Porphobilinogen Synthase from Pea: Expression from an Artificial Gene, Kinetic Characterization, and Novel Implications for Subunit Interactions†

Jukka Kervinen, Roland L. Dunbrack, Jr., Samuel Litwin, Jacob Martins, Robert C. Scarrow,‡ Marina Volin, Anthony T. Yeung, Erica Yoon, and Eileen K. Jaffe*

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111

Received March 17, 2000; Revised Manuscript Received May 16, 2000

ABSTRACT: Porphobilinogen synthase (PBGS) is present in all organisms that synthesize tetrapyrroles such as heme, chlorophyll, and vitamin B_{12}. The homooctameric metalloenzyme catalyzes the condensation of two 5-aminolevulinic acid molecules to form the tetrapyrrole precursor porphobilinogen. An artificial gene encoding PBGS of pea (*Pisum sativum* L.) was designed to overcome previous problems during bacterial expression caused by suboptimal codon usage and was constructed by recursive polymerase chain reaction from synthetic oligonucleotides. The recombinant 330 residue enzyme without a putative chloroplast transit peptide was expressed in *Escherichia coli* and purified in 100-mg quantities. The specific activity is protein concentration dependent, which indicates that a maximally active octamer can dissociate into less active smaller units. The enzyme is most active at slightly alkaline pH; it shows two pK_a values of 7.4 and 9.7. Atomic absorption spectroscopy shows maximal binding of three Mg(II) per subunit; kinetic data support two functionally distinct types of Mg(II) and the third appears to be nonphysiologic and inhibitory. Analysis of the protein concentration dependence of the specific activity suggests that the minimal functional unit is a tetramer. A model of octameric pea PBGS was built to predict the location of intermolecular disulfide linkages that were revealed by nonreducing sodium dodecyl sulfate−polyacrylamide gel electrophoresis. As verified by site-specific mutagenesis, disulfide linkages can form between four cysteines per octamer, each located five amino acids from the C-terminus. These data are consistent with the protein undergoing conformational changes and the idea that whole-body motion can occur between subunits.

Tetrapyrrole biosynthesis is an essential pathway in animals, plants, and microbes. The first common intermediate is 5-aminolevulinic acid (ALA).[1] Porphobilinogen synthase (PBGS, EC 4.2.1.24, aka 5-aminolevulinic acid dehydratase) catalyzes the formation of the tetrapyrrole precursor porphobilinogen from two molecules of ALA. The enzymatic reactions from ALA to protoporphyrin IX are common to tetrapyrrole biosynthesis in all organisms and this pathway leads to the biosynthesis of biologically important cofactors such as heme, chlorophyll, vitamin B_{12}, and cofactor F_{430} (*1*). PBGS are homooctameric metalloenzymes that utilize a variety of divalent and monovalent cations at catalytic and allosteric sites (*2*). Mammalian and yeast enzymes typically require Zn(II), some prokaryotic enzymes require either Mg(II) or Zn(II) or both for maximal activity, and plant enzymes seem to require only Mg(II) for enzymatic activity. The difference in the use of metal ions is reflected in a variation of residues in the primary structures in at least two metal-binding regions (*2*−*5*).

Mammalian and some microbial PBGS are well-characterized, whereas plant PBGS have been detected or purified from only a few sources in small quantities. These include enzymes from spinach (*6*), sunflower (*7*), soybean (*8*), and pea (*5*, *9*, *10*). A highly homologous PBGS has also been characterized from the green alga *Chlamydomonas reinhardtii* (*11*). At present, seven full-length or almost full-length gene sequences encoding plant PBGS are available (DDBJ/EMBL/GenBank Accession Nos.: soybean, U04525; pea, M71235; tomato, L31367; spinach, X57842; barley, X92402; Physcomitrella, X89886; and Selaginella, X75043); the primary structures typically consist of ∼425 amino acid residues. Due to a putative N-terminal chloroplast transit peptide of ∼50 residues, plant PBGS precursors are larger than the mammalian or microbial counterparts. Pea PBGS activity has been localized in plastids (*12*) where the dominant tetrapyrrole pigment in plants, chlorophyll, is also produced. Chlorophyll, protoheme, and phycobiline are the main end products of the plant tetrapyrrole pathway (for

† This work was supported by Grant ES03654 (E.K.J.) from the National Institute of Environmental Health Sciences, NIH, by NIH Grant CA06927 (Institute for Cancer Research), and by an appropriation from the Commonwealth of Pennsylvania. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute nor of the National Institute of Environmental Health Sciences.

* To whom correspondence should be addressed: Telephone 215-728-3695; fax 215-728-2412; e-mail EK_Jaffe@fccc.edu.

‡ Permanent address: Department of Chemistry, Haverford College, Haverford, PA 19041.

[1] Abbreviations: ALA, 5-aminolevulinic acid; βME, 2-mercaptoethanol; Bis-tris propane, 1,3-bis[tris(hydroxymethyl)methylamino]-propane; EDTA, ethylenediaminetetraacetic acid; IPTG, isopropyl β-D-thiogalactopyranoside; KP_i, potassium phosphate-KOH.; MALDI-TOF-MS, matrix-assisted laser desorption ionization time-of-flight mass spectroscopy; PBGS, porphobilinogen synthase; PCR, polymerase chain reaction; PMSF, phenylmethanesulfonyl fluoride; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi000620c CCC: $19.00
Published on Web 07/06/2000

recent reviews, see refs *13* and *14*). Although plant PBGS needs Mg(II) for the enzyme action, it is not known if Mg(II) is directly involved in catalysis. By analogy to the Mg(II) seen bound to *Pseudomonas aeruginosa* PBGS in the crystal structure (*15*) and kinetically characterized for *E. coli* PBGS (*16*) and *P. aeruginosa* PBGS (*17*), one role of Mg(II) in plant PBGS is presumably allosteric.

PBGS are present only in low amounts in plant tissues and thus a large-scale purification from original sources has not been practically feasible. Moreover, although PBGS has been cloned and sequenced from pea, its expression in *E. coli* yielded only a few milligrams of a pure protein, which limited the characterization of the enzyme (*9*). We recently developed a method to express human PBGS in *E. coli* in 100-mg quantities using an artificial gene that was constructed by recursive PCR from synthetic oligonucleotides that were designed for optimal codon usage (*18*). Here, we describe the design of an artificial gene for pea PBGS as well as a large-scale expression and purification of the recombinant enzyme. This enabled us to analyze the enzyme kinetics and the Mg(II) binding properties of the enzyme in detail. We also present evidence that the specific activity is protein concentration dependent and that a tetrameric structure is the minimal oligomer needed to form a catalytically active enzyme.

EXPERIMENTAL PROCEDURES

Materials. Oligonucleotides were synthesized in-house in the Fannie Rippel Biotechnology Center. Cloned *Pfu* DNA polymerase was obtained from Stratagene (La Jolla, CA) and used at 2.5 units per 50 μL of PCR reaction mixture. The plasmid pET3a as well as BLR(DE3) competent cells were obtained from Novagen (Madison, WI). The restriction enzymes *Bam*HI and *Nde*I were from New England Biolabs (Beverly, MA). Plasmids were purified with Qiagen (Valencia, CA) plasmid purification kits, and PCR products were extracted from 1−2% agarose with Qiaex II kits from Qiagen. DNA sequencing was carried out in-house by use of ABI sequencing technology. Site-specific mutagenesis to replace Cys-326 by alanine was carried out by the QuikChange method (Stratagene, La Jolla, CA). The sequence of the sense strand of the mutagenic primer was 5′-GCGCTGCAAGCTGCTCGCACTTTGGCAGGAGAGAAGCG-3′. Plasmids encoding mutant proteins were sequenced in both directions throughout the gene. Succinyl acetone (4,6-dioxoheptanoic acid) was from Sigma, St. Louis, MO.

Design and Synthesis of the Artificial Gene Encoding Pea PBGS. The design and synthesis of the artificial gene encoding pea PBGS followed our published rationale for an artificial gene encoding human PBGS (*18*). Approximately 10% of the codons used by the naturally occurring pea gene encoding PBGS, known as *ALAD*, are those for which *E. coli* has limited amounts of tRNA. Figure 1A highlights the amino acid locations corresponding to these codons and shows that many of them fall in clusters, which can hinder the efficiency and fidelity of heterologous expression (*19*). All of these codons were altered to those commonly used by *E. coli*. In addition, for all the amino acid identities between pea and *E. coli* PBGS where pea and *E. coli* used different codons, the codons for the artificial pea gene were changed to those used by *E. coli*. In all, 40% of the codons used by the naturally occurring pea *ALAD* gene were changed in the design of the artificial gene.

