(+) 由实验室传代冻存;克隆和表达大肠杆菌 *Transl-T1*、*Transetta* (DE3) 与无缝连接试剂盒 pEASY-Uni Seamless Cloning and Assembly Kit 均购自北京全式金生物技术有限公司; *Bam*H I 限制性内切酶购自 NEB 公司;植物多糖多酚总 RNA 提取试剂盒购自北京天根生化科技有限公司; KOD-Plus- 高保真酶购自日本 TOYOBO 公司; PrimeScript™ II 1st Strand cDNA Synthesis Kit、胶回收试剂盒、PrimeScript™ RT reagent Kit with gDNA Eraser、TB Green™ Premix Ex Taq™ II 均购自日本 TaKaRa 公司;引物合成及基因片段测序由广州擎科生物技术有限公司完成。

2 方法

2.1 何首乌总 RNA 的提取

取 -80 °C 冻存的根、茎、叶组织样品,按照植物多糖多酚总 RNA 提取试剂盒 (Tiangen) 操作说明提取何首乌总 RNA,利用 Nano-100 微量分光光度计检测 RNA 浓度和纯度,1.0% 琼脂糖凝胶电泳检测 RNA 的完整性。

2.2 何首乌糖基转移酶基因 (*PmUGTs*) 的全长 cDNA 克隆

本研究基于实验室已获得的何首乌转录组数据^[9]筛选了 4 条 *PmUGTs* 基因序列 (Gene ID 分别为 Unigene0051938、Unigene0017806、Unigene0027873 与 Unigene0034969),基于无缝克隆原理设计以上序列的特异性扩增引物 (引物序列见表 1)。按照 KOD-Plus-Neo 扩增试剂盒说明 (TOYOBO) 进行,以反转录产物为模板,扩增 *PmUGTs* 基因片段。扩增反应体系为: cDNA 1 μ L, KOD-Plus-Neo (1 U/ μ L) 1 μ L, 10 $\times$ PCR Buffer for KOD-Plus-Neo 5 μ L, dNTPs (2 mmol/L) 5 μ L, MgSO₄ (25 mmol/L) 3 μ L, 上、下游引物 (10 μ mol/L) 各 1.5 μ L, ddH₂O 补足至 50 μ L。反应程序: 94 °C 预变性 2 min; 94 °C 变性 30 s, 60 °C 退火 30 s, 72 °C 延伸 40 s, 35 个循环后; 72 °C 延伸 10 min。1.0% 琼脂糖凝胶电泳检测 PCR 产物,胶回收 PCR 产物。按照无缝克隆说明,将 PCR 产物与 *Bam*H I 限制性内切酶酶切后的 pET-28a 载体连接;转化到 *Transl-T1* 菌株中,在含卡那霉素抗性的平板上进行筛选,挑取单菌落进行菌落 PCR 检测,选取 5 个阳性菌液送广州擎科公司测序验证。

表 1 引物序列

Table 1 Primer sequences

作用	引物	引物序列 (5'-3')	预期长度/bp
PCR	UGT1-F	CAGCAAATGGGTCGCGGAATGGATTCTCCAACCTCCCATG	1545
	UGT1-R	ACGGAGCTCGAATTCGGATTACTTGTGATTGCTGGATTCTG	
	UGT2-F	CAGCAAATGGGTCGCGGAATGTCTCCTCCACCGGCAGCCA	1542
	UGT2-R	ACGGAGCTCGAATTCGGATTAAATTTGAGCGAGCACTAATA	
	UGT3-F	CAGCAAATGGGTCGCGGAATGGGATATTTATCAAGTGCCG	1500
	UGT3-R	ACGGAGCTCGAATTCGGATTACTTTGTTTGCGTGACGCTT	
	UGT4-F	CAGCAAATGGGTCGCGGAATGAAAAGCAGAACCTGTTCAAAG	1512
	UGT4-R	ACGGAGCTCGAATTCGGATCATAATCCGCGAAATAAAACG	
qRT-PCR	PP2A-F	GGACCAATGTGCGATCTCTTA	101
	PP2A-R	GCTGCTATGTCTTGTCCAAATG	
	UGT1-F	CTGCATGAATTGGCTCGACG	123
	UGT1-R	TTCTTGCTCCTGACCAACCC	
	UGT2-F	ACGAGAGTTGCATGGAGTGG	108
	UGT2-R	AGGCAATGCTCTCCTTGGTC	
	UGT3-F	AAGAGGAGGGCGTTGGAATG	120
	UGT3-R	TGTTTGCGTGACGCTTTGTT	
	UGT4-F	CGCGTCGTCAACATGGTTTT	105
	UGT4-R	TCCGTCTTCTCGAGTGGCTA	

2.3 PmUGTs 的生物信息学分析

使用 EMBOSS 在线软件与 CodonW 1.4.2 软件对 *PmUGTs* 核苷酸序列进行密码子偏好性参数分析, 包括同义密码子相对实用度 (relative synonymous codon usage, RSCU)、密码子使用频次、密码子适应指数 (codon adaptation index, CAI)、有效密码子数 (effective number of codons, ENC)、GC 含量百分比、密码子末尾 GC 含量百分比 (GC_{3S})。利用 ExPASy Proteomics Server 在线软件 ProtParam 对 4 个 *PmUGTs* 基因编码蛋白分子式组成、蛋白质分子质量、理论等电点及稳定性等理化性质进行分析; Cell-PLoc 2.0 进行亚细胞定位; 使用 SignalP 5.0 sever 预测跨膜结构域; 通过 EXPASY-SOPMA 在线预测蛋白质的二级结构; 利用 SWISS-MODEL 在线软件构建 *PmUGTs* 蛋白的三级结构模型; 使用 Autodock 4.2.3 软件对推测底物进行分子对接模拟实验; 将 *PmUGTs* 编码的氨基酸序列在 NCBI 中进行蛋白 BLAST 分析, 利用 DNAMAN 软件与其他物种的 UGT 氨基酸序列进行同源性分析; 并通过 MEGA 6.0 软件构建 neighbor-joining 系统进化树, 采用泊松校正法计算进化距离, Bootstrap 重复次数为 1000 次。

