***In vitro* expression and characterization of the translation start site of the *psbA* gene product (Q_B protein) from higher plants**

Bennett N.Cohen, Timothy A.Coleman, John J.Schmitt and Herbert Weissbach

Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110, USA

Received 9 February 1984; Revised and Accepted 9 July 1984

ABSTRACT

The *psbA* gene from higher plants, which codes for the atrazine herbicide binding protein of photosystem II (Q_B protein), has been recently sequenced by various laboratories. From these data there are two potential translation sites, one yielding a protein of 38,500 kd and another a protein of 34,500 kd. In the present study, cloned *psbA* gene sequences from maize, tobacco, and pea have been expressed in a highly defined *E. coli* *in vitro* transcription/translation system. In order to determine the start site of translation, we also have employed a simplified *E. coli* system designed to synthesize the first di- or tripeptide of the gene product. From these results, it is clear that the first ATG of the longest open reading frame of the *psbA* gene, that begins fMet-Thr, is not recognized *in vitro*. Instead, the next downstream Met at position 37 is the initiation site, since the expected dipeptide fMet-Ile is synthesized from all *psbA* clones. These data are in accord with the *in vivo* results that the gene product is a precursor protein of 34,500 kd.

INTRODUCTION

Studies of gene expression during the past ten years have centered on the regulation of molecular responses to biochemical stimuli. In higher plants and other photosynthetic eukaryotic organisms, light plays an important role in the regulation of both nuclear and chloroplast coded genes. It is now evident that the light stimulation of gene expression is characterized by the increase in unique populations of nuclear and chloroplast mRNA species after dark grown plants are exposed to light (1-8). However, before the mechanism of this effect can be elucidated, a suitable assay system for the identification of factors that affect gene expression has to be developed.

We have chosen to examine the effect of light on the expression of the Q_B protein or herbicide binding protein (*psbA* gene product) from higher plants. This gene is found in the large unique copy region of the chloroplast genome adjacent to an inverted repeat sequence and codes for proteins ranging from 33.5-34.5 kd depending on the species (9). Early studies on

the expression of the Q_B protein in vitro from total chloroplast DNA and mRNA has yielded limited results, with the chloroplast coded Q_B protein only being expressed in eukaryotic in vitro translation systems (10). The expression of this gene is light stimulated (11) and the Q_B protein mRNA constitutes a major portion of the mRNA synthesized by chloroplasts during light growth (12). The gene for this protein was first isolated by Bedbrook *et al.* (13) and since has been sequenced in Nicotiana debneyi, Spinacea oleracea, and Amaranthus hybridus (14,15).

Previously, we have utilized a highly defined E. coli based DNA-directed transcription/translation system to study the expression of various cloned bacterial genes (16-18). The chloroplast gene, rbcl, which codes for the large subunit (LS) of ribulose-1,5-bisphosphate carboxylase (RuBPC), has also been expressed in crude and defined E. coli expression systems (19,20). From our results and others, there is indication that E. coli RNA polymerase can initiate properly on chloroplast DNA templates and exercise some specificity of transcription (20,21).

In this paper we describe the in vitro expression of maize and tobacco Q_B proteins in a highly defined E. coli in vitro system. A simplified expression system based on the formation of the first dipeptide (22) of the gene product is also used to synthesize identical dipeptide products from maize, tobacco and pea psbA templates. From these latter experiments, we are able to determine that the start site of translation of the Q_B protein mRNA in an E. coli in vitro expression system is at the second AUG of the largest open reading frame of the Q_B protein mRNA.

MATERIALS AND METHODS

Chloroplast DNA Isolations and psbA Cloning

Maize chloroplast DNA was purified from 7 day seedlings by a modification of the method of Bovenberg *et al.* (23). After restriction digestion with endonuclease Bam H1, the Bam 8 fragment, which contains the psbA sequence (13), was isolated from low-melting point agarose (24). This fragment was ligated into the Bam H1 site of pBR322, transformed into E. coli RR1 and phenotypic transformants (Amp^r , Tet^s) isolated. Plasmids were screened by the method of Birnboim and Doly (25) and isolated by CsCl isopycnic centrifugation (26). Plasmids (pCMA) containing the psbA gene were selected by Southern hybridization to the purified maize Bam 8 fragment, which was ^{32}P radiolabeled by complementary strand synthesis (27). All plasmids were purified two times in CsCl gradients for in vitro expression

studies.

