

## 问题讨论 Discussion

## 植物水溶性蔗糖合成酶生物信息学分析初探

韩立敏 王喆之\*

陕西师范大学生命科学学院, 西安 710062

## A Preliminary Bioinformatics Analysis of Soluble Sucrose Synthase in Different Plants

HAN Li-Min, WANG Zhe-Zhi\*

College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China

**提要** 用生物信息学方法对已在 GenBank 上注册的黑麦草、绿竹、菜豆、马铃薯、颤杨等植物水溶性蔗糖合成酶基因的核苷酸序列以及推导的氨基酸序列、组成成分、氨基酸翻译后修饰、导肽、跨膜拓扑结构域、疏水性/亲水性、蛋白质二级结构以及功能结构域等进行分析预测和推断的结果表明, 这些植物的水溶性蔗糖合成酶位于线粒体中, 是非跨膜的亲水性蛋白,  $\alpha$ -螺旋和不规则卷曲是其蛋白质二级结构的主要结构元件,  $\beta$ -转角和延伸链散布于整个蛋白质中, 包含2个功能结构域, 即蔗糖合成功能域和糖基化合物转移功能域。

**关键词** 蔗糖合成酶; 水溶性; 生物信息学

糖在植物生长、发育和代谢中, 既是光合作用的产物, 又是呼吸作用的底物, 是植物生长发育过程中碳骨架和能量的源泉以及植物抗逆性的基础。糖代谢是整个生物代谢的中心, 它沟通着蛋白质代谢、脂类代谢、核酸代谢以及各类次生代谢。植物蔗糖合成酶(sucrose synthase, SuSy)是由分子量为 83~100 kDa 的亚基构成的四聚体(Moriguchi和Yamaki 1988), 在植物体中有3种存在方式, 即细胞质中的可溶性 SuSy、附着在细胞膜上的不溶性 SuSy 和与细胞骨架相结合的 SuSy (Winter 和 Huber 2000), 是植物蔗糖代谢过程中的关键酶, 其催化反应为: 果糖+UDPG  $\rightleftharpoons$  蔗糖+UDP, 既可以催化蔗糖合成, 又可以催化蔗糖分解, 但通常认为分解作用是主要的(张明方和李志凌 2002)。以往的研究(Ruan 等 2003; Tanase 和 Yamaki 2000; Komatsu 等 2002; Dejardin 等 1997)表明, 其分解产物尿苷二磷酸葡萄糖(UDPG)可为细胞壁构建和糖酵解代谢提供底物, 同时也是支链淀粉和直链淀粉合成的前体。此外, 植物 SuSy 与贮存器官(根、果实等)的贮存强度也有一定关系(Zrenner 等 1995; 吕英民和张大鹏 2000)。植物 SuSy 基因已分别从玉米、水稻、马铃薯、小麦、大豆、甜菜、甘蔗、拟南芥、胡萝卜、

Mokara Yellow 等植物中克隆(Werr 等 1985; Wang 等 1992; Salanoubat 和 Belliard 1987; Marana 等 1990; Barratt 等 2001; Hesse 和 Willmitzer 1996; Lingle 和 Dyer 2001; Chopra 等 1992; Sebkova 等 1995; Li 等 2004)。

核酸和蛋白质的计算机分析是分子生物学研究的一项新技术, 已逐渐形成一门由分子生物学、生物数学、比较生物学和计算机信息处理技术相结合的交叉学科——生物信息学(bioinformatics), 其基本内容是用数据库和各类分析软件对生物大分子的结构进行比较和统计分析, 推导核苷酸和氨基酸序列同源性并揭示分子的结构、功能和进化关系(叶雄和张楚富 2006)。本文以黑麦草(*Lolium perenne* L.)为重点, 对绿竹[*Dendrocalamopsis oldhamii* (Munro) Keng F.]、菜豆(*Phaseolus vulgaris* L.)、马铃薯(*Solanum tuberosum* L.)、颤杨(*Populus tremuloides* Michx.)等植物中水溶性 SuSy 基因的核苷酸序列和其推导氨基酸序列的组成成分、理化性质、结构特征、功能等进行了初步

