Plant Molecular Biology **13**: 139–150, 1989.

Isolation, characterization and sequence analysis of a full-length cDNA clone encoding NADH-dependent hydroxypyruvate reductase from cucumber

John McC. Greenler, James S. Sloan, Brian W. Schwartz and Wayne M. Becker*
Departments of Botany and Genetics, University of Wisconsin, Madison, WI 53706, USA (author for correspondece; telephone (608)262–5833)*

Received 21 March 1989; accepted 31 March 1989

Key words: cDNA, gene expression, hydroxypyruvate reductase, light regulation, peroxisomal enzyme

Abstract

A full-length cDNA encoding NADH-dependent hydroxypyruvate reductase (HPR), a photorespiratory enzyme localized in leaf peroxisomes, was isolated from a λgt11 cDNA library made by reverse transcription of poly(A)$^+$ RNA from cucumber cotyledons. *In vitro* transcription and translation of this clone yielded a major polypeptide which was identical in size, 43 kDA, to the product of *in vitro* translation of cotyledonary poly(A)$^+$ RNA and subsequent immunoprecipitation with HPR antiserum. *Escherichia coli* cultures transformed with a plasmid construct containing the cDNA insert were induced to express HPR enzyme activity. RNA blot analysis showed that HPR transcript levels rise significantly in the first eight days of light-grown seedling development. This closely resembles the pattern seen for HPR-specific translatable mRNA. DNA blot analysis indicated that a single HPR gene is likely present per haploid genome. Nucleotide sequence analysis revealed an open reading frame of 1146 bases which encodes a polypeptide with a calculated molecular weight of 41.7 kDa. The derived amino acid sequence from this open reading frame is 26% identical and 50% similar to the amino acid sequence of the *E. coli* enzyme phosphoglycerate dehydrogenase, which catalyzes a similar reaction and functions in a related pathway. Statistical analyses show that this similarity is significant ($z > 10$). The derived amino acid sequence for HPR also contains the characteristics of an NAD-binding domain.

Introduction

NADH-dependent hydroxypyruvate reductase (HPR, EC 1.1.1.29) is a leaf peroxisomal enzyme [44, 39] which catalyzes the reduction of hydroxypyruvate to glycerate with the simultaneous oxidation of NADH to NAD$^+$ [48, 36]. This reaction is part of the glycolate pathway involved in photorespiration [43]. Most of the enzymes involved in this pathway are localized in leaf peroxisomes, a specialized class of peroxisomes found in plant photosynthetic tissues [18].

The cotyledons of greening cucumber seedlings provide an excellent model system for the study of peroxisome biogenesis and the regulation of the corresponding genes [45]. During the first several

Sequence data will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number XI4609

days of development, the seedling functions heterotrophically, with the cotyledons serving as the site of protein and lipid catabolism. Upon emergence, the cotyledons green and become photosynthetically active, with a concomitant light-stimulated increase in both photosynthetic and photorespiratory enzymes.

The appearance of HPR in the cotyledons of cucumber seedlings is both developmentally and light-regulated [16]. This pattern is seen for other peroxisomal enzymes, including serine:glyoxylate aminotransferase (SGAT) [16] and glycolate oxidase [45]. Similar developmental profiles have been found for leaf peroxisomal enzymes in the cotyledons of sunflower [31], watermelon [23, 24], mustard [17, 33], and wheat [8]. The response of peroxisomal enzymes to light is regulated at least partially by phytochrome [33], although other photoreceptors have also been implicated [37, 38, 8].

Previous research in our laboratory has shown that the rise in HPR and SGAT activities in the cotyledons of greening cucumber seedlings is regulated at a pretranslational level [16]. Using an immunoprecipitation assay, HPR protein levels were found to rise in parallel with HPR enzyme activity. Data from *in vitro* translation of poly(A)$^+$ RNA followed by immunoprecipitation with HPR antiserum indicated that the rise in HPR protein depends on an increase in translatable HPR mRNA.

In this paper we report the cloning and characterization of a full-length cDNA clone encoding NADH-dependent HPR from cucumber cotyledons. Using this clone as a probe, we found that HPR-specific transcript levels in cucumber cotyledons rise during seedling development in the light. Sequence analysis revealed that HPR is significantly similar to the catalytically related enzyme phosphoglycerate dehydrogenase (PGDH) from *Escherichia coli*, and contains the characteristics of an NAD-binding protein.

Portions of this work have been published in abstract form [19].

Materials and methods

Sources

Cucumber seeds (*Cucumis sativus* cv. Improved Long Green) were purchased from L. L. Olds Seed Co., Madison, WI. Unless otherwise noted, all enzymes were from New England Biolabs, all *E. coli* strains and vectors were from Stratagene Cloning Systems, and all reagents were from Sigma Chemical Company.

Growth, harvesting and preparation of plant material

Cucumber seedlings were germinated and grown either in continuous white light (150 μE, micro-Einsteins) or in darkness at 24–26 °C on vermiculite saturated with Hoagland's solution [15]. Harvesting of cotyledons and preparations of tissue homogenates for RNA analysis were as described previously [16].

λgt11 cDNA library construction and immunological screening

Total RNA was prepared from cotyledons of seven-day light-grown cucumber seedlings by homogenization in 4 M guanidine thiocyanate and ultracentrifugation through a discontinuous CsCl gradient [16]. Poly(A)$^+$ RNA was isolated from total RNA by oligo(dT)-cellulose chromatography [21]. Double-stranded, blunt-end cDNA was synthesized using the Amersham cDNA synthesis system, which is based on the method of Gubler and Hoffman [13]. The cDNA product was methylated with *Eco* RI methylase, blunt-end ligated to phosphorylated decameric *Eco* RI linkers (New England Biolabs), digested with *Eco* RI endonuclease, and size-fractionated by Sepharose CL-4B chromatography [20]. cDNAs greater than 500 bp in length were ligated to λgt11 DNA which had been cut and dephosphorylated at the *Eco* RI restriction site. Ligation products were packaged with lambda head and tail proteins to produce viable phage particles. The resulting

library consisted of approximately 1×10^7 recombinant clones. One-third of the library was amplified on *E. coli* Y1090 (r^-, m^+) to allow multiple screening [28].

