An Artificial Gene for Human Porphobilinogen Synthase Allows Comparison of an Allelic Variation Implicated in Susceptibility to Lead Poisoning*

(Received for publication, August 16, 1999, and in revised form, October 15, 1999)

Eileen K. Jaffe†, Marina Volin, Colleen R. Bronson-Mullins, Roland L. Dunbrack, Jr., Jukka Kervinen, Jacob Martins, Jack F. Quinlan, Jr., Matthew H. Sazinsky, Erica M. Steinhouse, and Anthony T. Yeung

From the Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Porphobilinogen synthase (PBGS) is an ancient enzyme essential to tetrapyrrole biosynthesis (e.g. heme, chlorophyll, and vitamin B12). Two common alleles encoding human PBGS, K59 and N59, have been correlated with differential susceptibility of humans to lead poisoning. However, a model for human PBGS based on homologous crystal structures shows the location of the allelic variation to be distant from the active site with its two Zn(II). Previous microbial expression systems for human PBGS have resulted in a poor yield. Here, an artificial gene encoding human PBGS was constructed by recursive polymerase chain reaction from synthetic oligonucleotides to rectify this problem. The artificial gene was made to resemble the highly expressed homologous Escherichia coli hemB gene and to remove rare codons that can confound heterologous protein expression in E. coli. We have expressed and purified recombinant human PBGS variants K59 and N59 in 100-μg quantities. Both human PBGS proteins purified with eight Zn(II)/octamer; Zn(II) binding was shown to be pH-dependent; and Pb(II) could displace some of the Zn(II). However, there was no differential displacement of Zn(II) by Pb(II) between K59 and N59, and simple Pb(II) inhibition studies revealed no allelic difference.

Human genetic polymorphisms are increasingly being correlated with disease states through epidemiological analysis. These types of studies lay the groundwork for the extended practice of preventive medicine. A common genetic polymorphism in the human ALAD gene,1 encoding the enzyme porphobilinogen synthase (PBGS2; EC 4.2.1.24), has been correlated with susceptibility to the environmental toxin lead (1–6). As lead poisoning is the most common preventable childhood neurological environmental disease, correlation of the epidemiological results with protein structure/function studies is in order.

PBGS catalyzes the first common step in the biosynthesis of all tetrapyrroles (heme, chlorophyll, vitamin B12, cofactor F430, etc.). Human PBGS is a Zn(II) metalloenzyme unique in its sensitivity to inhibition by lead. Although all PBGSs appear to be metalloenzymes (7), metal ion usage varies dramatically between species (8, 9). One outcome of this variation is that microbial and plant PBGSs are poor models for studying the effect of lead on human PBGS function. Lead inhibition of human PBGS is one of the earliest physiological responses to lead intoxication and as such is believed to be related to the detrimental effects of low level lead poisoning. However, mapping the common human polymorphism onto the x-ray crystal structure of the related yeast PBGS protein does not indicate a structural variation that would obviously affect either metal binding or catalytic function. A model for human PBGS is presented in Fig. 1 and shows the location of amino acid 59, which is lysine in the ALAD1 gene product and asparagine in the ALAD2 gene product.

Production of human PBGS for study has been problematic. Purification of PBGS from human blood gives relatively low yield, is plagued by considerations of blood-borne diseases, and contains a mixture of the isozymes (10, 11). The human ALAD gene was cloned and sequenced more than a decade ago and found to express poorly in Escherichia coli (12). To generate a better expression system, the gene was cloned into yeast (13), but the levels of protein expression remain insufficient for thorough functional analysis.

For a novel approach to heterologous expression of human PBGS, we chose to mimic a construct used for overexpression of the E. coli hemB gene in E. coli (14, 15). This system can generate hundreds of milligrams of E. coli PBGS (up to 30% of the soluble protein). The reasons for high level expression are not fully understood. In this case, the hemB gene is apparently downstream from a naturally strong E. coli promoter, and other poorly understood aspects of gene structure may contribute to the phenomenon levels of constitutive expression observed. A second consideration in artificial gene design is codon usage. The human ALAD gene was analyzed and found to contain clusters of codons that are rarely used by E. coli (see Fig. 2A). In contrast, the E. coli hemB gene contains only one rare codon. Kane (16) has described how clusters of six specific rare codons can be detrimental to both the quality and quantity of heterologous proteins expressed in E. coli, and specific translational errors have been documented for such clusters (17). Hence, in the design of an artificial gene for human PBGS, the E. coli hemB gene structure was mimicked to the greatest extent possible, and rare codons were replaced.

Here we describe the design and synthesis of a PCR product

---

1 The naturally occurring human gene is called ALAD, encoding the enzyme porphobilinogen synthase.
2 The abbreviations used are: PBGS, porphobilinogen synthase; PCR, polymerase chain reaction; IPTG, isopropyl-β-D-thiogalactopyranoside; PMSF, phenylmethylsulfonyl fluoride.
Inc. The restriction enzymes BamHI, EcoRI, HpaI, and NdeI were from New England Biolabs Inc. Plasmids were purified using QIAGEN plasmid purification kits, and PCR products were extracted from 1–2% agarose using QIAEX II kits from QIAGEN Inc. DNA sequencing was carried out in-house using ABI sequencing technology. Mutagenesis of plasmid pMVhum was carried out using the QuikChange technology of Stratagene.