The artificial gene encoding pea PBGS was originally designed to encode a 358 amino acid protein. This places methionine in place of Ala41 of the pea *ALAD* gene (see Figure 1A). Based on an initial intent to mimic a constitutive expression system we had used for overexpressing *E. coli* PBGS in *E. coli*, the target final PCR product containing the artificial gene was 1229 nucleotides, about 12% longer than what we had used for synthesis of an artificial gene encoding human PBGS (*18*). Because the target PCR product was longer, a slightly different approach was used in the design of synthetic oligonucleotides as illustrated schematically in Figure 1B. The significant design difference was that the primers all extended beyond the 5′ end of the templates, thus providing additional sequence information. In addition, the eight initial 158−173 nucleotide-long synthetic templates were purified on DNA sequencing gels; in each case only the largest fragment was used for the PCR reactions. The PCR protocols, ligations, and plasmid propagation were as previously described (*18*). Following successful completion of the PCR reactions to yield the 1229 bp product, the 3′ and 5′ ends were changed by PCR to give a gene encoding a slightly larger 364 amino acid pea PBGS that starts at position 35, similarly to Cheung et al. (*9*), and contains an N-terminal *Nde*I site and a C-terminal *Bam*HI site for insertion into pET3a. This plasmid, pMVpea, which did not express well, was then altered by QuikChange mutagenesis to redesign the N-terminus to give a protein product that was identical in length to the mammalian and *E. coli* PBGS proteins (330 amino acids). The start site for the short pea PBGS gene corresponds to Ser69 of Figure 1A. The CCTTCA corresponding to Pro68-Ser69 was mutated to CATATG to form a 5′ *Nde*I site. This allowed digestion of the PCR product and ligation into the pET3a plasmid to prepare the expression plasmid (pMVpeashort).

Bacterial Expression and Purification of Pea PBGS. E. coli BLR(DE3) cells were freshly transformed with pMVpeashort and each liter of medium was inoculated with a single colony ∼24 h after the transformation. The growth was typically carried out in 6 × 1 L batches of Luria broth containing 100 μg mL^{-1} ampicillin, 12.5 μg mL^{-1} tetracycline, and 0.4% glucose at 37 °C in air shakers. After about 16 h (A_{600} typically 4−5), cells were harvested and resuspended in 6 × 1 L of Luria broth containing ampicillin, tetracycline, and 1 mM $MgCl_2$ at 15 °C. After preincubation at 15 °C for 1 h, 0.1 mM IPTG was added to induce expression. Following 48 h of expression at 15 °C, the culture typically yielded about 40 g of cells, which were stored at −80 °C.

Cells were thawed in a solution [2 mL (g cells)$^{-1}$] containing 50 mM potassium phosphate−KOH (KP_i), pH 8.0, 170 mM KCl, 5 mM EDTA, 10 mM 2-mercaptoethanol (βME), and 0.1 mM phenylmethanesulfonyl fluoride (PMSF). After addition of lysozyme (0.4 mg mL^{-1}), the suspension was stirred at room temperature for 20 min, after which time an equal volume of a solution (0.1 M KP_i, pH 7.0, 12 mM $MgCl_2$, 40 μM $ZnCl_2$, 10 mM βME, and 0.1 mM PMSF) with DNase I [∼65 units (g of cells)$^{-1}$] was added. The suspension was stirred for an additional 20 min. To enhance the release of the expressed protein, the cell suspension was frozen in liquid nitrogen and thawed at 37 °C. Freeze/thaw

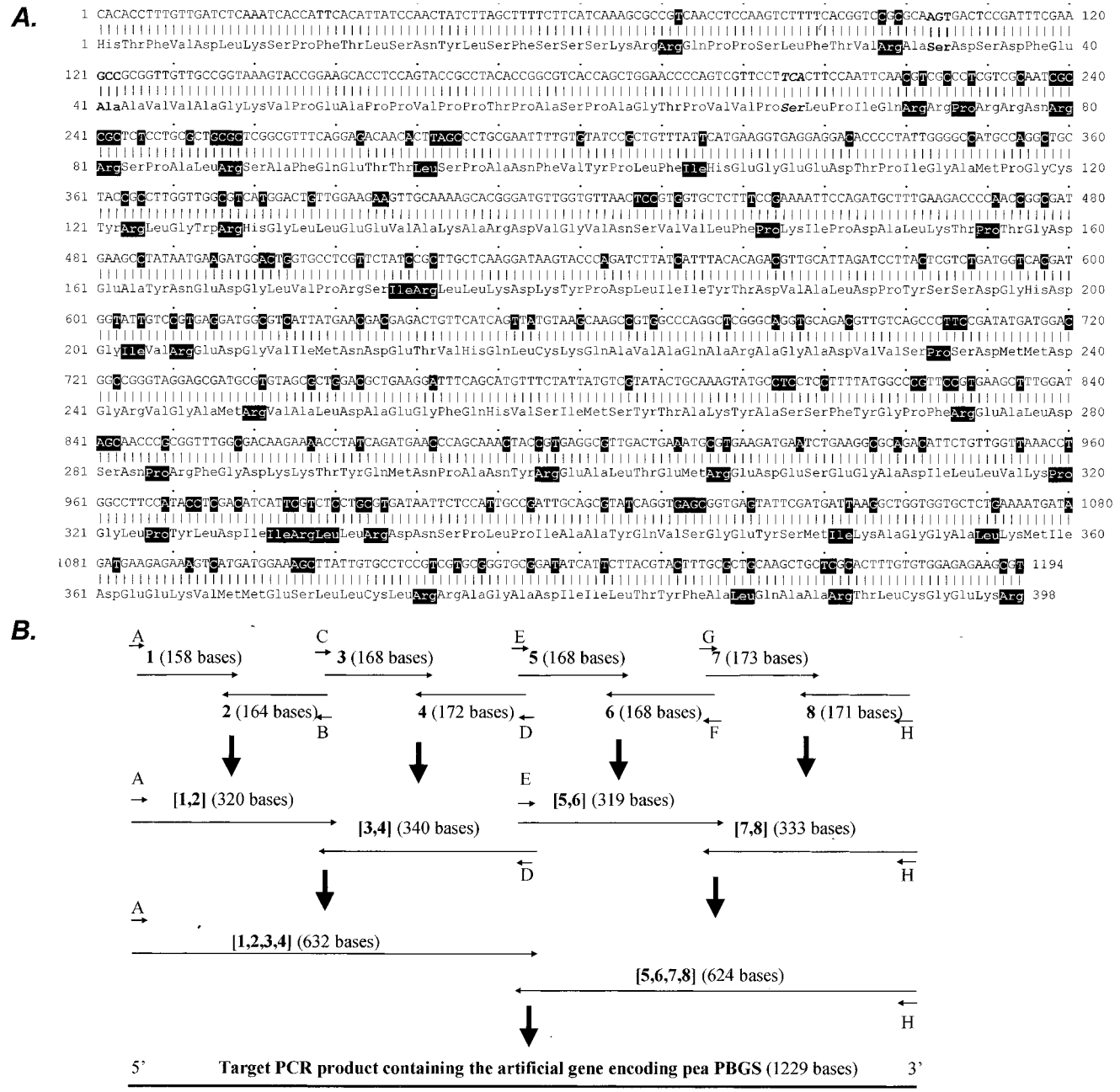

FIGURE 1: Design and synthesis of the artificial gene encoding pea PBGS. (A) The protein translation for the pea *ALAD* gene encoding pea PBGS (*5*) is aligned with the artificial gene sequence designed for optimal high-level low-error heterologous expression in *E. coli*. The amino acids shown in reverse type correspond to the locations of the six least favorable codons (for expression in *E. coli*) that are present in the pea *ALAD* gene. The nucleotides presented in reverse type correspond to all the locations where the artificial gene differs from the original pea *ALAD* gene. In some cases, these alterations removed these worst codons. In other cases, the changes simply altered the codons to those present in the *E. coli hem*B gene, which encodes *E. coli* PBGS. The rationale for making these types of changes is described in the text and in Jaffe et al. (*18*). The final expressed protein is 330 amino acids and starts at a methionine located at the site of Ser69 in this figure (shown in boldface italic type). (B) Recursive PCR strategy used to prepare a 1.2 kb PCR product containing the artificial gene for pea PBGS. The 3′ ends of the synthetic oligonucleotides are denoted by horizontal arrows. Oligonucleotide templates 1−8 are labeled with their length in parentheses. Oligonucleotide primers A−H were each 24−30 nucleotides long and complemented their respective templates over ~18 5′ terminal nucleotides, thus effectively adding ~6−7 nucleotides on the 5′ end of the PCR products [1,2], [3,4], [5,6], and [7,8]. Each downward arrow represents one PCR reaction with the two templates and two primers above the arrow. For example, the uppermost left downward arrow represents a PCR reaction with primers A and B and templates 1 and 2 to yield the 320 nucleotide product [1,2]. All PCR products were gel-purified prior to further use.

treatment was repeated twice and the resulting suspension was sonicated for 10 min.