The cDNA library was screened for HPR-specific clones using an antiserum previously prepared against native HPR from cucumber cotyledons [40]. Antibodies cross-reacting with *E. coli* and lambda antigens were removed from this serum by pseudoscreening [22]. To screen with this antiserum, 5×10^5 phage were plated out and overlaid with nitrocellulose filters saturated with isopropyl-β-D-thiogalactopyranoside (IPTG) [20]. HPR antigen was detected using the Western blot technique of Burnette [3] as modified by Tokuhisa *et al.* [42]. Three positive plaques were replated and rescreened (2–3 cycles) until all plaques yielded only positive signals. The size of the cDNA inserts in these three clones were determined by digestion of purified lambda DNA with *Eco* RI, and subsequent agarose gel electrophoresis [28].

In vitro *transcription and translation, and immunoprecipitation*

The largest cDNA (H18) was subcloned into the *Eco* RI site of the plasmid vector Bluescribe (pBS), such that it was flanked on either side by the T7 and T3 RNA polymerase promoters. Transcription reactions were performed using the circular plasmid, T7 or T3 RNA polymerase, and 5′7meGppp5′G according to Stratagene's recommended conditions. Transcripts generated *in vitro* and poly(A)$^+$ RNA were translated in a reticulocyte lysate system from Stratagene. HPR polypeptide was immunoprecipitated from the *in vitro* translation mixture by the addition of anti-native HPR serum and protein A-Sepharose [1]. Polypeptides labeled with [^{35}S]methionine were fractionated by electrophoresis through 10% (w/v) polyacrylamide gels in the presence of sodium dodecyl sulfate (SDS) [25].

Extraction of E. coli *proteins and HPR enzyme assay*

The H18 cDNA was subcloned in both orientations into the *Eco* RI site in the N-terminal portion of the *LacZ* gene fragment of the plasmid vector Bluescript (pKS). These constructs were then used to transform *E. coli* XL-1 Blue competent cells. 100-ml cultures were grown at 37 °C with vigorous shaking. At late log phase (A_{600} = 0.6), IPTG was added to a final concentration of 10 mM and the cultures were incubated for two more hours. Cells were then harvested by centrifugation at 2500 *g* for 10 min, and resuspended in 10 ml of STET (0.1 M NaCl, 10 mM Tris-HCl pH 8, 0.1 mM EDTA, 0.5% Triton X-100). Cells were lysed by the addition of lysozyme to a final concentration of 1.0 mg/ml, followed by two 10 sec sonication bursts [32]. Cellular debris and remaining intact cells were removed by centrifugation at 10000 *g* for 10 min. HPR enzyme activity in the resulting supernatant was measured as previously described [40].

RNA blot analysis

Total RNA was prepared from cotyledons of cucumber seedlings as described above, and poly(A)$^+$ RNA was isolated by poly(U)-Sephadex chromatography [16]. RNA was fractionated through 1.2% agarose gels containing 2.2 M formaldehyde, transferred to nitrocellulose or Nytran (Schleicher and Schuell) by blotting, hybridized at 42 °C in 50% (v/v) formamide, and washed, all according to the protocol of Schleicher and Schuell. These blots were probed with α[^{32}P]dATP-oligolabeled H18 fragment from the pBS subclone [9].

DNA blot analysis

Total cellular DNA greater than 50 kbp in length was isolated from frozen cucumber leaves according to the procedure of Dellaporta *et al.* [7]. The DNA was digested with *Eco* RI or *Hin*d III,

fractionated by agarose gel electroporesis (0.6% agarose, 1 × TBA at 25 V for approximately 20 h), and stained with ethidium bromide [28]. The stained gel was irradiated on a UV lightbox for five minutes to facilitate transfer of large DNA fragments. The DNA was denatured and transferred to a Nytran membrane according to the manufacturer's protocol.

The DNA blot was hybridized with a single-stranded DNA probe produced by primer extension in the presence of $\alpha[^{32}P]$dATP [5]. The template used for this reaction was a deletion derivative (H18.1-3) of the H18 clone lacking the poly(A) tail and 18 bases on the 3′ end of the cDNA (produced as described below 'DNA sequencing and sequence analysis'). Hybridization was at 42 °C overnight in hybridization solution containing 50% formamide. The blot was washed twice at room temperature in 2 × SSC, 0.1% SDS for 15 min, and once at room temperature in 0.1 × SSC, 0.1% SDS for 1.0h [28].

DNA sequencing and sequence analysis

The H18 cDNA was subcloned in both orientations into the *Eco* RI site of the M13 phage vector Phagescript. A series of nested deletions was produced for both orientations using Exonuclease III and S_1 nuclease as described by Henikoff [14] with the addition of a phenol-chloroform extraction and ethanol precipitation between the S_1 nuclease digest and the Klenow polymerase fill-in reaction. The S_1 reaction was terminated with 40 μl of stop buffer (200 mM NaCl 40 mM Tris-HCl 7 mM EDTA, pH 9.0). The fill-in reaction was carried out in 20 μl of Klenow mix (60 mM Tris-HCl, 10 mM $MgCl_2$ 0.25 μ/μl Klenow polymerase, pH 7.5) to which 30 μl of ligase mix was added. The dideoxynucleotide chain termination method was used for sequencing a series of overlapping clones in both orientations with either Sequenase polymerase or the Klenow fragment of DNA polymerase I, and 17-mer M13 primer (United States Biochemical) [30].

Sequence analysis was carried out with GCG sequence analysis software [6] or software developed by Pearson and Lipman [29]. Propensity measures for the formation of alpha-helixes and beta-sheets for the derived HPR polypeptide sequence were determined by the GCG program PepPlot, which uses the algorithms of Chou and Fasman [4]. The derived HPR amino acid sequence was compared with the National Biomedical Research Foundation (NBRF) protein sequence database using the GCG WordSearch program and with a fusion of the Genebank and EMBL nucleic acid sequence databanks using TFASTA [29]. An amino acid sequence dot-matrix plot was generated by a program written in our laboratory for IBM personal computers (program available upon request). An alignment of the derived amino acid sequences for HPR and PGDH was produced with the GCG program BestFit using a gap weight penalty of 3.0 and a gap length penalty of 0.1. Mismatches were scored based on the evolutionary distance between the amino acids as measured by Dayhoff and normalized by Gribskov and Burgess [12].

The program RDF2 [29] was used to determine the statistical significance of the similarity between the derived amino acid sequences for HPR and PGDH. The PGDH sequence was shuffled 100 times, either globally or locally in windows of 10. (A global shuffle scrambles the sequence by shuffling individual residues, while a local shuffle scrambles blocks of residues. The latter method negates the effect of locally biased amino acid composition which may be due to patches of hydrophobic or charged amino acid residues.) Each shuffle was then compared with the HPR sequence. The scores from these comparisons were then used to determine the significance of the genuine sequence similarity by calculating the z value [z = (genuine similarity score – mean of random similarity scores)/(standard deviation of random similarity scores) [26].

