Aminolevulinate Synthase: Functionally Important Residues at a Glycine Loop, a Putative Pyridoxal Phosphate Cofactor-Binding Site†

Jian Gong‡ and Gloria C. Ferreira*,§

Department of Biochemistry and Molecular Biology, College of Medicine, Institute for Biomolecular Science, and H. Lee Moffitt Cancer Center and Research Institute, University of South Florida, Tampa, Florida 33612

Received June 23, 1994; Revised Manuscript Received October 26, 1994®

ABSTRACT: 5-Aminolevulinate synthase catalyzes the first step of the heme biosynthetic pathway in nonplant higher eukaryotes. The enzyme functions as a homodimer and requires pyridoxal 5′-phosphate as its cofactor. Lysine-313 in murine erythroid aminolevulinate synthase has been identified as the residue involved in the Schiff base linkage with pyridoxal 5′-phosphate [Ferreira, G. C., Neame, P. J., & Dailey, H. A. (1993) *Protein Sci. 2*, 1959−1965]. However, other residues involved in binding and orienting the cofactor remain unknown. We studied the informational content of each residue within an 11 amino acid glycine-rich region, which we propose to be part of the phosphate-binding motif, based on amino acid sequence comparison with other pyridoxal 5′-phosphate-dependent enzymes and nucleotide-binding proteins. Partial random mutagenesis of this region in murine erythroid aminolevulinate synthase gene was followed by an efficient biological selection, using a *hemA*$^-$ *Escherichia coli* strain to recover functional unnatural enzymes. Among the total of 5444 variants produced, 283 were found to be functional. DNA sequencing results of 226 functional mutants indicated that most residues in this region contained a low informational content, being able to tolerate several other amino acid substitutions. However, three residues, namely, Arg-149, Gly-142, and Gly-144, were found to contain high informational content; Arg-149 was conserved in all of the functional mutants sequenced, while Gly-142 and Gly-144 could only tolerate alanine replacement. Two codon-specific random libraries of Arg-149, and Gly-142 and -144, respectively, were constructed to test further the stringency of these three positions. Interestingly, Arg-149 was conserved in all 74 functional variants identified from a population of 2250. Among the 29 functional mutants (out of 5965 variants) in the Gly-142 and -144 codon-specific library, G142C, G144A, G144S, and G144T substitutions were identified. The informational content analysis of this glycine-rich sequence provides a structural basis for pursuing the functional studies of aminolevulinate synthase.

5-Aminolevulinate synthase (ALAS)[1] (EC 2.3.1.37) catalyzes the condensation of glycine and succinyl-CoA to yield 5-aminolevulinic acid (ALA), CoA, and carbon dioxide (Jordan, 1991). This is the first reaction in the heme biosynthetic pathway in nonplant eukaryotes and some prokaryotes (Kikuchi et al., 1985; Jordan, 1991). The enzyme has pyridoxal 5′-phosphate (PLP) as an essential cofactor and functions as a homodimer (Jordan, 1991). It has been established that there are two isoforms of ALAS, a housekeeping form and an erythroid-specific form, which are encoded by distinct genes (Riddle et al., 1989). Significantly, the X-linked sideroblastic anemia syndrome, an erythropoietic disorder, has been associated with defects in erythroid ALAS (Cotter et al., 1992; Cox et al., 1994). In recent case reports of X-linked sideroblastic anemia patients, point mutations in erythroid ALAS were identified as the molecular defects (Cotter et al., 1992; Cox et al., 1994).

Recently, recombinant erythroid ALAS has been successfully overexpressed and subsequently purified from *Escherichia coli* with high yield (Ferreira and Dailey, 1993); this opens new avenues to study the structure and mechanism of the enzyme. An ϵ-amino group of a lysine residue has been identified to form the Schiff base linkage with PLP (Ferreira et al., 1993). However, other residues involved in binding and orienting the cofactor, or processing substrates, remain unknown. Sequence alignments of all known ALAS sequences indicate the presence of a conserved glycine-rich sequence (GAGAGG). A similar sequence motif (GXGXXG) has been proposed to constitute part of the cofactor binding site in some PLP-dependent enzymes (Weber et al., 1978; Marceau et al., 1988; Hyde et al., 1988) and in numerous nucleotide-binding proteins (Branden and Tooze, 1991; Swindells, 1993). This sequence motif forms a loop between a β-sheet and an α-helix, and since it is involved in binding of the nucleoside phosphate, it has been used as part of the fingerprint to predict dinucleotide-binding regions (Branden and Tooze, 1991; Swindells, 1993). In fact, the glycine-rich sequence is found in the region of (pyro)phosphate binding of many nucleotide-binding proteins (Saraste et al., 1990; Pai et al., 1990). Particularly, the crystal structure of *ras p21*, with the bound guanosine triphosphate, indicates that the glycine-rich loop is in close proximity to the phosphate moiety of the nucleotide (Pai et al., 1990). In addition, the glycine-rich sequence overlaps with the PLP

† This work was supported by the National Science Foundation (Grant MCB-9206574) (to G.C.F.).

* Author to whom correspondence should be addressed.

‡ American Heart Association Predoctoral Fellow, Florida Division (93GSF/9), and recipient of an IBS Summer Research Assistantship.

§ American Cancer Society Junior Faculty Awardee (JFRA-401).

® Abstract published in *Advance ACS Abstracts,* January 1, 1995.

[1] Abbreviations: ALA, 5-aminolevulinate; ALAS, 5-aminolevulinate synthase; PLP, pyridoxal 5′-phosphate; EDTA, disodium ethylenediaminetetraacetate; ATP, adenosine triphosphate; SDS, sodium dodecyl sulfate; SDS−PAGE, SDS−polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; bp, base pair(s); kb, kilobase pair(s).

0006-2960/95/0434-1678$09.00/0

cofactor in the three-dimensional structure of some PLP-dependent enzymes of known crystal structure (Weber et al., 1978; Oikonomakos et al., 1987; Hyde et al., 1988). Here, using limiting randomization, we analyze the informational content of each amino acid within the ALAS glycine-rich loop—the putative phosphate-gripper motif.