All the following purification steps were carried out at 4 °C. First, the cell suspension was subjected to a 20−45% saturated ammonium sulfate fractionation. The pellet from the 45% saturated ammonium sulfate fractionation was dissolved in ~30 mL of DEAE Bio-gel column equilibrium buffer (30 mM KP_i, pH 7.5, 1 mM $MgCl_2$, 10 mM βME, and 0.1 mM PMSF), and dialyzed against 4 L of the same buffer for 16 h. The dialyzed protein solution (38 mL) was clarified by centrifugation (20000 g for 30 min) and loaded onto a DEAE Bio-Gel column (length 35 cm, 5 cm i.d.). The column was washed with about 300 mL of an equilibrium buffer and was eluted by a 2 L linear gradient of 0−0.4

M KCl in the equilibrium buffer. Fractions of high constant specific activity ($\geq$148 μmol h^{-1} mg^{-1}) were pooled and concentrated to about 10 mg mL^{-1} by ultrafiltration (Amicon PM30 membrane, Millipore Corp., Bedford, MA). The concentrated protein was loaded (2 mL min^{-1}) onto a Sephacryl-300 gel-filtration column (length 90 cm, 5 cm i.d.) that was previously equilibrated with 0.1 M Tris-HCl, pH 8.5, 10 mM $MgCl_2$, and 10 mM βME. The fractions with high specific activity were combined, concentrated by ultrafiltration (Amicon, PM30 membrane) to about 14 mg mL^{-1}, flash-frozen in liquid nitrogen, and stored at -80 °C.

Activity Assays for Pea PBGS. Purified PBGS was preincubated [typically at 31 μg (mL of assay solution)$^{-1}$] in 0.1 M Bis-tris propane hydrochloride, pH 8.5, 10 mM βME, and 10 mM $MgCl_2$ for 10 min at 37 °C prior to the addition of substrate ALA-HCl to a final concentration of 10 mM. The reaction was allowed to proceed for 5 min prior to termination with a half volume of stop reagent (50% trichloroacetic acid and 0.1 M $HgCl_2$). Precipitated protein and βME were removed by centrifugation. The quenched assay solutions were diluted 1:25 with a 2:1 mixture of assay buffer and stop reagent and then 0.5 mL of diluted solutions were mixed with 0.5 mL of modified Ehrlich's reagent. Color development was allowed to proceed for about 8 min and the porphobilinogen formed was determined by absorbance at 555 nm. The extinction coefficient of the pink complex formed (ϵ_{555}) is 62 000 M^{-1} cm^{-1}. Specific activity is defined as micromoles of porphobilinogen formed per hour per milligram of protein. For the pH rate profile, pH values reflect the pH after the addition of ALA-HCl to 10 mM. In the determination of Mg(II) dependence on the activity at concentrations $\leq$0.1 mM Mg(II), the samples were first dialyzed against 0.1 M Bis-tris propane hydrochloride, pH 6.3, and 10 mM βME for at least 6 h to strip the existing Mg(II) from the enzyme. In the determination of the protein concentration dependence of the specific activity at very low protein concentration, the reaction time was extended up to 40 min. These reaction solutions were not diluted before the color reaction.

Inhibition by Succinyl Acetone. Activity of pea PBGS (0.8 μM subunit) was tested over a 0.001$-$1 mM range of succinyl acetone concentration. Three different preincubation times at 37 °C were tested before the specific activity was established by routine assay.

Dry Weight Analysis To Precisely Determine Pea PBGS Concentrations. Four labeled crucibles were placed in a drying oven at 55 °C. The mass of the crucibles was determined every day until two consecutive readings were identical. The OD_{280} of a PBGS solution ($\sim$10 mg/mL, 10-fold dilution before A_{280} determination), which had been dialyzed against 20 mM Bis-tris propane, pH 8.5, 10 mM $MgCl_2$, and 10 mM βME, was determined vs dialysis buffer. An apparent protein concentration was also determined relative to bovine serum albumin by use of the Pierce Coomassie reagent. One milliliter of $\sim$10 mg/mL protein solution was placed in two crucibles and 1 mL of dialysis buffer was placed in the others. The crucibles were held in the drying oven until the mass no longer changed. The weight of the protein was determined by subtraction.

Analysis of Mg(II) Binding by Atomic Absorption Spectroscopy. Mg(II) binding studies were carried out at room temperature. The purified PBGS ($\sim$12 mg mL^{-1}) was first dialyzed against 0.1 M Bis-tris propane hydrochloride, pH 6.3, and 10 mM βME for 6 h to strip the existing Mg(II). This was followed by dialysis against 0.1 M Bis-tris propane, pH 8.5, and 10 mM βME for 16 h. Dialyzed enzyme was concentrated to $\sim$31 mg/mL by ultrafiltration (Centricon-10, Amicon, Beverly, MA), and 0$-$14 mM $MgCl_2$ was added to concentrated samples with a total volume of 60 μL for each sample. After incubation for about 30 min, free and bound Mg(II) were separated by ultrafiltration in Microcon-10 concentrators. Both the effluent and the enzyme solutions were then diluted 75$-$250-fold with water to adjust Mg(II) concentration to the 10$-$60 μM range before determination of Mg(II) concentration with a Perkin-Elmer Analyst-100 flame atomic absorption spectrometer. The final dilutions contained protein at 3$-$11 μM subunits. This experiment was also carried out in the presence of 5 mM porphobilinogen, originally added to the enzyme sample as 10 mM ALA.

Model Building for Pea PBGS. A model for the structure of pea PBGS was constructed by the methods described in ref *20*. PSI-BLAST was used to construct a profile of the family of PBGS sequences by searching the nonredundant protein sequence database available from National Center for Biotechnology Information, NIH, Bethesda, MD, over four iterations. The resulting profile was used to search a sequence database of proteins in the PDB with PSI-BLAST. This produced alignments to both *P. aeruginosa* and yeast PBGS with 46% and 36% identity, respectively. The *P. aeruginosa* alignment had fewer gaps and the corresponding PDB entry, 1b4k (*15*), was used as the basis for modeling of pea PBGS. The alignment of pea PBGS with the *P. aeruginosa* sequence from PSI-BLAST was adjusted manually to shift some gaps into loop regions, which is where such sequence changes are most likely to occur. The backbone coordinates from 1b4k were renumbered and relabeled with the pea sequence according to the alignment, and the Cartesian coordinates for conserved residues were retained and numbered according to the pea sequence. We used the side-chain conformation prediction program SCWRL (*21*) to place all other side chains onto the pea structure model. SCWRL uses a backbone-dependent rotamer library (*22*) to place side chains on the backbone, followed by a combinatorial search to remove steric clashes that arise from the initial placement. Conserved side chains were kept fixed in their original conformations. SCWRL has proved more reliable than other methods in predicting side-chain conformations in blind trials (*20*).

Data Analysis. Most data were fitted to standard or cited equations by the program SigmaPlot (SPSS Inc., Chicago, IL). The protein dissociation data were fitted to the following equations. These are simplified versions of the equations originally used to fit the protein dissociation data for *B. japonicum* PBGS (*23*). The dynamic equilibrium between dimers, tetramers, and octamers is defined by the following reaction where S_2 is the dimer, S_4 is the tetramer, and S_8 is the octamer:

$$S_8 \overset{K_{d3}}{\rightleftharpoons} 2S_4 \qquad S_4 \overset{K_{d2}}{\rightleftharpoons} 2S_2$$

Current models dismiss the existence of a monomer. The total concentration of subunits, C, is defined as the sum of all subunits in all species as follows, where N_n is the

Table 1: Purification of Recombinant Pea PBGS

step	volume (mL)	protein		activity		yield (%)
		mg/mL	total (mg)	μmol h^{-1} mg^{-1}	total (μmol h^{-1})	
lysed cells	216	20.4	4406	na[a]	na	–
redissolved 45% $(NH_4)_2SO_4$ pellet	38	31.5	1197	93	111,321	100
pooled DEAE Bio-Gel peak	130	4.1	533	164	87,412	79
concentrated Sephacryl S-300 peak	25	11.7	293	213	62,409	56

[a] Not assayed.

concentration of *n*-multimers present:

$$C = 2N_2 + 4N_4 + 8N_8$$

Expressing C as a function of N_8 and the K_d values yields

$$C = 2(K_{d2}(K_{d3}N_8)^{1/2})^{1/2} + 4(K_{d3}N_8)^{1/2} + 8N_8$$

Newton's method is used to calculate N_8 as a function of C (*24*). The specific activity was calculated according to two models where α is the theoretical V_{max} for octamer:

model 1 specific activity $= \alpha[2N_4 + 8N_8]/C$

model 2 specific activity $= \alpha[4N_4 + 8N_8]/C$

These equations allowed us to determine specific activities for each concentration C at any set of constants K_{d2}, K_{d3}, and α. The space of these three constants was searched to fit the calculated specific activities to the observed values by the least-squares method.

Other Methods. Protein amounts were measured with Coomassie Plus protein assay reagent (Pierce, Rockford, IL) relative to the standard curve prepared with bovine serum albumin. Dry weight analyses indicated that the resultant protein concentration should be multiplied by 0.83 to give the correct PBGS value. SDS−PAGE was carried out on the PhastSystem (Amersham Pharmacia Biotech) and proteins were stained with Coomassie Brilliant Blue R-250. Matrix-assisted laser desorption ionization time-of-flight mass spectroscopy (MALDI-TOF-MS) was performed on a PerSeptive Biosystems Voyager DE mass spectrometer in linear mode at an accelerating voltage of 20 000 V. N-Terminal sequencing was carried out on an Applied Biosystems 477A gas-phase sequencer from a protein blotted onto a poly(vinylidene difluoride) (PVDF) membrane, and the phenylthiohydantoin amino acids were identified on-line with a model 120 analyzer (Applied Biosystems, Inc., Foster City, CA).