Results

Identification and isolation of hydroxypyruvate reductase cDNA clones

A λgt11 cDNA library was generated from poly(A)$^+$ RNA that was isolated from cotyledons of seedlings grown for seven days in continuous light. HPR cDNA clones in this library were identified by immunological screening with an anti-HPR polyclonal serum. Of 5×10^5 cDNA recombinant phage screened, approximately 100 plaques reacted positively with the anti-HPR serum. Since one-sixth of the cDNA inserts would be expected to have the proper orientation and frame for expression, we estimate that approximately 0.12% of the phage in the library contained an HPR cDNA insert. The 20 plaques which gave the strongest signals were picked, and three of these were plaque-purified to homogeneity. These three clones contained inserts with sizes of 0.8, 1.0 and 1.5 (kbp).

The identity of the largest clone, λH18, was confirmed in two ways. For initial confirmation, the H18 insert was subcloned into the plasmid vector Bluescribe (pBSH18) and used as a template for *in vitro* transcription and translation. Figure 1 shows an SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of the polypeptide products from these reactions. Using T7 RNA polymerase, the reactions produced a major polypeptide with an apparent molecular weight (M_r) of approximately 43 kDa and a minor polypeptide of 33 kDa (lane 4). Reactions using T3 polymerase did not produce any detectable polypeptide (not shown). The products of the T7 polymerase reaction were identical in size to those produced by *in vitro* translation of poly(A)$^+$ RNA from cotyledons of seven-day light-grown seedlings followed by an immunoprecipitation with anti-HPR serum (lane 3). In addition, the polypeptide products from the T7 polymerase reactions were precipitated with anti-HPR serum (lane 6), but not with the corresponding preimmune serum (lane 5).

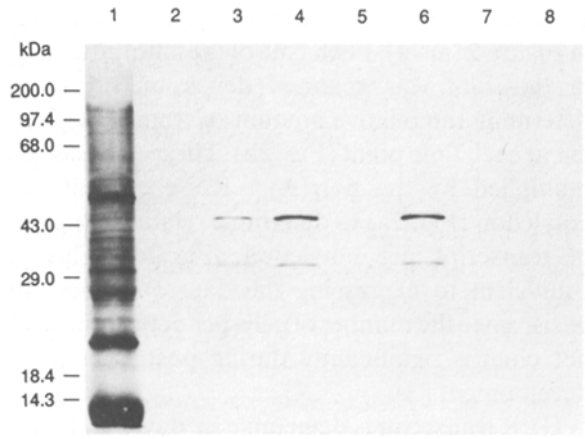

Fig. 1. Comparison of the H18 cDNA transcription-translation product with hydroxypyruvate reductase polypeptide. [^{35}S]methionine-labeled polypeptides were synthesized by *in vitro* translation using cotyledonary poly(A)$^+$ RNA from seedlings grown for seven days in the light (lanes 1–3), RNA transcribed from the H18 cDNA (lanes 4–6), RNA transcribed from pBS vector (lane 7), or no RNA (lane 8). Products of translation were immunoprecipitated with anti-HPR serum (lanes 3 and 6), or with preimmune serum (lanes 2 and 5). Samples were subjected to electrophoresis through 10% polyacrylamide gels in the presence of SDS.

For further confirmation, the H18 insert was subcloned into the plasmid vector Bluescript in both orientations (pKSH18-1 and pKSH18-2). *E. coli* cultures transformed with these constructs were induced to produce fusion protein containing the HPR polypeptide and the amino terminus of β-galactosidase. Table 1 shows the HPR enzyme activities associated with these cultures. Induced cultures containing the H18 insert in the

Table .1 HPR activities in H18 transformed *E. coli* cultures.

Culture[a]	Induction[b]	HPR activity[c] (units/ml)
pKSH18-1	+	0.25 ± 0.01
pKSH18-1	−	0.09 ± 0.01
pKSH18-2	+	0.08 ± 0.00
pKS	+	0.02 ± 0.00

[a] Cultures of *E. coli* were transformed with wild-type plasmid (pKS) or with plasmid containing the H18 insert in either of the two orientations (pKSH18-1 or pKSH18-2).
[b] Cultures were either induced with IPTG at late log phase (+) or left uninduced (−).
[c] The supernatants from lysates of these cultures were assyed for HPR activity.