EXPERIMENTAL PROCEDURES

Materials

Restriction enzymes were obtained from New England Biolabs and Promega and were used according to the suppliers' instructions. Sequenase was from United States Biochemical Corp. T_4 DNA ligase and Klenow fragment of *E. coli* DNA polymerase were purchased from New England Biolabs. Deoxy- and dideoxynucleotide triphosphates were from United States Biochemical Corp. [α-^{35}S]-dATP was from Dupont/NEN Research Products. The oligonucleotide-directed *in vitro* mutagenesis kit was purchased from Amersham. GeneClean II and Mermaid GeneClean kits were products of Bio 101 Inc. The bicinchoninic acid protein assay reagents were obtained from Pierce Chemical Co. All other chemicals were of the highest purity available. The pKL plasmid, a derivative of pUC-19, was a gift of Dr. Keith Lunnen (New England Biolabs). The bacterial strain HU227 (Li et al., 1989) was a gift from Dr. C. S. Russell (City University of New York). The oligonucleotide primers were synthesized in the DNA Synthesis Core Laboratory, University of Florida.

Methods

Engineering of an AvrII Site in M13-ALAS. An *Avr*II site was engineered as a unique restriction enzyme site upstream of the glycine-rich loop-encoding sequence. This site was engineered by changing the codon of Arg-129 [in mature erythroid ALAS (Schoenhaut and Curtis, 1986)] from CGT to AGG. Oligonucleotide-directed mutagenesis, using M13mp18 as the cloning vector, was carried out as described in the Amersham oligonucleotide-directed *in vitro* mutagenesis kit directions. The mutagenesis primer used was 5′-TGC AAG ACC CTA GGG TG-3′. The mutation was verified by the susceptibility to *Avr*II digestion, and by DNA sequencing (Sanger et al., 1977).

Construction of Plasmid pJG3. The pJG3 plasmid contains the sequence for the mature erythroid ALAS (Schoenhaut and Curtis, 1986), with two unique restriction sites (i.e., *Avr*II and *Kpn*I) flanking the glycine-rich loop-encoding sequence. A DNA fragment encoding the engineered *Avr*II site was retrieved from the site-directed mutated M13-ALAS upon digestion with *Sal*I and *Kpn*I. This fragment was then subcloned in the original ALAS expression plasmid, pGF23 (Ferreira and Dailey, 1993), previously digested with *Sal*I and *Kpn*I. The construct was verified by its susceptibility to *Avr*II digestion, and by DNA sequencing (Sanger et al., 1977). The introduced *Avr*II site, although it led to a change in codon, did not lead to a change in the amino acid encoded.

Construction of Plasmid pGF33—An Inactive Vector for the Insertion of Random Nucleotide Sequences. The pGF33 plasmid, which expresses totally inactive ALAS, was used as the vector plasmid throughout the construction of random libraries (Figure 1). pGF33 was constructed by replacing the glycine loop encoding sequence with a non-ALAS-

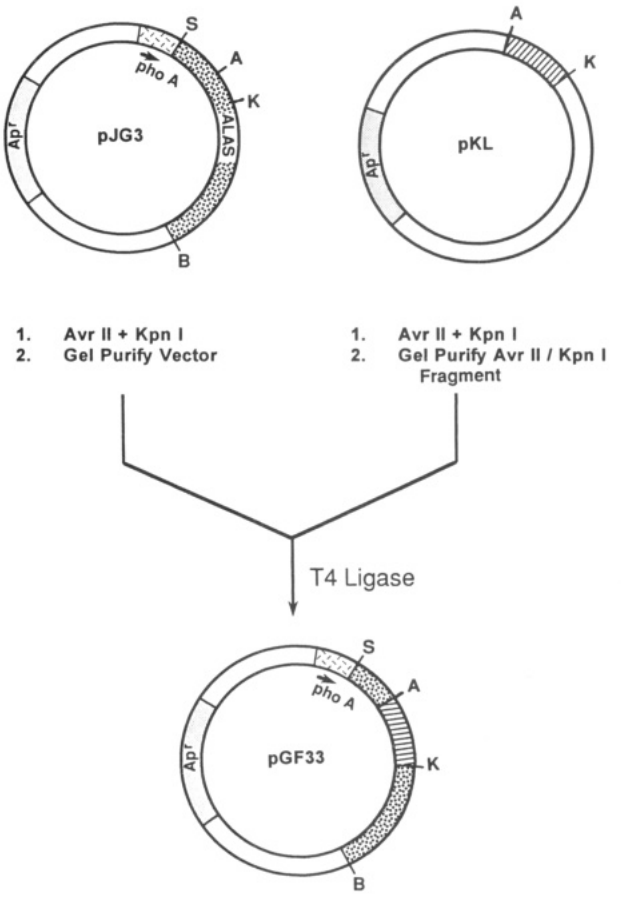

FIGURE 1: Construction of the pGF33 plasmid. Shown are the steps performed to substitute a non-ALAS-related fragment with the *Avr*II/*Kpn*I fragment in the ALAS expression plasmid (pJG3), yielding the pGF33 plasmid, which does not express a functional ALAS (see Methods for details). *phoA*, alkaline phosphatase promoter; *Ap^r*, ampicillin resistance gene; S, *Sal*I; A, *Avr*II; K, *Kpn*I; B, *Bam*HI.

encoding sequence. The pKL plasmid was digested with *Avr*II and *Kpn*I, and the generated fragment was purified from an agarose gel and ligated into the pJG3, previously digested with the same two enzymes. The construct was verified by sequencing according to the dideoxy chain termination method (Sanger et al., 1977).