RESULTS

Protein Expression and Purification. Production of human (*3*, *25*) and plant (*9*) PBGS in bacterial expression systems has been difficult, in part due to suboptimal codon usage. This is one general reason eukaryotic proteins may express poorly in bacterial expression systems. In the case of human PBGS, poor expression has recently been overcome by altering the triplet codons in the sequence so that the codons would better accommodate the expression machinery present in *E. coli* (*18*). In the present study, a similar protocol was followed. An artificial gene encoding pea PBGS was constructed by recursive PCR from synthetic oligonucleotides so that the triplet codons resembled as much as possible the highly expressed homologous *E. coli hem*B gene. The rare codons (CTA, ATA, AGG, CCC, AGA, and AGT) (*19*), numbering 31 in the original pea *ALAD* gene, were also replaced by codons that can be easily recognized by *E. coli* (see Figure 1). A construct designed to start at S35M did not express well. An *Nde*I site was inserted at a position that eliminated the N-terminal 68 amino acids such that the resulting peaPBGS(short) is 330 amino acids, identical in length to *E. coli* and human PBGS. This resultant plasmid is pMVpeashort.

Pea PBGS(short), below simply stated as pea PBGS, was expressed from the artificial gene from pMVpeashort in *E. coli* strain BLR(DE3). To optimize cell growth, fidelity of expression, and optimal folding, (1) glucose was included in the first growth to repress expression of the genes under the control of the *lac* promoters, (2) the cells were transferred to fresh medium with no glucose prior to induction, (3) IPTG induction was added at 0.1 mM concentration, and (4) expression was at 15 °C for 2 days. The primary purification steps for pea PBGS were ammonium sulfate fractionation, ion-exchange chromatography on a DEAE Bio-Gel column, and gel filtration on Sephacryl S300. Purification from 6 L batches typically yielded around 300 mg of PBGS protein (Table 1) with the purity being >95% as determined by SDS−PAGE (Figure 2A). SDS−PAGE carried out in the absence of βME showed a significant amount of covalent dimer formation, which was explored experimentally, with modeling, and with mutagenesis (see below). Mass spectral analysis for the sample in 10 mM βME gave a molecular mass of 36 805 ±10 Da (Figure 2B), which is in good agreement with the predicted mass (36 725 Da) of the 330 amino acid protein. N-Terminal sequencing revealed unambiguously a single sequence, Met-Leu-Pro-Ile-Gln-Arg-Arg, which indicated that *E. coli* cells effectively removed the formyl group but did not remove the N-terminal Met from the expressed protein. Dry weight analysis showed the $A_{280}^{0.1\%}$ to equal 0.80 AU and determined that the protein concentration, as determined by Pierce Coomassie assay relative to bovine serum albumin, had to be multiplied by 0.83 to get an accurate protein concentration for the 330 amino acid form of pea PBGS.

Specific Activity Is Protein Concentration Dependent. All the known PBGS proteins exist as oligomers, and they are probably all homooctamers (*2*). Thus, it has been hypothesized that some kind of communication between individual monomers is needed for the completion of enzymatic catalysis. For example, a protein concentration dependence for the specific activity has been seen for *B. japonicum* (*23*) and *P. aeruginosa* PBGS (*17*). However, protein concentration dependence is not a universal characteristic of PBGS. It has not been documented for PBGS from *E. coli*, yeast, or from mammalian sources. We tested here if the protein

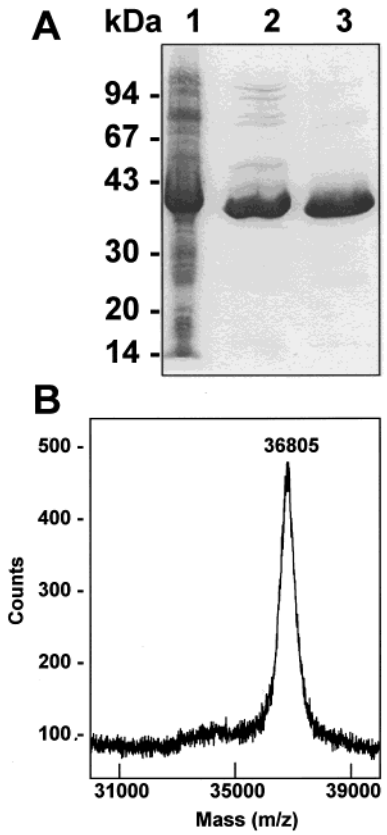

FIGURE 2: Protein expression from the pMVpeashort plasmid in BLR(DE3) cells and the purified protein product. (A) Sample from *E. coli* strain BLR(DE3)pMVpeashort cell lysate after 2 days of expression of the artificial gene at 15 °C (lane 1); purified pea PBGS (6 μg) from the DEAE column (lane 2); and from the Sephacryl S300 column (lane 3). Samples were analyzed in an SDS−12.5% polyacrylamide gel and stained with Coomassie Blue. (B) MALDI-TOF-MS analysis of the Sephacryl S300-purified pea PBGS.

concentration of pea PBGS affects its specific activity. The routine enzyme assay was used, which measures the formation of porphobilinogen from the asymmetric condensation of two molecules of ALA. The results clearly show that the specific activity is protein concentration dependent and about 0.8 μM PBGS subunit [31 μg of protein (mL of assay solution)$^{-1}$] is needed to obtain the maximal activity (Figure 3A). Thus, at lower enzyme concentration, the optimally active octameric form dissociates to less active or inactive units. In the case of *B. japonicum* PBGS, similar data were fitted to a model where a monomer was considered inactive, a dimer was the minimal functional unit, and the activity of tetramers and octamers were considered simply proportional to the numbers of dimers they contained. This model must be updated to conform to the more recently determined crystal structure of PBGS (see Discussion). All further kinetic characterization was carried out at 31 μg mL^{-1} PBGS.

Activity Optimization for pH and Mg(II). The effect of pH and the dependence of the enzymatic activity on Mg(II) were tested to optimize conditions for activity assays. Previous studies have shown that the optimal pH for plant PBGS activity is typically in the range of 8−9 (*9*, *10*). Purified pea PBGS showed a similar pH rate profile at 10 mM ALA and 10 mM Mg(II). The enzyme was inactive below pH 6.7, and the specific activity increases sharply above pH 7 and remains constant at about a maximal activity of ∼210 μmol h^{-1} mg^{-1} from pH 7.7 to 9.2 (Figure 3B). The apparent pK_a values are 7.4 and 9.7, and both pK_a transitions are sharper than expected for a simple one-proton transition and may reflect an element of cooperativity.

Mg(II) was an essential cofactor for the enzymatic activity (Figure 3C). The Mg(II) activation profile was fitted to a model developed by Petrovich et al. (*23*), where there is a required Mg(II) and activating Mg(II). The equation describing the Mg(II) dependence is

$$\text{specific activity} = \frac{V_o[\text{Mg}]}{K_d\text{req} + [\text{Mg}]} + \frac{\{(V_o)(\text{activation factor}) - V_o\}[\text{Mg}]}{K_d\text{act} + [\text{Mg}]}$$

In this case, the maximal velocity associated with the required Mg(II), V_o, is 97 μmol h^{-1} mg^{-1} and the $K_d$req for this Mg-(II) is 35 μM. The allosteric Mg(II) appears to bind with a $K_d$act of 2.5 mM and the activation factor is a 2.3-fold activation. The solid line describes the nonlinear best fit to this equation for the indicated points. Figure 3C also shows the effect of an additional inhibitory Mg(II) that becomes dominant at >30 mM Mg(II).

Mg(II) Binding Studies. Equilibrium dialysis and atomic absorption spectroscopy was used to determine Mg(II) binding stoichiometries for the pea PBGS. The binding curve, presented in Figure 3D for millimolar concentrations of Mg(II) and enzyme subunits fits most cleanly to a single K_d model (2.6 mM) at a stoichiometry of 3 Mg(II)/subunit. Models with a tight catalytic Mg(II) binding site did not fit the data. The presence of product porphobilinogen in the Mg(II) binding buffer had no effect on the Mg binding affinity or stoichiometry under these high concentration conditions (data not shown). This result is inconsistent with the kinetic data presented in Figure 3C, because it does not support a tight Mg(II) site. It may be significant that the data in Figure 3C was acquired at lower protein concentration in the presence of excess substrate. One possible explanation is that a tight catalytic site only exists in the presence of the substrate ALA but not the product porphobilinogen. As precedent, the substrate has been shown to cause the disproportionation of substoichiometric Mn(II) from the allosteric site of *B. japonicum* PBGS to the catalytic site (*26*).

Modeling of the Pea PBGS Structure. The crystal structure of one Mg(II)-dependent PBGS, from the human pathogen *Pseudomonas aeruginosa*, has been solved at 1.67 Å resolution (*15*). The primary structure of *P. aeruginosa* PBGS is 46% identical to that of pea PBGS over the 330 amino acid core (Figure 4A) and thus it is assumed that these two proteins share a similar fold. A molecular model for pea PBGS was built from a profile-based sequence alignment from *P. aeruginosa* PBGS (ID code 1b4k) according to the alignment shown in Figure 4A. The deletions in the alignment result in two chain breaks at positions 40−41 and 212−213 that are artifacts of the modeling process. The insertion at D293 causes this residue to be omitted from the model. The N-terminal two amino acids and the C-terminal three amino acids are also missing in the model because they are disordered in 1b4k.