2.4 PmUGTs 基因的原核表达

挑选测序验证后的转化子, 过夜培养, 提取

pET-28a-*PmUGTs* 重组质粒, 转化 *Transtetta* (DE3) 感受态大肠杆菌中培养, 在含卡那霉素抗性的平板上进行筛选, 挑取单菌落进行菌落 PCR 检测, 选取 5 个阳性菌液送广州擎科公司测序验证。挑选序列正确菌株, 按 1:100 比例加入到适量含有 50 μg/mL 卡那霉素的 LB 液体培养基中, 37 °C, 200 r/min 振荡培养至 600 nm 吸光度 (A_{600nm}) 达 0.6~1.0; 加 IPTG 至终浓度 0.4 mmol/L, 20 °C, 200 r/min 培养 12 h, 诱导产生 *PmUGT* 重组蛋白。经诱导表达的培养物, 以 5 500 r/min 离心 15 min 收集菌体, 用 ddH₂O 清洗菌体 2 次后, 将菌体悬浮于缓冲液 (50 mmol/L Tris-HCl, 1 mmol/L EDTA, 10%甘油, 1 mmol/L PMSF) 中, 超声破碎 (750 W 功率, 超声 5 s, 间歇 5 s, 工作 5 min)。细胞破碎液于 4 °C, 13 000 r/min, 离心 15 min, 去除细胞碎片。取上清加入 6×loading buffer 混匀, 沸水浴 5 min, 13 000 r/min 离心 5 min, 上样 10 μL 进行 10% SDS-PAGE 电泳, 电泳结束后将胶置于考马斯亮蓝中染色 1 h, 用脱色液将背景色脱去, 检测蛋白表达情况。

2.5 PmUGTs 基因的组织特异性表达分析

利用 qRT-PCR 方法检测何首乌 *PmUGTs* 基因在不同组织中的表达水平。取-80 °C 冻存的根、茎、叶组织样品, 每个器官取 3 个生物学重复, 提取其总 RNA, 用反转录试剂盒 PrimeScript™ RT reagent

Kit with gDNA Eraser (TaKaRa 公司) 将以上总 RNA 反转录生成单链 cDNA。使用 TB Green™ Premix Ex Taq™ II (TaKaRa), 在 CFX96 Touch (BioRad) 仪上进行扩增。选取何首乌 *PP2A* 基因^[9]作为目标基因定量表达的内参基因, 通过 primer BLAST 在线设计扩增引物, 引物序列见表 1。扩增体系中含 10 μ L SYBR® Green Pro Taq HS Premix (2 $\times$)、上下游引物 (10 μ mol/L) 各 0.4 μ L、cDNA 2 μ L, ddH₂O 补至总体积为 20 μ L, 每个体系设 3 次技术重复。反应程序: 95 $^{\circ}$ C 预变性 30 s; 95 $^{\circ}$ C、5 s, 60 $^{\circ}$ C、30 s, 40 个循环后, 进行溶解曲线分析。采用 $2^{-\Delta\Delta C_t}$ 方法进行相对定量分析, 用 SPSS 23.0 进行各组间方差分析, 并用 GraphPad Prism 7.0 作图。

3 结果与分析

3.1 *PmUGTs* 基因全长 cDNA 克隆

以逆转录 cDNA 为模板进行扩增后, 4 条基因的扩增产物在 2000~1000 bp 内, 如图 1 所示。构建的 pET-28a-*PmUGTs* 重组载体经测序结果显示, 克隆的目的片段与何首乌转录组数据中的基因序列一致, 基因分别命名为 *PmUGT1*、*PmUGT2*、*PmUGT3*、*PmUGT4*。

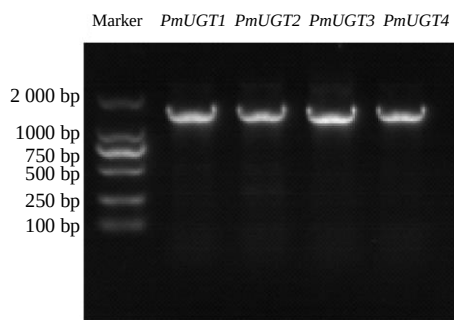


图 1 *PmUGTs* 基因 PCR 扩增产物

Fig. 1 PCR amplified production *PmUGTs*

3.2 *PmUGTs* 基因的生物信息学分析

3.2.1 *PmUGTs* 密码子偏好分析及其蛋白理化性质分析、亚细胞定位、跨膜区域分析 运用 EMBOSS 在线软件与 CodonW 1.4.2 分析 *PmUGTs* 进行密码子偏好参数分析, 结果见表 2、3, 得出 *PmUGT1*、

PmUGT2、*PmUGT3*、*PmUGT4* 的 ENC 值分别为 57.242、49.211、51.597、56.674, 均值为 53.681 (远大于 20 而偏向 61), 表明该密码子偏好性较弱; 此外, 所得的 CAI 值分别为 0.350 0、0.351 0、0.408 3、0.324 2, 均值为 0.358 4 (远低于 1.0), 进一步说明 *PmUGTs* 对密码子的偏好性较弱; *PmUGTs* 的 GC 与 GC_{3s} 含量均大于 50%, 表明该开放阅读框 (ORF) 序列中 AT 含量小于 GC 含量, 且 *PmUGTs* 偏好使用以 G/C 结尾的密码子。*PmUGTs* 密码子 RSCU 值分析如表 3 所示, 其中 24 个密码子 RSCU 值大于 1, 是 *PmUGTs* 的偏好密码子; GCC、AUC、AGG、AGC、ACC、UAC 的 RSCU 值大于 1.5, 即为高频率密码子; 其中 ACC (2.12) 的 RSCU 值大于 2, 表明 *PmUGTs* 对该密码子具有极强偏好性。