The 5' region of the pea psbA gene (Pisum alaska) in pBR322 was kindly supplied by Dr. J. Palmer and is referred to as pJP501 (28). Plasmid pNT32 was a gift of Dr. L. Bogorad. This plasmid contains the coding sequence for the tobacco psbA gene as a 3.2 kb fragment inserted into the EcoRI site of pBR322. Purified cloned chloroplast DNA fragments were isolated for in vitro studies from low-melting point agarose by ElutipTM-d column chromatography (29).

In Vitro Protein Synthesis and Dipeptide Systems to Study psbA Gene Expression

Plasmid clones of the psbA gene from maize (pCMA) and tobacco (pNT32), were expressed in an E. coli directed coupled transcription/translation system (16-18) and the [³⁵S]labeled gene products analyzed by 12.5% NaDodSO₄/polyacrylamide gel electrophoresis and fluorography (30,31). Similarly, [³⁵S]methionine, [³H]lysine, and [³H]isoleucine incorporation into total protein was determined by CCl₃COOH precipitable counts on GF/C glass fiber filters. Due to the similarity in molecular weights between the Q_B proteins (~34 kd) and the plasmid coded β -lactamase, the maize psbA (pCMA) plasmid was either linearized by a partial Pvu I restriction digestion to prevent synthesis of full-length β -lactamase or purified chloroplast DNA fragments were isolated from appropriate clones as described above. In addition, plasmids or DNA fragments containing the psbA gene were tested for expression in a simplified in vitro expression system, in which the N-terminal di- or tripeptide of the gene products is synthesized (22,32). The assay for di- or tripeptide formation was based on extraction of the formylated peptide from acidic solution into ethyl acetate (33). Purified tRNA isoacceptor species were obtained from Subriden RNA, Rolling Bay, Washington.

RESULTS

Table I summarizes the templates that are used in these experiments. The original clone that was constructed contained the maize chloroplast psbA gene present on the Bam 8 fragment as previously described by Bedbrook et al. (13). The sequence of the maize gene was utilized to reconfirm the identities of pJP501 and pNT32 as plasmids that code for the pea (partial product) and the tobacco psbA gene product, respectively, by Southern hybridization (data not shown).

Figure 1 is a fluorograph of a 12.5% NaDodSO₄/polyacrylamide gel

Table 1: Summary of *psbA* Clones Used in the *In Vitro* Studies

Clone	Source	Chloroplast DNA fragment	Size (kb)
pCMA	Maize	Bam 8	4.3
pJP-501	Pea	Pst I	5.0
pNT32	Tobacco	EcoRI	3.2

All fragments were cloned into pBR322 as described in Materials and Methods. Plasmid pJP-501 contains the 5' control region and only part of the coding region of the pea *psbA* gene.

analysis of the protein products synthesized *in vitro* from the cloned *psbA* genes. Lane 1 shows the proteins synthesized using as template a Pvu I linearized pCMA clone. A major *in vitro* product has a molecular weight of approximately 34,000, which corresponds to the unprocessed 34.5 kd maize

Fig. 1: In Vitro Protein Synthesis Analysis on NaDodSO₄ Polyacrylamide Slab Gel Electrophoresis Carried Out at 50 v for 16 hr. Equal Cl_3CCOOH precipitable counts (5×10^4) were loaded in each lane. Lane 1, proteins synthesized from 1 μg Pvu I linearized pCMA; Lane 2, proteins synthesized from 0.4 μg 3.2 kb tobacco *psbA* DNA; Lane 3, no DNA template (17.5 μl reaction mix used as sample); and Lane 4, [^{14}C]labeled carbonic anhydrase.

Table 2: Incorporation of Labeled Amino Acids into CCl₄COOH Insoluble Material Directed by DNA Fragments Containing³the *psbA* Gene

	Labeled Amino Acid		
	[³⁵ S]Met	[³ H]Lys	[³ H]Ile
	pmol incorporated		
pCM-A	2.4	0	8.5
pNT-32	2.9	0.1	8.6
pJP501	2.2	---	6.7
pBR322	4.6	9.3	17.7

A defined *in vitro* DNA-directed system was used (16-18). The incorporation of radioactivity into acid insoluble material was performed as described in Materials and Methods utilizing 0.05 µg of purified *psbA* insert DNA as template from the clones pCM-A, pNT-32, and pJP501. pBR322 was a control plasmid to show that [³H]lysine could be incorporated into acid insoluble material under the conditions used. Gel analysis of the products synthesized from the *psbA* containing DNA fragments showed that the Q_B protein was labeled after incubation with [³⁵S]-Met and [³H]Ile.

protein seen *in vivo* (11). Lane 2 shows the *in vitro* products that are synthesized from a 3.2 kb fragment of pNT32 that contains the entire tobacco *psbA* gene sequence (L. Bogorad, personal communication). Again a product of about 34.5 kd is observed plus a lower molecular weight protein that has not yet been identified. Lane 3 is the coupled transcription/translation system without exogenous template and lane 4 is a [¹⁴C]labeled 30 kd protein standard. Analysis of the protein products from pJP501 was not performed because this template contains only a partial sequence of the pea *psbA* gene.

