

## 高粱叶绿体 *psaA* 基因的克隆和核苷酸序列分析

龚小松\* 阎景智 吴乃虎

(中国科学院发育生物学研究所, 北京)

阎隆飞

(北京农业大学, 北京)

我们克隆了含高粱叶绿体 *psaA* 基因的 6.7kb *EcoRI* 酶切片段, 进行了限制酶图谱分析和核苷酸序列结构测定, 所测核苷酸序列总长为 3080bp, 其中高粱叶绿体 *psaA* 基因序列为 2253bp, 编码一个含 750 个氨基酸残基的蛋白质, 其分子量为 83000Da。高粱 *psaA* 基因的核苷酸序列与玉米、水稻、菠菜的同源性分别为 99.4%、96.3% 和 89.4%; 推导出的氨基酸序列的同源性分别为 99.3%、98.0% 和 95.7%。 $C_4$  植物与  $C_3$  植物比较, 由于  $P_{700}$  脱辅基蛋白的第 493 位氨基酸性质的不同 (Gly  $\rightarrow$  Arg, Ser), 而导致了它们的多肽在疏水性图谱上的差异。

**关键词** 高粱; *psaA* 基因; 核苷酸序列

光合作用的光反应是在叶绿体的类囊体膜上发生的, 它由光系统 I、光系统 II、细胞色素 b/f、ATP 酶等复合体协同完成。通过光能的捕获、电荷分离、电子传递和质子转位等一系列光化学反应, 把光能转化成化学能, 为生命活动提供能量和物质来源。光系统 I 是光能转变成化学能的重要反应场所之一。

*psaA* 和 *psaB* 是编码光系统 I 反应中心蛋白的两个基因, 其中由 *psaA* 基因编码的反应中心蛋白 A, 结合了反应中心色素  $P_{700}$ , 故称  $P_{700}$  脱辅基蛋白。*psaB* 基因位于 *psaA* 基因的 3'-端下游, 两者之间相距仅 25bp。它们由叶绿体基因组编码, 位于大单拷贝区中央, 以共转录的方式合成一个长度约为 4.9kb 的 mRNA。目前, 有关 *psaA* 基因的核苷酸序列已有报道<sup>[1-3]</sup>。不同物种之间 *psaA* 基因表现出较高的同源性。在白色体发育成叶绿体的过程中,  $C_4$  植物的高粱和玉米的 *psaA* 基因表现出明显的光诱导调控特征<sup>[10,11]</sup>,

这一性质与  $C_3$  植物不同<sup>[1,2]</sup>。 $C_4$  植物和  $C_3$  植物在光合效率上也存在差别。为进一步探讨  $C_4$  植物光系统 I 反应中心蛋白的结构及其表达调控, 我们克隆了高粱叶绿体 *psaA* 基因, 并进行了核苷酸序列结构分析。

### 材料和方法

#### (一) 材料

高粱 (*Sorghum bicolor* L.) 选用秋冀 5 号品种, 购自河北省平泉县种子分公司, 温室中发芽, 三叶期时采其叶片。

限制酶购自华美生物工程公司或 Boehringer 公司。pUC19 DNA, M13RF DNA 购自华美生物工程公司。M13 序列分析试剂盒购自 BRL 公司。

本文于 1991 年 7 月 24 日收到。

\* 北京农业大学研究生

本研究系国家高科技“863-101-09-37”基金资助。

## (二) 方法

1. 高粱叶绿体DNA的提取与纯化: 参照龚小松等发展的高盐低 pH 法<sup>[13]</sup>。

2. DNA 的限制酶酶解: 按产品说明进行。

3. Southern杂交、缺口转移、<sup>32</sup>P 同位素标记探针, 重组片段的克隆、质粒 DNA 的提取等: 均参照 Maniatis<sup>[14]</sup> 的方法。

4. DNA 核苷酸序列分析: 参照 BRL 产品说明书的程序进行。DNA 序列资料的分析采用MicroGene序列分析软件(MG-1M-6.0)。

## 结果与讨论

### (一) 高粱叶绿体 *psaA* 基因的克隆和限制酶图谱

EcoR I 降解的高粱叶绿体 DNA, 经 1% 琼脂糖凝胶电泳分离后, 转移到硝酸纤维素滤膜上, 用 <sup>32</sup>P 同位素标记的水稻 *psaA* 基因<sup>[15]</sup> 为探针, 进行 Southern 杂交, 检测含有高粱 *psaA* 基因的限制片段。结果显示: EcoR I 酶切后的 6.7kb 片段与探针杂交。用低熔点琼脂糖凝胶电泳法回收该片段, 与载体 pUC19DNA 连接后转化 *E. coli* JM 101 菌株, 并在含有 IPTG 和 X-gal 的平板上筛选。从所获得的 15 个阳性克隆分离的质粒中, 经 Southern 杂交分析证实有 4 个带有 *psaA* 基因的 6.7kb 插入片段。我们将其中的一个质粒命名为 pSB *psaA*。

