

拟南芥类结瘤素基因 *At1g01070* 的生物信息学分析

王丽,袁园园,赵盼盼,常卫东,王俐锋,王林嵩

(河南师范大学 生命科学学院,河南 新乡 453007)

摘要:类结瘤素基因在非结瘤植物生长发育中承担着重要的功能,但仅少数基因的功能被鉴定.本研究对拟南芥类结瘤素 MtN21(*Medicago truncatula* NODULIN 21)家族成员 *At1g01070* 进行了生物信息学分析及亚细胞定位验证.结果表明,*At1g01070* 有两种可变剪切体,编码两种蛋白 *At1g01070.1* 和 *At1g01070.2*,前者比后者在 N 端多 47 个氨基酸,它们的分子量分别为 40 KD 和 35 KD,等电点分别为 9.2 和 8.9;均含有 2 个 MtN21 蛋白的特征结构域 EamA(药物/代谢物转运结构域),*At1g01070.1* 比 *At1g01070.2* 多一个 PLN00411 结构域,且其靠近 N 端的 EamA 结构域较后者长;三级结构均由 7 个 α -螺旋和一些不规则卷曲组成,分别含有 10 个和 8 个疏水跨膜区,均含 13 个磷酸化修饰位点,亚细胞定位预测主要定位在质膜;进化上,与琴叶拟南芥(*Arabidopsis lyrata*)亲缘关系最近;亚细胞定位结果显示 *At1g01070.1* 定位在质膜,与预测结果一致.推测 *At1g01070* 在植物体内可能参与物质运输.

关键词:类结瘤素基因;*At1g01070*;MtN21;生物信息学分析;亚细胞定位

中图分类号:Q945.31

文献标志码:A

结瘤素基因(Nodulin Genes)是豆科植物在根瘤形成过程中所特异表达或表达增强的一类基因,主要参与根瘤的形态发生、发育、固氮活性的维持和固氮产物的代谢等过程^[1,2].此外,在非结瘤植物中也发现了其同源物.非结瘤植物中结瘤素基因的同源基因被称为类结瘤素基因(Nodulin-like genes),它们在拟南芥(*Arabidopsis thaliana*)、水稻(*Oryza sativa*)、玉米(*Zea mays* L.)和杨属(*Populus*)等植物中均以基因家族的形式存在^[3-5],在植物生长发育中可能发挥着更为广泛的功能.

拟南芥中类结瘤素基因共 132 个,被分成 7 个家族,虽只有少数成员的功能被鉴定,但已鉴定的类结瘤素蛋白大多具有运输功能^[3].*AtWAT1* 和 *AtSIAR1* 是拟南芥类结瘤素 MtN21(*Medicago truncatula* NODULIN 21)家族中仅被功能鉴定的两个成员.其中,*AtWAT1*(WALLS ARE THIN1)是定位在液泡膜上的生长素转运体,可从液泡向胞质运输生长素,其编码基因突变可引起植株的矮化,同时提高植株对导管类病原菌的免疫力^[6-8].*AtSIAR1*(Siliques Are Red1)是定位在质膜上的氨基酸双向转运体,在拟南芥发育过程中有机氮的分配尤其是维持氨基酸的平衡方面发挥重要的作用^[9].*At1g01070* 也是 MtN21 家族成员之一,有研究表明其可能与生长素的代谢或转运有关^[6,10],但具体功能不清楚.为此,本研究利用生物信息学的方法对 *At1g01070* 基因和蛋白的结构、理化性质、磷酸化位点和系统进化进行分析,并构建了 *At1g01070.1* 与 GFP 融合的植物表达载体,进行亚细胞定位验证,为深入研究该基因的生物功能奠定基础.

1 材料与方法

1.1 植物材料

Col-0 野生型拟南芥(*Arabidopsis thaliana*)和本氏烟草(*Nicotiana glauca*)种子、大肠杆菌

收稿日期:2016-11-25;修回日期:2017-03-20.

基金项目:国家自然科学基金(31601241);河南省高校科技创新团队(15IRTSTHN020);河南师范大学博士科研启动基金(39970774).

作者简介:王丽(1980-),女,河南新乡人,河南师范大学讲师,博士,研究方向为植物生长发育调控,E-mail:wangli1015@htu.cn.

通信作者:王林嵩,教授,博士生导师,E-mail:wls@htu.cn.