收稿 2006-06-06 修定 2006-09-21

\* 通讯作者(E-mail: zzwang@snnu.edu.cn, Tel: 029-85308317)。

分析, 为该基因的进一步研究提供一些资料。

## 材料与方法

数据资料来源于National Center for Biotechnology Information (NCBI) 核酸及蛋白质数据库中已注册的植物 SuSy 的核苷酸序列及其对应的氨基酸序列, 包括黑麦草 (*Lolium perenne* L., AB232656、BAE79815)、绿竹 [*Dendrocalamopsis oldhamii* (Munro) Keng F., AF412038、AAL50571]、菜豆 (*Phaseolus vulgaris* L., AF315375、AAN76498)、马铃薯 (*Solanum tuberosum* L., AY205302、AA067719)、颤杨 (*Populus tremuloides* Michx., AY341026、AAR03498)、大麦 (*Hordeum vulgare* L., X65871、CAA46701)、小麦 (*Triticum aestivum* L., AJ001117、CAA04543)、水稻 (*Oryza sativa* L., AK100546、CAA46017)、甘蔗 (*Saccharum officinarum* L., AF263384、AAM68126)、玉米 (*Zea mays* L., X02400、P04712)、芦苇 [*Phragmites australis* (Cav.) Trin. ex Steud., DQ296674]、郁金香 (*Tulipa gesneriana* L., CAA65639)。

依据DNASTar软件和<http://www.ncbi.nlm.nih.gov/>、<http://www.ebi.ac.uk/>、<http://www.cbs.dtu.dk/>、<http://cn.expasy.org/>等网站提供的各类生物信息学软件进行在线分析。核酸及氨基酸序列的组成成分分析、理化性质分析、开放阅读框(open

reading frame, ORF) 的查找和翻译均用DNASTar软件进行; 核苷酸和氨基酸序列的同源性比对用Blast在线工具完成; 氨基酸翻译后修饰、跨膜结构域和亲水性/疏水性的分析用在线工具NetPhosK 1.0 Server、NetPhos 2.0 Server、TargetP 1.1 Server、TMHMM 2.0 Server、ProtScale完成; 用SOPMA在线工具预测蛋白质的二级结构。

## 结果与讨论

### 1 核苷酸序列、氨基酸序列的组成成分和理化性质分析

用DNASTar软件的ORF Finder、Editseq、Protean程序分析黑麦草(AB232656、BAE79815)、绿竹(AF412038、AAL50571)、菜豆(AF315375、AAN76498)、马铃薯(AY205302、AA067719)、颤杨(AY341026、AAR03498)水溶性 SuSy 的核苷酸序列及氨基酸序列的结果(表1)表明, 除黑麦草的ORF、分子量、理论等电点高于或大于其它几种植物外, 绿竹、菜豆、马铃薯、颤杨水溶性 SuSy 的ORF、分子量、理论等电点均基本上一致; 终止密码子TAA、TAG、TGA出现频率高; Leu、Glu、Val、Ala、Gly是这几种植物 SuSy 氨基酸序列中含量最为丰富的氨基酸, 其平均含量分别为10.90%、7.84%、6.79%、6.46%、6.27%; 在颤杨 SuSy 氨基酸组成中, 酸性和碱性氨基酸的比例最高, 分别

表1 不同植物SuSy的核酸序列以及推导的氨基酸序列组成成分和理化性质分析

植物	ORF/bp	终止密码子	推导的氨基酸残基数	分子量/kDa	pI	氨基酸比例/%				
						碱性氨基酸	酸性氨基酸	亲水氨基酸	疏水氨基酸	带电氨基酸
黑麦草	2 658	TAA	885	100.388	6.81	11.53	12.20	24.07	36.27	32.54
绿竹	2 427	TAG	808	92.252	6.29	11.14	12.87	24.01	36.14	32.55
菜豆	2 418	TAA	805	92.023	6.03	11.18	13.66	22.48	37.02	32.92
马铃薯	2 436	TGA	811	92.775	6.23	11.22	12.95	24.54	34.90	32.43
颤杨	2 418	TAA	805	92.522	6.27	11.93	13.79	22.73	35.65	34.78

植物	含量最丰富的氨基酸/%									
	Leu	Glu	Val	Gly	Ser	Ala	Lle	Arg	Lys	
黑麦草	23.00	6.89	6.44	6.55	6.44					
绿竹	11.01	7.05	6.06	6.55		6.81	6.06			
菜豆	11.55	9.19	7.70	6.09		5.96		6.21		
马铃薯	10.60	7.64	6.78	5.80	6.04	5.84		5.80		
颤杨	10.93	8.45	6.96	6.34						6.21

为13.79%、11.93%；在马铃薯和菜豆总氨基酸中，亲水的(hydrophilic)氨基酸和疏水的(hydrophobic)氨基酸的比例最高，分别为24.54%、37.02%。