In order to further establish that the 34.5 kd protein synthesized *in vitro* is the Q_B protein, differential incorporation of radiolabeled amino acids has been used. *In vivo* labeling studies had indicated that the Q_B protein does not contain lysine (34). The lack of lysine in the protein has subsequently been verified from the DNA sequence of the gene (14,15). Therefore, isolated chloroplast DNA fragments were used as templates in the defined *in vitro* expression system containing either [³⁵S]methionine, [³H]lysine, or [³H]isoleucine. From the sequence data, the protein contains 11 and 28 residues of Met and Ile respectively, but as noted above no Lys.

Table 3: DNA-Directed Di- and Tripeptide Synthesis of the *psbA* Gene

DNA	fMet-Thr	fMet-Ile	fMet-Ile-Pro
		pmo1	
pJP-501	0.11	2.6	0.48
pNT32	0.09	1.2	0.31
pCMA	0.02	1.1	---

All assays were performed with 0.5 μ g of purified DNA fragments from the above clones. For fMet-Thr synthesis, the results are mean values from assays using either crude *E. coli* tRNA or purified isoacceptor species of tRNA^{Thr}. For the experiments with isoleucine, the purified isoacceptor species tRNA^{Ile} was used. Di- and tripeptide synthesis were performed as described previously (22,32).

The data presented in Table II show that no [³H]lysine is incorporated *in vitro* into acid insoluble material directed by fragments containing the *psbA* gene, although both [³⁵S]Met and [³H]Ile are. As a control for these experiments, it was shown that [³H]lysine is incorporated into protein when plasmid pBR322 is used as template.

The sequence data of Zurawski *et al.* (14) on *N. debneyi* and *S. oleracea psbA* genes and similar data on *A. hybridus* by Herschberg and McIntosh (15) show that the largest open reading frame of all the *psbA* genes begins with the amino acid sequence Met-Thr (ATG ACT). However, downstream there is another potential start site in the same reading frame that begins Met-Ile (ATG ATC). This latter sequence is also present in the maize *psbA* gene (personal communication, L. Bogorad). Table III is a compilation of the data obtained from experiments on the formation of the first dipeptide of the *psbA* product directed by the DNA fragments of the various clones. Using either crude *E. coli* tRNA or three different purified isoacceptor species of tRNA^{Thr}, aminoacylated with [³H]Thr, there is only a barely detectable level of fMet-Thr synthesis. However, as shown in Table III, there is excellent formation of fMet-Ile in the dipeptide system directed by the *psbA* clones from pea, tobacco, and maize. To further verify that the start site of translation is at the second AUG of the *psbA* genes, pea and tobacco DNA fragments were also used as templates for tripeptide synthesis. If the methionine at position 37 initiates, proline would be the expected third amino acid (14,15). As shown in Table II, the predicted tripeptide, fMet-Ile-Pro is formed. Figure 2 shows the effect of template

Fig. 2: DNA Dependence of fMet-Ile Dipeptide Synthesis from Various Cloned *psbA* DNA Fragments. Templates are as follows: 5.1 kb DNA from pea clone pJP501 (Δ), and 3.2 kb DNA from tobacco clone pNT32 (□). Details of the incubations are described elsewhere (22,32,33).

concentration on the synthesis of the dipeptide, fMet-Ile, using as templates purified DNA fragments obtained from the pea and tobacco *psbA* clones.

Experiments similar to these described above were performed on the *psbA* gene from *Euglena gracilis* as described by Hallick (35). Little or no synthesis of protein was observed in the coupled transcription/translation system. These negative results would be expected from the recent data on the *psbA* and *rbcl* gene sequences from *E. gracilis* which showed that they contain introns within the coding region (R. Hallick, personal communication (36)).