**Design of the Nucleotide Sequence of the Artificial Gene and Plasmid pEJhum**—The *E. coli* gene construct pCR261 is an outstanding constitutive expression system (14). The pUC119 version, denoted pLM1228, is illustrated in Fig. 2B. The artificial gene for human PBGS was placed in the same context (Fig. 2C), the context is denoted pCR261 (Fig. 2D). Four steps were included in the design process. First, the sequence of pLM1228 was determined from original sequence data and put in the context of pUC119 (u07650.gb_sgy) and the 6′–8′ region of the *E. coli* chromosome (u73857.gb_new). The region of pLM1228 containing *E. coli* DNA is illustrated in Fig. 2B; nucleotides 1–271 are the first 271 nucleotides of pUC119. Nucleotides 272–1744 are identical to a region of the *E. coli* chromosome (the reverse of nucleotides 97935 to 95921 of u73857.gb_new); these include 75 nucleotides of 5′-untranslated sequence from upstream from *hemB* and the entire *hemB* gene, followed by 364 bases 3′ to *hemB*, which include a part of the coding sequence of the *yaiG* gene. The remaining sequence derives from pUC119. A schematic of the target PCR product is shown in Fig. 2C. Plasmid pEJhum was designed to retain all of the sequence of pLM1228 that is upstream and downstream of the *E. coli* *hemB* gene as illustrated in Fig. 2D.

The second step was to alter the codons of the human ALAD gene to mimic those of the *E. coli hemB* gene in all cases where the amino acids of the two proteins are identical. This was done manually and was assisted by the Genetics Computer Group programs FRAMEALIGN (Fig. 2A) and DESTPFIT.

The third step removed any remaining rare codons (CTA, ATA, AGG, CCC, AGA, and AGT) that have been identified as particularly problematic in heterologous expression (16). The sequence of the human ALAD gene (GenBank™ accession number M13928) contains 20 of these codons, which Fig. 2A shows are relatively frequent and that some fall in clusters. The Genetics Computer Group programs CODONFREQUENCY and FRAMEALIGN were used to assist the identification of these codons, which were arbitrarily altered to common *E. coli* codons. In addition, the sequence was searched for possibly detrimental “over-represented” codons as described by Irwin et al. (18). Three codon pairs were found with a CHISQ3 value of >50; these were altered to common codons to generate the final artificial gene. The *E. coli hemB* gene was also found to contain three over-represented codon pairs (one CTGAGC and two CTGGTG). The significance of altering over-represented codons remains untested.

The fourth step assembled and analyzed the sequence of the target plasmid pEJhum using the Genetics Computer Group program ASSEMBLE. This file contained the first 347 nucleotides of pLM1228, followed by the coding region of the artificial gene (990 nucleotides) and nucleotides 1320–4636 of pLM1228. The resulting plasmid contained the beginning of the *lacZ* gene in the first reading frame (starting at nucleotide 217), no extended translatable sequence in the second reading frame, and the artificial gene in the third reading frame (starting at nucleotide 348). To insert a stop site in the *lacZ* fusion without interfering with any of the possibly important transcriptional regions, we altered nucleotide 333 from A to T.

**Generation of PCR Fragments to Yield EJhum**—The synthetic target *EJhum*, illustrated schematically in Fig. 2C and in detail in Fig. 3, contains nucleotides 264–1371 of pEJhum and extends from a unique *BamHI* site to a unique *HpaI* site. Additional nucleotides were added to the 5′- and 3′-ends to serve as handles. Using the program Oligo (National Biosciences), eight templates and eight primers were designed for the total synthesis of *EJhum* using recursive PCR. The templates were each 155–175 nucleotides in length with at least a 25-base pair overlap so that adjacent oligonucleotides would prime each other. The end primers were each 30–35 nucleotides in length. All were matched as well as possible in melting temperatures.

**Design of the Synthetic Route for EJhum**—The synthetic target *EJhum*, illustrated schematically in Fig. 2C and in detail in Fig. 3, contains nucleotides 264–1371 of pEJhum and extends from a unique *BamHI* site to a unique *HpaI* site. Additional nucleotides were added to the 5′- and 3′-ends to serve as handles. Using the program Oligo (National Biosciences), eight templates and eight primers were designed for the total synthesis of *EJhum* using recursive PCR. The templates were each 155–175 nucleotides in length with at least a 25-base pair overlap so that adjacent oligonucleotides would prime each other. The end primers were each 30–35 nucleotides in length. All were matched as well as possible in melting temperatures.