(*Escherichia coli*) 菌株 DH5 α 和根癌农杆菌 (*Agrobacterium tumefaciens*) 菌株 GV3101 均由本实验室保存。载体 pMD18-T 购自大连宝生物公司,改造的表达载体 pCAMBIA1300 由中国农业大学植物生理学与生物化学国家重点实验室惠赠。

将拟南芥种子用含 0.5% NaClO 和 0.001% Triton X-100 的溶液消毒处理 10 min,再用无菌蒸馏水洗种子,并播种在 MS 培养基上,4 °C 冷处理 3 d 后,将其放于人工气候室中培养。培养条件:光照强度 150 $\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$,相对湿度 70%~80%,温度为(22 \pm 2) °C,16 h 光照/8 h 黑暗,培养两周的幼苗用于总 RNA 提取。本氏烟种子播于中性营养土中,置于人工气候室中,培养条件同拟南芥,培养 5 周左右用于瞬时表达分析。

1.2 亚细胞定位表达载体的构建

用植物 RNA 提取试剂盒(艾德莱,北京)提取两周的拟南芥幼苗总 RNA,用逆转录酶(M-MLV Reverse Transcriptase, Takara)合成 cDNA。根据 *At1g01070.1* (NM_099989.2) 的 ORF 序列设计特异性引物(上游引物:5'-CTGCAGATGGCTGGAGATATGCAAGGAG-3',下游引物:5'-GGTACCAACGGGCGACT-TAGAGTCGTT-3',下划线分别表示 *Pst* I 和 *Kpn* I 酶切位点,下游引物不含终止密码子),以 cDNA 为模板,用 KOD-Plus 高保真酶(TOYOBO, Japan)进行 RT-PCR 扩增。之后将扩增的 *At1g01070.1* 开放阅读框序列连接到含 p-super promoter 和 GFP 的 pCAMBIA1300 植物表达载体上,测序并确认末端融合后转化农杆菌并侵染植物。

1.3 农杆菌介导的烟草瞬时表达和亚细胞定位观察

将携带了 *At1g01070.1*-GFP 的农杆菌在 28 °C 条件下,含有 50 $\mu\text{g} \cdot \text{mL}^{-1}$ 卡那霉素和 50 $\mu\text{g} \cdot \text{mL}^{-1}$ 利福平的 LB 液体培养基中培养至对数生长期,取 1 mL 菌液 12 000 r/min 离心 1 min,收集菌体后用悬浮液(10 mmol \cdot L $^{-1}$ MES pH 5.6, 10 mmol \cdot L $^{-1}$ MgCl $_2$, 和 150 $\mu\text{mol} \cdot$ L $^{-1}$ 乙酰丁香酮)悬浮,调整菌液浓度 OD_{600} =0.6~0.8,室温放置 3 h。用 1 mL 注射器吸取适量菌液,用无针头的针管从本氏烟叶片背部将菌液慢慢注入叶片组织空隙中,浸润接种的植株于 25 °C 暗培养 48 h,剪取幼嫩的茎制备组织切片,在激光共聚焦显微镜下观察 GFP 信号,GFP 的激发光波长为 488 nm,发射光波长为 498~551 nm。

FM4-64 细胞膜共定位染料的使用方法参考 Jelinkova 等^[11]。将制备好的含 *At1g01070.1*-GFP 的烟草幼茎用 2 μM FM4-64 暗处浸润 2~6 min,之后用无菌水清洗几次,用激光共聚焦显微镜观察 FM4-64 红色荧光,激发光波长为 561 nm,发射光波长为 610~671 nm。

1.4 *At1g01070* 基因的生物信息学分析

基因结构信息通过 NCBI (<http://www.ncbi.nlm.nih.gov/gene/>) 获取^[12]。蛋白质结构域组织信息通过 NCBI Conserved Domain Search (CDD, <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) 平台获取^[13]。利用 ProtParm 程序(<http://web.expasy.org/protparam>) 获取各蛋白质的相对分子质量及等电点预测信息^[14]。亲水/疏水性分析利用在线工具 ProtScale (<http://www.expasy.ch/tools/protscale.html>) 完成^[15]。蛋白质的跨膜螺旋结构域利用在线工具 TMHMM v2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0>) 完成^[16]。通过 Phyre 2.0 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) 和 SWISS-MODEL(<https://swissmodel.expasy.org/>) 分析蛋白的三级结构^[17,18]。亚细胞定位预测通过拟南芥 SUBA (TAIR The SubCellular Proteomic Database) 数据库获得。通过 NCBI 查找 *At1g01070* 同源蛋白序列,利用 Clustal W 对序列进行完全比对,用 MEGA5.0 软件,通过邻接法(Neighbor-Joining,简称 NJ) 构建基因分子进化树。