我们构建了带有高粱 *psaA* 基因的插入片段的限制酶图谱(图1), 并以 M13mp18 和 M13mp19 为载体进行亚克隆, 用于核苷酸序列分析。

### (二) DNA 序列

高粱 *psaA* 基因的核苷酸序列及其推

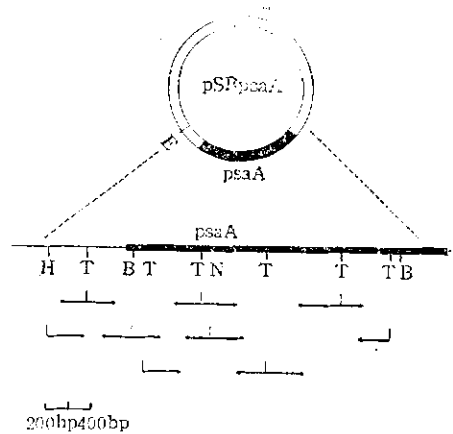


图1 高粱叶绿体 *psaA* 基因限制酶图谱及序列测定策略图

Fig.1 Restriction map and strategy of sequencing the sorghum chloroplast *psaA* gene and its flanking regions

The sorghum *psaA* coding region is represented by the dark box. The horizontal arrows indicate the direction and extent of sequencing

E: EcoR I; B: BamH I; H: Hind III; N: Nru I; T: Taq I

导的氨基酸序列如图2所示。我们共测定了 3080bp 的核苷酸序列, 其中高粱 *psaA* 基因的编码区长 2253bp, 编码一种 750 个氨基酸的蛋白质, 分子量为 83kDa。5' 端上游区为 715 bp, 其中靠近 *psaA* 基因的约 180bp 区域为富含 AT 的非转译区。3' 端下游区为 112bp。

### (三) 原核的启动子特征

在 *psaA* 基因的下游为 *psaB* 基因, 中间间隔的 25 bp 序列在玉米、水稻、豌豆、高粱之间高度保守。在 *psaA* 基因上游约 160 bp 处, 推测的启动子序列, -10 区为 CATAAT (原核为 TATAAT), -35 区为 TTAGGC (原核为 TTGACA), 它们都不同程度地表现出类似原核生物的启动子元件的结构特征。其中 -10 区的保守序列仅相差一个核苷酸。在距离第一个转译起始密码子 ATG 上游 -6 到 -11 区, 有与原核生物一致的核糖体结合位点 AGGAGG。