Construction of Random Library. The overall strategy for construction of the 20% random mutant library was similar to that described by Munir et al. (1992) (Figure 2). Two long oligonucleotides with 12 complementary base pairs at their 3′ ends were synthesized. One of the oligomers, oligo A (58-mer), contained random nucleotides from codons 141 to 151, and a sequence for a *Kpn*I site: 5′-TGC TGG TAC CTG ANN NNN NNN NNN NNN NNN NNN NNN NNN NNN NAT TCT TCA GGG T-3′, where N = 80% wild-type nucleotides and a 20% mixture of the other three. The other oligomer, oligo B (52-mer), encoded the wild-type sequence from codons 124 to 140 and the sequence for an *Avr*II site: 5′-CAT AAG CAG ACA CCC TAG GGT CTT GCA GGC CAT AGA GGA GAC CCT GAA GAA T-3′. A 6 μL annealing reaction containing 60 pmol of each oligo was incubated at 65 °C for 2 min and then slowly cooled down to room temperature. The hybrid was then extended with 2.5 units of Klenow fragment of *E. coli* DNA polymerase I, in the presence of all four dNTPs at a final concentration of 33 μM each. The reaction was left at room temperature for 20 min and was terminated by adding EDTA

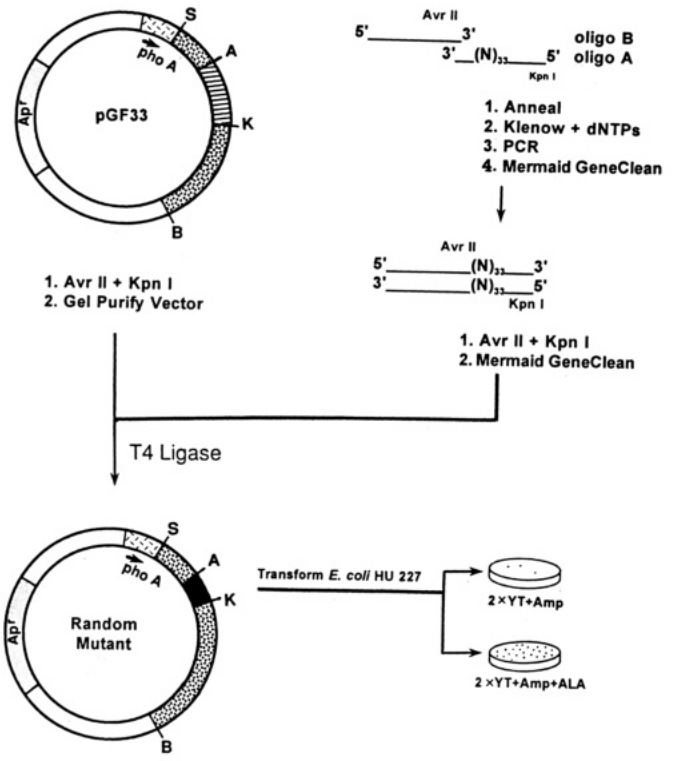

FIGURE 2: Overall strategy for construction of random nucleotide-containing library (see Methods for details). *phoA*, alkaline phosphatase promoter; *Ap*^{r}, ampicillin resistance gene; S, *Sal*I; A, *Avr*II; K, *Kpn*I; B, *Bam*HI; N, 80% wild-type nucleotides and a 20% mixture of the other three.

to a final concentration of 10 mM and heating at 75 °C for 10 min. The resulting double-stranded DNA was used as a template for the polymerase chain reaction (PCR) on an MJ Research MiniCycler, according to the manufacturer's instructions. The two synthetic primers used were as follows: primer a, 5′-CAT AAG CAG ACA CCC TAG GGT-3′, which covered the first 21 bp of the 5′-terminus of oligo A; and primer b, 5′-TGC TGG TAC CTG A-3′, which corresponded to the 13-base sequence of the 5′-terminus of oligo B. The reaction mixture (100 μL) contained 2 pmol of DNA template, 200 μM dideoxynucleotide triphosphates, 50 pmol of each primer, 2.5 units of *Taq* DNA polymerase, and reaction buffer (Boehringer Mannheim). A total of 30 cycles of 94 °C for 1 min, 34 °C for 2 min, and 72 °C for 7 min was followed by an extension of 10 min at 72 °C. After Mermaid GeneClean to remove excess primers and enzymes, the amplified DNA was digested with *Avr*II and *Kpn*I, purified by Mermaid GeneClean, and ligated into the pGF33 inactive vector, previously digested with *Avr*II and *Kpn*I. The 10−20 μL ligation reaction contained the random sequence insert and pGF33 vector fragment at 10:1 molar ratio, 1 mM ATP, and 200 units of T4 DNA ligase (New England Biolabs). After 16−20 h at 16 °C, the reaction was terminated by phenol/chloroform extraction and ethanol precipitation. The dry pellet was then dissolved in 10 μL of water and used to transform competent *E. coli* HU227 cells by electroporation, using a Gene-pulser electroporator (Bio-Rad). Competent HU227 cells (40 μL) were electroporated with 1−2 μL of ligated product at 2.5 kV, 25 μF, and 200 Ω. Immediately after electroporation, 1 mL of SOC medium (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) was added. The transformation solution was then incubated at 37 °C for 2 h with continuous shaking. One-tenth of each transformation solution was spread onto 2×YT-agar medium (1.6% Bacto tryptone, 1% Bacto yeast extract, 0.5% NaCl, 1.5% Agar) containing 50 μg/mL ampicillin and 10 μg/mL ALA, to score the total number of colonies produced. The rest of transformation solution was plated onto 2×YT-agar medium containing 50 μg/mL ampicillin, to select for active ALAS mutants. The plates were incubated at 37 °C for 16−20 h.