2 结果

2.1 *At1g01070* 基因序列及其编码蛋白保守结构域分析

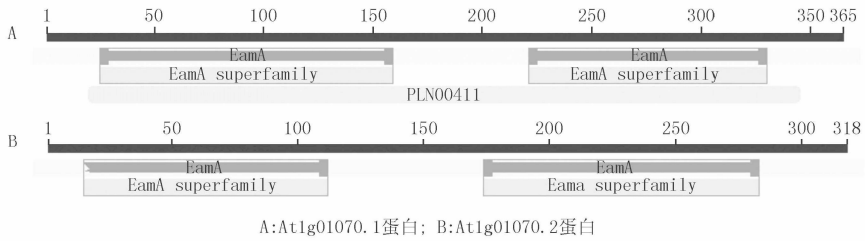
At1g01070 基因位于拟南芥 1 号染色体上,全长 2193bp,有 5 个内含子和 6 个外显子,存在两种可变剪切体 *At1g01070.1* 和 *At1g01070.2*,它们的编码区全长分别为 1098bp 和 957bp(图 1),这是由于拼接时第 1 个外显子(203bp)缺失和第二个外显子的 5' 剪接位点上移了 62bp 引起的,由此, *At1g01070.1* 蛋白和

Atlg01070.2 蛋白分别由 365 个和 318 个氨基酸组成,前者比后者多 47 个氨基酸,此外,N 端差异氨基酸有 20 个.

利用 NCBI CDD 数据库对 Atlg01070.1 和 Atlg01070.2 蛋白保守结构域进行分析(图 2),发现前者含有 2 个 EamA 和 1 个 PLN00411 保守结构域,而后者仅含有 2 个 EamA 保守结构域,并且与 Atlg01070.2 蛋白相比,Atlg01070.1 蛋白靠近 N 端的 EamA 结构域较长.



图1 拟南芥Atlg01070基因结构

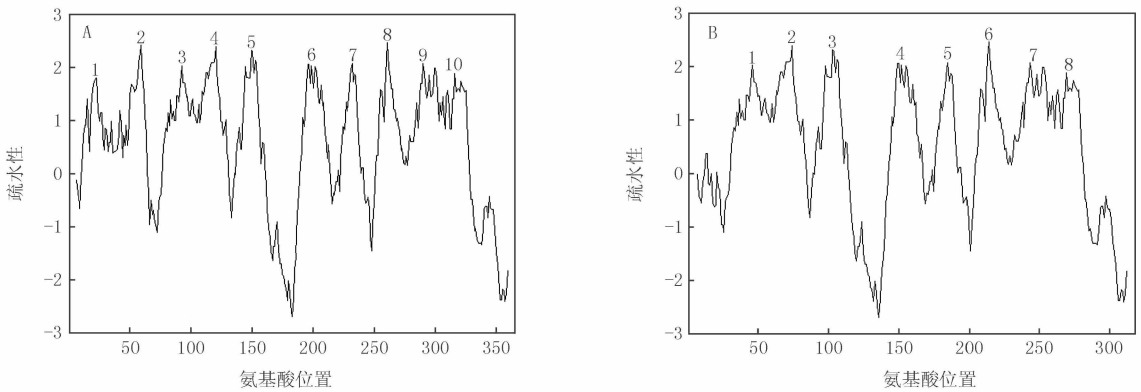


A:Atlg01070.1蛋白; B:Atlg01070.2蛋白

图2 拟南芥Atlg01070蛋白保守结构域预测

2.2 Atlg01070 蛋白的理化性质分析

利用在线工具 ProtParam 预测,发现 Atlg01070.1 蛋白和 Atlg01070.2 蛋白的理论等电点 pI 分别为 9.2 和 8.9,分子量分别为 40 KD 和 35KD. 利用 ProtScale 在线工具分析它们的疏水性/亲水性(图 3),发现两种蛋白均表现出较强的疏水性,经 TMHMM v2.0 对其跨膜区分析,发现 Atlg01070.1 蛋白有 10 个跨膜区,每个 EamA 结构域有 5 个跨膜区,而 Atlg01070.2 蛋白有 8 个跨膜区,靠近 N 端的 EamA 结构域含有 3 个跨膜区,比 Atlg01070.1 蛋白少 2 个跨膜区,而靠近 C 端的 EamA 结构域有 5 个跨膜区,与 Atlg01070.1 编码蛋白相同.