使用在线软件 ProtParam 分析 *PmUGTs* 的蛋白理化性质, 结果如表 4 所示, 得出 *PmUGT1*、*PmUGT2*、*PmUGT3*、*PmUGT4* 的相对分子质量分别在 56 940、54 780、54 305 和 54 620; 等电点 (PI) 分别在 5.97、5.30、5.73、5.91; 且均为亲水性蛋白; 除 *PmUGT3* 外 (不稳定系数=32.84<40), 其他蛋白皆为不稳定蛋白 (不稳定系数>40)。使用亚细胞定位软件 Cell-PLoc 2.0 预测出 *PmUGT1*、*PmUGT4* 定位于细胞膜中, *PmUGT2*、*PmUGT3* 则定位于细胞膜与叶绿体中。SignalIP 5.0 sever 在线软件分析得出 *PmUGTs* 均不存在跨膜区域。

3.2.2 *PmUGTs* 蛋白二级和三级结构分析 预测结果如表 5 所示, *PmUGTs* 蛋白的二级结构由大部分的无规则卷曲 (random coil) 与 α -螺旋 (α -helices) 组成, 推测两者为蛋白主要二级结构元件; 延伸链 (extended strand) 和 β -折叠 (β -turn) 则分布于整个蛋白中。根据同源建模的原理, 使用 SWISS-MODEL 对蛋白三维结构进行预测, 三维结构如图 2 所示。*PmUGT1*、*PmUGT2*、*PmUGT3*、*PmUGT4* 蛋白同源建模使用的模板分子及其同源性分别为 2pq6.1.A^[10], 36.19%; 6l8x.1.A^[11], 33.93%; 2pq6.1.A, 52.22%; 6lzx.1.A^[12], 27.86%。

表 2 *PmUGTs* 核苷酸序列特征分析

Table 2 Analysis of nucleotide sequence of *PmUGTs* from *P. multiflora*

基因名称	Unigene 基因号	ORF 长度/bp	密码子选择偏好性			
			GC/%	GC _{3s} /%	ENC	CAI
<i>PmUGT1</i>	0051938	1512	51.65	57.74	57.242	0.3500
<i>PmUGT2</i>	0017806	1512	57.21	69.44	49.211	0.3510
<i>PmUGT3</i>	0027873	1467	55.90	70.76	51.597	0.4083
<i>PmUGT4</i>	0034969	1479	50.03	55.17	56.674	0.3242

表 3 *PmUGTs* 密码子 RSCU 值分析

Table 3 RSCU value analysis for *PmUGTs* from *P. multiflora*

氨基酸	密码子	频次	RSCU	氨基酸	密码子	频次	RSCU
丙氨酸	GCA	18	0.55	天冬酰胺	<u>AAU</u>	33	1.06
	<u>GCC</u>	52	1.58	脯氨酸	CCA	23	0.79
	<u>GCG</u>	33	1.00		<u>CCC</u>	30	1.03
	GCU	29	0.88		<u>CCG</u>	35	1.20
半胱氨酸	<u>UGC</u>	25	1.28		CCU	29	0.99
	UGU	14	0.72	谷氨酰胺	<u>CAA</u>	37	1.25
天冬氨酸	<u>GAC</u>	62	1.24		CAG	22	0.75
	GAU	38	0.76	精氨酸	<u>AGA</u>	24	1.36
谷氨酸	GAA	43	0.61		<u>AGG</u>	29	1.64
	<u>GAG</u>	99	1.39		CGA	13	0.74
苯丙氨酸	<u>UUC</u>	48	1.26		CGC	17	0.96
	UUU	28	0.74		CGG	15	0.85
甘氨酸	<u>GGC</u>	37	1.08		CGU	8	0.45
	GGA	26	0.76	丝氨酸	<u>AGC</u>	39	1.58
	<u>GGG</u>	48	1.40		AGU	12	0.49
	GGU	26	0.76		UCA	22	0.89
组氨酸	CAC	33	0.71		<u>UCC</u>	35	1.42
	<u>CAU</u>	18	1.29		UCG	23	0.93
异亮氨酸	AUA	26	0.81		UCU	17	0.69
	<u>AUC</u>	48	1.50	苏氨酸	ACA	20	0.68
	AUU	22	0.69		<u>ACC</u>	62	2.12
赖氨酸	AAA	35	0.74		ACG	21	0.72
	<u>AAG</u>	59	1.26		ACU	14	0.48
亮氨酸	<u>CUA</u>	35	1.00	缬氨酸	GUA	19	0.48
	<u>CUC</u>	61	1.74		GUC	39	0.98
	CUG	28	0.80		<u>GUG</u>	59	1.48
	CUU	31	0.89		<u>GUU</u>	42	1.06
	UUA	18	0.51	酪氨酸	<u>UAC</u>	23	1.53
	<u>UUG</u>	37	1.06		UAU	7	0.47
天冬酰胺	AAC	37	0.94				

下划线表示 *PmUGTs* 基因密码子使用频率较高

The underlines mean codon usage with high frequency in *PmUGTs*

表 4 *PmUGTs* 蛋白理化性质分析

Table 4 Physicochemical property of *PmUGTs* protein from *P. multiflora*.

蛋白名称	氨基酸数	分子式	相对分子质量	等电点	不稳定系数	亲水性	亚细胞定位
PmUGT1	503	C ₂₅₂₇ H ₃₉₈₈ N ₇₀₀ O ₇₄₄ S ₂₇	56 940	5.97	46.48	-0.256	细胞膜
PmUGT2	502	C ₂₄₄₆ H ₃₈₃₉ N ₆₅₉ O ₇₃₁ S ₁₉	54 780	5.30	40.89	-0.069	细胞膜、叶绿体
PmUGT3	488	C ₂₄₃₉ H ₃₈₀₀ N ₆₅₂ O ₇₀₆ S ₂₅	54 350	5.73	32.84	-0.134	细胞膜、叶绿体
PmUGT4	492	C ₂₄₂₉ H ₃₈₈₅ N ₆₆₁ O ₇₁₅ S ₂₆	54 620	5.91	47.10	-0.079	细胞膜

表 5 *PmUGTs* 蛋白的二级结构

Table 5 Secondary structure of *PmUGTs* protein from *P. multiflora*.