## 2 核酸及氨基酸序列的比对分析

用Blast程序比对黑麦草与其它植物水溶性SuSy的核酸及氨基酸序列同源性的结果表明，黑麦草水溶性SuSy的核酸序列与其它植物的核酸序列如大麦(X65871)、小麦(AJ001117)、绿竹(AF412038)、水稻(AK100546)、甘蔗(AF263384)、玉米(X02400)和芦苇(DQ296674)等有较高的同源性，分别为98%、93%、89%、87%、87%、86%、86%；对蛋白质-蛋白质的Blastp (protein-protein blast)比对来说，同源谱的相似性更大，即黑麦草水溶性SuSy的氨基酸序列与其它植物的氨基酸序列相似性更大，如与大麦(CAA46701)、小麦(CAA04543)、绿竹(AAL50571)、水稻(CAA46017)、甘蔗(AAM68126)、玉米(P04712)、郁金香(CAA65639)的一致性分别为98%、95%、92%、91%、92%、92%、80%。与绿竹、菜豆、马铃薯、颠杨中水溶性SuSy的核酸及氨基酸序列的比对结果与此类似，即不同植物的氨基酸序列比核酸序列有更大的相似性。

## 3 疏水性/亲水性的预测和分析

氨基酸是蛋白质的构件分子，自然界中存在的成千上万种蛋白质的结构和功能的多样性归根结底是由20种常见氨基酸的内在性质造成的。疏水

性是20种氨基酸都具有的特性，即氨基酸远离周围水分子，将自己包埋进蛋白质核心的相对趋势。这一趋势加上空间立体条件和其它一些因素，决定了一个蛋白质最终折叠成的三维空间构象。此外，通过了解肽链中不同肽段的疏水性，可以对跨膜蛋白的跨膜结构域进行预测。因此，在预测蛋白质次级结构和分析功能结构域时，可以参考疏水性/亲水性的预测和分析的结果进行。为此，用ProtScale预测黑麦草水溶性SuSy氨基酸序列的疏水性/亲水性(Kyte和Doolittle 1982)的结果(图1)表明，多肽链第613位的天冬氨酸(Asp)具有最低的分值-3.400；第846位的丙氨酸(Ala)具有最高的分值2.989。依据氨基酸分值越低亲水性越强和分值越高疏水性越强(参考软件ProtScale)的规律，可以看出，在第613位的Asp亲水性最强，第846位的Ala疏水性最强，而就整体来看(图1)，亲水性氨基酸均匀分布在整条肽链中，且多于疏水性氨基酸。因此，整个多肽链表现为亲水性，没有明显的疏水区域，可认为黑麦草SuSy是亲水性蛋白。对绿竹、菜豆、马铃薯、颠杨水溶性SuSy氨基酸序列的疏水性/亲水性分析，也得到类似的结果。