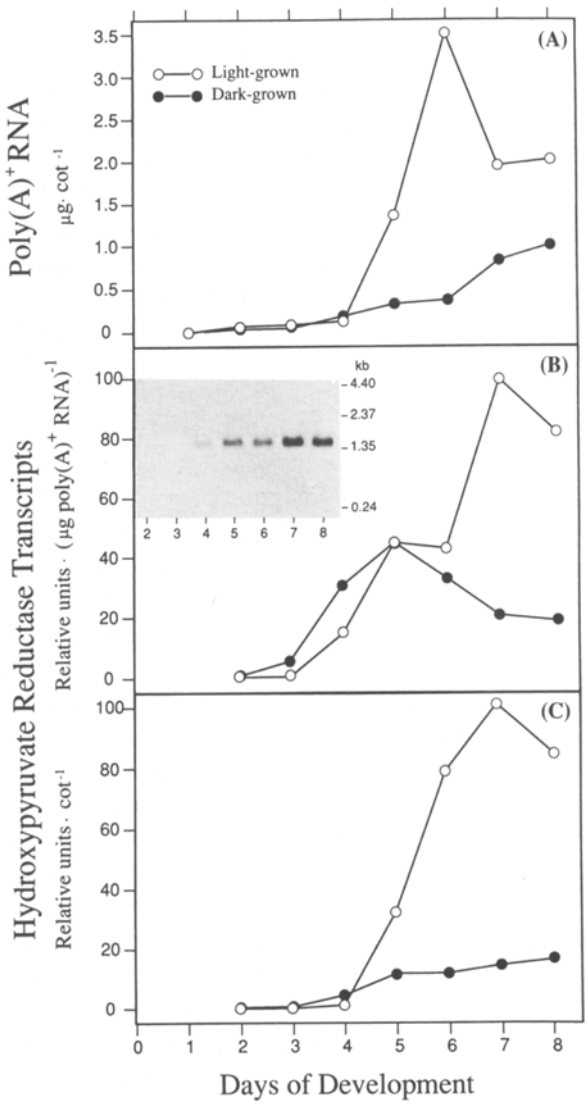

Fig. 2. Developmental accumulation of poly(A)$^+$ RNA and HPR transcript in the cotyledons of light-grown and dark-grown cucumber seedlings. A. Amounts of poly(A)$^+$ RNA per cotyledon for light-grown and dark-grown cucumber seedlings. B. Relative amounts of HPR transcript per μg of poly(A)$^+$ RNA in cotyledons from light-grown and dark-grown seedlings. Cotyledonary poly(A)$^+$ RNA (1.5 μg) was subjected to electrophoresis through agarose/formaldehyde, blotted to Nytran, and probed with α[^{32}P]dATP-labeled H18 DNA. The RNA blot for light-grown seedlings is shown in the inset; poly(A)$^+$ RNA from dark-grown seedlings yielded a single band with the same molecular weight (blot not shown). The band in each lane of both blots was scanned densitometrically, and all values were normalized to the peak value (day 7 = 100 relative units). C. Relative amounts of HPR transcript per cotyledon from light-grown and dark-grown seedlings. Each of the densitometric values described in (B) was multiplied by the appropriate poly(A)$^+$ RNA content per cotyledon and normalized to the peak value (day 7 = 100 relative units).

orientation which would yield fusion protein from the β-gal promoter (pKSH18-1) had levels of HPR enzyme activity 12.5 times greater than background levels for induced cultures transformed with the pKS vector alone. Uninduced cultures of pKSH18-1 and induced cultures with the H18 insert in the opposite orientation (pKSH18-2) had HPR activity levels that were only 4.0 and 4.5 times greater than background, respectively. The HPR activity found in these two cultures may be due to low levels of transcription from the uninduced β-galactosidase promoter or to transcription of the fusion gene from other promoters on the plasmid.

Accumulation of HPR mRNA during seedling development

The appearance of HPR mRNA during seedling development in both the light and dark was quantified by RNA blot analysis using the HPR cDNA fragment H18 as a probe. The approximate size of the HPR transcript in cotyledons is 1.5 kb as determined from the RNA blot shown in Figure 2 (inset). Each lane of the autoradiogram of this blot was scanned densitometrically to determine the relative amount of transcript present at each time point (Fig. 2b). These values were multiplied by the poly(A)$^+$ RNA content per cotyledon (Fig. 2a) to determine relative amounts of transcript per cotyledon (Fig. 2c). This is equivalent to expressing the data on a per-cell basis, since the number of cells per cotyledon does not change significantly during postgerminative development [2].

HPR transcript is detectable by day 4 and rises rapidly thereafter in the light, while remaining at low levels in the dark. The increase in transcript abundance in the light is due in part to the effect of light on the level of poly(A)$^+$ RNA. However, light also triggers a rise in the abundance of HPR-specific transcript relative to poly(A)$^+$ RNA.

Similar results were found when these blots were probed with a cDNA isolated with an antibody against SGAT, while probes corresponding to malate synthase and isocitrate lyase dedected transcripts whose levels peaked and declined beyond detection within the same time period [19].

HPR genomic fragments

Figure 3 shows a DNA blot of total cellular DNA from cucumber leaves, digested with either *Eco* RI or *Hin*d III endonuclease, and probed with H18.1-3 DNA. Digestion with *Eco* RI resulted in detection of a single major fragment, 14 kbp in length, while *Hin*d III digestion produced two detectable fragments, 4.4 kbp and 1.2 kbp. Washes at higher stringency (65 °C instead of room temperature) produced results similar to those shown. Digestion with other restriction

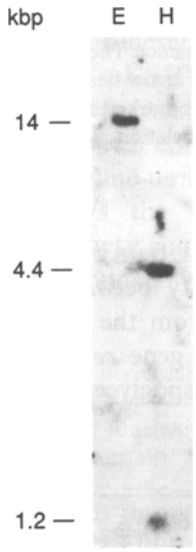

Fig. 3. Hybridization of HPR cDNA to endonuclease-digested cucumber leaf DNA. DNA from cucumber leaves was digested with either *Eco* RI (E) or *Hin*d III (H) endonuclease, subjected to agarose gel electrophoresis, blotted to Nytran and probed with $\alpha[^{32}P]$dATP-labeled H18.1-3 DNA. The final wash was at 25 °C in 0.1 × SSC, 0.1% SDS for one hour. Washing at higher stringencies produced similar results (data not shown). Molecular weights of DNA fragments were determined by DNA standards run in parallel (not shown).

nucleases which also recognize six-base oligomers yielded only one or two fragments (data not shown). These results suggest that a single gene encoding HPR is present per haploid genome. Sequence analysis (described below) of the cDNA predicted the presence of a 0.3 kbp *Hin*d III fragment which was not detected on gels designed to resolve DNA of this size (data not shown). This finding suggests the presence of at least one intron located between the two *Hin*d III sites of the cDNA.

HPR cDNA and derived polypeptide sequences

A series of nested deletions was created through the H18 insert in the M13 vector Phagescript, and a subset was sequenced. Figure 4 shows the nucleotide sequence and derived amino acid sequence. The cDNA contains an open reading frame of 1146 nucleotides, flanked at the 5′ end by 40 bases of noncoding sequence, and at the 3′ end by 207 bases of noncoding sequence and a 44 base poly(A) tail. All but the first two residues of sequence at the putative translation start site (TCCAATGGC; underlined in Fig. 4) match the reported consensus sequence for this region in plants (AACAATGGC) as described by Lutcke *et al.* [27]. The derived amino acid sequence from the open reading frame codes for a polypeptide of 382 amino acid residues with a calculated M_r of 41.7 kDa. This is in good agreement with the range of apparent M_r from 39 to 44 kDa found for the HPR polypeptide by SDS-PAGE [40].