In the predicted pea PBGS structure, each monomer forms a $(\alpha\beta)_8$-barrel fold typical for all PBGS that have been

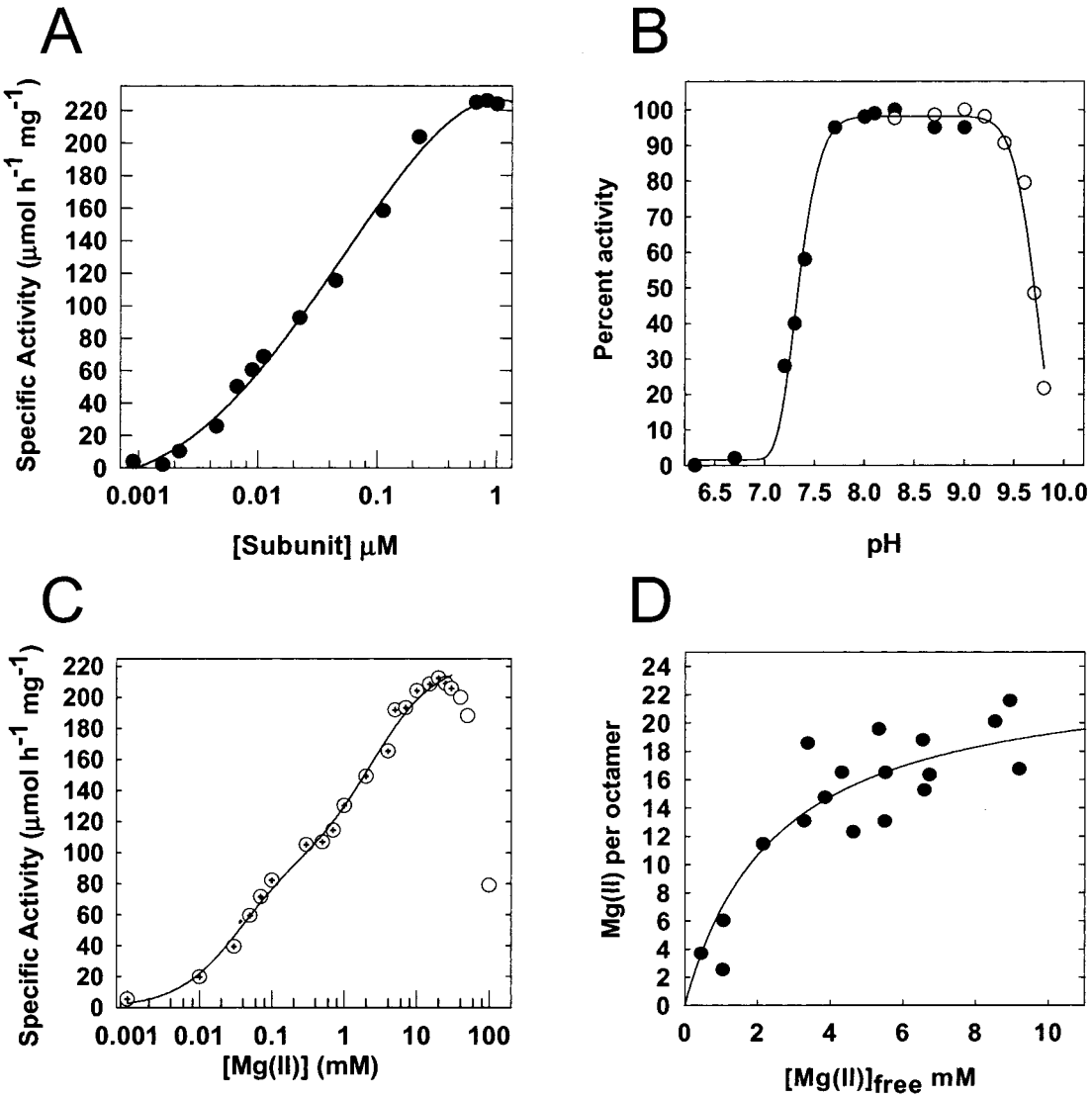

FIGURE 3: (A) Protein concentration dependent specific activity of pea PBGS. Specific activity is presented as a function of subunit concentration in 0.1 M Bis-tris propane hydrochloride, pH 8.5, 10 mM $MgCl_2$, 10 mM βME, and 10 mM ALA and was determined by a fixed-time assay. (B) Activity profile for pea PBGS in (●) Bis-tris propane hydrochloride or (○) Ches-NaOH as a function of pH. (C) Effect of Mg(II) concentration on pea PBGS activity. Specific activity was assayed at various Mg(II) concentrations in a solution containing 0.1 M Bis-tris propane hydrochloride, pH 8.5, 10 mM βME, and 10 mM ALA. The enzyme concentration was 31 μg mL^{-1}. The line is a nonlinear best fit to a model where one kind of Mg(II) is required and a second kind of Mg(II) is allosteric (see text). Data included in the fit are indicated by ⊕. The fit does not consider the inhibitory Mg(II) that appears to bind above 30 mM Mg(II). (D) Mg(II) binding to pea PBGS in 0.1 M Bis-tris propane hydrochloride, pH 8.5, and 10 mM βME. The line represents the maximal binding of 24 Mg(II)/octamer (3/subunit) with a single K_d of 2.5 mM.

crystallized (*15*, *27*, *28*) and contains an extended N-terminal arm that forms several noncovalent interactions with the neighboring subunits. The crystal structure of the *P. aeruginosa* PBGS shows an asymmetric dimer as the fundamental structural unit and this feature was included in the pea PBGS dimer model (Figure 4B). The asymmetry is related to the general thermal factors of each monomer, the order/disorder of an active-site lid, and the presence of Mg(II) on only four of the eight subunits of *P. aeruginosa* PBGS (*15*, *17*). Published yeast and *E. coli* PBGS structures do not show this asymmetry and we have no independent data that could assign the pea PBGS structure as symmetric or asymmetric. However, we proposed that 1b4k is the best template for modeling pea PBGS since the 46% sequence identity between the 330-residue core region of pea PBGS and that of *P. aeruginosa* exceeds that of pea PBGS relative to yeast (36%). The sequence identity between pea and *E. coli* PBGS is 49%. However, *E. coli* PBGS is a Zn(II)-requiring enzyme. The packing of dimers in the octameric structure is presented in Figure 4C. The most distinctive structural difference in the pea PBGS model based on 1b4k compared to PBGS structures from yeast (1aw5) and *E. coli* (1b4n) PBGS is somewhat looser packing of dimers in the octamer and thus fewer contacts between monomers in neighboring dimers. This feature may be the discriminating factor between those PBGS that exhibit protein concentration dependence of specific activity and those that do not.

The model was searched for possible inter- and intrasubunit disulfide bonds that might account for the dimers that were detected by SDS−PAGE under nonreducing conditions. In the pea PBGS structure, each monomer contains four Cys residues, illustrated in Figure 4B, that are >18 Å apart from each other within one monomer; this argues against the formation of intramolecular disulfide