Construction of Codon-Specific Random Libraries: Gly Library and Arg Library. Two codon-specific libraries, namely, Gly library and Arg library, were constructed, using the same strategy as described above. To construct the Gly library, the following two oligonucleotides, oligo C and oligo D, were annealed and subsequently amplified by PCR. Both of the two oligomers contain randomized nucleotide sequences at codons 142 and 144. Oligo C: 5′-AAT CAT MMX GCT MMX GCT GGG GGC ACT CGC AAT ATC TC-3′; oligo D: 5′-GAG TGC CCC CAG CYM MAG CYM MAT GAT TCT TCA GGG TCT-3′; where M = an equimolar concentration of all four nucleotides, X = 50% T + 50% G, and Y = 50% A + 50% C. The annealed oligonucleotides were extended with Klenow polymerase, and the generated double-stranded DNA product was PCR-amplified. The 5′ end PCR primer is oligo B as described above. The 3′ end PCR primer (5′-TTG CTG GTA CCT GAG ATA TT-3′) contained the sequence for a *Kpn*I site and a sequence complimentary to the 3′ end 8 nucleotides of oligo C. To construct the Arg library, oligo B (described above) was annealed to oligo E, in which only codon 149 was randomized, 5′-TGC TGG TAC CTG AGA TAT TYM MAG TGC CCC CAG CTC CAG CTC CAT GAT TCT TCA GGG T-3′, where M = an equimolar concentration of all four nucleotides and Y = 50% A + 50% C. The annealed oligonucleotides were extended with Klenow polymerase, and the generated double-stranded fragment was PCR-amplified using primers a and b (described above).

DNA Sequencing Templates, Protein Concentration, ALAS Purification, and Activity Determination. Double-stranded DNA sequencing templates were prepared from 4 mL of bacterial overnight culture by alkaline lysis method (Sambrook et al., 1989). The DNA sequence was determined by the dideoxynucleotide chain-termination method (Sanger et al., 1977). The primer used was 5′-AAG AGA GTA GAA TCA TT-3′. Protein concentrations were determined by the bicinchoninic acid assay according to the manufacturer's instructions; bovine serum albumin was used as the standard. ALAS wild-type and selected random mutants were overexpressed in *E. coli* cells and purified by the published method (Ferreira and Dailey, 1993). ALAS enzyme activity was measured by a continuous spectrophotometric assay as described by Hunter and Ferreira (1994). Lineweaver−Burk double-reciprocal plots were employed to obtain K_m and V_{max} values. Apparent K_m and V_{max} were determined for glycine and succinyl-CoA at a constant concentration of 20 μM for succinyl-CoA and 100 mM for glycine, respectively.

RESULTS

Identification of Conserved Glycine-Rich Sequence in All Known ALASs. Comparison of the amino acid sequences of all known ALASs revealed a conserved glycine-rich sequence, ranging from bacterial to the human enzyme (141-

Mouse Erythroid ALAS	141	H	G	A	G	A	G G T R N I (Schoenhaut & Curtis, 1986)
Human Hepatic ALAS	219						(Bawden et al., 1987)
Human Erythroid ALAS	170						(Cox et al., 1991)
Rat Hepatic ALAS	219						(Srivastava et al., 1991)
Chicken Hepatic ALAS	210						(Urban-Grimal et al., 1986)
Chicken Erythroid ALAS	119						(Riddle et al., 1989)
Saccharomyces cerevisiae ALAS	121	Y	C				(Urban-Grimal et al., 1986)
Bradyrhizobium japonicum ALAS	74	V	T				(McClung et al., 1987)
Rhodobacter sphaeroides ALAS *(HemA)*	76	T		S			(Neidle & Kaplan, 1993)
Rhodobacter sphaeroides ALAS *(HemT)*	76	S		T			(Neidle & Kaplan, 1993)
Rhodobacter capsulatus ALAS	76	V		S			(Biel et al., 1988)
Agrobacterium radiobacter ALAS *(HemA)*	76	C					(Drolet & Sasarman, 1991)

FIGURE 3: Presence of a glycine-rich sequence in all known ALASs. Sequence alignment of the mouse erythroid ALAS glycine loop with other ALASs of known sequence. Amino acids that differ from mouse erythroid ALAS are shown. The mouse erythroid ALAS amino acids in bold type correspond to the conserved glycines identified in some other PLP-dependent enzymes and nucleotide-binding proteins (Marceau et al., 1988; Branden & Tooze, 1991; Swindells, 1993). Note that this sequence motif is present in the conserved C-domain (catalytic domain) of all known ALASs. The numbers indicate the position in the primary structure of the first residues shown for each sequence.

HGAGAGGTRNI in mouse erythroid ALAS) (Figure 3). Significantly, the glycine loop consensus sequence (GXGXXG) has been implicated as part of the PLP-binding site in some PLP-dependent enzymes (Weber et al., 1978; Hyde et al., 1988; Marceau et al., 1988), and as the phosphate-binding motif in nucleotide-binding proteins (Branden and Tooze, 1991; Swindells, 1993).

Construction of Random Library and Selection of Active Mutants. In order to assess the contribution of the conserved glycine loop to the proper functioning of ALAS, we employed random sequence mutagenesis to analyze the informational content of each residue in the target region. The overall scheme is illustrated in Figure 2. A continuous block of 11 codons, spanning the entire glycine loop sequence, from His-141 to Ile-151, was randomized simultaneously. With a mixture of 80% wild-type nucleotide and 20% of other nucleotides at each position, the bias is toward the wild-type sequence, and the chances for recovering active mutants are therefore increased. This approach permits the efficient determination of the spectrum of permissible substitutions within a defined region and the subsequent assessment of the importance of each residue, based on its tolerance to substitutions.

To accomplish these goals, we followed a three-step strategy. First, we engineered a unique site, *Avr*II site, at the 5′ end of the glycine loop-encoding region. By having the *Kpn*I site present at the 3′ end of the glycine loop-encoding region, the DNA sequencing effort was minimized, since the PCR-amplified regions are limited to inserts flanked by the *Avr*II and *Kpn*I sites. Second, we constructed a totally inactive vector, pGF33, as the acceptor for insertion of random nucleotide sequences. As pGF33 cannot express a functional ALAS protein, the selection of active mutants will not be affected by background from the vector. Finally, we used an efficient biological selection system to identify the active mutants. We took advantage of an *E. coli hemA*$^-$ strain, HU 227, which can only grow when ALA is added to the medium, or when the strain is transformed with an active ALAS expression plasmid.