**Generation of PCR Fragments to Yield EJhum**—Fig. 3 illustrates the sequence of EJhum with notations for the location of the eight templates and the eight primers. Sequential primer pairs and their cognate templates were used to synthesize four fragments, each containing at least 300 nucleotides. Fragments A–D were flanked by primer pairs 1/2, 3/4, 5/6, and 7/8, respectively. The fragments were agarose gel-purified and used as templates for the recursive PCR synthesis of two >580-base pair fragments. These fragments, AB and CD, were flanked by primer pairs 1/4 and 5/8, respectively. Finally, the 1.1-kilobase PCR product *EJhum* arose from the recursive PCR of fragments AB and CD using.
primer pair 1/8. The product EJhum theoretically contains the artificial gene encoding human PBGS. However, this synthetic strategy did not generate an error-free version of EJhum, and a second strategy was needed (see below).

Creation of pEJhum and Its Transformation into a Host Strain—Plasmid pLM1228 and the PCR product EJhum were digested with BamHI and HpaI. The vector was dephosphorylated, and the vector and the 1.1-kilobase PCR product were gel-purified and ligated with T4 DNA ligase to form pEJhum (Fig. 2D). If correct, this plasmid, when used to transform E. coli strain RP523 (hemB mutant) (19), should complement the hemin auxotrophy. pEJhum-containing transformants were not obtained under conditions where pLM1228 could successfully be used to prepare complementing transformants of RP523. However, transformants of pEJhum in E. coli strain HB101 were obtained. Ninety-two transformants were screened by colony PCR using primers 3 and 6 (Fig. 3). Fourteen colonies were selected, but none overexpressed human PBGS. EJhum was amplified by PCR from each of these 14 colonies using primers 1 and 8, and the PCR products were sequenced and uniformly found to contain multiple randomly located deletions (1–30 bases) and some errors.

To determine whether the errors derived from the unpurified templates or from the PCR process, each fragment (A–D) was blunt end-ligated into pPCR-Script Amp SK(+) at the SrfI site and transformed into Epicurian Coli XL1-Blue MRF for plasmid purification and

![Diagram](image)

**Fig. 2.** From ALAD to the artificial gene for human PBGS. A, the cDNA sequence encoding human PBGS (GenBank™ accession number M13928) is aligned with the encoded amino acid sequence (12). Regions of amino acid identity to E. coli PBGS are shaded (131 out of 330 codons). The six most detrimental E. coli codons (16) with their encoded amino acids are in large type (20 out of 330 codons). Concerns for the quantity or quality of the expressed protein stem from clusters of these rare codons. All codons shaded and/or in large type were subject to change in the design of the artificial gene according to the rationale described under “Experimental Procedures.” B–F, shown are maps of the DNA constructs used in this study. B, plasmid pLM1228 was originally prepared for expression of site-directed mutants of E. coli PBGS. The E. coli hemB gene and its flanking DNA were cloned into the EcoRI site of pUC119 (14). Section I is part of the natural hemB promoter region; section II is the coding region of hemB; section III is 3′-untranslated DNA; and section IV is part of the yatG gene. For unknown reasons, this construct gives high constitutive expression of E. coli PBGS in a variety of E. coli hosts. C, the PCR target EJhum resembles its cognate portion of pLM1228 except that the artificial gene encodes human PBGS (see “Experimental Procedures”). D, the plasmid pEJhum, designed for constitutive expression of the artificial gene, did not yield stable constitutive expression. E, the PCR target MVhum contains the coding region of the artificial gene alone plus the NdeI and BamHI sites (and some flanking DNA) for insertion into pET3a or pET11a. F, the final plasmid pMVhum contains the artificial gene under control of T7 polymerase in a pET3 background.
Allelic Variation of Human PBGS

Creation of the PCR Product MVhum and Plasmid pMVhum, a Controlled Expression System—To control the expression of human PBGS in E. coli, Edhum was reengineered 1) to remove the 5’- and 3’-flanking regions derived from the E. coli genome and 2) to allow incorporation into one of the pET plasmids for T7 polymerase-directed expression. Two oligonucleotides (AAAACUGAAGGCGGCTGCATATGCAGCCTCAGTCCGTTC and GGCTGAGAGGAGACCGGCTGTTGCT) were used as PCR primers to excise the artificial gene from pEJhum, to align the start site with the Ndel site of pET3a or pET11a, and to add a BamHI site past the termination codon (Figs. 2F and 3). The PCR product MVhum was gel-purified, blunted end-ligated into the pPCR-Script Amp SK (+) vector, and transformed into XL1-Blue MRF’ Kan. White transformants were screened by colony PCR using the above-mentioned primers. Following sequence confirmation, the artificial gene was excised and ligated into the Ndel and BamHI sites of both pET3a and pET11a to yield plasmids pMVhum3 and pMVhum11, respectively. Because of high sequence identity (~60%) between the artificial gene and the E. coli hemB gene, the recA1 host strains BLR(DE3) and BLR(DE3) pLM1228 were used as potential hosts for both pMVhum3 and pMVhum11. In all cases, transformants were obtained, and overexpression was observed upon induction with IPTG using SDS Phastgels as the analytical tool. BLR(DE3)(pMVhum3) showed the lowest basal expression and the highest induced expression; thus all further work was done with this strain, renamed BLR(DE3)(pMVhum). An illustration of pMVhum is included in Fig. 2F.