蛋白名称	无规则卷曲/%	α-螺旋/%	延伸链/%	β-折叠/%
PmUGT1	38.97	41.55	14.12	5.37
PmUGT2	42.43	37.05	14.54	5.98
PmUGT3	35.66	44.06	13.93	6.35
PmUGT4	38.41	42.89	12.40	6.30

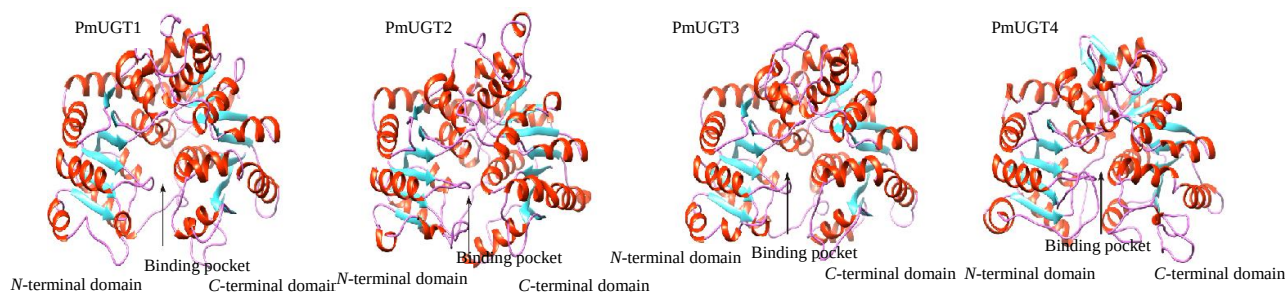


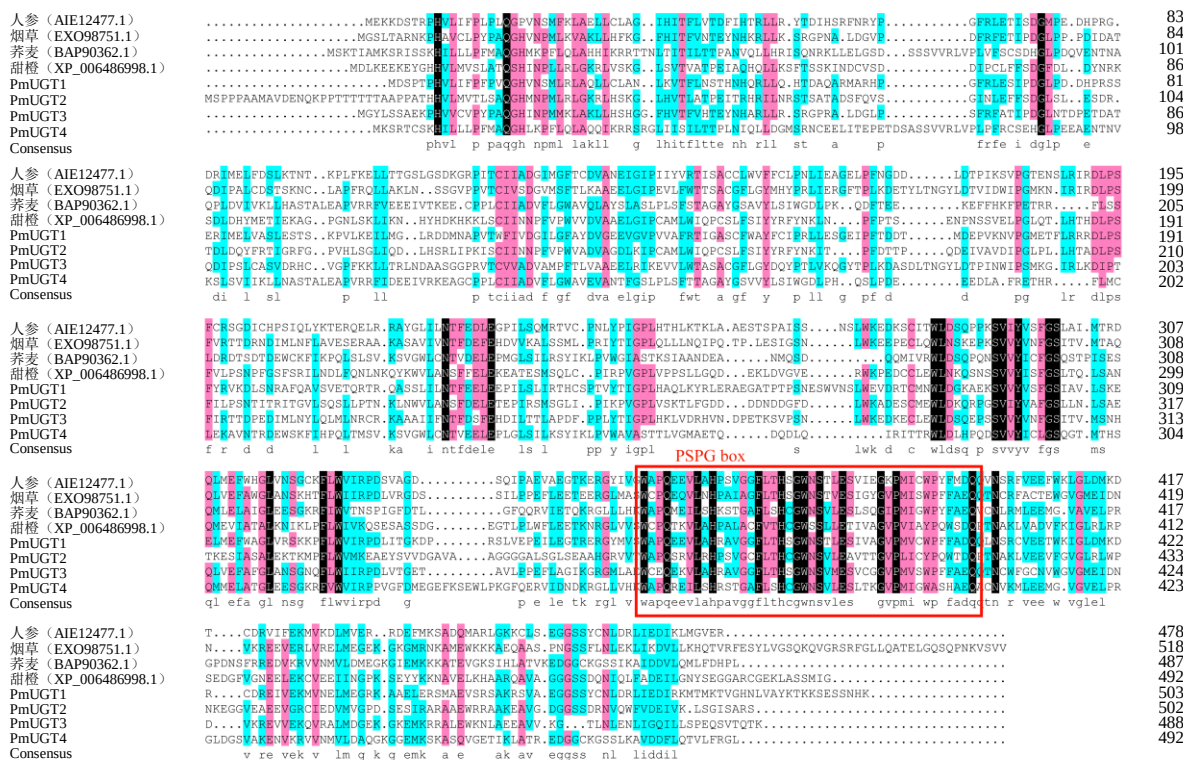
图 2 PmUGTs 蛋白的三级结构

Fig. 2 Tertiary structure prediction of PmUGTs

3.2.3 PmUGTs 氨基酸序列与系统进化树分析 将 PmUGTs 氨基酸序列在 BLASTP 中进行比对后, 选择相似性高的氨基酸序列在 DNAMAN 8.0 软件中进行多重序列比对, 结果如图 3 所示, PmUGTs 与其他物种 UGT 相似性为 42.26%, N 端同源性较 C 端低。供比对的 UGTs 都存在一段由 44 个氨基酸组成的 PSPG 盒 (plant secondary product glycosyltransferase box), 其中还包含一段高度保守的氨基酸序列 HCGWNS。选择不同植物的 UGT 构建系统发育树分析, 结果如图 4 所示, PmUGT1、