The 33 kDa polypeptide product of *in vitro* transcription and translation of the H18 insert (Fig. 1, lane 4) is consistent with secondary initiation of translation at a downstream internal ATG initiation codon. Six of the nine residues at this putative translation start site (TAATATGGC; underlined with dashes in Fig. 4) match the consensus sequence. A polypeptide of 290 residues with a calculated M_r of 31.5 kDa is encoded by this reading frame.

```
                                                                                                    GGTGGAGAAAAGAGCTTCCAAGCTTTCAAGGTTAGCTCCA    40

Met Ala Lys Pro Val Gln Ile Glu Val Trp Asn Pro Asn Gly Lys Tyr Arg Val Val Ser Thr Lys Pro Met Pro Gly Thr Arg Trp Ile
ATG GCG AAA CCA GTT CAA ATC GAA GTA TGG AAT CCA AAT GGA AAG TAC AGA GTT GTC AGC ACT AAG CCT ATG CCT GGA ACT CGC TGG ATC   130

Asn Leu Leu Ile Glu Gln Asp Cys Arg Val Glu Ile Cys Thr Glu Lys Lys Thr Ile Leu Ser Val Glu Asp Ile Leu Ala Leu Ile Gly
AAT CTC TTG ATC GAA CAA GAT TGC CGA GTC GAA ATT TGT ACT GAG AAA AAG ACG ATA CTC TCG GTT GAG GAT ATT CTT GCT CTC ATC GGC   220

Asp Lys Cys Asp Gly Val Ile Gly Gln Leu Thr Glu Asp Trp Gly Glu Val Leu Phe Ser Ala Leu Ser Arg Ala Gly Gly Lys Ala Phe
GAT AAG TGC GAC GGC GTC ATT GGA CAG TTG ACT GAA GAT TGG GGA GAG GTG CTG TTT TCT GCG TTA AGC AGA GCG GGA GGG AAA GCT TTT   310

Ser Asn Met Ala Val Gly Tyr Asn Asn Val Asp Val Asn Ala Ala Asn Lys Tyr Gly Val Ala Val Gly Asn Thr Pro Gly Val Leu Thr
AGT AAT ATG GCT GTT GGT TAC AAT AAT GTA GAT GTT AAT GCT GCT AAT AAG TAC GGT GTT GCT GTA GGA AAT ACT CCT GGA GTA CTT ACA   400

Glu Thr Thr Ala Glu Leu Ala Ala Ser Leu Ser Leu Ala Ala Ala Arg Arg Ile Val Glu Ala Asp Glu Phe Met Arg Ala Gly Arg Tyr
GAG ACT ACA GCA GAG TTG GCA GCT TCA CTT TCT CTT GCA GCT GCA AGA AGG ATA GTT GAA GCA GAT GAG TTT ATG AGG GCA GGT CGA TAT   490

Asp Gly Trp Leu Pro Asn Leu Phe Val Gly Asn Leu Leu Lys Gly Glu Thr Val Gly Val Ile Gly Ala Gly Arg Ile Gly Ser Ala Tyr
GAT GGA TGG CTT CCA AAT TTG TTT GTT GGG AAC TTG CTG AAA GGA CAG ACT GTT GGT GTG ATT GGA GCT GGT CGT ATT GGA TCT GCT TAT   580

Ala Arg Met Met Val Glu Gly Phe Lys Met Asn Leu Ile Tyr Phe Asp Leu Tyr Gln Ser Thr Arg Leu Glu Lys Phe Val Thr Ala Tyr
GCA AGA ATG ATG GTA GAA GGG TTT AAG ATG AAC CTG ATC TAC TTC GAT CTT TAC CAG TCA ACT AGA CTC GAA AAG TTC GTT ACA GCC TAT   670

Gly Glu Phe Leu Lys Ala Asn Gly Glu Ala Pro Val Thr Trp Arg Arg Ala Ser Ser Met Asp Glu Val Leu Arg Glu Ala Asp Val Ile
GGT GAG TTC CTG AAA GCC AAC GGT GAG GCT CCT GTG ACA TGG AGA AGA GCA TCA TCA ATG GAT GAG GTG CTT CGA GAA GCA GAT GTG ATA   760

Ser Leu His Pro Val Leu Asp Lys Thr Thr Phe His Leu Val Asn Lys Glu Ser Leu Lys Ala Met Lys Lys Asp Ala Ile Leu Ile Asn
AGT CTT CAT CCC GTA CTG GAT AAA ACT ACC TTC CAT CTT GTG AAC AAA GAA AGT CTT AAA GCG ATG AAG AAG GAT GCA ATT CTC ATT AAC   850

Cys Ser Arg Gly Pro Val Ile Asp Glu Ala Ala Leu Val Asp His Leu Arg Asp Asn Pro Met Phe Arg Val Gly Leu Asp Val Phe Glu
TGC AGT AGG GGA CCT GTG ATC GAT GAA GCA GCC CTT GTT GAT CAT TTG AGA GAT AAT CCC ATG TTT CGA GTT GGC CTT GAT GTT TTT GAG   940

Asp Glu Pro Tyr Met Lys Pro Gly Leu Ala Asp Met Lys Asn Ala Ile Ile Val Pro His Ile Ala Ser Ala Ser Lys Trp Thr Arg Glu
GAT GAA CCA TAC ATG AAA CCT GGA CTT GCT GAT ATG AAG AAC GCG ATC ATT GTT CCT CAC ATT GCT TCT GCT TCC AAG TGG ACT CGC GAA  1030

Gly Met Ala Thr Leu Ala Ala Leu Asn Val Leu Gly Lys Ile Lys Gly Tyr Pro Val Trp Ser Asp Pro Asn Arg Val Glu Pro Phe Leu
GGA ATG GCA ACA CTG GCT GCT CTT AAT GTC TTG GGA AAG ATT AAA GGA TAT CCT GTT TGG TCC GAT CCA AAC CGA GTA GAA CCG TTC CTC  1120

Asp Glu Asn Val Ser Pro Pro Ala Ala Ser Pro Ser Ile Val Asn Ala Lys Ala Leu Gly Asn Ala  *
GAC GAG AAT GTT TCA CCT CCA GCT GCA TCA CCT AGC ATT GTG AAT GCC AAA GCC TTG GGT AAC GCC TAA CAGAAACCTCCCTAACTGAAAAGGATC  1216

GTCTACTTCAAAGCTGTGAAGCAGGAGGGGATGGCTAAAAAATGTATGAGCACGAATTGGAAGCATTATAGATATGGGGGTAAATGGAGGTTGTATTTCATGAATGTACAACTCTGGTC  1335

AATTGATTATAATGTGTATCTTATCTTCTTCCTTCAATGCAATTCCATTACTTGGTTTAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAATAAAAAAAAAAAA
```

Fig. 4. Nucleotide sequence of the HPR cDNA H18 and the derived HPR amino acid sequence. The derived amino acid sequence is shown above the nucleotide sequence. The two internal *Hind* III sites are marked by arrow heads. The translation start site for the reading frame coding for HPR is underlined, and the corresponding stop codon is marked with an asterisk. The putative internal translation start site for the minor polypeptide produced upon *in vitro* translation is underlined with dashes.

Amino acid sequence comparisons

The derived amino acid sequence from the H18 cDNA open reading frame described above was compared with sequences in the NBRF and Genebank/EMBL data bases. Both comparisons revealed similarity between the derived amino acid sequence from the HPR cDNA and that from the *E. coli* gene *serA*, which encodes the enzyme phospoglycerate dehydrogenase

Table 2. Statistical significance of the HPR/PGDH protein sequence similarity.