Expression of Human PBGS from BLR(DE3)(pMVhum)—BLR(DE3) (pMVhum) was grown in 1-liter batches of Luria broth, 100 μg/ml ampicillin, and 0.4% glucose at 37 °C in an air shaker, starting the inoculation from a single colony from a fresh transformation. After 16 h, the cells typically reached an A600 of 4–5, at which point they were spun down and resuspended in 1 liter of Luria broth at 45 °C. After 30 min to 1 h of shaking, the flasks were cooled to 37 °C, and IPTG was added to a concentration of 10 μM. Overnight growth at 37 °C under these conditions in an air shaker led to a final A590 of 6–11, and 10–20% of the total PBGS (K59 variant) was in the soluble fraction of the cell lysate. The N59 variant partitions more favorably into the soluble fraction (~60%). An alternate expression procedure that improves protein solubility involved a 42 °C heat shock and used 100 μM IPTG to induce expression at 15 °C for a period of 48 h in the presence of 20 μM Zn(II).

Purification of Recombinant Human PBGS from BLR(DE3) (pMVhum)—Frozen BLR(DE3)(pMVhum) cells were suspended (2 ml/g) in 50 mM potassium phosphate, pH 8.0, 170 mM KCl, 5 mM EDTA, 10 mM 2-mercaptoethanol, and 0.1 mM phenylmethylsulfonyl fluoride (PMSF). After the cells were thawed and dispersed, lysozyme was added to the suspending buffer to a concentration of 0.4 mg/ml, and the suspension was stirred at room temperature for 1.5 h. At this point, an equal volume of 0.1 mM potassium phosphate, pH 7.0, 12 mM MgCl2, 40 μM ZnCl2, 10 μM β-mercaptoethanol, and 0.1 mM PMSF was added with DNase I (~60 units/g of cells), and the suspension was stirred an additional 20 min. All of the subsequent procedures were carried out at 4 °C. The cell mixture was passed through an 85-micron French press at 20,000 p.s.i. The resulting cell lysate was centrifuged at 20,000 × g for 20 min. The supernatant was subjected to a 20–45% saturated ammonium sulfate fractionation. The 45% saturated ammonium sulfate pellet was dissolved in 30 mM potassium phosphate, pH 7.0, 10 μM ZnCl2, 10 μM β-mercaptoethanol, 0.1 mM PMSF, and 20% saturated ammonium sulfate and applied to a phenyl-Sepharose column that had been equilibrated with the same buffer. The column was washed with ~1 column volume of the same buffer and then subjected to a linear 2-column volume gradient that ended at 2× column volume of potassium phosphate, pH 7.0, 10 μM ZnCl2, 10 μM β-mercaptoethanol, and 0.1 mM PMSF. Human PBGS eluted after the gradient in a 1-column volume wash of the ending buffer and was pumped directly onto a DEAE-Bio-Gel column that had been equilibrated with 30 mM potassium phosphate, pH 7.5, 10 μM ZnCl2, 10 μM β-mercaptoethanol, and 0.1 mM PMSF. The DEAE column was washed with 1 column volume of this buffer and was subjected to a 2-column volume linear gradient in the same buffer with a final salt concentration of 0.4 M KCl. The tail end of the peak off the DEAE column was checked for contamination with E. coli PBGS as shown by a stimulation of activity in those fractions by Mg(II) (15). Fractions of high constant specific activity were pooled and concentrated under pressure to ~10 mg/ml. The concentrated protein
was passed down a 1-m-long S-300 column (column volume = 70 × sample volume) at a flow rate of 0.1 cm/min. The S-300 buffer contained 0.1 mM potassium phosphate, pH 7, 10 mM β-mercaptoethanol, and 10 mM Zn(II).

Zn(II) Binding by Equilibrium Dialysis—Aliquots (1.0 ml) of purified recombinant human PBGS (K59, 0.56 mg/ml) were placed in Slidealyzer cassettes and equilibrated against 250 ml of buffer (0.1 mM potassium phosphate, pH 7, and 10 mM β-mercaptoethanol) containing various initial concentrations of Zn(II) (0–30 μM). Following 24 h of dialysis at 4 °C, the protein was removed from the cassette and analyzed for both protein concentration by Pierce Coomassie assay and for total Zn(II) concentration by direct reading at 213.9 nm using a flame atomic absorption spectrometer. Dialysis buffers were read directly using the same method to determine free Zn(II). Bound Zn(II)/octamer was calculated from the protein samples as follows: ([Zn,oct]/[Zn,free] = 280 mg (mg of octamer)/protein concentration (mg/ml)). The apparent binding constants were determined by nonlinear best fit using a model for two sites of equal stoichiometry as we found best fit the Zn(II) binding data for bovine PBGS obtained previously (21).

Additional Zn(II) binding data were obtained at room temperature following overnight dialysis in 50 mM sodium acetate and 10 mM β-mercaptoethanol, pH 5.0, which served to strip the Zn(II) from the protein. The Slidealyzer cassettes were then moved to 0.1 M potassium phosphate, pH 7, and 10 mM β-mercaptoethanol, and 10 mM Zn(II). Enzyme Activity Assays—Enzymes were preincubated in 0.1 M potassium phosphate, pH 9, and 10 mM β-mercaptoethanol prior to harvest and the 42 °C heat shock (lane A); 2 h after induction with 10 μM IPTG at 37 °C (lane B); 4 h after IPTG induction (lane C); and harvested cells (lane D). Lane E shows purified protein.