DISCUSSION

A review of data compiled on chloroplast gene expression *in vitro* shows the limited expression of various *psbA* mRNAs in reticulocyte lysate or wheat germ extracts (10). However, the data in this paper indicate that the cloned *psbA* gene sequences of maize and tobacco can be expressed in a highly defined *E. coli* based DNA directed transcription/translation system. The data in Figure 1 show that proteins of identical molecular weights as the Q_B protein precursor have been synthesized *in vitro* from the maize and tobacco *psbA* genes. Recently, antibodies have been raised to the Q_B

protein by using a synthetic peptide of the Q_B protein. This antiserum immunoprecipitated a single product from the in vitro synthesis reactions that corresponded to the Q_B protein precursor and also immunoprecipitated a 34.5 kd protein from maize and pea thylakoid membranes (Felix, A.M., Heimer, E.P. and Cohen, B.N., manuscript in preparation). The in vitro synthesis of higher plant Q_B proteins in the E. coli expression system should make it possible to reclone these various psbA gene sequences into suitable E. coli-based expression vectors for the overproduction of the Q_B proteins for studies on their structure and function. The use of the E. coli in vitro expression system represents a heterologous approach to examine chloroplast gene expression. The system relies on the components of E. coli gene expression (e.g. RNA polymerase, ribosomes, etc.) to function in place of the corresponding chloroplast components. Because of this assumption, not only were the products examined for size but also for labeled amino acid incorporation. The results in Table II are indicative of those obtained in vivo and seem to indicate that the E. coli factors are faithfully expressing the chloroplast DNA sequences. Recent evidence comparing proteins expressed in a crude E. coli expression system with a homologous tobacco in vitro system also shows good correlation (Dr. D. Bourque, personal communication).

Furthermore, the use of the DNA-directed synthesis of the N-terminal dipeptide, fMet-Ile, should facilitate investigations on the role of chloroplast derived factors in regulating the expression of the Q_B protein by providing a rapid and reproducible assay system. In addition to the results from the in vitro studies described here, we have been able to show that the proposed 38,500 kd translation product described by Zurawski et al. (14) is not synthesized in vitro. This protein would require that translation initiation begin at the first ATG of the largest open reading frame, resulting in the formation of the dipeptide fMet-Thr. However, the dipeptide studies described here show that fMet-Thr is not formed in this in vitro system. Instead, the second AUG is utilized for initiation of translation as shown by the formation of fMet-Ile and fMet-Ile-Pro as seen in Table III. Our results are also in agreement with the reported size of the in vivo synthesized Q_B protein precursors, i.e. 34.5 kd and the reported maize sequence data (L. Bogorad, personal communication). Finally, Hirschberg and McIntosh (15) have concluded, based on sequence data, that the second AUG is the start site of translation in A. hybridus.

It should be noted that a similar situation with regard to the start of translation is seen in the mRNA for the maize rbcl gene. There are two

potential translation start sites in the same reading frame, 18 base pairs apart (37) that would yield the dipeptide, fMet-Ser. Both of these sequence regions also contain potential ribosomal binding sites that are homologous to the 16S rRNA from maize (38). Experiments in this laboratory, using different tRNA isoacceptor species, have shown that the first dipeptide of the protein is synthesized from the second fMet-Ser sequence of the maize rbcl gene and that the other possible start site 18 bp upstream is not recognized (39).

ACKNOWLEDGEMENT

We would like to thank Dr. L. Bogorad and Dr. J. Palmer for providing psbA clones and some unpublished data. We would also like to thank Dr. R. Hallick for his data on the E. gracilis psbA gene and Dr. M. Edelman for his helpful discussions.