Characterization of the Random Library. A total of 5444 transformants grew on medium containing ALA. Of these, 283 were identified as functional mutants, based on the ability to grow on ALAS selection medium (2×YT+Amp). We sequenced 226 functional mutants and 6 nonfunctional mutants. Substitutions at an average of 3 positions/mutant were found in nonfunctional mutants, whereas the number dropped to 1.2 in functional ones. Indeed, among the 226

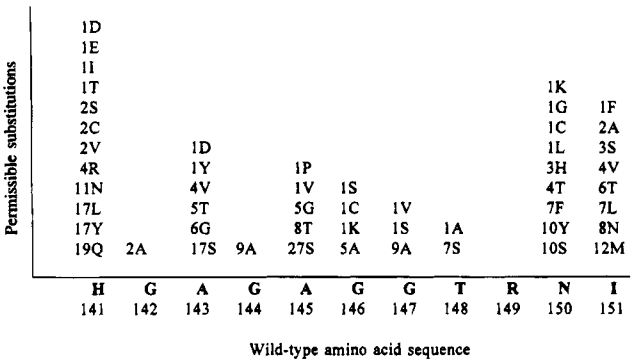

FIGURE 4: Functionally acceptable amino acid substitutions at each position in the glycine loop. Wild-type amino acid sequence is shown below the line. The number in front of each amino acid substitution indicates the total number of substitutions obtained for that particular position.

functional mutants sequenced, only 1 mutant could tolerate substitutions at 5 positions, while 2 mutants had 4 substitutions, 11 mutants had 3 substitutions, 70 mutants had 2 substitutions, 87 mutants had 1 substitution, and 55 were the same as the wild-type sequence. The total number of nucleotides mutated was 683, of which 32% were replaced by A, 27% by T, and 20% each by C and G. The distribution of substitutions for the four bases was similar, indicating that no bias was introduced and that it indeed represented random mutagenesis.

Functionally Permissible Amino Acids. Figure 4 shows the spectrum of functionally permissible amino acids at each position in the glycine loop. Most positions within this motif are quite flexible in accommodating different amino acids. Actually, 8 out of 11 wild-type residues can be substituted by amino acids with different physical properties (Figure 5). However, Arg-149 was conserved in all the functional mutants, while Gly-142 and Gly-144 could only be replaced by alanine. At the nucleotide sequence level, a total of 41 mutants had Arg codons different from the wild-type (data not shown). In contrast, among the 43 mutants with non-wild-type codons identified at Asn-150, only 5 encoded the wild-type amino acid. An average of 54 non-wild-type codons was found at each amino acid position. In nonfunctional ALAS mutants, Arg-149 was not conserved. Likewise, substitutions other than alanine were identified at Gly-142 and Gly-144 positions in nonfunctional ALAS mutants (data not shown).

Characterization of the Codon-Specific Glycine Library. To examine further the stringency of the two glycine

AA Class \ Total	78	2	34	9	42	8	11	8	0	38	43
Acidic	2	0	1	0	0	0	0	0	0	0	0
Basic	4	0	0	0	0	1	0	0	0	4	0
Polar	35	0	22	0	35	2	1	7	0	15	17
Hydrophobic	37	2	11	9	7	5	10	1	0	19	26

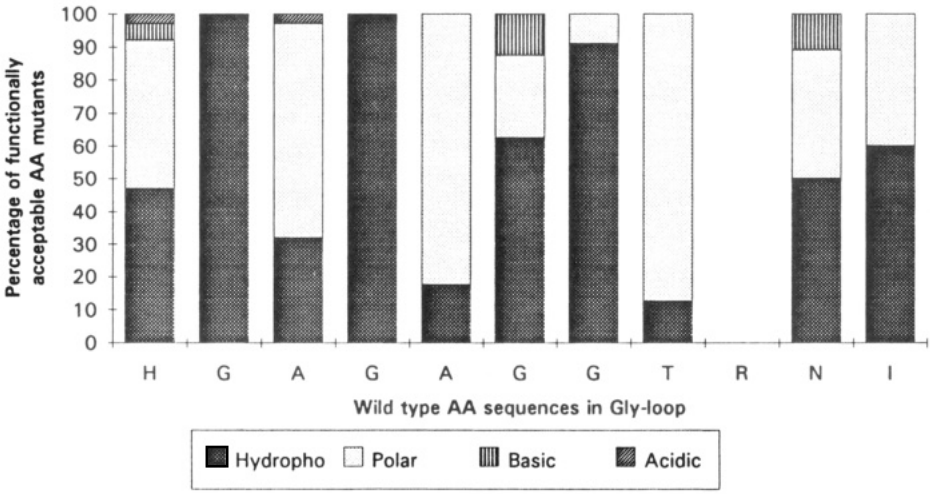

FIGURE 5: Percentage of amino acid substitutions at each position in functional glycine loop mutants. The table at the top indicates the number of substitutions with different physical properties obtained at any given position. Wild-type amino acid residues are shown on the *x*-axis. Neutral and hydrophobic amino acids include G, A, V, L, I, M, F, W, Y, and P; neutral and polar amino acids include C, S, T, N, and Q; basic amino acids include K, R, and H; and acidic amino acids include D and E.

Table 1: Mutants Obtained from Glycine Loop Codon-Specific Libraries

glycine library		arginine library	
permissible substitutions	nonpermissible substitutions	permissible substitutions	nonpermissible substitutions
			Glu
	Ser/Tyr		Gln
Cys/Gly	Asp/Asp		His
Gly/Ala	Pro/Asp		Ala
Gly/Ser	Leu/Val		Val
Gly/Thr	Ser/Val		Ser
Gly/Gly	Gly/His	Arg	Thr
Gly/Gly	Gly/Gly	Arg	Arg
142[a]/144	142/144	149	149

[a] Number indicates the sequence position in wild-type ALAS.

positions, namely, Gly-142 and Gly-144, we constructed a codon-specific glycine library containing 100% random nucleotides at these two positions and wild-type sequence at other positions. Of the 5965 colonies produced, only 29 were identified as active mutants. We sequenced 16 functional mutants and 40 nonfunctional ones. Only single point mutations were obtained in functional mutants, whereas most nonfunctional mutants had the two glycines changed simultaneously (Table 1). The overall stringency of the two glycine positions was less than that obtained in the random library. Gly-142 could be changed to cysteines and the mutants remained functional. Similarly, serine, threonine, and alanine were permissible substitutions for Gly-144.