**RESULTS**

**Design and Synthesis of an Artificial Gene Encoding Human PBGS**—Recursive PCR (27) was used to prepare an artificial gene encoding the protein formed by the more frequent allele of the human gene for porphobilinogen synthase. The gene was designed to resemble the homologous *E. coli* hemB gene to the greatest extent possible and to remove codons rarely used by *E. coli*. As described under “Experimental Procedures,” the synthetic process and constitutive expression were problematic. We conclude that it was unwise to use unpurified synthetic oligonucleotide templates 150–175 bases in length for recursive PCR. The gene toxicity problem may be related to alternative protein folding/degradation functions reported for PBGS proteins (28, 29). The artificial gene encoding human PBGS in pMVhum is optimized for controlled T7-directed expression in an *E. coli* BL21-type host and has been shown to yield high levels of expression as determined by SDS-polyacrylamide gel electrophoresis (Fig. 4).

**Optimized Expression of Human PBGS in the Soluble Extract of BLR(DE3)/(pMVhum3)—**Standard conditions for expressing genes from pET vectors in BL21 or its derivatives often yield high expression of proteins that are found aggregated in inclusion bodies, as we have found for human PBGS. The inclusion body might be considered as a partially purified form of the protein that can sometimes be denatured, purified, and refolded into an active protein. Human PBGS is a homooligomeric enzyme containing four active sites, reactive cysteines, and two different types of Zn(II)-binding sites and thus would have made the refolding exercise a challenge. Instead, we used active enzyme purified from the soluble extract of the cells in the characterization of human PBGS expressed from pMVhum. The rationale for the growth/expression protocol is based on the following points. Glucose was included in the first growth to repress expression of genes under the control of the lac promoter such as the T7 RNA polymerase in BL21(DE3). Prior to induction, the cells were transferred to fresh medium that did not contain glucose. To increase the basal level of chaperones in the cells, a 42 °C heat shock preceded induction. Finally, low level IPTG induction was used to promote slow expression from pMVhum so as to optimize the opportunity of the protein to fold correctly (20). Fig. 4 illustrates the low level of expression seen after the first glucose-supplemented growth and good expression following the IPTG induction. This procedure can be performed on an 8-liter scale using air shakers and yielding up to 50 g of cells. The 15 °C procedure gives equivalent expression with somewhat better partitioning of the protein into the soluble fraction.

**Purification of Recombinant Human PBGS from BLR(DE3)** (pMVhum)—The method for purifying human PBGS from the soluble fraction of BLR(DE3)/(pMVhum) drew on our experience and published purifications of bovine, human, *E. coli*, and *Bradyrhizobium japonicum* PBGSs (8, 15, 30–32). In all cases, human PBGS behaved as expected from prior experience with mammalian PBGS. Following a 20–45% ammonium sulfate fractionation, K59 constituted ~11% of the total protein, and N59 constituted ~30% of the total protein. The phenyl-Sepharose column removed the majority of the UV-absorbing components and yielded a 2-fold or greater increase in specific activity. Human PBGS eluted near the end of the DEAE column gradient and yielded protein of high specific activity (60–85% of the maximal value). The final pure protein pool from the Sephacryl column step (Fig. 4, lane E) for the protein product of pMVhum (peak at ~58% of the column volume) typically had a specific activity of ~45 μmol/h/mg, which is ~50% larger than the highest value reported for any mammalian PBGS (31). The less common N59 protein had a specific activity of ~24 μmol/h/mg when expressed under identical conditions. The ~2-fold difference in the specific activity of K59 and N59 was highly reproducible between growths and preparations. The overall...
yield of purified protein was \(-10\) mg/liter for K59 and \(-35\) mg/liter for N59; the difference relates to the differential partitioning of these isozymes between the soluble fraction and the inclusion bodies.

**Zn(II) Interactions with Recombinant Human PBGS**—The protein purified in the presence of \(10\) mM Zn(II) and \(10\) mM \(\beta\)-mercaptoethanol was found by atomic absorption spectroscopy to contain eight Zn(II)/octamer, as was found earlier for bovine PBGS (30). A generally consistent model for mammalian PBGS is a homo-octameric protein with four functional active sites, each of which contains two zinc ions denoted ZnA and ZnB (33, 34). One piece of data supporting this model is the binding of Zn(II) to bovine PBGS, which shows tight binding of four Zn(II)/octamer (\(K_d = 0.1\) mM) and looser binding of a second four Zn(II)/octamer (\(K_d \approx 5\) mM) when holoenzyme is dialeded at 4 °C against 0–30 mM Zn(II) (21). In the case of recombinant human PBGS, Fig. 5A illustrates that this model for Zn(II) binding gives a good fit to data obtained under these conditions. Here the fit shows a tight Zn(II) at \(n_1 = 4.2\) and \(K_d = 0.1\) mM and a looser Zn(II) at \(n_2 = 4.2\) and \(K_d \approx 5\) mM (Table I). Fig. 5A and Table I include data obtained previously on bovine PBGS for comparative purposes. Also consistent with the four-ZnA and four-ZnB model is that only four Zn(II)/octamer are required for full activity of bovine PBGS in the presence of \(\beta\)-mercaptoethanol (30, 35). Fig. 5B shows a Zn(II) activation curve for N59, where the initial slope corresponds to full activation upon addition of 0.5 Zn(II)/subunit. The apparent concentration of Zn(II) available from assay components is \(0.2\) \(\mu\)M, as shown by the \(x\) intercept in Fig. 5B. The activity of the purified protein was found to be insensitive to the addition