Comparison[a]	HPR/PGDH Score[b]	Mean of Scores	Standard deviation of scores	Statistical significance (z value)[c]
Initial scan	117	25.6	6.6	13.8
Randomized:				
Global	148	38.4	7.8	14.1
Local	148	40.8	10.7	10.0

[a] Similarity scores for the HPR/PGDH match were compared to the scores for all other comparisons in the initial data-base scan, or to the scores from 100 comparisons of HPR with either global or local randomizations of PGDH.

[b] The scores are expressed as initial values for the initial scan, and as optimized values for the randomized comparisons.

[c] z value = (HPR/PGDH score − mean score)/standard deviation. $z > 10$ denotes statistical significance [26].

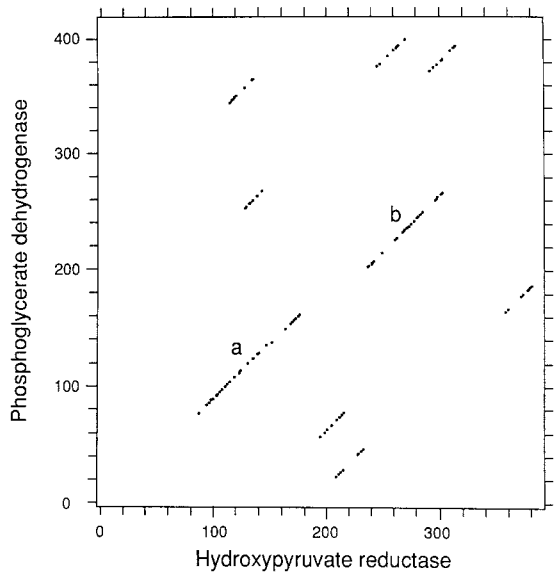

Fig. 5. Dot-matrix plot comparison of the derived amino acid sequences for HPR and PGDH. The derived amino acid sequence for cucumber HPR (residues 1–382; abscissa) is compared with that for *E. coli* PGDH (residues 1–410; ordinate). Each dot represents an identical match between a single amino acid residue in HPR and a residue in PGDH. A match was plotted only if it occurred within a window of 26 residues containing at least eight matches. Diagonals of dots labeled a and b correspond to the alignment of HPR and PGDH shown in Fig. 6.

```
HPR-S   KPVSIEVYNPSGKL 14
HPR-C MAKPVQIEVWNPNGKYRVVSTKPMPGTRWINLLIEQDCRVEICTEKKTIL
PGDH  MAK.....VSLEKDKIKFLLVEGVHQKALESLRAAGYTNIEF...HKGAL

   51 SVEDILALIGDKCDGVIGQLTEDWGEV.LFSALSRAGGK..AFSNMAVGY
   43 DDEQLKESIRD......AHFIGLRSRTHLTEDVINAAEKLVAIGCFCIGT

   98 NNVDVNAANKYGVAVGNTPGVLTETTAELAASLSLAAARRIVEADEFMRA
   87 NQVDLDAAAKRGIPVFNAPFSNTRSVAELVIGELLLLLRGVPEANAKAHR

                        △□□G G G □ □      □□○
  148 GRYDGWLPNLFVGNLLKGQTVGVIGAGRIGSAYARMMVEGFKMNLIYFDL
  137 GVWNKLAAGSFEA...RGKKLGIIGYGHIGTQLG.ILAESLGMYVYFYDI

  198 YQSTRLEKFVTAYGEFLKANGEAPVTWRRASSMDEVLREADVISLHPVLD
  183 ENKLPL.............GNAT....QVQHLSDLLNMSDVVSLHVPEN

  248 KTTFHLVNKESLKAMKKDAILINCSRGPVIDEAALVDHLRDNPMFRVGLD
  215 PSTKNMMGAKEISLMKPGSLLINASRGTVVDIPALCDALASKHLAGAAID

  298 VFEDEPYMKPG.....LADMKNAIIVPHIASASKWTREGMATLAALNVLG
  265 VFPTEPATNSDPFTSPLCEFDNVLLTPHIGGSTQEAQENIGLEVAGKLIK

  343 KIKGYPVWSDPN.............RVEPFLDENVSPPAASPSIVNAKALG
  315 YSDNGSTLSAVNFPEVSLPLHGGRRLMHIHENRPGVLTALNKIFAEQGVN

  381 NA
      |
  365 IA
```

Fig. 6. Alignments of the derived amino acid sequences for HPR and PGDH. The derived amino acid sequence for cucumber HPR (HPR-C) is aligned with the derived amino acid sequence for *E. coli* PGDH and with the N-terminal sequence for spinach HPR (HPR-S). Gaps inserted for optimal alignment are represented by dots in the sequence. Identical amino acid residue matches between two sequences are connected by a solid line (|). Mismatches were scored based on the evolutionary distance between two amino acids with a score of -1.5 for the greatest distance and 1.5 for the least distance. Matches with a score of $\geq 0.5 < 1.5$ are connected by double dots (:), while matches with a score of $\geq 0.1 < 0.5$ are connected by a single dot (.). The HPR and PGDH sequences believed to be involved in binding of the ADP moiety of NAD are boxed. The reported amino acid sequence fingerprint for this binding site is shown above the cucumber HPR sequence. A triangle represents a basic or hydrophilic residue, a square represents a small and hydrophobic residue, a circle represents an acidic residue, and G represents glycine.

(PGDH) [41]. Figures 5 and 6 show a dot-matrix plot comparison and alignment, respectively, of these two sequences. Using this alignment, the derived sequences for HPR and PGDH were found to be 25.7% identical and 50.0% similar. Table 2 presents the results of a statistical analysis of the similarity between the derived amino acid sequences. This analysis shows the similarity between HPR and PGDH to be significant ($z > 10$) by the criteria of Lipman and Pearson [26].

Crystallographic analyses of NAD-dependent oxidoreductases have shown that these proteins are functionally divisible into two domains, the NAD-binding domain and the catalytic domain [41]. Eleven amino acids in the NAD-binding domain have been found to be conserved and are believed to function in binding of the ADP moiety of NAD [47]. The sequences in HPR and PGDH (G. Grant, personal communication) corresponding to this binding site are boxed in Fig. 6, and the conserved amino acid sequence fingerprint is indicated above the HPR and PGDH sequences. The HPR sequence matches ten out of the eleven fingerprint residues. The predicted secondary

structure for this sequence in HPR (data not shown) also matches the known $\beta\alpha\beta$-fold structure for NAD-binding domains [47].