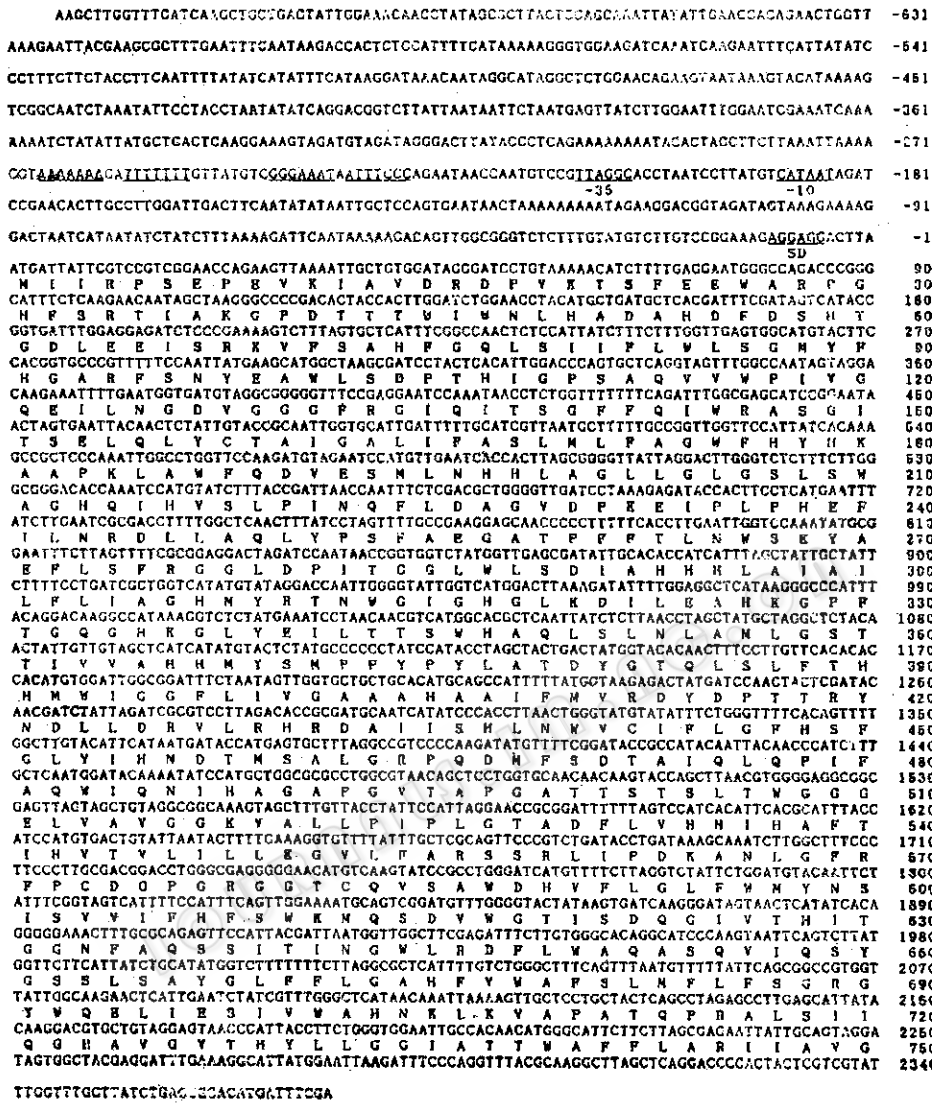


图2 高粱叶绿体psaA基因及其非编码区序列和推导的蛋白质序列  
 Fig.2 Nucleotide sequence of the sorghum chloroplast psaA gene and its flanking regions

The deduced amino acid sequence is also given. The pribnow box, the -35 box and the ribosome binding site are indicated. The inverted repeat sequences are underline

在psaB基因的起始密码子的上游-15-20区也存在推测的类似原核核糖体结合位点。同样在-116处和-136处(psaA基因内部)也存在类似原核的启动子-10区CATAAC和-35区TTGAAT。因此psaB基因可能单独转录成mRNA,但是有证

据表明:psaA基因与psaB基因以共转录的方式合成mRNA,在玉米和菠菜中,转录本长度分别为4.9和6.0kb<sup>[7,8]</sup>。尚未发现单独的psaB基因转录本<sup>[9,10]</sup>。基因表达调控系统在叶绿体内是如何控制psaB基因转录的问题有待于深入研究。

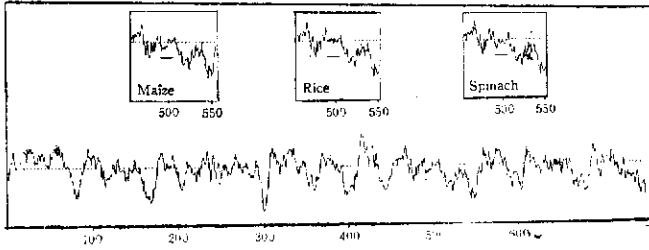


图3 高粱叶绿体psaA基因疏水性图谱, 以及与玉米、水稻和菠菜在第450到第550位氨基酸残基间疏水性图谱的比较(窗口为9)

Fig.3 The plot of hydropathy distribution of sorghum chloroplast psaA gene and comparison between sorghum and maize, rice and spinach in the region from 450th to 550th amino acid

表1 高粱与玉米、水稻、菠菜的psaA基因的同源性

Table 1 Sequence homology of the psaA gene and its deduced polypeptide between sorghum and maize, rice and spinach

	Maize	Rice	Spinach
Nucleotide sequences (%)	99.4	96.3	89.4
Polypeptide sequences (%)	99.3	98.0	95.7
5'-upstream region (%)	97.2	73.0	76.8
5'-215bp upstream region (%)	98.2	95.3	—