| pH | Studies done at RT\(^a\) starting from apo-PBGS | Studies done at 4 °C starting from holo-PBGS | Bevine |
|----|---------------------------------------------|---------------------------------------------|--------|
|    | K59 | N59 | K59 | N59 | K59 | N59 | K59 | N59 |
| 5  | \(44.1\) | \(1.1\) | \(0.54\) | \(3.9\) | \(9.5\) | \(0.99\) | \(0.001\) | \(5.8\) | \(4.2\) | \(<0.1\) | \(5.1\) | \(4.0\) |
| 6  | \(4.9\) | \(12.1\) | \(0.99\) | \(3.9\) | \(9.5\) | \(0.99\) |                    |        |        |        |        |        |
| 7  | \(0.71\) | \(8.2\) | \(0.91\) | \(0.90\) | \(7.6\) | \(0.95\) |                    |        |        |        |        |        |
| 8  | \(0.82\) | \(7.5\) | \(0.95\) | \(0.67\) | \(7.8\) | \(0.91\) |                    |        |        |        |        |        |

\(^a\) RT, room temperature.
of Mg(II), as expected for human PBGS based on prior studies (15).

PBGS is often purified with tightly bound product (15), and one active-site model for Zn(II)-containing PBGS includes the amino group of porphobilinogen as a Zn(II) ligand (33). Mildly acidic pH can be used to strip both divalent metals and tightly bound product from a variety of microbial PBGSs (8, 36).

Hence, we elected to investigate Zn(II) binding to human PBGS following dialysis at pH 5, which was shown to strip all preexisting Zn(II) (Fig. 5C). By analogy to prior studies, low pH is also presumed to strip all enzyme-bound product. The results illustrated in Fig. 5C for K59 show pH-dependent Zn(II) binding that did not distinguish between the two different types of Zn(II) sites that are illustrated in Fig. 1 and that are apparent in Fig. 5A. The data for N59 are not illustrated, but the binding parameters are included in Table I. There was no significant difference in Zn(II) binding between human PBGS K59 and N59. At pH 5, virtually no Zn(II) bound to the apoenzyme; at pH 6, the apparent $K_d$ is 4–5 μM, and the number of Zn(II) sites fits eight or more sites/octamer; and at pH 7 and pH 8, the apparent $K_d$ is 0.7–0.9 μM, and the number of Zn(II) sites is approximately eight/octet. We conclude that the low pH treatment causes a loss of asymmetry in human PBGS, but the structural basis for the difference between Fig. 5A and Fig. 5C is unknown.

Equilibrium Dialysis with Pb(II) in Competition for the Zn(II) Sites—Lead has been shown to be a slow-binding inhibitor of mammalian PBGS (37). Precise analytical studies using controlled metal ion buffers have shown that the $K_d$ for Pb(II) is ~20-fold tighter than the $K_d$ for Zn(II) (38). As a first step in determining the relative ability of Pb(II) to displace Zn(II) from the two isozymes of human PBGS, equilibrium dialysis studies were carried out at pH 7 at initial Zn(II) concentrations of 1 and 10 μM and Pb(II) concentrations of 0 and 20 μM. These experiments started with apoenzyme prepared by low pH dialysis. The results illustrated in Fig. 5D show that Pb(II) competed effectively for about one-half of the Zn(II) sites. There was no differential displacement of Zn(II) by Pb(II) for K59 relative to N59.

Similar results were seen in simple fixed-time inhibition studies. When the holoenzymes were assayed for the standard 5 min, 100% activity was seen at 10 μM Zn(II); 75–80% activity was seen at 10 μM Zn(II) plus 20 μM Pb(II); ~80% activity was seen with no added Zn(II); and 10–12% activity was seen with no added Zn(II) plus 20 μM Pb(II). Because Pb(II) is a slow-binding inhibitor (37), a more thorough investigation of Pb(II) inhibition may still reveal some differences between K59 and N59.

DISCUSSION

Design of Artificial Genes—Advances in recombinant DNA technology have yielded spectacular results in heterologous protein expression in E. coli. In some cases such as PBGS, the E. coli homolog can be expressed at very high levels, whereas expression of the heterologous human protein is poor. Part of the problem with the poorly expressed protein may be clusters of codons rarely used in the problem with the poorly expressed protein may be clusters of codons rarely used in the E. coli gene to the greatest extent possible while still encoding the human protein and avoiding rarely used E. coli codons. In the case of human PBGS, the approach was successful and is recommended for others facing similar problems in protein expression.