## 4 翻译后修饰的预测和分析

多肽链在核糖体上合成释放后，一般要经过翻译后修饰如糖基化、甲基化、磷酸化等才能正确折叠形成有效的三维构象，并运输到特定场所发挥功能。翻译后修饰的预测和分析，对

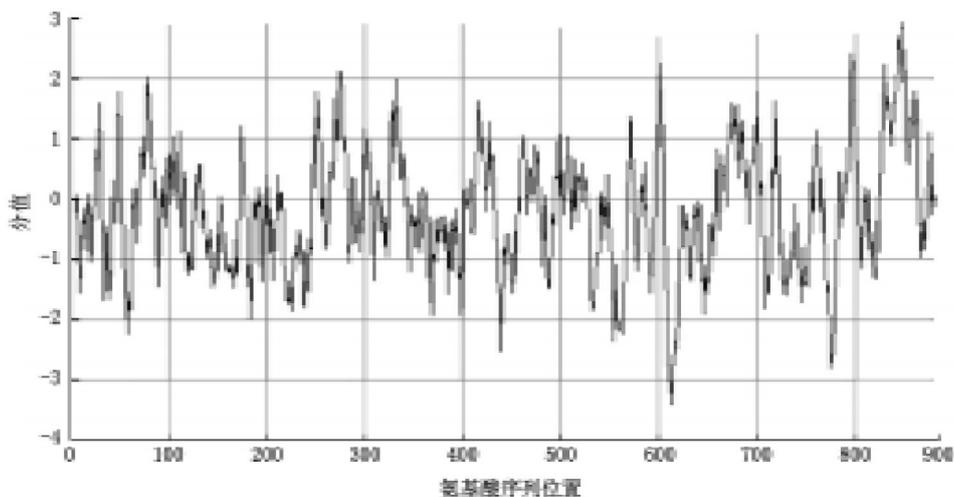


图1 ProtScale对黑麦草SuSy序列疏水性/亲水性的预测

正确认识和理解蛋白质的细胞定位和功能划分很重要。用NetPhosK 1.0 Server和NetPhos 2.0 Server对黑麦草水溶性 SuSy 的翻译后修饰预测(Blom等2004)的结果(图2)表明,整个多肽链中分值在0.5以上的氨基酸位点有48个,依据分值大于0.5(threshold)的氨基酸位点都是磷酸化位点(参考软件NetPhos 2.0 Server),可知黑麦草SuSy的磷酸化位点有48个,而且均匀分布于整个多肽链中(图

2)。Winter和Huber(2000)认为,磷酸化作用有利于SuSy可溶性构象的形成。据此可以推断,黑麦草SuSy是水溶性的。对绿竹、菜豆、马铃薯、颤杨的SuSy氨基酸序列分析,也得到同样的结果。

### 5 导肽的预测和分析

在核糖体上合成的蛋白质或者留在细胞质中,或者运送到其它细胞器,并形成有功能的空间构象后才能发挥其功能。导肽是一段引导新合

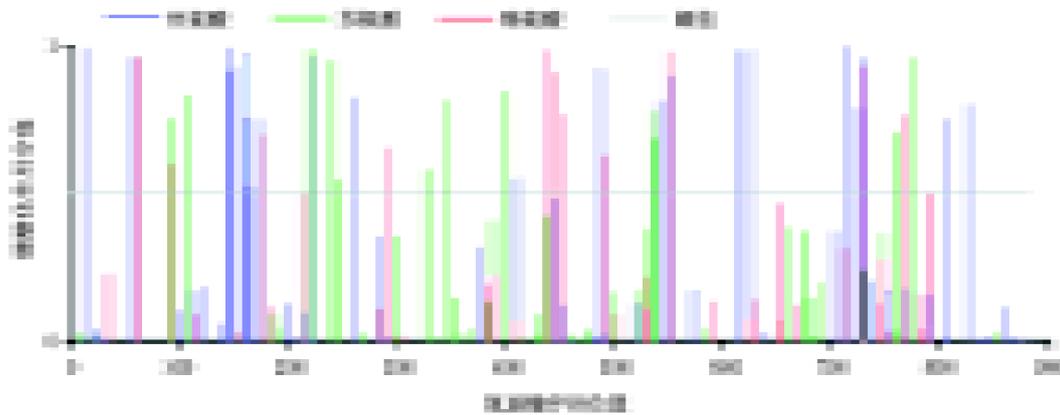


图2 NetPhos 2.0对黑麦草SuSy氨基酸序列翻译后的磷酸化修饰预测

成的肽链进入细胞器的识别序列。因此,导肽的预测和分析,对正确认识蛋白质的亚细胞定位与功能作用途径和机制有一定的意义。为此,用TargetP 1.1 Server预测黑麦草水溶性SuSy氨基酸序列的导肽(Emanuelsson等2000)的结果表明,此序列含有线粒体导肽的分值最高(0.854),预测的可靠性等级为3,线粒体导肽包含33个氨基酸残基。对绿竹、菜豆、马铃薯、颤杨水溶性SuSy氨基酸序列分析,也得到相同的结果,即这些植物的水溶性SuSy存在导肽酶切位点和线粒体导肽。