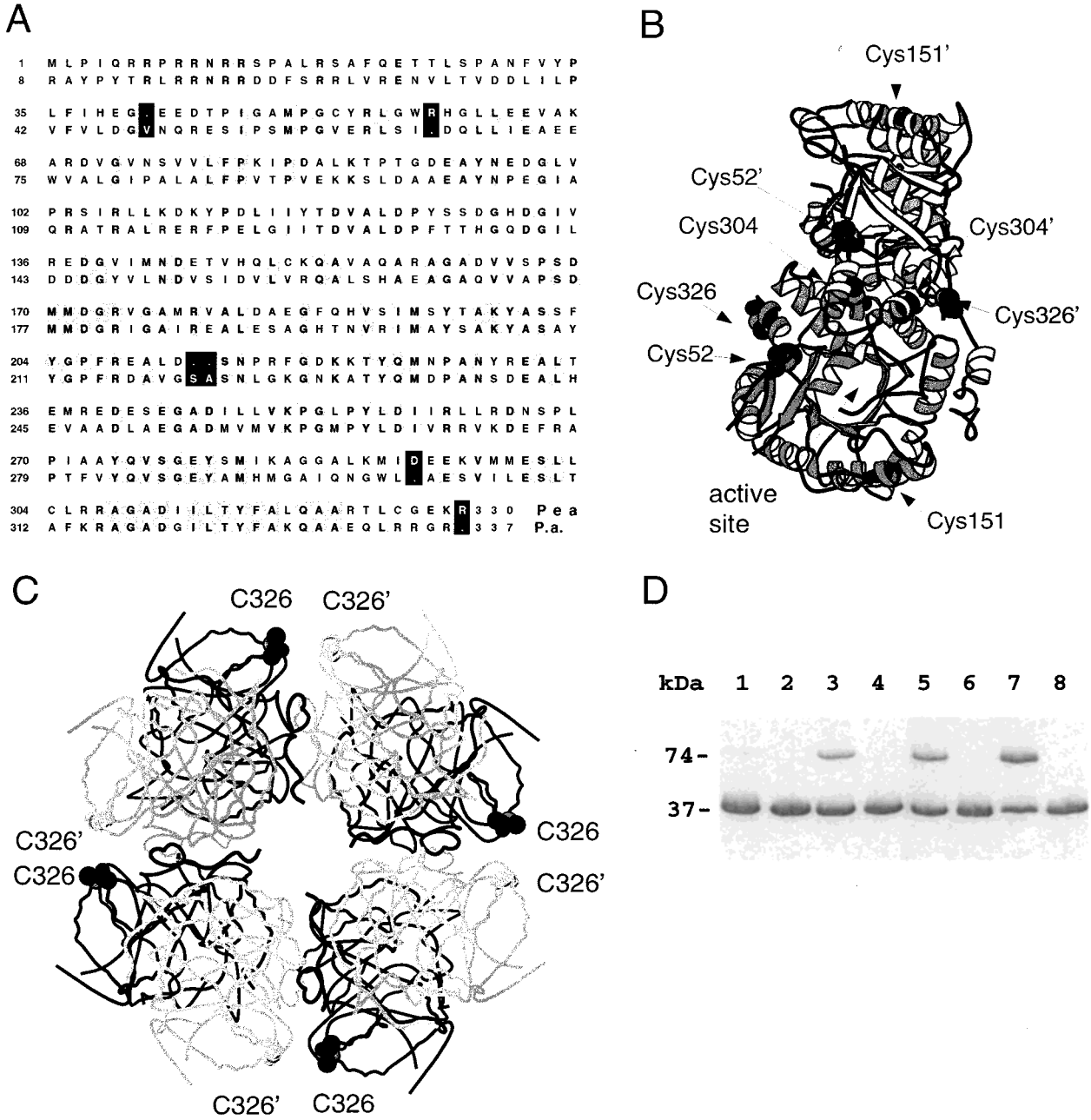

FIGURE 4: A model for pea PBGS based on the crystal structure of *Pseudomonas aeruginosa* PBGS was built according to the amino acid alignment shown in panel A. Identical residues are shaded. Black background highlights insertions and deletions that were not modeled. (B) A tightly associated dimer is depicted as a ribbon diagram with all cysteines space-filled in black. This homodimer is the asymmetric unit in the *P. aeruginosa* PBGS crystal structure. The location of the active site is marked in the lower monomer. (C) A Cα trace of the monomers in the octamer is shown at a 90° vertical rotation relative to panel B. This view shows that the sulfur of Cys326 of each dimer is within 11.3 Å of its counterpart Cys326′ on another dimer. (D) Appearance of a covalently cross-linked form of pea PBGS. Native PBGS (lanes 1, 3, 5, and 7) and a mutant C326A (lanes 2, 4, 6, and 8), both 5 mg mL^{-1} (total 2 mL), were dialyzed at room temperature against 1 L of 0.1 M Tris-HCl, pH 8.5, and 10 mM $MgCl_2$ without βME. The samples were collected after 0 h (lanes 1 and 2), 2 h (lanes 3 and 4), 4 h (lanes 5 and 6), and 24 h (lanes 7 and 8). Sample proteins treated with SDS in the absence of βME were analyzed in an SDS−12.5% polyacrylamide gel and stained with Coomassie Blue.

linkages. There are also no Cys locations that appear likely to form intermolecular disulfide linkages between the monomers of the asymmetric dimer (Figure 4B). However, if one considers the octamer, it appears that a conformational change in the C-terminal region and/or a minor whole-body motion of the dimers toward each other might allow Cys-326 to form intermolecular disulfide linkages between asymmetric dimers. In principle, each octamer can maximally form four such disulfides, thus covalently linking each subunit to a neighboring subunit of a different asymmetric dimer. Site-specific mutagenesis was used to test this hypothesis.

Removal of Cys-326 by Site-Specific Mutagenesis. Cys-326 was replaced with alanine to test whether the C326A mutation would disallow the formation of intermolecular disulfide linkages. The C326A mutant of pea PBGS was expressed and purified as described above including βME in column buffers. The specific activity for purified C326A is 233 μmol h^{-1} mg^{-1}, which indicates that removal of Cys-326 does not affect the enzymatic activity. To test for the

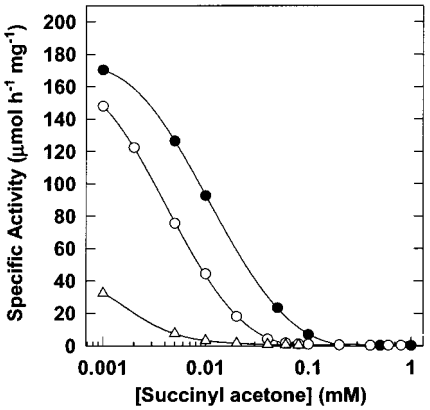

FIGURE 5: Effect of succinyl acetone on the specific activity of pea PBGS. The samples (0.8 μM subunit) were preincubated with succinyl acetone for 10 min (●), 30 min (○), or overnight (△), and the activity was measured under standard assay conditions (0.1 M Bis-tris propane hydrochloride, pH 8.5, 10 mM $MgCl_2$, 10 mM βME, and 10 mM ALA).

appearance of covalent-bound dimers, the formation of disulfide linkages was enhanced by dialyzing both the native and C326A enzyme preparations in a pH 8.5 buffer solution that did not contain βME. Samples were removed at various time points and the formation of multimeric forms linked by covalent bonding were analyzed by SDS−PAGE in sample buffer that did not contain βME (Figure 4D). For the native pea PBGS, formation of intermolecular disulfide linkages started immediately, as was detected by the appearance of a dimeric protein (size 74 kDa) in the 2 h time point (lane 3). The amount of 74 kDa protein was pronounced in the 4 h time point (lane 5), and after 24 h of dialysis (lane 7) around 50% of the sample had undergone dimer formation. No higher molecular weight complexes were detected. For the C326A protein, no formation of dimers or higher molecular weight forms were detected during the entire time course (lanes 2, 4, 6, and 8). Thus, it is evident that, in the native PBGS, disulfide linkages can form between Cys-326, which results in a covalently bound 74 kDa dimer. It is significant that the ratio between the formed 74 kDa dimer and the decreased 37 kDa monomeric form never exceeded 1:1 although the dialysis was continued for 4 additional days. Having covalently linked two of the four dimers apparently leaves insufficient mobility to form the final two disulfides (see Figure 4C). Thus, upon denaturation, only four of the eight subunits are cross-linked. Dialysis at pH 9.4 or addition of oxidized and reduced glutathione in varied concentrations did not enhance dimerization.

Inhibition of Pea PBGS with Succinyl Acetone. Some studies on the inactivation of plant PBGS with various inhibitors have been carried out (*6*, *9*, *10*). One of the potent compounds found was succinyl acetone, a substrate analogue believed to bind at the enzyme active site and known to inactivate the protein (*9*, *29*). Figure 5 shows enzyme activity over a wide range of succinyl acetone concentrations following enzyme−inhibitor incubation times of 10 min, 30 min, and 16 h. Pea PBGS exhibited a time-dependent inactivation with succinyl acetone that prevented the determination of a K_i value for the inhibitor. However, a practically inactive enzyme/inhibitor complex was obtained with around 5 μM succinyl acetone after 16 h of preincubation. If the preincubation time was shortened to 10 min, the lowest inhibitor concentration for a complete inhibition

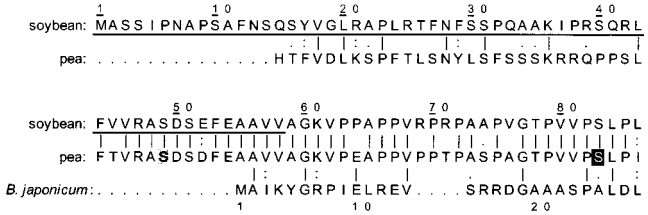

FIGURE 6: Alignment of the N-terminal sequences of soybean, pea, and *Bradyrhizobium japonicum* PBGS. The putative chloroplast transit peptide predicted for soybean PBGS by PSORT and ChloroP programs (see text) is underlined. In pea PBGS, starting residues (mutated to Met) for the enzymes used in Cheung et al. (*9*) and in this study are marked by gray and black backgrounds, respectively.

was around 0.1 mM. Thus, these experiments show that if the preincubation time is extended >16 h, about a 6-fold excess of succinyl acetone completely inhibits pea PBGS. Pea PBGS crystallizes fairly easily in a complex with this inhibitor (unpublished result) and thus it may be a good alternative to levulinic acid in the structural studies of the plant PBGS/inhibitor complexes.