REFERENCES

1. Rosner, A., Jakob, K.M., Gressel, J. and Sagher, D. (1975) *Biochem. Biophys. Res. Commun.* **67**, 383-391
2. Gressel, J. (1978) *Photochem. Photobiol.* **27**, 167-169
3. Tobin, E. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4749-4753
4. Sasaki, Y., Ishiye, M., Sakihama, T. and Kanikubo, T. (1981) *J. Biol. Chem.* **256**, 2315-2320
5. Tobin, E.M. (1981) *Plant Molec. Biol.* **1**, 35-51
6. Shinozaki, K., Sasaki, Y., Sakihama, T. and Kamikubo, T. (1982) *FEBS Lett.* **144**, 73-76
7. Gallagher, T.F. and Ellis, R.J. (1982) *EMBO Journal* **1**, 1493-1498
8. Sasaki, Y., Sakihama, T., Kamikubo, T. and Shinozaki, K. (1983) *Eur. J. Biochem.* **133**, 617-620
9. Edelman, M., Marder, J.B. and Mattoo, A.K. (1983) in *Structure and Function of Plant Genomes*, Cifleri, O. and Dure, L. Eds., Plenum Press, New York (in press)
10. Edelman, M. (1981) in *The Biochemistry of Plants*, Vol. 6, Marcus, A. Ed., pp. 249-301, Academic Press, New York
11. Bogorad, L., Jolly, S.O., Link, G., McIntosh, L., Poulsen, C., Schwarz, Z. and Steinmetz, A. (1980) in *Genome Organization and Expression in Plants*, Leaver, C.Y. Ed., pp. 291-304, Plenum Press, New York
12. Reisfeld, A., Gressel, J., Jakob, K.M. and Edelman, M. (1978) *Photochem. Photobiol.* **27**, 161-165
13. Bedbrook, J.R., Link, G., Coen, D.M. and Bogorad, L. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3060-3064
14. Zurawski, G., Bohnert, H.J., Whitfield, P.R. and Bottomley, W. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7699-7703
15. Hirschberg, J. and McIntosh, L. (1983) *Science* **222**, 1346-1349
16. Zarucki-Schulz, T., Jerez, C., Goldberg, G., Kung, H.-F., Huang, K.H., Brot, N. and Weissbach, H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 6115-6119
17. Kung, H.-F., Redfield, B. and Weissbach, H. (1979) *J. Biol. Chem.* **254**, 8404-8408

18. Kung, H.-F., Redfield, B., Treadwell, B.V., Eskin, B., Spears, C. and Weissbach, H. (1977) *J. Biol. Chem.* **252**, 6889-6894
19. Bottomley, W. and Whitfeld, P.R. (1979) *Eur. J. Biochem.* **93**, 31-39
20. Erion, J., Tarnowski, J., Peacock, S., Caldwell, P., Redfield, B., Brot, N. and Weissbach, H. (1983) *Plant Molec. Biol.* **2**, 279-290
21. Zech, M., Hartley, M.R. and Bohnert, H.J. (1981) *Current Genetics* **4**, 37-46
22. Robakis, N., Meza-Basso, L., Brot, N. and Weissbach, H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4261-4264
23. Bovenberg, W.A., Kool, A.J., Nijkamp, H.J.J. (1981) *Nucleic Acids Res.* **9**, 503-517
24. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in *Molecular Cloning*, p. 170, Cold Spring Harbor Laboratory, New York
25. Birnboim, H.C. and Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523
26. Kopecko, D.J. and Cohen, S.N. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1373-1377
27. Summers, J. (1975) *J. Virol.* **15**, 946-953
28. Palmer, J.D. and Thompson, W.F. (1981) *Gene* **15**, 21-26
29. Schmitt, J.J. and Cohen, B.N. (1983) *Anal. Biochem.* **133**, 462-464
30. Laemmli, U.K. (1970) *Nature* **227**, 680-685
31. Bonner, W.M. and Laskey, R.A. (1974) *Eur. J. Biochem.* **46**, 83-88
32. Cenatiempo, Y., Robakis, N., Reid, B.R., Weissbach, H. and Brot, N. (1982) *Arch. Biochem. Biophys.* **218**, 572-578
33. Peacock, S., Cenatiempo, Y., Robakis, N., Brot, N. and Weissbach, H. (1982) *Proc. Natl. Acad. Sci.* **79**, 4609-4612
34. Edelman, M. and Reisfeld, A. (1980) in *Organization and Expression in Plants*, C.Y. Leaver, Ed., pp. 353-362, Plenum Press, New York
35. Hallick, R.B. (1983) in *The Biology of Euglena*, Vol. 4, D.E. Buetow, Ed., Academic Press, New York (in press).
36. Hallick, R.B., Hollingsworth, M.J., Karabin, G.D., Nickoloff, J.A., Passavant, C. and Stiegler, G.L. (1983) *J. Cell Biochem.* **7B**, 295
37. McIntosh, L., Poulsen, C. and Bogorad, L. (1980) *Nature* **288**, 556-560
38. Krebbs, E.T., Larrinua, I.M., McIntosh, L. and Bogorad, L. (1982) *Nucleic Acids Res.* **10**, 4985-5002
39. Bloom, M., Brot, N., Cohen, B.N. and Weissbach, H. (1984) *Plant Molecular Biology*, in press.