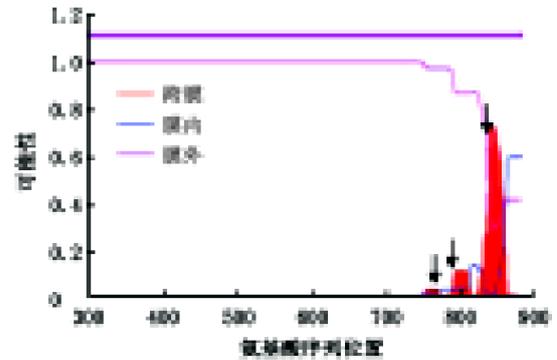


图3 TMHMM 2.0 Server对黑麦草SuSy序列跨膜结构域(箭头所示)的预测

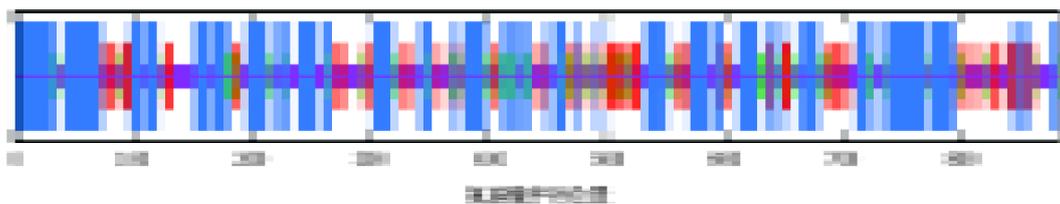


图4 黑麦草SuSy二级结构的预测

蓝色:  $\alpha$ -螺旋; 红色: 延伸链; 绿色:  $\beta$ -转角; 紫色: 不规则卷曲。

## 6 跨膜结构域的预测和分析

跨膜结构域是膜中蛋白与膜脂结合的主要部位,一般由20个左右的疏水氨基酸残基组成,形成 $\alpha$ -螺旋,固着于细胞膜上起“锚定”作用。跨膜结构域的预测和分析,对认识蛋白质的结构、功能、分类以及在细胞中的作用部位均有一定的意义。用TMHMM 2.0 Server预测(Ikeda等2002)黑麦草水溶性SuSy氨基酸序列的跨膜结构域的结果(图3)表明,整条肽链跨膜结构域的可能性均小于1,而软件默认可能性大于1才具有跨膜结构,因此,预测的可能存在的3个跨膜结构域(箭头所示)并不跨膜,黑麦草的水溶性SuSy不存在跨膜结构域,属于非跨膜蛋白类。对绿竹、菜豆、马铃薯、颤杨水溶性SuSy氨基酸序列的跨膜结构域分析,也得到类似的结果。

## 7 二级结构的预测和分析

蛋白质的生物学功能是蛋白质分子天然构象所具有的性质,由其高级结构所决定,高级结构则是由一级结构即氨基酸顺序决定,而氨基酸顺序是由遗传物质DNA的碱基顺序规定的。氨基酸的多肽链借助氢键排列成沿一维方向而呈现有规则的重复构象的二级结构,是氨基酸顺序与三维构象之间的桥梁。二级结构借助范德华力、氢键、静电和疏水等相互作用形成蛋白质的三级结构,

从而发挥正常的生物学功能。据此,用SOPMA预测黑麦草水溶性SuSy氨基酸序列的二级结构(Geourjon和Deleage 1995)的结果(图4)显示, $\alpha$ -螺旋和不规则卷曲是黑麦草水溶性SuSy最大量的结构元件,而 $\beta$ -转角和延伸链则散布于整个蛋白质中。统计表明,黑麦草水溶性SuSy的二级结构由47.80%的 $\alpha$ -螺旋、28.02%的不规则卷曲、15.93%的延伸链和8.25%的 $\beta$ -转角组成。这些结果在绿竹、菜豆、马铃薯和颤杨SuSy二级结构中也有相同的表现。

## 8 功能域的预测和分析

蛋白酶由多个模体组成,组成这些模体的氨基酸区段行使特异的功能,同时还蕴含着各自的遗传进化信息。在分析黑麦草水溶性SuSy氨基酸序列功能域(图5)中,得到的结果与卢合全等在棉花中的SuSy功能结构特征一致(卢合全等2005),即黑麦草水溶性SuSy具有2个功能域:N端6~550肽段与蔗糖合成(sucrose\_synth)功能区段配对几率为100%(NCBI计算值),起催化尿苷二磷酸葡萄糖和果糖合成蔗糖的功能;C端554~742肽段与糖基转移(glycos\_transf\_1)功能区段配对几率为98.8%(NCBI计算值),行使糖基化合物(UDP、ADP、GDP或CMP)转移功能,可将具备活性的糖基转移至糖原、果糖-6-P、脂多糖。



图5 黑麦草SuSy氨基酸序列功能域分析

sucrose\_synth: 蔗糖合成功能区段; Glycos\_transf\_1: 糖基转移功能区段。

总之,本文从生物信息学角度,以黑麦草为主要分析对象,对不同植物中水溶性SuSy基因的核苷酸序列及其推导的氨基酸序列结构特征进行分析,同时,对其整个肽链的亲水性、疏水性、可能的亚细胞定位、二级结构进行了预测,并对其功能结构域进行了分析。上述虽属推断和预测,但为进一步研究蛋白质的高级结构及探寻水溶性SuSy的潜在功能和了解整个糖代谢,可能有一定的参考价值。