DISCUSSION

Chloroplast Transit Sequence. Here we describe a short form of pea PBGS that is devoid of all N-terminal sequence beyond that found in mammalian and some microbial PBGS. The characteristics of this protein resemble those determined for a recombinant pea PBGS that contained an additional 34 N-terminal amino acids and was produced in much smaller quantities (*9*). Some of the additional N-terminal residues might normally be part of a chloroplast transit sequence, which is present in many nuclear encoded proteins destined to chloroplasts (*30*, *31*). The pea *ALAD* cDNA clone, originally isolated from etiolated leaves (*5*), was used as a starting material in Cheung et al. (*9*). This clone does not encode the full-length pea PBGS and thus prevented predictions for the possible length or type of the putative chloroplast transit peptide. Another full-length plant PBGS gene (U04525) has been isolated from a closely related leguminous plant soybean (*Glycine max* L.), and the amino acid sequence identity for the whole sequences of pea and soybean PBGS is 88% (Figure 6). PSORT (*32*) and ChloroP (*33*) analyses for soybean PBGS sequence revealed a putative N-terminal chloroplast transit sequence of 58 residues (Figure 6). Thus, it seems probable that the pea PBGS used in Cheung et al. (*9*) contained approximately 10 amino acid residues belonging to the transit peptide. Here the artificial *MVpeashort* gene is missing the entire transit peptide and additionally 24 residues from the mature N-terminus. However, on the basis of the high specific activity (>210 μmol h^{-1} mg^{-1}) of the expression products used here and in Cheung et al. (*9*), we conclude that the additional N-terminal residues of pea PBGS used in Cheung et al. (*9*) are unrelated to enzymatic activity.

We cannot exclude the possibility that the protein concentration dependence for the specific activity observed here for the pea PBGS is a result of the N-terminally truncated protein used in the assays. However, in support of our assumption that it is not, *Bradyrhizobium japonicum* PBGS, which shows a protein concentration dependence of the specific activity almost identical to the data presented here (see below), contains 24 N-terminal residues beyond the N-terminus of pea PBGS used here (Figure 6). In addition,

the published activity assays for the 34 amino acid longer pea PBGS in Cheung et al. (*9*) were reported to be carried out at high protein concentration.

Mg(II) of Pea PBGS. Plant PBGS functions in the chloroplast, where Mg(II) concentration can be as high as 10 mM (*34*) and free Zn(II) must be relatively low to prevent its spontaneous insertion into protoporphyrin IX (*35*). Prior work (*9*) established that pea PBGS binds two metal ions per subunit, as tested with Co(II) and Mn(II), which bind with a higher affinity than Mg(II). However, the binding data with Mg(II) presented here suggests that pea PBGS can bind up to 3 Mg(II)/subunit. The kinetic data support a required Mg(II), an allosteric Mg(II), and an inhibitory Mg(II), but the current data are insufficient to address the individual stoichiometries of these three types of Mg(II). One might intuitively set these stoichiometries each to eight per octamer. We also lack data to address whether pea PBGS exhibits half-site reactivity, i.e., of the eight spatially distinct active sites, only four are catalytically functional at any one point in time. Half-site reactivity may or may not be a universal characteristic of the PBGS proteins.

Minimal Functional Unit of PBGS. The protein dependence to the specific activity of pea PBGS closely resembles that documented for *B. japonicum* PBGS (*23*). One necessary component to the analysis of this type of data is the assumption that there is a dissociated unit that is inactive, or far less active, than one or more oligomeric units. For PBGS, in general, the possible active oligomers are octamers, tetramers, dimers, or any combination thereof. The possible inactive species are monomers, dimers, tetramers, or any combination thereof. The data on the protein dependence of the specific activity of *B. japonicum* PBGS fit several of these models equally well. Consequently the published analyses used the simplest model; monomers were considered inactive and all larger oligomers were considered to have activity directly proportional to the number of subunits (*23*). This model, published before the release of any PBGS crystal structures, must now be revisited in light of the new structural data.

The PBGS dimer, shown in Figure 4B, is tightly associated and corresponds to the asymmetric unit in the crystal structure of *P. aeruginosa* PBGS. Due to numerous noncovalent interactions between monomers, the dimer is deemed unlikely to reversibly dissociate into monomers. Hence, the existence of monomers is omitted from new models used to analyze the protein concentration dependence of PBGS specific activity. If monomers are omitted, the required active species are limited to octamers or tetramers or both, while the possible inactive species are tetramers or dimers or both. An analysis of subunit interactions can be used to further limit the possible models. Published crystal structures (*15, 27, 28*) have shown that, in tetramers and octamers, the N-terminal arm of one subunit fits into the base of the $(\alpha\beta)_8$ barrel of a subunit diagonally situated across the tetramer as shown in Figure 7A for the tetramer. This interaction is unique to the tetramer and octamer and is not present in the dimer structure. Two models are considered in fitting the kinetic data; in both the dimer is assumed to be inactive. In the first model, it is proposed that the relative specific activity is proportional to the number of the unique subunit interactions illustrated in Figure 7A. In one tetramer, two subunits contain these unique interactions; thus, only half the active sites are functional. In the octamer, all eight subunits contain this interaction and all the active sites are functional. Thus,

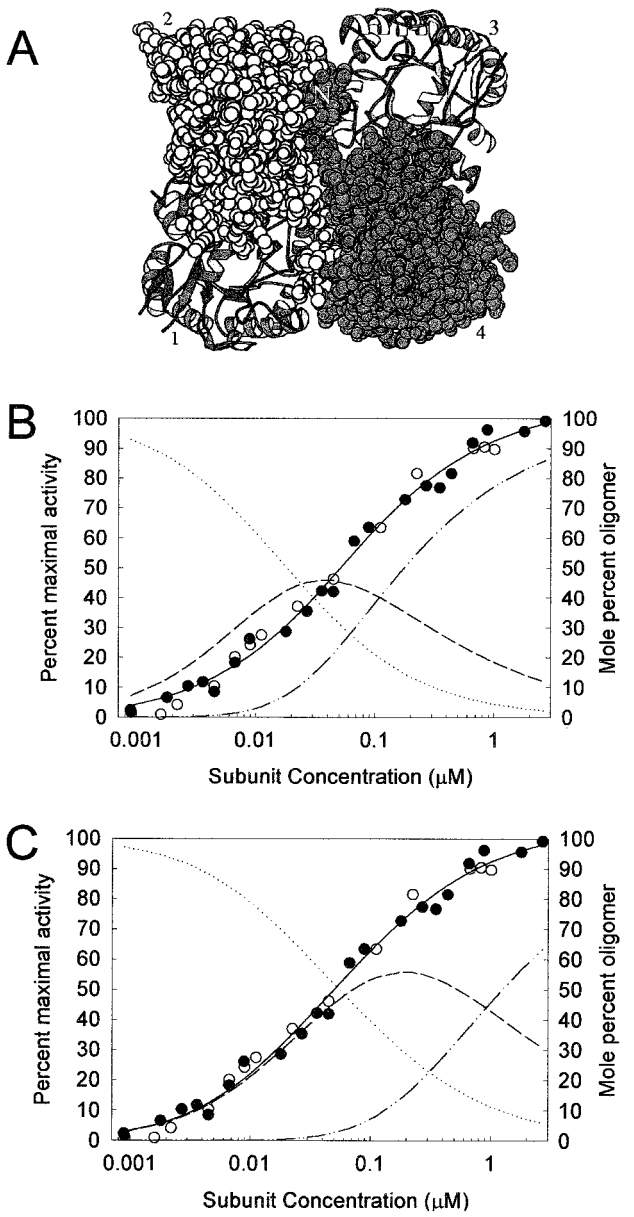

FIGURE 7: Analysis of the effect of multimerization on the specific activity of pea PBGS. (A) View of the tetramer (half of the octamer) for the pea PBGS model. Monomers are labeled 1−4. Monomers 1 and 3 are depicted as ribbon diagrams; monomers 2 and 4 are shown as space-filling models. The orientation of the tightly associated dimer consisting of monomers 1 and 2 is approximately the same as presented in Figure 4B. Monomers 4 and 3 form an identical dimer. The interaction that is present in the tetramer but not present in the dimer is the N-terminal arms of monomers 2 and 4 fitting into the base of the $(\alpha\beta)_8$ barrel in the diagonally situated monomers 4 and 2. These N-termini are marked N. In the octamer, the second tetramer (not illustrated) is located behind the one illustrated. In the octamer, the N-terminal arms fitting into the base of the $(\alpha\beta)_8$ barrel of monomers 1 and 3 come from the second tetramer. (B) Fit of the protein concentration dependence to a model where the dimer is inactive, half of the asymmetric units of the tetramer are active, and all the asymmetric units of the octamer are active. (○) Data for pea PBGS; (●) data for *B. japonicum* PBGS (*23*); percent protein as (•••) dimer, (− − −) tetramer, and (−••−) octamer is also shown. (C) Fit of the same data to a model where the dimer is inactive and the relative activity of the tetramer to the octamer is simply proportional to the number of subunits. Symbols and lines are as in panel B.

for this model, one tetramer has one-quarter of the active sites as one octamer. It is possible that at least some PBGS enzymes exhibit half-site reactivity. If one invokes an asymmetric dimer as seen in *P. aeruginosa* PBGS, then the tetramer has one functional active site and the octamer has four. However, whether the enzyme exhibits half-site reactivity, the mathematical model is the same because the ratio of active sites is 1:4 for tetramer:octamer. In the second model, the relative activity of the octamer to the tetramer is a simple direct proportion of the number of subunits and thus one tetramer has half the number of active sites as one octamer.