PmUGT3 聚在同一支, PmUGT2 独在一支, 三者与拟南芥的糖基转移酶蛋白距离较近; 而 PmUGT4 与蒺藜苜蓿距离则较近。

3.2.4 PmUGTs 蛋白与底物分子对接 根据 SWISS-MODEL 进行同源建模可知, PmUGT1 和 PmUGT3、PmUGT2、PmUGT4 分别以蒺藜苜蓿的 UGT85H2、罗汉果的 UGT74AC1、垂序商陆的 PaGT3 作为模板构建蛋白三维模型。使用 autodock 4.2.3 对 PmUGT1 和 PmUGT3、PmUGT2、PmUGT4 分别与其分子模板主要的反应底物(鹰嘴豆芽素 A、



黑色: 相似性等于 100%; 粉红色: 75% ≤ 相似性 < 100%; 浅蓝色: 50% ≤ 相似性 < 75%; 红色框区域代表 UGT 高度保守的 PSPG 盒结构域
black: homology=100%; pink: 75% ≤ homology < 100%; light blue: 50% ≤ homology < 75%; the red frame indicates the PSPG box which is highly conserved in UGT

图 3 PmUGTs 与其他植物糖基转移酶序列比对结果

Fig. 3 Multiple sequence alignment of PmUGTs and UGT from other plant species

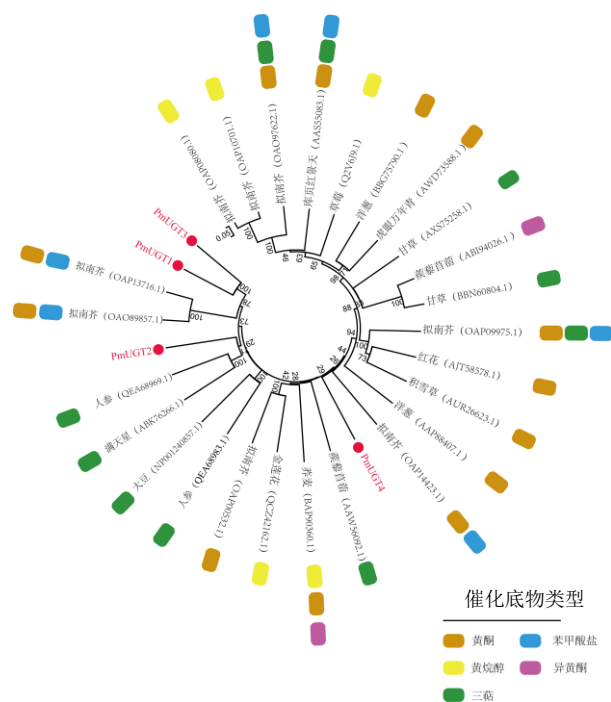


图 4 PmUGTs 蛋白氨基酸序列的系统发育树

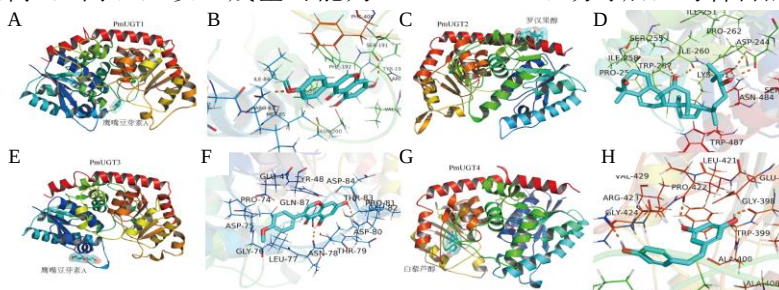
Fig. 4 Phylogenetic tree of amino acid sequence of PmUGTs

罗汉果醇以及白藜芦醇) 进行分子对接模拟实验。结果表明(图 5), 鹰嘴豆芽素 A 与 PmUGT1 的 Ser191、Ile84 [(抑制常数, K_i) = 44.07 $\mu\text{mol/L}$, (结合自由能, ΔG_{bind}) = -5.94] 和 PmUGT3 的 ASN78、ASP80 (K_i = 8.59 $\mu\text{mol/L}$, ΔG_{bind} = -6.91) 通过氢键对接实现互相作用(氢键键程都在 0.25 nm 以内), 推测以上残基可能是 PmUGT1、PmUGT3 识别鹰嘴豆芽素 A 并作用的关键残基; 罗汉果醇对 PmUGT2 的 ILE260、ASP244 与 ASN484 (K_i = 9.94 $\mu\text{mol/L}$, ΔG_{bind} = -6.82) 通过 5 个氢键对接(氢键键程在 0.28 nm 以内), 同理, 以上残基可能为

PmUGT2 识别罗汉果醇的关键残基; 白藜芦醇对 PmUGT4 的 PRO422、GLY398 与 GLY424 (K_i = 22.48 $\mu\text{mol/L}$, ΔG_{bind} = -6.34) 通过 3 个氢键连接(氢键键程在 2.3A 以内), 即以上残基可能为白藜芦醇与 PmUGT4 相互作用的关键残基。

3.3 PmUGTs 的原核表达

使用 pET-28a 作为原核表达载体, 利用同源重组原理构建 pET-28a-PmUGTs 重组表达载体, 转化到 *Transetta* (DE3) 感受态大肠杆菌中诱导相关表达蛋白。SDS-PAGE 结果如图 6 所示, pET-28a-PmUGTs 经诱导后, 与含目的基因的未诱导蛋白以