A 14-residue amino-terminal polypeptide sequence for HPR from spinach was provided by Dr C.-H. Chen (personal communication). This sequence was found to match an amino terminal cucumber HPR sequence at 11 out of 14 positions (Fig. 6). This match provides additional evidence that the identified reading frame in the H18 fragment codes for HPR.

Discussion

Immunological screening of the cDNA library constructed from poly(A)$^+$ RNA isolated from the cotyledons of seven-day light-grown cucumber seedlings revealed that approximately 0.12% of the phage in this labrary contained an HPR cDNA insert. This is in good agreement with our findings that HPR-specific mRNA represents 0.13% of total translatable mRNA from the same tissue [16]. One of the clones from this library, λH18, contains an insert which is similar in size (1.5 kbp) to the corresponding transcript as determined by RNA blot analysis. The identity of this clone was verified by *in vitro* transcription and translation, induction of HPR enzyme activity in *E. coli* cultures transformed with the HPR cDNA, and comparison of the N-terminal sequence of spinach HPR with the corresponding derived amino acid sequence from the cucumber HPR cDNA.

HPR transcript levels in cucumber cotyledons are regulated both developmentally and by light (Fig. 2). After four days of seedling development, HPR transcripts become detectable in cotyledons. At this point, they either rise rapidly in the light or remain at low levels in the dark. This pattern is similar to that seen for HPR enzyme activity, HPR protein, and HPR translatable mRNA [16]. These findings substantiate our earlier suggestion that both the initial appearance of HPR enzyme and the subsequent enhancement of its abundance in the light may be regulated at the level of transcription [16]. To test this hypothesis further, we plan to perform nuclear run-on experiments.

A similar developmental profile was found using a cDNA clone isolated with a serum against another leaf peroxisome-specific protein, SGAT [19]. Earlier results have shown that HPR and SGAT protein synthesis are regulated in a parallel fashion [16]. In contrast with the developmental profile seen for HPR and SGAT, the transcript levels for the glyoxysomal enzymes malate synthase (MS) and isocitrate lyase rise rapidly during seedling development in the light with a peak at day five, and then decline below the limits of detection by day eight [19]. These results are similar to those reported by Smith and Leaver [35] for MS in the cotyledons of germinating cucumber seedlings and serve as positive controls for RNA samples isolated from early time points.

Sequencing of the H18 cDNA fragment revealed that it contains an 1146-base open reading frame which codes for a polypeptide with an M_r of 41.7 kDa. This size is in good agreement with SDS-PAGE determination of the apparent M_r of the HPR polypeptide. The similarity between the translation start site for this open reading frame and the reported consensus sequence for this region in plants [27] is consistent with the hypothesis that it codes for the HPR polypeptide.

A computer search of sequence data bases found a statistically significant similarity between the derived amino acid sequence from the HPR cDNA and the derived amino acid sequence from the *E. coli* gene *SerA*, which encodes the enzyme PGDH. PGDH and HPR have similar enzymatic activities: HPR catalyzes the reduction of hydroxypyruvate to glycerate, while PGDH catalyzes the oxidation of phosphoglycerate to phosphohydroxypyruvate. Both enzymes function in pathways involved in serine metabolism, and both utilize NADH/NAD$^+$ as a cofactor. A putative NAD-binding site was identified in the derived HPR peptide sequence at the same location as the proposed NAD-binding site for PGDH. These sequence characteristics for HPR provide additional evidence that the identified open reading frame in the H18 fragment codes for HPR and

should aid in the structure-function analysis of HPR and NAD-dependent oxidoreductases in general.

Putative peroxisomal targeting signals have recently been identified by several laboratories. Targeting signals of 27 amino acids or less have been located at the carboxy termini of five different animal peroxisomal proteins [10, 11]. Each of these signals contains the tripeptide sequence Ser-Lys/His-Leu. This sequence is not found in HPR. However, it is known that for at least one other protein, acyl-CoA oxidase from *Candida tropicalis*, this sequence is not necessary for import into peroxisomes [34]. Volokita and Somerville [46] reported a sequence similarity among the plant peroxisomal proteins glycolate oxidase, MS, and uricase II, which they speculate may be a peroxisomal 'sorting sequence'. This sequence is also not present in HPR.

DNA blot analysis indicated that the HPR gene in cucumber is likely present in a single copy per haploid genome. We are in the process of cloning genomic fragments containing the HPR gene for use in the study of *cis*-acting elements and *trans*-acting factors in the regulation of HPR protein synthesis. We are also using the HPR cDNA clones to generate HPR polypeptide *in vitro* for the study of targeting and transport of HPR into peroxisomes.

Acknowledgements

We thank Drs Gregory A. Grant and Chong-maw Chen for permission to cite their unpublished work, and Ann Bakken and Amy Hasseldahl for their technical assistance. This work was supported in part by NSF grants DCB-8509852 and DCB-8801955 to WMB, and by NIH traineeships to JS and BWS.

References

1. Anderson DJ, Blobel G: Immunoprecipitation of proteins from cell-free translations. Meth Enzymol 96: 111–120 (1983).
2. Becker WM, Leaver CJ, Weir EM, Riezman H: Regulation of glyoxysomal enzymes during germination of cucumber I. Developmental changes in cotyledonary protein, RNA and enzyme activities during germination. Plant Physiol 62: 542–549 (1978).
3. Burnette WN: 'Western blotting': electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radiolabeled protein A. Anal Biochem 112: 195–203 (1981).
4. Chou PY, Fasman GD: Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47: 45–148 (1978).
5. Church GM, Gilbert W: Genomic sequencing. Proc Natl Acad Sci USA 81: 1991-1995 (1984).
6. Devereux J, Haeberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395 (1984).
7. Dellaporta SL, Wood J, Hicks JB: A plant DNA minipreparation: version II. Plant Mol Biol Reporter 1: 19–21 (1983).
8. Feierabend J: Developmental studies on microbodies in wheat leaves. III. On the photocontrol of microbody development. Planta 123: 63–77 (1975).
9. Feinberg AP, Vogelstein B: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132: 6–13 (1983).
10. Gould SJ, Keller G-A, Subrimani S: Identification of a peroxisomal targeting signal at the carboxy terminus of firefly luciferase. J Cell Biol 105: 2923–2931 (1987).
11. Gould SJ, Keller G-A, Subrimani S: Identification of peroxisomal targeting signals located at the carboxy terminus of four peroxisomal proteins. J Cell Biol 107: 897–905 (1988).
12. Gribskov M, Burgess RR: Sigma factors from *E. coli*, *B. subtilis*, phage SP01 and phage T4 are homologous proteins. Nucleic Acids Res 14: 6745–6763 (1986).
13. Gubler U, Hoffmann BJ: A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269 (1983).
14. Henikoff S: Unidirectional digestion with exonuclease III in DNA sequence analysis. Meth Enzymol 155: 156-165 (1987).
15. Hoagland DR, Snyder WC: Nutrition of strawberry plant under controlled conditions. (a) Effects of deficiencies of boron and certain other elements. (b) Susceptibility to injury from sodium salts. Proc Am Soc Hort Sci 30: 288–293 (1933).
16. Hondred D, Wadle D-M, Titus DE, Becker WM; Light-stimulated accumulation of the peroxisomal enzymes hydroxypyruvate reductase and serine : gloxylate aminotransferase and their translatable mRNAs in cotyledons of cucumber seedlings. Plant Mol Biol 9: 259–275 (1987).
17. Hong YN, Schopfer P: Control by phytochrome of urate oxidase and allantoinase activities during peroxisome development in the cotyledons of mustard (*Sinapis alba L.*) seedlings. Planta 152: 325–335 (1981).