Figure 7B,C presents the normalized data describing the protein concentration dependence for both pea and *B. japonicum* PBGS and fits the data according to these two alternative models as described by the equations in Data Analysis above. A comparison of Figure 7 panels B,C shows that the data do not discriminate between the two models; both fit equally well. The difference lies in the distribution of the species, dimer, tetramer, and octamer, and in the K_d values. The model illustrated in Figure 7B is preferred over the model illustrated in Figure 7C because it considers a functional importance to the unique subunit associations present in the tetramer and octamer (Figure 7A). It also predicts that a high proportion of the protein is octameric under standard assay conditions and that the dimer−tetramer K_d is approximately equivalent to the tetramer−octamer K_d, 11 nM and 22 nM, respectively. The fit to the model in Figure 7C is less satisfying because the dimer−tetramer K_d (29 nM) is an order of magnitude stronger than the tetramer−octamer K_d (202 nM). For an alternative model where the octamer is considered the only active species, the predicted concentration of the octamer would simply mimic the activity curve, and there is no rational basis for calculating the proportions of the protein present as inactive dimer and tetramer. This latter model was not considered, in part because an N-terminal variant of *P. aeruginosa* PBGS purified as an active tetramer (*17*).

Specific Activity of Pea PBGS. One striking feature for pea PBGS is its much higher catalytic rate compared to animal and microbial PBGS. As presented here and in Cheung et al. (*9*), the specific activity for pea PBGS exceeds 210 μmol h^{-1} mg^{-1}, which is severalfold higher than the values reported for PBGS from human (25 μmol h^{-1} mg^{-1}) (*18*), bovine (24 μmol h^{-1} mg^{-1}) (*36*), *E. coli* (50 μmol h^{-1} mg^{-1}) (*16*, *37*), *Bradyrhizobium japonicum* (45 μmol h^{-1} mg^{-1}) (*23*), and *Pseudomonas aeruginosa* (60 μmol h^{-1} mg^{-1}) (*38*). Although a high sequence homology exists in the active-site region of all PBGS (*2*), it is not unusual for a few changes in the active site to have a measurable effect in the enzymatic efficiency. In addition, subunit−subunit interactions and molecular flexibility, for which we present evidence in this work, may partly explain the high catalytic rate. Whether or not it is physiologically significant for plant PBGS to produce porphobilinogen at an elevated rate for the large-scale synthesis of chlorophyll during germination and for vegetative growth remains to be elucidated because PBGS concentration in chloroplasts is not known.

Conclusion. We have prepared an artificial gene for pea PBGS that allows production of 100-mg quantities of protein. The current data (1) suggest that the N-terminal extension (part of it belonging to the chloroplast transit peptide) seen in plant PBGS is not catalytically important for the enzyme, (2) confirm that the plant protein has a specific activity nearly an order of magnitude higher than the mammalian protein, (3) provide evidence for the PBGS tetramer as the minimal functional unit, and (4) establish some validity to the modeled structure through mutagenesis of Cys326. The current data do not establish the precise stoichiometry of the three different types of Mg(II) that can bind to plant PBGS, the presence or absence of half-site reactivity for this protein, or the locations of the Mg(II) binding sites in the structure. These are the subject of ongoing studies.

ACKNOWLEDGMENT

We thank Steven H. Seeholzer for carrying out the MALDI-TOF-MS analysis, Catherine A. Oleykowski for the purification of the oligonucleotide templates, and J. Michael Sauder for assistance in preparing Figure 7A. Terry D. Copeland, NCI, Frederick Cancer Research and Development Center, MD, is thanked for carrying out N-terminal sequencing.

REFERENCES

1. Shoolingin-Jordan, P. M. (1998) *Biochem. Soc. Trans. 26*, 326−336.
2. Jaffe, E. K. (2000) *Acta Crystallogr. D 56*, 115−128.
3. Wetmur, J. G., Bishop, D. F., Cantelmo, C., and Desnick, R. J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7703−7707.
4. Mitchell, L. W., and Jaffe, E. K. (1993) *Arch. Biochem. Biophys. 300*, 169−177.
5. Boese, Q. F., Spano, A. J., Li, J., and Timko, M. P. (1991) *J. Biol. Chem. 266*, 17060−17066.
6. Liedgens, W., Lutz, C., and Schneider, H. A. (1983) *Eur. J. Biochem. 135*, 75−79.
7. Schneider, H. A. (1976) *Z. Naturforsch. 31C*, 55−63.
8. Kaczor, C. M., Smith, M. W., Sangwan, I., and O'Brian, M. R. (1994) *Plant Physiol. 104*, 1411−1417.
9. Cheung, K. M., Spencer, P., Timko, M. P., and Shoolingin-Jordan, P. M. (1997) *Biochemistry 36*, 1148−1156.
10. Senior, N. M., Brocklehurst, K., Cooper, J. B., Wood, S. P., Erskine, P., Shoolingin-Jordan, P. M., Thomas, P. G., and Warren, M. J. (1996) *Biochem. J. 320*, 401−412.
11. Matters, G. L., and Beale, S. I. (1995) *Plant Mol. Biol. 27*, 607−617.
12. Smith, A. G. (1988) *Biochem. J. 249*, 423−428.
13. Reinbothe, S., and Reinbothe, C. (1996) *Eur. J. Biochem. 237*, 323−343.
14. Grimm, B. (1998) *Curr. Opin. Plant Biol. 1*, 245−250.
15. Frankenberg, N., Erskine, P. T., Cooper, J. B., Shoolingin-Jordan, P. M., Jahn, D., and Heinz, D. W. (1999) *J. Mol. Biol. 289*, 591−602.
16. Jaffe, E. K., Ali, S., Mitchell, L. W., Taylor, K. M., Volin, M., and Markham, G. D. (1995) *Biochemistry 34*, 244−251.
17. Frankenberg, N., Jahn, D., and Jaffe, E. K. (1999) *Biochemistry 38*, 13976−13982.
18. Jaffe, E. K., Volin, M., Bronson-Mullins, C. R., Dunbrack, R. L., Jr., Kervinen, J., Martins, J., Quinlan, J. F., Jr., Sazinsky, M. H., Steinhouse, E. M., and Yeung, A. T. (2000) *J. Biol. Chem. 275*, 2619−2626.
19. Kane, J. F. (1995) *Curr. Opin. Biotechnol. 6*, 494−500.
20. Dunbrack, R. L., Jr. (1999) *Proteins Suppl. 3*, 81−87.
21. Bower, M. J., Cohen, F. E., and Dunbrack, R. L., Jr. (1997) *J. Mol. Biol. 18*, 1268−1282.
22. Dunbrack, R. L., Jr., and Cohen, F. E. (1997) *Protein Sci. 6*, 1661−1681.
23. Petrovich, R. M., Litwin, S., and Jaffe, E. K. (1996) *J. Biol. Chem. 271*, 8692−8699.
24. Hamming, R. W. (1962) in *Numerical Methods for Scientists & Engineers*, McGraw-Hill Books Co., Inc., New York, pp 81−82.
25. Schauer, W. E., and Mattoon, J. R. (1990) *Curr. Genet. 17*, 1−6.

26. Petrovich, R. M., and Jaffe, E. K. (1997) *Biochemistry 36*, 13421−13427.
27. Erskine, P. T., Senior, N., Awan, S., Lambert, R., Lewis, G., Tickle, I. J., Sarwar, M., Spencer, P., Thomas, P., Warren, M. J., Shoolingin-Jordan, P. M., Wood, S. P., and Cooper, J. B. (1997) *Nat. Struct. Biol. 4*, 1025−1031.
28. Erskine, P. T., Norton, E., Cooper, J. B., Lambert, R., Coker, A., Lewis, G., Spencer, P., Sarwar, M., Wood, S. P., Warren, M. J., and Shoolingin-Jordan, P. M. (1999) *Biochemistry 38*, 4266−4276.
29. Tschudy, D. P., Hess, R. A., and Frykholm, B. C. (1981) *J. Biol. Chem. 256*, 9915−9923.
30. Keegstra, K., and Cline, K. (1999) *Plant Cell 11*, 557−570.
31. Chen, X., and Schnell, D. J. (1999) *Trends Cell Biol. 9*, 222−227.
32. Nakai, K., and Horton, P. (1999) *Trends Biochem. Sci. 24*, 34−36.
33. Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999) *Protein Sci. 8*, 978−984.
34. Walker, C. J., and Weinstein, J. D. (1994) *Biochem. J. 299*, 277−284.
35. Baum, S. J., and Plane, R. A. (1966) *J. Am. Chem. Soc. 88*, 910−913.
36. Bevan, D. R., Bodlaender, P., and Shemin, D. (1980) *J. Biol. Chem. 255*, 2030−2035.
37. Spencer, P., and Jordan, P. M. (1993) *Biochem. J. 290*, 279−287.
38. Frankenberg, N., Heinz, D. W., and Jahn, D. (1999) *Biochemistry 38*, 13968−13975.

BI000620C