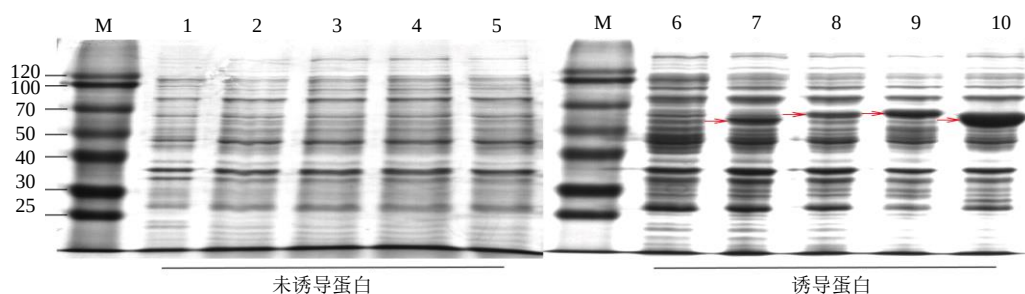


A、B-鹰嘴豆芽素 A 与 PmUGT1 对接的整体、局部构象 C、D-罗汉果醇与 PmUGT2 对接的整体、局部构象 E、F-鹰嘴豆芽素 A 与 PmUGT3 对接的整体、局部构象 G、H-白藜芦醇与 PmUGT4 对接的整体、局部构象; 蛋白与配体的连接在局部构象图中以虚线表示

A, B-the overall and local conformation of PmUGT1 docking with biochanin A C, D-the overall and local conformation of PmUGT2 docking with mogrol E, F-the overall and local conformation of PmUGT3 docking with biochanin A G, H-the overall and local conformation of PmUGT4 docking with resveratrol; The binding of proteins to ligands is shown by dotted lines in the local conformation

图 5 PmUGTs 与相应化合物配体的分子对接

Fig. 5 Molecular dockings of PmUGTs and relevant ligands



M-Marker 1~5 未诱导蛋白 6~10-诱导蛋白 1、6-pET-28a 空白载体蛋白 2、7-PmUGT1 重组蛋白 3、8-PmUGT2 重组蛋白 4、9-PmUGT3 重组蛋白 5、10-PmUGT4 重组蛋白 红色箭头所指即为相应蛋白条带

M-Marker 1—5 Uninduced protein 6—10 induced protein 1, 6-pET-28a empty vector protein 2, 7-PmUGT1-pET-28a recombinant protein 3, 8-PmUGT2-pET-28a recombinant protein 4, 9-PmUGT3-pET-28a recombinant protein 5, 10-PmUGT4-pET-28a recombinant protein. PmUGT1, PmUGT2, PmUGT3, PmUGT4 protein bands were pointed out by the red arrow, respectively

图 6 PmUGTs 重组蛋白 SDS-PAGE 电泳图

Fig. 6 SDS-PAGE analysis of the expression of recombinant PmUGTs

及不含目的基因的空白载体诱导蛋白进行对比, 在 50 000~70 000 诱导样品 6~10 皆存在明显的蛋白条带, 与所预测 PmUGT1、PmUGT2、PmUGT3、PmUGT4 的相对分子质量 56 940、54 780、54 350 和 54 620 结果相近, 所以图 6 中红色箭头所指 4 个蛋白条带分别为 PmUGT1、PmUGT2、PmUGT3 与 PmUGT4 的重组蛋白。

3.4 组织特异性表达分析

利用 qRT-PCR 分析 *PmUGTs* 基因在何首乌根、茎、叶中的相对表达水平。结果如图 7 所示, *PmUGT1*、*PmUGT3* 与 *PmUGT4* 中均呈现叶表达水平最高, 茎次之, 而根则最低的趋势。根据显著性差异分析得出, *PmUGT1* 与 *PmUGT3*

基因在茎与根的相对表达量相对于叶均下调, 且具有显著性差异 ($P < 0.05$); *PmUGT4* 基因的根相对于叶、茎的相对表达量均下调, 分别呈现极显著性差异 ($P < 0.01$) 与显著性差异的结果; 而在 *PmUGT2* 中, 茎相对于根与叶的相对表达量均下调, 分别有极显著性差异与显著性差异的结果。

4 讨论

何首乌中存在着丰富的糖苷类化合物, 如二苯乙烯苷类、黄酮苷类以及蒽醌苷类成分等, 暗示着其中具有多种相应催化的糖基转移酶。至今仍未发现关于何首乌糖基转移酶的报道, 为此, 本研究以前期实验室获得的转录组测序数据为基础, 成功从