18. Huang AHC, Trelease RN, Moore TS Jr: Plant Peroxisomes. Academic Press, New York (1983).
19. Hunter JM, Schwartz BW, Sloan JS, Becker WM: Postgerminative regulation of genes encoding peroxisomal proteins in cucumber cotyledons. J Cell Biol 12C: 192 (1988).
20. Huynh TV, Young RA, Davis RW: Construction and screening cDNA libraries in λgt10 and λgt11. Glover DM (ed), DNA Cloning, Volume I, 49–78. IRL Press, Washington, DC (1985).
21. Jacobson A: Purification and fractionation of poly(A)$^+$ RNA. Meth Enzymol 152: 254–261 (1987).
22. Johnson LM, Snyder M, Chang LMS, Davis RW, Campbell JL: Isolation of the gene encoding yeast DNA polymerase I. Cell 43: 369–377 (1985).
23. Kagawa T, Beevers H: The development of microbodies (glyoxysomes and leaf peroxisomes) in cotyledons of germinating watermelon seedlings. Plant Physiol 55: 258–264 (1975).
24. Kagawa T, McGregor DI, Beevers H: Development of enzymes in the cotyledons of watermelon seedlings. Plant Physiol 51: 66–71 (1973).
25. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
26. Lipman D.J, Pearson WR: Rapid and sensitive protein similarity searches. Science 227: 1435–1441 (1985).
27. Lutcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA: Selection of AUG initiation codons differs in plants and animals. EMBO J 6: 43–48 (1987).
28. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
29. Pearson RW, Lipman DJ: Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85: 2444–2448 (1988).
30. Sanger F, Coulson AR, Barrell BG, Smith AJH, Roe BA: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J Mol Biol 143: 161–178 (1980).
31. Schnarrenberger C, Oeser A, Tolbert NE: Development of microbodies in sunflower cotyledons and castor bean endosperm during germination. Plant Physiol 48: 566–574 (1971).
32. Schnaitman CA: Cell fractionation. In: Gerhardt P (ed) Manual of Methods for General Bacteriology, 52–61. American Society for Microbiology, Washington, DC (1981).
33. Schopfer P, Bajracharya D, Bergfeld R, Falk H: Phytochrome-mediated transformation of glyoxysomes into peroxisomes in the cotyledons of mustard (*Sinapis alba L.*) seedlings. Planta 133: 73–80 (1976).
34. Small GM, Szabo LJ, Lazarow PB: Acyl-CoA oxidase contains two targeting sequences each of which can mediate protein import into peroxisomes. EMBO J 7: 1167–1173 (1988).
35. Smith SM, Leaver CJ: Glyoxysomal malate synthase of cucumber: molecular cloning of a cDNA and regulation of enzyme synthesis during germination. Plant Physiol 81: 762–767 (1986).
36. Stafford HA, Magaldi A, Vennesland B: The enzymatic reduction of hydroxypyruvic acid to D-glyceric acid in higher plants. J Biol Chem 207: 621–629 (1954).
37. Tchang F, Lecharny A, Mazliak P: Photostimulation of hydroxypyruvate reductase activity in peroxisomes of *Pharbitis nil* seedlings: I. Action spectrum. Plant and Cell Physiol 25: 1033–1037 (1984).
38. Tchang F, Lecharny A, Mazliak P: Photostimulation of hydroxypyruvate reductase activity in peroxisomes of *Pharbitis nil* seedlings: II. Photoreceptors in blue light. Plant and Cell Physiol 25: 1039–1043 (1984).
39. Titus DE, Becker WM: Investigation of the glyoxysome-peroxisome transition in germinating cucumber cotyledons using double-label immunoelectron microscopy. J Cell Biol 101: 1288–1299 (1985).
40. Titus DE, Hondred D, Becker WM: Purification and characterization of hydroxypyruvate reductase from cucumber cotyledons. Plant Physiol 72: 402–408 (1983).
41. Tobey KL, Grant GA: The nucleotide sequence of the *SerA* gene of *E. coli* and the amino acid sequence of the encoded protein, D-3-phosphoglycerate dehydrogenase. J. Biol Chem 261: 12179–12183 (1986).
42. Tokuhisa JG, Daniels SM, Quail PH: Phytochrome in green tissue: Spectral and immunochemical evidence for two distinct molecular species of phytochrome in light-grown *Avena sativa* L. Planta 164: 321–332 (1985).
43. Tolbert NE: Metabolic pathways in peroxisomes and glyoxysomes. Ann Rev Biochem 50: 133–157 (1981).
44. Tolbert NE, Yamazaki RK, Oeser A: Localization and properties of hydroxypyruvate and glyoxylate reductases in spinach leaf particles. J Biol Chem 245: 5129–5136 (1970).
45. Trelease RN, Becker WM, Gruber PJ, Newcomb EH: Microbodies (glyoxysomes and peroxisomes) in cucumber cotyledons. Plant Physiol 48: 461–475 (1971).
46. Volokita M, Sommerville CR: The primary structure of spinach glycolate oxidase deduced from the DNA sequence of a cDNA clone. J Biol Chem 262: 15825–15828 (1987).
47. Wierenga RK, Terpstra P, Hol WGJ: Prediction of the occurrence of the ADP-binding $\beta\alpha\beta$-fold in proteins, using an amino acid sequence fingerprint. J Mol Biol 187: 101–107 (1986).
48. Zelitch I: Oxidation and reduction of glycolic and glyoxylic acids in plants II. Glyoxylic acid reductase. J Biol Chem 201: 719-726 (1953).