高粱psaA基因核苷酸序列及推导的氨基酸序列与玉米、水稻、菠菜的相应序列比较如表1所示。

#### (四) psaA基因5'端上游

由表1可以看出,高粱与玉米、水稻、菠菜相比,在psaA基因5'端上游的非编码区的差异较之编码区的要大,在非编码区中,靠近非编码区的部分,也包括推测的启动子的215bp区域,其同源性程度比更上游的区域要高些。玉米和高粱的psaA基因上游区同源性较高,只在高粱的一些区域发生了核苷酸缺失,而高粱和水稻相比,虽然推测的启动子-10及-35区较保守,但在启动子上游却出现了较显著的差异。高粱启动子区上游第-229位-243位核苷酸间,第-252-267位

核苷酸间存在反向重复序列。类似的结构特点在玉米中也可观察到<sup>[9]</sup>,但在水稻中却没有。这是否与C<sub>4</sub>植物psaA基因较为特殊的表达调控模式有关值得进一步探讨。

#### (五) 密码子的使用

在高粱的psaA基因中,密码子的第3位偏向使用A或T,高达65%,这在其他叶绿体基因中也已观察到<sup>[17]</sup>。而且,密码子的第二位核苷酸56.3%也为A或T。因此,这造成了在psaA基因中偏向使用第二、三位为A或T的密码子。编码Arg的CGG密码子在基因中没有出现,碱性和酸性氨基酸的比率为12.5%,疏水性氨基酸的比率为39.1%(表2)。

#### (六) psaA基因及其编码的蛋白质

psaA基因在高粱、玉米、水稻、菠菜中保守性很强,氨基酸序列的同源性与其核苷酸序列的同源性相比更强,这表明P<sub>700</sub>脱辅基蛋白的功能是很保守的。高粱与其它单子叶植物(玉米、水稻)之间的同源程度要高于双子叶植物(菠菜)的同源性。玉米psaA基因编码751个氨基酸,而高粱、水稻、菠菜都是750个氨基酸,它们都缺失了玉米的第40位氨基酸天冬酰胺。

在氨基酸序列第530位—570位之间

表 2 高粱叶绿体 *psaA* 基因密码子使用表  
Table 2 Codon usage of the sorghum chloroplast *psaA* gene

TTT	Phe	30	TCT	Ser	13	TAT	Tyr	13	TGT	Cys	3
TTC	Phe	21	TCC	Ser	12	TAC	Tyr	7	TGC	Cys	1
TTA	Leu	25	TCA	Ser	3	TAA	End	0	TGA	End	0
TTG	Leu	17	TCG	Ser	5	TAG	End	1	TGG	Trp	28
CTT	Leu	19	CCT	Pro	13	CAT	His	26	CGT	Arg	6
CTC	Leu	5	CCC	Pro	9	CAC	His	16	CGC	Arg	6
CTA	Leu	11	CCA	Pro	8	CAA	Gln	21	CGA	Arg	6
CTG	Leu	4	CCG	Pro	2	CAG	Gln	7	CGG	Arg	0
ATT	Ile	31	ACT	Thr	12	AAT	Asn	11	AGT	Ser	15
ATC	Ile	14	ACC	Thr	14	AAC	Asn	7	AGC	Ser	5
ATA	Ile	16	ACA	Thr	14	AAA	Lys	15	AGA	Arg	6
ATG	Met	15	ACG	Thr	3	AAG	Lys	2	AGG	Arg	2
GTT	Val	10	GCT	Ala	32	GAT	Asp	26	GGT	Gly	25
GTC	Val	5	GCC	Ala	11	GAC	Asp	6	GGC	Gly	19
GTA	Val	21	GCA	Ala	18	GAA	Glu	13	GGA	Gly	21
GTG	Val	2	GCG	Ala	7	GAG	Glu	13	GGG	Gly	9

出现较为保守的亮氨酸拉链结构。这种结构除了在 DNA 结合蛋白上观察到外, 在膜蛋白(如 *psaA*) 中亦存在。据推测这可能与反应中心蛋白二聚体和  $F_x$  位点的形成有关<sup>[18]</sup>, 被认为结合  $P_{700}$  反应中心色素的六肽段(Asp-Pro-Thr-Thr-Arg-Tyr)<sup>[19]</sup> 在高粱中也很保守、位于第 405 位—第 420 位氨基酸之间。

$C_4$  植物(高粱、玉米)和  $C_3$  植物(水稻、菠菜)相比较, *psaA* 基因编码的脱辅基蛋白的变异主要集中在 N-末端和第 487 位—第 494 位氨基酸之间(表 3)。而在这些位置上,  $C_3$  植物间或  $C_4$  植物均不存在任何差异。值得注意的是在第 493 位氨基酸