Characterization of the Codon-Specific Arginine Library. A codon-specific arginine library with 100% randomization of only Arg-149 was constructed to challenge the stringency of this residue. From a total of 2250 mutants, 82 were able to grow on the ALAS selection medium. Sequencing results of 74 functional mutants revealed that Arg-149 was conserved in all of these mutants (Table 1), although Arg-149 in the different mutants could be coded by different codons (data not shown). We also sequenced 39 nonfunctional mutants and identified a list of nonpermissible amino acids with different physical properties (including hydrophobic, polar, basic, and acidic residues) (Table 1).

Enzymatic Activity of Wild-Type and Selected Mutant ALAS Proteins. The enzymatic activity of the mutant ALAS proteins was compared to that of the wild-type ALAS (Table 2). Both the parent vector pGF33 and the nonfunctional mutant R110 exhibited no detectable activity, while all of the functional mutants tested showed decreased activity, when compared to that of the wild-type. All the functional mutants displayed a decreased apparent V_{max} , from 6−83-fold lower than that of the wild-type value. The apparent K_m for glycine was not changed for R101, but was increased 4−10-fold for the other four selected functional mutants. As to the apparent K_m for succinyl-CoA, most mutants showed a slightly higher value, from 1.4−6-fold of that of the wild-type; while a 21-fold increase was detected with R303. The results showed no correlation between enzymatic activity and the number of substitutions in the mutants, indicating that some substitutions might exhibit more impact on enzyme function than others or that possible combinatorial effects among multisubstitutions might compensate for the decrease in enzymatic activity.

DISCUSSION

ALAS has been known to be a PLP-dependent enzyme for more than thirty years (Neuberger, 1961); however, with the exception of the newly identified lysine residue involved in the Schiff base linkage with PLP (Ferreira et al.,1993), active site residues involved in binding and orienting the cofactor are still largely unknown. This is the first report

Table 2

(A) ALAS Activity of Selected Glycine Loop Mutants[a]

enzyme	app K_m^{Gly} (mM)	app K_m^{Suc} (μM)	app V_{max} (nmol mg^{-1} h^{-1})
wild-type ALAS	12.3	1.15	2.1×10^4
R101	11.6	1.63	2.5×10^2
R284	106.8	1.92	3.7×10^3
R303	100.7	24.3	1.7×10^3
R408	118.9	3.78	9.5×10^2
R500	48.7	6.68	6.7×10^2
R110	ND	ND	ND
pGF33	ND	ND	ND

(B) Glycine Loop Sequence of the Selected Mutants

R284					A							
R500			A				S					
R408		L					S		A			
R303		Y		V			T					L
R101				D				K	V		K	F
R110				V	Q					S		L
WT	141-	H	G	A	G	A	G	G	T	R	N	I

[a] Abbreviations: app K_m^{Gly} apparent K_m for glycine; app K_m^{Suc}, apparent K_m for succinyl-CoA; app V_{max}, apparent V_{max}; ND, not detectable. R101, R284, R303, R408, and R500: selected functional glycine loop mutants; R110: nonfunctional glycine loop mutant; pGF33: inactive vector.

on the analysis of the informational content of a conserved glycine loop sequence, which is probably present at the cofactor binding site of ALAS.

Comparison of all known ALAS sequences indicated that a glycine-rich sequence, GAGAGG, is a conserved motif in all ALASs (Figure 3). Significantly, a consensus sequence of GXGXXG, where X can be any residue, has been found in other PLP-dependent enzymes (Weber et al., 1978; Hyde et al., 1988; Marceau et al., 1988) and in some nucleotide-binding proteins (Scrutton et al., 1990; Branden and Tooze, 1991; Swindells, 1993). This motif constitutes a tight turn that connects a β-sheet to an α-helix (Branden and Tooze, 1991). The first glycine residue is essential for the tightness of the turn (Scrutton et al., 1990). In many nucleotide-binding proteins, this highly conserved glycine loop is in close proximity to the phosphate group of the bound nucleotide cofactor or substrate (Wierenga et al., 1983, 1985). In two PLP enzymes of known crystal structures, this motif has been found in the active site and to be juxtaposed with the cofactor (Weber et al., 1978; Oikonomakos et al., 1987; Hyde et al., 1988). In *E. coli* D-serine dehydratase, a PLP-dependent enzyme, natural variants with point mutations in the glycine loop sequence have been identified (Marceau et al., 1988). With the first, or the second, glycine residue mutated to aspartate, the variants became catalytically inactive (Marceau et al., 1988).

All of these observations prompted us to investigate further the glycine loop presented in ALASs. In addition, recently it was suggested that the glycine-rich sequence ALAS might be involved in PLP or CoA binding (Neidle and Kaplan, 1993). We replaced the entire glycine loop (11 codons) with 20% random nucleotide sequences and screened, from a large population of random mutants, to select those biologically functional sequences. DNA sequencing results permitted us to establish the spectrum of functionally permissible substitutions at each position in the glycine loop (Figure 4). We found that eight out of the eleven amino acid residues could be replaced by more than one amino acid with different physical properties (Figure 5), indicating a high degree of plasticity in the substitutions of the wild-type amino acids, and suggesting that these residues are not crucial for the proper function of ALAS.