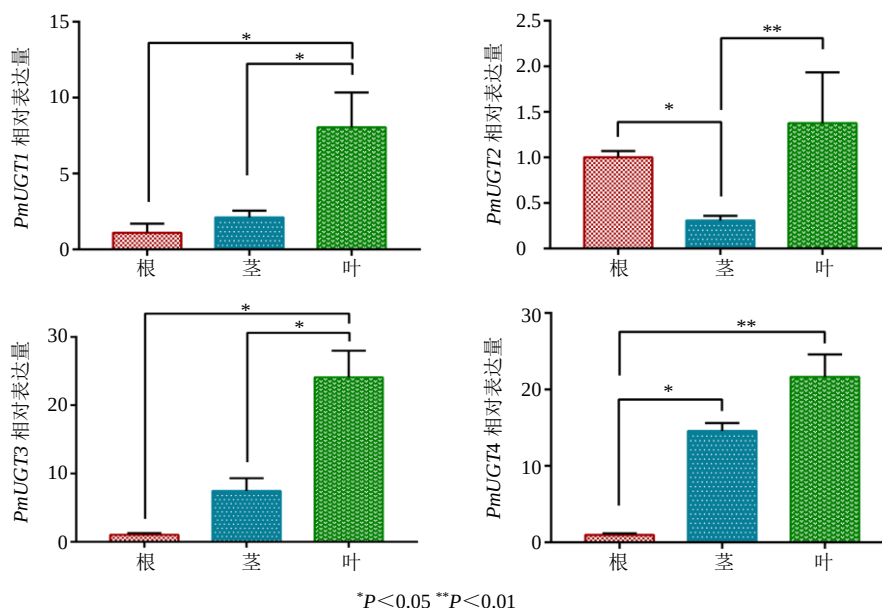


图 7 *PmUGTs* 在何首乌不同组织中的相对表达水平

Fig. 7 Relative expression level of *PmUGTs* gene in different tissues

何首乌叶片中克隆出了 4 条糖基转移酶基因,并分别命名为 *PmUGT1*、*PmUGT2*、*PmUGT3*、*PmUGT4*。对 *PmUGTs* 的密码子偏好模式进行分析,发现其 CAI 与 ENC 均值为 0.358 4 与 53.681,表明 *PmUGTs* 的基因密码子偏好性较弱;但对 ACC 密码子具有极强偏好 (RSCU>2.00);且其 GC_{3s} 均>50%,该结果不符合大部分双子叶植物密码子偏好使用 A/U 结尾的特征,可能是在进化过程中产生突变,但是其具体原因还有待进一步研究。*PmUGTs* 基因编码的蛋白 C 端结构域均可发现一段含有糖基转移酶识别 (WAPQV)、结合供体分子位点 (HCGWNS) 的 PSPG box 保守序列^[13-16],且该 PSPG box 的第 22、23 与 44 位氨基酸种类一般会影响 GT 的供体分子类型^[17-18],*PmUGTs* 在以上位点皆分别为色氨酸、天冬氨酸与谷氨酰胺,由此可推测 *PmUGTs* 的糖基供体极可能为 UDP-葡萄糖。GT 根据其结构特性一般可分为 GT-A 和 GT-B 2 种折叠类型^[19]。其中 GT-B 型是由 2 个正对的 $\beta/\alpha/\beta$ 类 Rossmann 折叠区域构成^[20],一个为 N 端结构域,另一个则为 C 端结构域,且 N 端同源性较 C 端相比低,二者之间还有一个结合底物的口袋^[20-21],观察本实验所预测的 4 个 *PmUGTs* 蛋白三维构型均与 GT-B 型相符合。另外,本研究通过将 *PmUGTs* 基因与 pET-28a 质粒构建原核表达载体,诱导重组蛋白的表达,经过 SDS-PAGE 验证得 pET-28a-*PmUGTs* 重组蛋白成功表达。何首乌中的糖苷类成分生物合成仍存在许多未知,因此,为探究何首乌糖苷类成分的主要合成部位及可能存在的作用,本实验采用 qRT-PCR 的方法研究 *PmUGTs* 基因在何首乌各组织中的相对表达情况。结果显示 *PmUGTs* 在何首乌各组织中存在特异性表达,总体而言,*PmUGTs* 在根中皆表现出较低的表达量。张伦等^[22]研究表明,何首乌块根中二苯乙烯苷含量在 6 到 10 月逐渐升高,10 月达到峰值,然后一直到 12 月都逐渐降低,本研究 qRT-PCR 实验所使用的样品采收于 11 月,所得结果与同时期二苯乙烯苷的含量逐渐减少相呼应,可能由于盛花期由 10 月开始,随后进入果熟期,此时营养生长已停止,糖基转移酶基因的表达量也随之降低。

由于糖苷类化学成分的重要性以及相关研究技术的发展,截止 2020 年 11 月,Carbohydrate Active enZymes (CAZy) 数据库中已分类有 111 个糖基转移酶家族 (GT1~111)^[16,24],其中 GT1 为最大

族^[24],可以催化萜类、生物碱类、黄酮类、苯丙素类以及木脂素类的糖基化^[25]反应。一般通过分析 GTs 对底物偏好的生化数据,使用系统发育树可以很好地对新发现的 GTs 的底物特异性进行预测^[26-27]。不仅如此,随着 GT 酶结构的研究深入,还可以将已知的模板蛋白所能催化的受体底物,通过计算机模拟进行目标蛋白的底物预测^[28]。经文献报道^[10-12],UGT85H2 是莢藜苜蓿的一种重要的糖基转移酶,是 *PmUGT1*、*PmUGT3* 蛋白同源建模所用分子模板,能够对鹰嘴豆芽素 A 等化合物反应生成糖基化产物;同时,UGT74AC1 与 PaGT3 分别对罗汉果醇与白藜芦醇都具有糖基化作用,分别为 *PmUGT2*、*PmUGT4* 蛋白同源建模所用分子模板。根据该结果,本实验利用 autodock4 软件与 *PmUGTs* 的三维构象分子模板的反应底物分别进行模拟对接实验,结果得出 *PmUGT1* 与 *PmUGT3*、*PmUGT2*、*PmUGT4* 上的残基分别可以与鹰嘴豆芽素 A、罗汉果醇以及白藜芦醇以氢键连接发生互相作用,且所得 K_i 、 ΔG_{bind} 指数均在合理范围之内;且根据前文可知 *PmUGTs* 均以 UDP-葡萄糖作为糖基供体。由此可推测:*PmUGT1* 可催化鹰嘴豆芽素 A 的 5 位羟基,4'位甲氧基生成糖基,形成 5,7-二羟基-4'-甲氧基异黄酮-5-O- β -D-葡萄糖苷与 5,7-二羟基-4'-甲氧基异黄酮-4'-O- β -D-葡萄糖苷 2 种产物;*PmUGT3* 可催化鹰嘴豆芽素 A 的 5,7 位羟基生成糖基,形成 5,7-二羟基-4'-甲氧基异黄酮-5-O- β -D-葡萄糖苷与 5,7-二羟基-4'-甲氧基异黄酮-7-O- β -D-葡萄糖苷 2 种产物;而 *PmUGT2* 可催化罗汉果醇的 11、23、24 位羟基生成糖基,形成罗汉果醇-11-O- β -D-葡萄糖苷、罗汉果苷 IA1 (CAS: 88901-46-6)、罗汉果醇-23-O- β -D-葡萄糖苷;*PmUGT4* 则可催化白藜芦醇 3,5,4' 位羟基生成糖基,形成虎杖苷 (CAS: 27208-80-6)、白藜芦醇-5-O- β -D-葡萄糖苷与白藜芦醇-4'-O- β -D-葡萄糖苷 3 种产物。以上结果可为后续进一步对 *PmUGTs* 功能机制的研究提供理论基础。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] McArthur J B, Chen X. Glycosyltransferase engineering for carbohydrate synthesis [J]. *Biochem Soc Trans*, 2016, 44(1): 129-142.
- [2] de Bruyn F, Maertens J, Beauprez J, et al. Biotechnological advances in UDP-sugar based glycosylation of small molecules [J]. *Biotechnol Adv*,