## 参考文献

卢合全,沈法富,刘凌霄,孙维方(2005). 棉花蔗糖合成酶(SuSy)

分子结构特征与功能预测分析. 西北植物学报, 25 (7): 1372~1376

吕英民, 张大鹏(2000). 果实发育过程中糖的积累. 植物生理学通讯, 36 (3): 258~265

叶雄, 张楚富(2006). 生物信息学在植物谷氨酰胺合成酶同工酶基因研究中的应用. 植物生理学通讯, 42 (1): 115~121

张明方, 李志凌(2002). 高等植物中与蔗糖代谢相关的酶. 植物生理学通讯, 38 (3): 289~295

Barratt DHP, Barber L, Kruger NJ, Smith AM, Wang TL, Martin C (2001). Multiple, distinct isoforms of sucrose synthase in pea. *Plant Physiol*, 127: 655~664

Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence.

- Proteomic, 4 (6): 1633~1649
- Chopra S, Delfavero J, Dolferus R, Jacobs M (1992). Sucrose synthase of *Arabidopsis thaliana*: genomic cloning and sequence characterization. *Plant Mol Biol*, 18: 131~134
- Dejardin A, Rochat C, Wuillem S, Bouthin J-P (1997). Contribution of sucrose synthase, ADP-glucose pyrophosphorylase and starch synthase to starch synthesis in developing pea seeds. *Plant Cell Environ*, 20: 1421~1430
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*, 300: 1005~1016
- Geourjon C, Deleage G (1995). SOPMA: significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*, 11 (6): 681~684
- Hesse H, Willmitzer L (1996). Expression analysis of a sucrose synthase gene from sugar beet (*Beta vulgaris* L.). *Plant Mol Biol*, 30: 863~872
- Ikeda M, Arai M, Lao DM, Shimizu T (2002). Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *InSilicoBiol*, 2 (1): 19~33
- Komatsu A, Moriguchi T, Koyama K, Omura M, Akihama T (2002). Analysis of sucrose synthase genes in citrus suggests different roles and phylogenetic relationships. *J Exp Bot*, 53: 61~71
- Kyte J, Doolittle RF (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, 157 (6): 105~132
- Li CR, Zhang XB, Huang CH, Hew CS (2004). Cloning, characterization and tissue specific expression of a sucrose synthase gene from tropical epiphytic CAM orchid *Mokara* Yellow. *J Plant Physiol*, 161: 87~94
- Lingle SE, Dyer JM (2001). Cloning and expression of sucrose synthase-1 cDNA from sugarcane. *Plant Physiol*, 158: 129~131
- Marana C, Garcia OF, Carbonero P (1990). Different expression of two types of sucrose synthase-encoding genes in wheat in response to anaerobiosis, cold shock and light. *Gene*, 88: 167~172
- Moriguchi T, Yamaki S (1988). Purification and characterization of sucrose synthase from peach (*Prunus persica*) fruit. *Plant Cell Physiol*, 29: 1361~1366
- Ruan YL, Llewellyn DJ, Furbank RT (2003). Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. *Plant Cell*, 15: 952~964
- Salanoubat M, Belliard G (1987). Molecular cloning and sequencing of sucrose synthase cDNA from potato (*Solanum tuberosum* L.): preliminary characterization of sucrose synthase mRNA distribution. *Gene*, 60: 47~56
- Sebkova V, Unger C, Hardegger M (1995). Biochemical, physiological, and molecular characterization of sucrose synthase from *Daucus carota*. *Plant Physiol*, 108: 75~83
- Tanase K, Yamaki S (2000). Purification and characterization of two sucrose synthase isoforms from Japanese pear fruit. *Plant Cell Physiol*, 41: 408~414
- Wang AY, Yu WP, Juang RH, Sung HY, Su JC (1992). Presence of three rice sucrose synthase genes revealed by cloning and sequencing of cDNA. *Plant Mol Biol*, 18: 1191~1194
- Werr W, Frommer WB, Maas C, Starlinger P (1985). Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L. *EMBO J*, 4: 1373~1380
- Winter H, Huber SC (2000). Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes. *Crit Rev Plant Sci*, 19: 31~67
- Zrenner R, Salanoubat M, Willmitzer L, Sonnewald U (1995). Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (*Solanum tuberosum* L.). *Plant J*, 7: 97~107