The design of artificial genes is not a novel concept and has previously been used, for instance, to engineer in specific restriction sites (41). Our considerations in mimicking the homologous E. coli gene and minimizing rare codons were intended to increase both protein expression levels and protein quality by decreasing translational errors. Production of heterologous proteins of questionable quality is a significant concern for the biotechnology industry, and specific examples of errors in translations are appearing in the literature (17). Although the current studies do not prove the efficacy of the following additional factors, expression in fresh medium, induction with low levels of IPTG, and induction following a heat shock to increase natural chaperone levels can also be used to enhance translational fidelity or proper protein folding.

Model for Human PBGS—PBGS is an ancient and highly conserved protein with specific phylogenetic sequence variations in regions that recent crystal structures have shown to be essential to metal ion binding. Three different types of divalent metal ion-binding sites have been delimited that correspond to metals that have been called ZnA, ZnB, and MgC (42). The two Zn(II) are apparent in the yeast PBGS structure 1AWS (22), and the Mg(II) is apparent in the P. aeruginosa PBGS structure 1B4K (26). Yeast PBGS crystallizes as a symmetric octamer with eight somewhat disordered active-site regions, and the occupancy of the Zn(II) sites is not stoichiometric. In contrast, the P. aeruginosa PBGS octamer is composed of four asymmetric dimers wherein only one of the monomers of each dimer contains MgC, and only this monomer has a well ordered lid over the active site. Human PBGS has a 53% identity (61% similarity) to the yeast sequence and contains the ligands seen to bind to ZnA and ZnB in 1AW5. Because human PBGS displays half-site reactivity, implying that a dimer is needed to make one functional active site, these two Zn(II) are shown in the model to be bound to only one of the monomers.

It is interesting to note that the ligands to the four tight Zn(II), as determined by extended x-ray absorption fine structure (ZnA, with mostly oxygen and/or nitrogen ligands) (34), are not the same as the ligands to the highly populated Zn(II) in the crystal structure of yeast PBGS (ZnB, with mostly sulfur ligands) (22). We have shown that factors such as pH and substrate binding can control the disproportionation of metal ions between the Mg(II)-binding sites of B. japonicum PBGS (43), and similar factors may confound reconciliation of different studies on the various Zn(II)-binding PBGSs. These factors also confound drawing functional conclusions on Pb(II) inhibition of PBGS based strictly on where Pb(II) binds in the absence of substrate.

Conclusion—We have prepared an expression system for human PBGS that is designed for optimal heterologous expression with a low translational error frequency. This system was used for the purification of human PBGS encoded by the two common alleles (K59 and N59) found in human populations. The purified proteins exhibit characteristics consistent with other mammalian PBGSs and can be used for physical, chemical, and structural analysis of human PBGS and mutants thereof. The only significant difference seen between the proteins encoded by the two alleles is an ~2-fold variation in specific activity. There is no differential effect of the isozymes on Zn(II) binding, Pb(II) competition for the Zn(II) sites, or inhibition of activity by Pb(II). Further studies will probe deeper into differential effects of lead on Zn(II) binding, activity, and protein folding to determine a more subtle basis for the
reported genetic susceptibility of N59-expressing individuals toward lead poisoning. Finally, although statistically significant epidemiological data correlate the allelic variation in human PBGS with lead poisoning parameters (1, 2, 5), such data do not establish a causal relationship, and one may not exist.