A:Atlg01070.1蛋白,疏水区编号从1至10;B:Atlg01070.2蛋白,疏水区编号从1至8;Windows size=13.

图3 Atlg01070蛋白疏水性

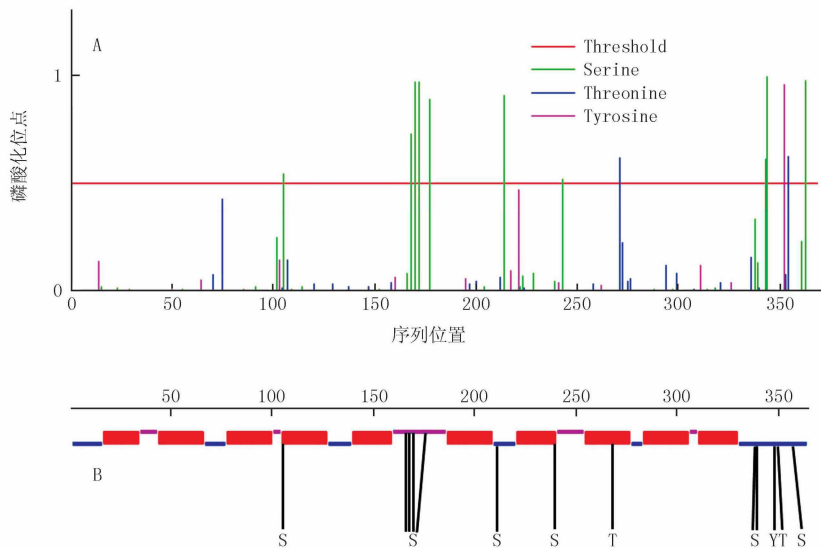
2.3 Atlg01070 蛋白磷酸化位点分析

利用在线软件 NetPhos 2.0 Server 预测磷酸化位点(图 4),发现 Atlg01070.1 蛋白有 1 个酪氨酸(Tyr)、10 个丝氨酸(Ser)、2 个苏氨酸(Thr)磷酸化位点. 结合蛋白跨膜结构域预测分析可知,Atlg01070.1 蛋白第 214、343、344、363 位的 Ser,354 位的 Thr 以及 352 位的 Tyr 均位于细胞膜内,这些氨基酸主要位于 C 端. Atlg01070.2 蛋白总的磷酸化位点和膜内侧的磷酸化位点数目均与 Atlg01070.1 蛋白相同.

2.4 Atlg01070 蛋白三级结构和亚细胞定位预测

利用 phyre 2 和 SWISS-MODEL 软件分别分析了 Atlg01070 蛋白的三级结构,结果一致,即

Atlg01070.1 和 *Atlg01070.2* 的空间结构均由 7 个 α 螺旋和一些不规则卷曲经缠绕折叠形成. 利用拟南芥 SubCellular Proteomic Database (SUBA) 预测 *Atlg01070.1* 和 *Atlg01070.2* 编码蛋白定位在质膜的可能性分别为 92.6% 和 98.1%.

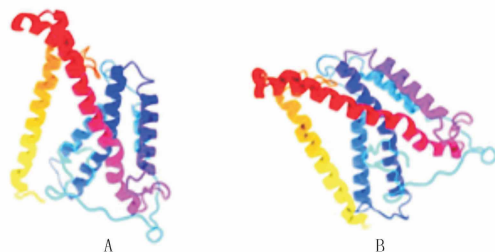


A: *Atlg01070.1* 蛋白的磷酸化位点; B: *Atlg01070.1* 蛋白磷酸化位点在二级结构上的分布, 其中上方细线代表膜外侧, 下方细线代表膜内侧, 粗线代表跨膜螺旋, T, S, Y 分别代表 Thr, Ser 和 Tyr.