However, three positions in the glycine loop, namely, codons 142, 144, and 149, were found to contain high informational content. Arg-149 was conserved, while Gly-142 and Gly-144 could tolerate only alanine substitutions. Two codon-specific libraries with 100% randomization at codons 142 and 144, and codon 149, were then constructed to challenge the stringency of these three positions. Since 100% randomization was introduced to each of these three positions, unlike the case in 20% random library, there is no bias toward the wild-type sequence in the codon-specific libraries; therefore, the probability is much higher to recover all the functionally permissible substitutions from a reasonable size of population. As a result, in the codon-specific glycine library, the two glycine positions became less stringent, and G142C, G144S, G144T, and G144A mutants were identified to be functionally active. However, in the codon-specific arginine library, arginine continued to be the only residue present in all functional ALAS construct; it could be encoded by three different Arg codons. In the nonfunctional arginine library mutants, Arg-149 was replaced by many different amino acids (Table 1).

Recently, a computer program was used to analyze the glycine loop sequences in nucleotide-binding proteins (Swindells, 1993). It revealed that only the first glycine is completely conserved, while the second two glycine positions can be replaced by alanine as well as serine substitutions (Swindells, 1993). This result is consistent with our findings in that glycine residues in this loop structure may be preferred rather than absolutely required. Indeed, although the G144A mutant was still functional, its apparent maximum enzymatic activity dropped to 18% of that of the wild-type (Table 2), suggesting the structure of the mutant was no longer optimal for the proper function of the enzyme. However, the precise role of the glycine loop in the ALAS active site remains unclear. The selected glycine loop functional mutants displayed lower V_{max} values and higher K_m values for the substrates, as compared to those of the wild-type ALAS (Table 2). This indicates that both the catalysis and the binding of substrates were affected in the mutants, suggesting that perturbations in the glycine loop may induce conformational changes and/or alterations of cofactor orientation, which affect residues involved in both catalysis and substrate binding.

Crystallographic studies on other PLP-dependent enzymes provide some clues as to the nature of the interactions of the glycine loop and the cofactor (Weber et al., 1978; Oikonomakos et al., 1987; Hyde et al., 1988). In the PLP-binding site of glycogen phosphorylase *b*, there is a glycine loop with the sequence 130-GLGNGGLG, that associates with the phosphate group of the cofactor through hydrogen bonding and multiple van der Waals contacts. The main chain NH of Gly-135 is linked to the second oxygen of the PLP phosphate group (O22) via a water; while Gly-135 also makes van der Waals interactions with C5′ of PLP (Oikonomakos et al., 1987). Gly-134 makes van der Waals contacts with N1 and C6 of PLP (Oikonomakos et al., 1987). A similar glycine-rich sequence, 232-GGGSNAIG, exists in the

cofactor-binding region of tryptophan synthase, where the phosphate group of PLP forms hydrogen bonds with the main chains of Gly-232, Gly-233, Gly-234, Ser-235, and Ala-237, and with the side chains of Ser-235 and Asn-236 (Hyde et al., 1988). The authors also pointed out that there is a highly conserved glycine rich region in several threonine dehydratases as identified by Datta et al. (1987), and all these may suggest that glycine loop is a common PLP phosphate-binding motif in some PLP-dependent enzymes (Hyde et al., 1988). The glycine loop sequence in ALAS is similar to those identified in phosphorylase and tryptophan synthase, and it would be of interest to know whether Gly-142 and Gly-144 in ALAS play analogous roles to those of Gly-134 and Gly-135 in phosphorylase and Gly-232, Gly-233, and Gly-234 in tryptophan synthase. But, this will have to wait until the determination of the ALAS three-dimensional structure.

Arg-149 did not tolerate any other substitutions in all the functional mutants identified. The high informational content of this residue strongly suggests that it has an important functional role. Out of the 2250 mutants of the codon-specific arginine library, lysine was not found as a functional substitution. This suggests that either a positive charge is not enough or the lysine substitute is not protonated at the active site. Although there is not an analogous arginine residue in the glycine loop sequences in other PLP enzymes of known structure, crystal structures of several other PLP enzymes have indicated the presence of arginine residues in the active sites (Ford et al., 1980; Oikonomakos et al., 1987; Toney et al., 1993; Almo et al., 1994). In glycogen phosphorylase *b*, Arg-569 and Arg-649 are linked via hydrogen bonds and van der Waals contacts to the PLP phosphate group, N1, and C2′ atoms (Oikonomakos et al., 1987). Also the guanidinium groups of Arg-386 and Arg-292 protrude toward the cofactor in aspartate aminotransferase (Ford et al., 1980). Arginine residues are also reported to bind the carboxylate group of the substrates in some structurally available PLP-dependent enzymes (Ford et al., 1980; Toney et al., 1993; Almo et al., 1994). Arg-406 is involved in binding the L-isovaline carboxylate group in dialkylglycine decarboxylase (Toney et al., 1993), while Arg-292 and Arg-386 interact with the substrate at the ω- and α-carboxyl group, respectively, in aspartate aminotransferase (Almo et al., 1994). In addition, evidence from chemical modification and site-directed mutagenesis studies has implicated the involvement of arginine residues in the active site of several PLP-dependent enzymes (Kazarinoff and Snell, 1977; Cheung and Fonda, 1979; Tunnicliff, 1980; Schnackerz and Snell, 1983; Tanizawa and Miles, 1983; Marceau et al., 1989). In *E. coli* D-serine dehydratase, a single arginine residue near the active site was shown to be important for both the cofactor affinity and catalytic activity; however, this residue does not exert these effects through binding with the substrate or the cofactor (Marceau et al., 1989). Whether Arg-149 contributes to the catalysis and/or cofactor or substrate binding in mouse erythroid ALAS still needs further investigation.

In summary, through sequence alignments of all known ALASs, a conserved glycine-rich sequence has been identified. This motif might constitute part of the cofactor binding site as it does in some other PLP-dependent enzymes and numerous nucleotide-binding proteins. We studied the essentiality of each residue within this motif. While most positions in the glycine loop could accommodate different functionally permissible substitutions, the first two glycines and the arginine, namely, Gly-142, Gly-144, and Arg-149 in mouse erythroid ALAS, were found to contain high informational content. Future work is targeted to investigate the relation of these residues to the proper function of ALAS.