- 2015, 33(2): 288-302.
- [3] Bowles D, Lim E K, Poppenberger B, et al. Glycosyltransferases of lipophilic small molecules [J]. *Annu Rev Plant Biol*, 2006, 57: 567-597.
- [4] 郭淑, 罗红梅, 宋经元, 等. 糖基转移酶在植物次生代谢途径中的研究进展 [J]. 世界科学技术—中医药现代化, 2012, 14(6): 2126-2130.
- [5] 刘潇斐, 张良, 冯旭东, 等. 偶联尿苷二磷酸循环体系的天然产物体外糖基化修饰 [J]. 化工进展, 2020, 39(1): 329-340.
- [6] 中国药典 [S]. 一部. 2020: 183.
- [7] 黄世琼, 张毅, 杨军宣. 何首乌主要成分二苯乙烯苷的研究进展 [J]. 海峡药学, 2016, 28(6): 37-39.
- [8] 任红微, 魏静, 高秀梅, 等. 何首乌及其主要化学成分药理作用及机制研究进展 [J]. 药物评价研究, 2018, 41(7): 1357-1362.
- [9] 王浩, 蔡启忠, 刘露, 等. 何首乌实时荧光定量 PCR 内参基因筛选 [J]. 中国中药杂志, 2021, 46(1): 80-85.
- [10] Li L N, Modolo L V, Escamilla-Trevino L L, et al. Crystal structure of *Medicago truncatula* UGT85H2: Insights into the structural basis of a multifunctional (Iso)flavonoid glycosyltransferase [J]. *J Mol Biol*, 2007, 370(5): 951-963.
- [11] Li J, Yang J G, Mu S C, et al. Efficient O-glycosylation of triterpenes enabled by protein engineering of plant glycosyltransferase UGT74AC1 [J]. *ACS Catal*, 2020, 10(6): 3629-3639.
- [12] Maharjan R, Fukuda Y, Nakayama T, et al. Crown-ether-mediated crystal structures of the glycosyltransferase PaGT3 from *Phytolacca americana* [J]. *Acta Crystallogr D Struct Biol*, 2020, 76(Pt 6): 521-530.
- [13] Bairoch A. PROSITE: a dictionary of sites and patterns in proteins [J]. *Nucleic Acids Res*, 1992, 20(Suppl): 2013-2018.
- [14] Hughes J, Hughes M A. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (*Manihot esculenta* Crantz) cotyledons [J]. *DNA Seq*, 1994, 5(1): 41-49.
- [15] Altschul S F, Madden T L, Schäffer A A, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs [J]. *Nucleic Acids Res*, 1997, 25(17): 3389-3402.
- [16] Campbell J A, Davies G J, Bulone V V, et al. A classification of nucleotide-diphospho-sugar glycosyl transferases based on amino acid sequence similarities [J]. *Biochem J*, 1998, 329(Pt 3): 719.
- [17] 吕鹤书, 薛飞燕, 柳春梅, 等. 植物尿苷二磷酸糖基转移酶超家族晶体结构 [J]. 生物工程学报, 2014, 30(6): 838-847.
- [18] Wang Y L, Huang W J, Wang Y. Cloning and expression analysis of the EsUF3GT Gene in *Epimedium sagittatum* (Sieb. and Zucc.) Maxim. [J]. *Plant Science Journal*, 2014, 32(6): 602-611.
- [19] Bourne Y, Henrissat B. Glycoside hydrolases and glycosyltransferases: Families and functional modules [J]. *Curr Opin Struct Biol*, 2001, 11(5): 593-600.
- [20] Lairson L L, Henrissat B, Davies G J, et al. Glycosyltransferases: structures, functions, and mechanisms [J]. *Annu Rev Biochem*, 2008, 77: 521-555.
- [21] Hu Y N, Walker S. Remarkable structural similarities between diverse glycosyltransferases [J]. *Chem Biol*, 2002, 9(12): 1287-1296.
- [22] 张伦, 谭凯丽, 廖海民, 等. 何首乌生长过程中二苯乙烯苷的含量变化 [J]. 西北植物学报, 2010, 30(7): 1481-1484.
- [23] Coutinho P M, Deleury E, Davies G J, et al. An evolving hierarchical family classification for glycosyltransferases [J]. *J Mol Biol*, 2003, 328(2): 307-317.
- [24] Le Roy J, Huss B, Creach A, et al. Glycosylation is a major regulator of phenylpropanoid availability and biological activity in plants [J]. *Front Plant Sci*, 2016, 7: 735.
- [25] Wen C, Huang W, He M M, et al. Cloning and characterization of a glycosyltransferase from *Catharanthus roseus* for glycosylation of cardiotonic steroids and phenolic compounds [J]. *Biotechnol Lett*, 2020, 42(1): 135-142.
- [26] Vogt T, Jones P. Glycosyltransferases in plant natural product synthesis: Characterization of a supergene family [J]. *Trends Plant Sci*, 2000, 5(9): 380-386.
- [27] Lim E K, Baldauf S, Li Y, et al. Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis* [J]. *Glycobiology*, 2003, 13(3): 139-145.
- [28] Singh S, Patel K A, Sonawane P D, et al. Enhanced activity of *Withania somnifera* family-1 glycosyl transferase (UGT73A16) via mutagenesis [J]. *World J Microbiol Biotechnol*, 2018, 34(10): 150.

[责任编辑 时圣明]