图4 *Atlg01070.1* 蛋白磷酸化位点分析

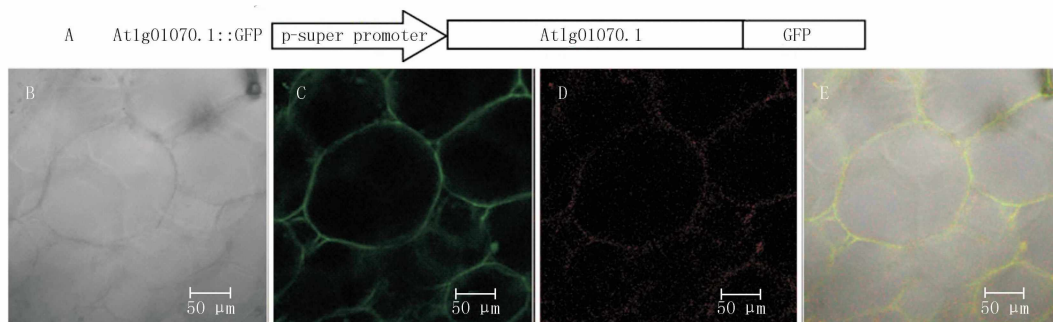
2.5 *Atlg01070.1* 的亚细胞定位验证

与 *Atlg01070.2* 蛋白相比, *Atlg01070.1* 蛋白跨膜区较多, 且两者亚细胞定位预测均为质膜, 由此, 本研究仅对 *Atlg01070.1* 的亚细胞定位进行了验证. 将 *Atlg01070.1* 基因与 GFP 融合后转化入农杆菌并侵染本氏烟观察该基因的亚细胞定位, 如图 6 所示, *Atlg01070.1::GFP* 融合蛋白的 GFP 绿色荧光出现在转基因烟草茎细胞的细胞膜或细胞壁上(图 6C), 用细胞膜染料 FM4-64 染转基因烟草茎的细胞显现红色荧光(图 6D), 且与 *Atlg01070.1::GFP* 融合蛋白的 GFP 绿色荧光完全重叠, 表现为黄绿色(图 6E), 表明 *Atlg01070.1* 定位在细胞质膜上.



A: *Atlg01070.1* 蛋白; B: *Atlg01070.2* 蛋白

图5 *Atlg01070* 蛋白三级结构预测



A: *Atlg01070.1::GFP* 融合载体构建; B: 明场; C: *Atlg01070.1::GFP* 融合蛋白的绿色荧光; D: 细胞膜染料 FM4-64 的红色荧光; E: C 和 D 叠加.

图6 *Atlg01070.1* 在烟草茎中的亚细胞定位

2.6 *Atlg01070* 基因系统进化树分析

使用 NCBI 中的 Blastp 程序, 搜索拟南芥 *Atlg01070* 的同源性蛋白, 并采用 MEGA5.0 软件构建系统进化树, 由图 7 可知, *Atlg01070.1* 蛋白(NP_563617)与琴叶拟南芥(*Arabidopsis lyrata*)、亚麻芥(*Cameli-*

na sativa)、甘蓝(*Brassica oleracea*)、盐芥(*Eutrema salsugineum*)、荠菜(*Capsella rubella*)、油菜(*Brassica napus*)、白菜(*Brassica rapa*)和 高山南芥(*Arabis alpine*)蛋白的氨基酸序列一致性均达到 80% 以上,而与 麻疯树(*Jatropha curcas*)、巨桉(*Eucalyptus grandis*)、大豆(*Glycine max*)和雷蒙德氏棉(*Gossypium raimondii*)亲缘关系较远。