ACKNOWLEDGMENT

We thank Dr. P. Gunaratne, Mr. D. Tan, and Mrs. U. Vajapey for discussions and suggestions throughout this work.

REFERENCES

Almo, S. C., Smith, D. L., Danishefsky, A. T., & Ringe, D. (1994) *Protein Eng. 7*, 405−412.

Bawden, M. J., Borthwick, I. A., Healy, H. M., Morris, C. P., May, B. K., & Elliott, W. H. (1987) *Nucleic Acids Res. 1*, 8563.

Biel, S. W., Wright, M. S., & Biel, A. J. (1988) *J. Bacteriol. 170*, 4832−4834.

Branden, C., & Tooze, J. (1991) in *Introduction to Protein Structure* (Branden, C., & Tooze, J., Eds.) pp 141−159, Garland, New York.

Cheung, S. T., & Fonda, M. (1979) *Arch. Biochem. Biophys. 198*, 541−547.

Cotter, P. D., Baumann, M., & Bishop, D. F. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 4028−4032.

Cox, T. C., Bawden, M. J., Martin, A., & May, B. K. (1991) *EMBO J. 10*, 1891−1902.

Cox, T. C., Bottomley, S. S., Wiley, J. S., Bawden, M. J., Matthews, C. S., & May, B. K. (1994) *N. Engl. J. Med. 330*, 675−679.

Datta, P., Goss T. J., Omnaas J. R., & Patil R. V. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 393−397.

Drolet, M., & Sasarman, A. (1991) *Mol. Gen. Genet. 226*, 250−256.

Ferreira, G. C., & Dailey, H. A. (1993) *J. Biol. Chem. 268*, 584−590.

Ferreira, G. C., Neame, P. J., & Dailey, H. A. (1993) *Protein Sci. 2*, 195−1965.

Ford, G. C., Eichele, G., & Jansonius, J. N. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2559−2563.

Hunter, G. A., & Ferreira, G. C. (1994) *Anal. Biochem.* (submitted for publication).

Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W., & Davies, D. R. (1988) *J. Biol. Chem. 263*, 17857−17871.

Jordan, P. M. (1991) in *Biosynthesis of Tetrapyrroles* (Jordan, P. M., Ed.) pp 1−66, Elsevier, Amsterdam.

Kazarinoff, M. N., & Snell, E. E. (1977) *J. Biol. Chem. 252*, 7598−7602.

Kikuchi, G., Kumar, A., Talmage, D., & Shemin, D. (1985) *J. Biol. Chem. 223*, 1214−1219.

Li, J.-M., Brathwaite, O., Cosloy, S., & Russell, C. S. (1989) *J. Bacteriol. 171*, 2547−2552.

Marceau, M., McFall, E., Lewis, S. D., & Shafer, J. A. (1988) *J. Biol. Chem. 263*, 16926−16933.

Marceau, M., Lewis, S. D., Kojiro, C. L., & Shafer, J. A. (1989) *J. Biol. Chem. 264*, 2753−2757.

McClung, C. R., Somerville, J. E., Guerinot, M. L., & Chelm, B. K. (1987) *Gene 54*, 133−139.

Munir, K. M., French, D. C., Dube, D. K., & Lawrence, A. L. (1992) *J. Biol. Chem. 267*, 6584−6589.

Neidle, E. L., & Kaplan, S. (1993) *J. Bacteriol. 175*, 2292−2303.

Neuberger, A. (1961) *Biochem. J. 78*, 1−10.

Oikonomakos, N. G., Johnson, L. N., Acharya, K. R., Stuart, D. I., Barford, D., Hajdu, J., Varvill, K. M., Melpidou, A. E., Papageorgiou, T., Graves, D. J., & Palm, D. (1987) *Biochemistry 26*, 8381−8389.

Pai, E. F., Krengel, U., Petsko, G., Goody, R. S., Kabsch, W., & Wittinghofer, A. (1990) *EMBO J. 9*, 2351−2359.

Riddle, R. D., Yamamoto, M., & Engel, J. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 792−796.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, pp 1.25−1.32, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
Saraste, M., Sibbald, P. R., & Wittinghofer, A. (1990) *Trends Biochem. Sci. 15*, 430−434.
Schnackerz, K. D., & Snell, E. E. (1983) *J. Biol. Chem. 258*, 4839−4841.
Schoenhaut, D. S., & Curtis, P. J. (1986) *Gene (Amsterdam) 48*, 55−63.
Scrutton, N. S., Berry, A., & Perham, R. N. (1990) *Nature 343*, 38−43.
Srivastava, G., Borthwick, I. A., Maguire, D. J., Elferink, C. J., Bawden, M. J., Mercer, J. F. B., & May, B. K. (1991) *J. Biol. Chem. 263*, 5202−5209.
Swindells, M. B. (1993) *Protein Sci. 2*, 2146−2153.
Tanizawa, K., & Miles, E. W. (1983) *Biochemistry 22*, 3594−3603.
Toney, M. D., Hohenester, E., Cowan, S. W., & Jansonius, J. N. (1993) *Science 261*, 756−759.
Tunnicliff, G. (1980) *Biochem. Biophys. Res. Commun. 97*, 160−165.
Urban-Grimal, D., Volland, C., Garnier, T., Dehoux, P., & Labbe-Bois, R. (1986) *Eur. J. Biochem. 156*, 511−519.
Weber, I. T., Johnson, L. N., Wilson, K. S., Yeates, D. G. R., Wild, D. L., & Jenkins, J. A. (1978) *Nature 274*, 433−437.
Wierenga, R. K., Drenth, J., & Schulz, G. E. (1983) *J. Mol. Biol. 167*, 725−739.
Wierenga, R. K., De Maeyer, M. C. H., & Hol, G. J. (1985) *Biochemistry 24*, 1346−1357.

BI9413903