上,  $C_3$  植物存在与  $C_4$  植物性质不同的氨基酸, 即水稻和菠菜的 Arg 和 Ser 分别代替高粱和玉米的 Gly。我们采用 Kyte 及 Doolittle 的方法对高粱、水稻和菠菜的氨基酸序列进行了疏水性作图和比较(图 3), 结果表明:  $C_3$  植物和  $C_4$  植物间由于在第 493 位氨基酸性质的不同, 使这一附近区域的疏水性质出现明显差异, 这说明: 第 493 位氨基酸允许性质不同的氨基酸的取代, 而不影响  $P_{700}$  脱辅基蛋白的基本功能, 但是否会在它们执行功能的动力学性质上出现差异, 尚有待于进一步研究。

表 3 氨基酸突变分析  
Table 3 Analysis of amino acid differences between  $C_4$  plants (sorghum and maize) and  $C_3$  plants (rice and spinach)

Plant	Position of amino acids							
	5	6	13	487	493	494	625	
Sorghum	Pro	Ser	Ala	Ile	Gly	Val	Ile	
Maize	Ser	Ser	Ala	Ile	Gly	Val	Ile	
Rice	Ser	Pro	Val	Val	Arg	Leu	Val	
Spinach	Ser	Pro	Leu	Thr	Ser	Ala	Val	

## 参 考 文 献

- [1] Cantrell, A. and Bryant, D. A., *Plant Mol. Biol.*, 9: 453—468, 1987.  
 [2] Kuck, U., et al., *EMBO J.*, 6: 2185—2195, 1987.  
 [3] Cushman, J. C., et al., *Curr. Genet.*, 13: 159—171, 1988.  
 [4] Ohyama, K., et al., *Nature* 322: 572—574, 1986.  
 [5] Shinozaki, K., et al., *EMBO J.*, 5: 2043—2049, 1986.  
 [6] Lehmbeck, J., et al., *Plant Mol. Biol.*, 7: 3—10, 1986.  
 [7] Kirsh, W., et al., *Curr. Genet.*, 10: 843—855, 1986.  
 [8] Hiratsuka, H., et al., *Mol. Gen. Genet.*, 217: 185—194, 1989.  
 [9] Fish, L. E., et al., *J. Biol. Chem.*, 260: 1413—1421, 1985.  
 [10] Rodermel, S. R. and Bogorad, L., *J Cell Biol.*, 100: 463—476, 1985.  
 [11] Schrubar, H., et al., *Planta*, 183: 101—111, 1990.  
 [12] Mullet, J. E., et al., *EMBO J.*, 6: 1571—1579, 1987.  
 [13] 龚小松, 阎隆飞, *科学通报*, 7: 467—469, 1991.  
 [14] Maniatis, T., et al., *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor NY, 1982.  
 [15] Wu, N. H., et al., *Chinese J. Bot.*, 1: 2—8, 1989.  
 [16] Woodbury, N. W., et al., *Curr. Genet.*, 16: 433—445, 1989.  
 [17] Wu, N. H., et al., *Dev. Genet.*, 8: 339—350, 1987.  
 [18] Webber, A. N., et al., *FEBS lett.*, 264: 1—4, 1990.  
 [19] Fish, L. E., et al., in "Molecular Biology of the Photosynthetic Apparatus", Cold Spring Harbor Laboratory, Cold Spring Harbor NY, P: 111—120, 1985.

## Molecular Cloning and Nucleotide Sequence of Chloroplast psaA Gene from Sorghum

Gong Xiaosong Yan Jingzhi Wu Naihu

(Institute of Developmental Biology, Beijing)

Yan Lungfei

(Beijing Agricultural University, Beijing)

The EcoRI-6.7 kb fragment containing the sorghum chloroplast psaA gene was cloned. The restriction endonuclease map and the nucleotide sequence were determined. The length of the determined nucleotide sequence was 3080 bp, of which the psaA coding sequence was 2253 bp, which encoded a 750 amino acid polypeptide with a predicted molecular weight of 83 000. The homologies of this psaA gene with maize, rice and spinach were 99.4%, 96.3% and 89.4%, respectively, and the homologies of the putative polypeptide sequences were 99.3%, 98.0%, 95.7% respectively. We found that a discrepancy in the plot of hydropathy distribution of P<sub>700</sub> apoprotein existed between C<sub>4</sub> and C<sub>3</sub> plants due to a change occurred at the 493th amino acid (C<sub>4</sub> plant: Gly-C<sub>3</sub> plants: Arg or Ser).

### Key words

Sorghum; psaA gene; nucleotide sequence