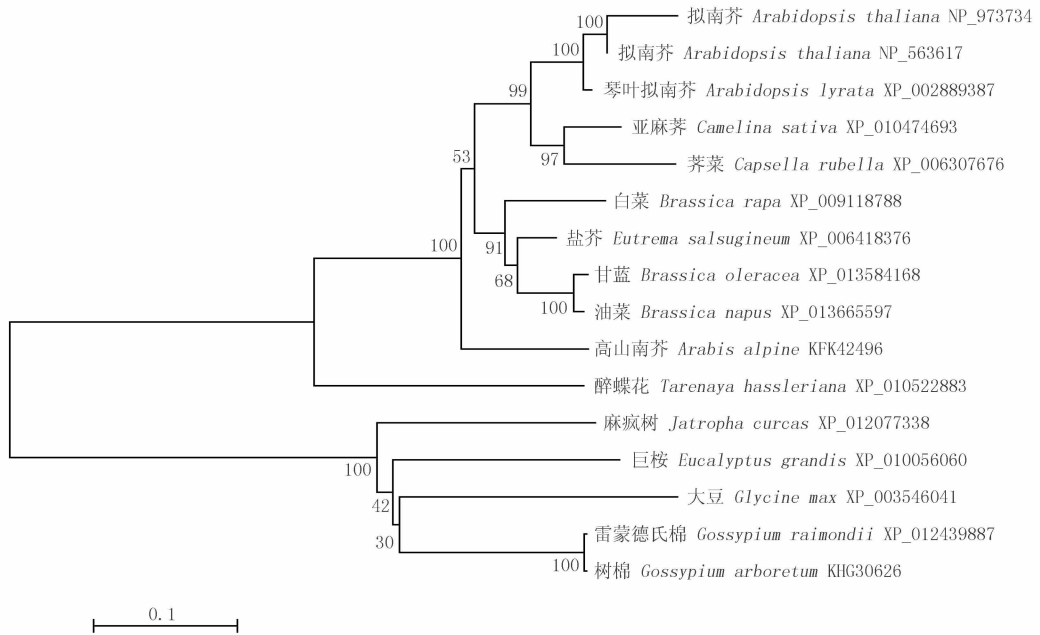


图7 拟南芥At1g01070基因编码序列进化分析

3 讨 论

随着类结瘤素基因的发现,它们在非结瘤植物中的生物功能越来越受到关注.本研究对拟南芥类结瘤素 MtN21 家族成员之一 *At1g01070* 进行生物信息学分析,并对其亚细胞定位进行了验证,为进一步揭示其生物功能奠定基础.

选择性剪接是重要的转录后调控机制,通过选择性剪接可以提高蛋白质的多样性以及影响 mRNA 的稳定性^[19].拟南芥中超过 42% 的基因涉及选择性剪接^[20].本研究中 *At1g01070* 存在两种可变剪切体 *At1g01070.1* 和 *At1g01070.2*,后者比前者少了第一个外显子并增加了第一个内含子的部分片段,导致后者编码的蛋白序列比前者编码的蛋白序列短,因而 *At1g01070.2* 的分子量和等电点均比 *At1g01070.1* 低,但 *At1g01070.2* 蛋白第 20 位之后的氨基酸序列与 *At1g01070.1* 蛋白第 68 位之后的氨基酸序列完全一致,因此,两者的一级结构仅在 *At1g01070.1* 蛋白 N 端的 67 个氨基酸存在差异.蛋白质的一级结构决定其空间结构.*At1g01070* 编码的两种蛋白保守结构域较为相似,均有 2 个 MtN21 家族的特征结构域 EamA(也称 DUF6, domain unknown function 6)^[3],EamA 命名源于对大肠杆菌氨基输出蛋白如 EmaA(O-乙酰胆碱和半胱氨酸)、YdeD(半胱氨酸,天冬酰胺和谷氨酰胺)、RhtA(苏氨酸和高丝氨酸)等的特征结构域(药物/代谢物转运结构域)的分析预测^[3, 21].但 *At1g01070.1* 蛋白和 *At1g01070.2* 蛋白在结构域组成上也存在差异,前者比后者多了一个 PLN00411 结构域,且其靠近 N 端的 EamA 结构域比后者长,这与两种蛋白在 N 端氨基酸序列组成上的差异相一致. PLN00411 结构域被认为是生育酚甲基转移酶的特征结构域^[22],表明 *At1g01070.1* 蛋白功能更为丰富.

蛋白质的空间结构影响其理化性质.*At1g01070* 编码的两种蛋白在三级结构上主要由 α -螺旋和不规则卷曲组成.疏水/亲水性分析表明两种蛋白均为疏水蛋白,具有多个跨膜区,但 *At1g01070.1* 蛋白在靠近 N 端的 67 个氨基酸内比 *At1g01070.2* 蛋白多了 2 个跨膜区,是由于后者刚好缺少了前者 N 端的这段氨基酸.拟南芥中类结瘤素蛋白可靶向不同的生物膜如质膜或液泡膜,因而承担不同的功能,但均与物质转运有关,

如营养物质、溶质、氨基酸和激素等的转运^[3,7,23]。亚细胞定位预测 *At1g01070* 编码的两种蛋白位于质膜的可能性最大,本研究构建了 *At1g01070.1*-GFP 融合表达载体并瞬时表达于本氏烟草中,发现 *At1g01070.1* 定位在质膜上,进一步验证了 *At1g01070.1* 蛋白为跨膜蛋白,并与文献报道的类结瘤素蛋白为跨膜蛋白相一致^[3]。植物中 nodulin 26-like 蛋白主要运输不带电荷的物质,它可发生翻译后磷酸化修饰,从而提高其转运的活性^[24],蛋白质发生磷酸化和去磷酸化的酶主要位于细胞内^[24],因此膜内侧的磷酸化位点对于调控蛋白活性更为重要。本研究中 *At1g01070* 编码的两种蛋白含有完全相同的磷酸化位点,表明两种蛋白转运活性的调控机制相似。