REFERENCES

1. Wetmur, J. G., Lehnert, G., and Desnick, R. J. (1991) *Environ. Res.* **56**, 109–119
2. Schwartz, B. S., Lee, B. K., Stewart, W., Ahn, K. D., Springer, K., and Kelsey, K. (1995) *Am. J. Epidemiol.* **142**, 738–745
3. Smith, C. M., Wang, X., Hu, H., and Kelsey, K. T. (1998) *Environ. Health Perspect.* **106**, 248–253
4. Bergdahl, I. A., Gerhardsson, L., Schutz, A., Desnick, R. J., Wetmur, J. G., and Skerfving, S. (1997) *Arch. Environ. Health* **52**, 91–96
5. Bergdahl, I. A., Grubb, A., Schultz, A., Desnick, R. J., Wetmur, J. G., Sassa, S., and Skerfving, S. (1997) *Pharmacol. Toxicol.* **81**, 153–158
6. Fleming, D. E., Chettle, D. R., Wetmur, J. G., Desnick, R. J., Robinson, P. J., Boulay, D., Richard, N. S., Gordon, C. L., and Webber, C. E. (1998) *Environ. Res.* **77**, 49–61
7. Shemin, D. (1972) in *The Enzymes* (Boyer, P., ed) Vol. VII, 3rd Ed., pp. 232–237, Academic Press, New York
8. Petrovich, R. M., Litwin, S., and Jaffe, E. K. (1996) *J. Biol. Chem.* **271**, 8692–8699
9. Frankenberg, N., Jahn, D., and Jaffe, E. K. (1999) *Biochemistry* **38**, 13976–13982
10. Anderson, P. M., and Desnick, R. J. (1979) *J. Biol. Chem.* **254**, 6924–6930
11. Gibbs, P. N., Chaudhry, A. G., and Jordan, P. M. (1985) *Biochem. J.* **230**, 25–34
12. Wetmur, J. G., Bishop, D. F., Cantelmo, C., and Desnick, R. J. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 7703–7707
13. Schauer, W. E., and Mattoon, J. R. (1990) *Curr. Genet.* **17**, 1–6
14. Roessner, C. A., Spencer, J. B., Ozaki, S., Min, C., Atshaves, B. P., Nayar, P., Anousis, N., Stolowich, N. J., Holderman, M. T., and Scott, A. I. (1995) *Protein Expression Purif.* **6**, 155–163
15. Mitchell, L. W., and Jaffe, E. K. (1993) *Arch. Biochem. Biophys.* **300**, 169–177
16. Kane, J. F. (1995) *Curr. Opin. Biotechnol.* **6**, 494–500
17. Forman, M. D., Stack, R. F., Masters, P. S., Hauer, C. R., and Baxter, S. M. (1998) *Protein Sci.* **7**, 500–503
18. Irwin, B., Heek, J. D., and Hatfield, G. W. (1995) *J. Biol. Chem.* **270**, 22801–22806
19. Li, J. M., Mannanoff, H., Proença, R., Russell, C. S., and Cosloy, S. D. (1988) *J. Bacteriol.* **170**, 1021–1025
20. Yang, Q. H., Wu, C. L., Lin, K., and Li, L. (1997) *Protein Expression Purif.* **10**, 320–324
21. Jaffe, E. K., Abrams, W. R., Kaempfen, H. X., and Harris, K. A., Jr. (1992) *Biochemistry* **31**, 2113–2123
22. Erskine, P. T., Senior, N., Awan, S., Lambert, R., Lewis, G., Tickle, I. J., Sarwar, M., Spencer, P., Thomas, P., Warren, M. J., Shoeling-Jordan, P. M., Wood, S. P., and Cooper, J. B. (1997) *Nat. Struct. Biol.* **4**, 1025–1031
23. Dunbrack, R. L., Jr., and Cohen, F. E. (1997) *Protein Sci.* **6**, 1661–1681
24. Bower, M. J., Cohen, F. E., and Dunbrack, R. L., Jr. (1997) *J. Mol. Biol.* **267**, 1268–1282
25. Dunbrack, R. L., Jr. (1999) *Proteins Struct. Funct. Genet. Suppl.* 3, 81–87
26. Frankenberg, N., Erskine, P. T., Cooper, J. B., Shoeling-Jordan, P. M., Jahn, D., and Heinz, D. W. (1999) *J. Mol. Biol.* **289**, 591–602
27. Prodromou, C., and Pearl, L. H. (1992) *Protein Eng.* **5**, 827–829
28. Guo, G. G., Gu, M., and Ethinger, J. D. (1994) *J. Biol. Chem.* **269**, 12399–12402
29. Gross, M., Hessefort, S., and Olin, A. (1999) *J. Biol. Chem.* **274**, 3125–3134
30. Jaffe, E. K., Salowe, S. P., Chen, N. T., and DeHaven, P. A. (1984) *J. Biol. Chem.* **239**, 5032–5036
31. Jordan, P. M., and Seehra, J. S. (1986) *Methods Enzymol.* **123**, 427–434
32. Schlosser, M., and Beyersmann, D. (1987) *Biol. Chem. Hoppe-Seyler* **368**, 1469–1477
33. Jaffe, E. K. (1995) *J. Bioenerg. Biomembr.* **27**, 169–179
34. Dent, A. J., Beyersmann, D., Block, C., and Hasnain, S. S. (1990) *Biochemistry* **29**, 7222–7228
35. Bevan, D. R., Boddalender, P., and Shemin, D. (1980) *J. Biol. Chem.* **255**, 2030–2035
36. Frankenberg, N., Heinz, D., and Jahn, D. (1999) *Biochemistry* **38**, 13968–13975
37. Jaffe, E. K., Bagla, S., and Michini, P. A. (1991) *Biol. Trace Elem. Res.* **28**, 223–231
38. Simon, T. J. (1995) *Eur. J. Biochem.* **234**, 178–183
39. Zimmerman, K. K., Scholten, J. D., Huang, C. C., Fierke, C. A., and Hupe, D. J. (1998) *Protein Expression Purif.* **14**, 395–402
40. Hill, A. M., Cane, D. E., Mau, C. J. D., and West, C. A. (1996) *Arch. Biochem. Biophys.* **336**, 283–289
41. Graham, R. W., Atkinson, T., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. (1993) *Nucleic Acids Res.* **21**, 4923–4928
42. Jaffe, E. E. (1989) *Comments Inorg. Chem.* **15**, 67–83
43. Petrovich, R. M., and Jaffe, E. K. (1997) *Biochemistry* **36**, 13421–13427