利用 BlastP 在 NCBI 数据库中搜索 *At1g01070* 同源蛋白,发现与其同源性较高的物种均属双子叶十字花科,表明十字花科中此类蛋白参与的代谢途径较为接近,而其他亲缘关系较远的双子叶植物中此类蛋白可能参与其他代谢途径。目前这些同源蛋白功能均未知,被命名为 WAT1-related 蛋白。拟南芥中 *AtWAT1* 与 *At1g01070* 蛋白同属 MtN21 家族,是定位在液泡膜上的生长素转运蛋白,与野生型拟南芥相比,其缺失突变体 *wat1* 的茎中色氨酸和生长素含量下降,同时发现 *At1g01070* 的表达下调^[6-8],表明 *At1g01070* 与 *AtWAT1* 关系密切,可能与生长素代谢有关。Li 等^[10]报道的拟南芥 *ptl-D* 功能获得性突变体生长素平衡受损,突变体中 *At1g01070* 的表达水平高于野生型。此外,拟南芥 MtN21 蛋白家族中与 *At1g01070* 编码蛋白同源性最高的 *At1g11450* 可参与植株生长调控^[3,25]。拟南芥 MtN21 蛋白家族中另一个成员 *AtSIAR1* 也定位在质膜上,参与氨基酸的输出^[7]。结合本研究对 *At1g01070* 基因的结构、理化性质、磷酸化位点和进化等的生物信息学分析及对其亚细胞定位的验证,推测 *At1g01070* 可能参与氨基酸的运输,并与生长素代谢有关,从而参与植物的生长发育调控,进一步的工作是对 *At1g01070* 具体功能进行研究。

参 考 文 献

- [1] Hohnjec N, Küster H, Albus U, et al. The broad bean nodulin VIENOD18 is a member of a novel family of plant proteins with homologies to the bacterial MJ0577 superfamily [J]. *Mol Gen Genet*, 2000, 264(3): 241-250.
- [2] Sousa C, Johansson C, Charon C, et al. Translational and structural requirements of the early nodulin gene *enod40*, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex [J]. *Mol Cell Biol*, 2001, 21(1): 354-366.
- [3] Denancé N, Szurek B, Noël L D. Emerging functions of nodulin-like proteins in non-nodulating plant species [J]. *Plant Cell Physiol*, 2014, 55(3): 469-474.
- [4] Jack D L, Yang N M, Saier M II Jr. The drug/metabolite transporter superfamily [J]. *Eur J Biochem*, 2001, 268(13): 3620-39.
- [5] Guillaumie S, Goffner D, Barbier O, et al. Expression of cell wall related genes in basal and ear internodes of silking brown-midrib-3, caffeic acid O-methyltransferase (COMT) down-regulated, and normal maize plants [J]. *BMC Plant Biol*, 2008, 8: 71.
- [6] Ranocha P, Denancé N, Vanholme R, et al. Walls are thin1 (WAT1), an Arabidopsis homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers [J]. *Plant J*, 2010, 63(3): 469-483.
- [7] Ranocha P, Dima O, Nagy R, et al. Arabidopsis WAT1 is a vacuolar auxin transport facilitator required for auxin homeostasis [J]. *Nat Commun*, 2013, 4: 2625.
- [8] Denancé N, Ranocha P, Oria N, et al. Arabidopsis *wat1* (*walls are thin1*)-mediated resistance to the bacterial vascular pathogen, *Ralstonia solanacearum*, is accompanied by cross-regulation of salicylic acid and tryptophan metabolism [J]. *Plant J*, 2013, 73(2): 225-239.
- [9] Ladwig F, Stahl M, Ludewig U, et al. Siliques Are Red1 from Arabidopsis acts as a bidirectional amino acid transporter that is crucial for the amino acid homeostasis of siliques [J]. *Plant Physiol*, 2012, 158(4): 1643-1655.
- [10] Li X, Qin G, Chen Z, et al. A gain-of-function mutation of transcriptional factor PTL results in curly leaves, dwarfism and male sterility by affecting auxin homeostasis [J]. *Plant Mol Biol*, 2008, 66(3): 315-327.
- [11] Jelinkova A, Malinska K, Simon S, et al. Probing plant membranes with FM dyes: tracking, dragging or blocking [J]? *Plant J*, 2010, 61(5): 883-892.
- [12] Reddy P S, Rao T S, Sharma K K, et al. Genome-wide identification and characterization of the aquaporin gene family in *Sorghum bicolor* (L.) [J]. *Plant Gene*, 2015, 1(1): 18-28.
- [13] Aron M B, Lu S, Anderson J B, et al. CDD: a Conserved Domain Database for the functional annotation of proteins [J]. *Nucleic Acids Res*, 2011, 39: 225-229.
- [14] Gasteiger E, Gattiker A, Hoogland C, et al. ExPASy: the proteomics server for in-depth protein knowledge and analysis [J]. *Nucleic Acids Res*, 2003, 31(13): 3784-3788.
- [15] Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein [J]. *J Mol Biol*, 1982, 157(1): 105-132.

- [16] Krogh A, Larsson B, von Heijne G, et al. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes [J]. *J Mol Biol*, 2001, 305(3): 567-580.
- [17] Kelley L A, Sternberg M J. Protein structure prediction on the Web: a case study using the Phyre server [J]. *Nat Protoc*, 2009, 4(3): 363-371.
- [18] Biasini M, Bienert S, Waterhouse A, et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*, 2014, 42: 252-259.
- [19] Lareau L F, Green R E, Bhatnagar R S, et al. The evolving roles of alternative splicing [J]. *Curr Opin Struct Biol*, 2004, 14(3): 273-282.
- [20] Filichkin S A, Priest II D, Givan S A, et al. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana* [J]. *Genome Res*, 2010, 20(1): 45-58.
- [21] Livshits V A, Zakataeva N P, Aleshin V V, et al. Identification and characterization of the new gene *rhtA* involved in threonine and homoserine efflux in *Escherichia coli* [J]. *Res Microbiol*, 2003, 154(2): 123 - 135.
- [22] 石鹏, 曹红星, 李东霞, 等. 油棕等植物 γ -生育酚甲基转移酶的生物信息学分析 [J]. *热带作物学报*, 2015, 36(2): 308-315.
- [23] Xu W, Dai W, Yan H, et al. *Arabidopsis* NIP3;1 plays an important role in arsenic uptake and root-to-shoot translocation under arsenite stress conditions [J]. *Mol Plant*, 2015, 8(5): 722-733.
- [24] Wallace IS, Choi WG, Roberts DM. The structure, function and regulation of the nodulin 26-like intrinsic protein family of plant aquaglyceroporins [J]. *Biochim Biophys Acta*, 2006, 1758(8): 1165-1175.
- [25] Zhang H, Kim M S, Krishnamachari V, et al. Rhizobacterial volatile emissions regulate auxin homeostasis and cell expansion in *Arabidopsis* [J]. *Planta*, 2007, 226(4): 839-851.

Bioinformatic Analysis of *Nodulin-like* Gene *At1g01070* in *Arabidopsis thaliana*

Wang Li, Yuan Yuanyuan, Zhao Panpan, Chang Weidong, Wang Lifeng, Wang Linsong

(College of Life Sciences, Henan Normal University, Xinxiang 453007, China)

Abstract: Recent studies highlight the importance of *nodulin*-like genes during development of plants that are unable to nodulate. However, only few gene functions were identified. In this study, *At1g01070*, a member of MtN21 gene family in *Arabidopsis thaliana*, was analyzed with bioinformatics methods and the subcellular localization of it was determined. The results showed that *At1g01070* undergoes alternative splicing to produce two splice variants encoding *At1g01070.1* and *At1g01070.2* respectively. The molecular weights of *At1g01070.1* and *At1g01070.2* are 40 KD and 35 KD respectively. The isoelectric points of them are 9.2 and 8.9 respectively. The two proteins both contain two drug/metabolite domains named Eam Λ , a characteristic domain of MtN21 protein family. *At1g01070.1* has a specific PLN00411 domain and its Eam Λ domain at N terminal is longer than that of *At1g01070.2*. The two proteins have similar tertiary structure containing seven α -helicals and some random coils. They exhibit strong hydrophobicity with ten and eight α -helical transmembrane-spanning domains respectively. There were thirteen phosphorylation sites in *At1g01070* protein. The two proteins are predicted to localize to the plasma membrane most likely. To confirm this, *At1g01070.1::GFP* fusion protein was transiently expressed in tobacco. The observed results were in agreement with the prediction. Phylogenetic analysis showed that *At1g01070* had the highest evolutionary relationship with predicted protein in *Arabidopsis lyrata*. The results suggested that *At1g01070* may be involved in the transport of solutes in plants.

Keywords: *nodulin*-like gene; *At1g01070*; MtN21; bioinformatic analysis; subcellular localization

[责任编辑 王凤产]