A Rapid Screen of Active Site Mutants in Glycinamide Ribonucleotide Transformylase†

Mark S. Warren, Ariane E. Marolewski, and Stephen J. Benkovic*

Department of Chemistry, 152 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802

Received December 5, 1995; Revised Manuscript Received March 13, 1996⊗

ABSTRACT: Specific and saturation site-directed mutageneses have been used to alter each polar residue within 6 Å of the catalytic center of glycinamide ribonucleotide transformylase (EC 2.1.2.2). These mutants were rapidly screened for catalytic activity using functional complementation of auxotrophic cells. This screen allows a rapid qualitative estimate of enzyme activity for each of these mutants. These results have shown that none of the polar residues close to the catalytic center of the enzyme are irreplaceable, although several are important for full catalytic activity, namely, Asn106, His108, Ser135, and Asp144. A mechanism is proposed in which a fixed water molecule mediates the required proton transfers between substrate and cofactor, while the formyl group is transferred from 10-formyltetrahydrofolate by direct nucleophilic attack by the amine of glycinamide ribonucleotide. The active site polar residues may act to alter the pK_a values of the attacking and leaving amino groups within a putative tetrahedral intermediate in order to facilitate the transfer of the formyl group.

Glycinamide ribonucleotide transformylase (GAR transformylase,[1] EC 2.1.2.2) is the third enzyme in the *de novo* purine biosynthetic pathway. This enzyme catalyzes the conversion of β-glycinamide ribonucleotide (GAR) to formyl-β-glycinamide ribonucleotide (fGAR) and is the first of two formyl transfer reactions in the *de novo* purine biosynthetic pathway. Two enzymes possessing this activity are known to exist in *Escherichia coli*. The first of these to have been characterized is the *E. coli pur*N gene product (Inglese et al., 1990a). The *pur*N transformylase is a monomer of molecular weight 23 241 that uses 10-formyltetrahydrofolate as a cofactor. Recently, a second GAR transformylase using formate and ATP in place of 10-formyltetrahydrofolate and encoded by the *E. coli pur*T gene has been described (Marolewski et al., 1994; Nygaard et al., 1993). While little is known about the structure of the *pur*T enzyme, high-resolution X-ray crystal structures of the *pur*N GAR transformylase have been reported (Chen et al., 1992; Almassy et al., 1992; Klein et al., 1995). These structures have revealed several amino acids that could be involved in substrate binding and catalysis. Although the crystal structures have identified which residues are available to participate in the catalytic mechanism, they do not determine whether these residues are essential. In order to assess the importance of the individual active site residues, we have used site-directed mutagenesis to alter the side chains of every conserved polar amino acid within 6 Å of the catalytic center of the enzyme.

Although mutagenesis techniques such as "alanine scanning mutagenesis" (Cunningham & Wells, 1989) are certainly useful to identify residues that are essential for catalytic function, the limitation to a single mutation at any given residue may give misleading results about its mechanistic importance (Santi et al., 1990). In order to study thoroughly each active site residue, we have used saturation site-directed mutagenesis to create amino acid replacement sets at three positions (N106, H108, D144) while also constructing more constrained alterations of three additional residues (H119, S135, H137). In order to screen rapidly the numerous mutants for catalytic activity, we have applied functional complementation in auxotrophic cells as a screening technique.

MATERIALS AND METHODS

Materials. Restriction enzymes were obtained from Promega, New England Biolabs, and Boehringer Mannheim. T4 DNA ligase was purchased from Boehringer Mannheim. *Taq* polymerase and nucleotides were obtained from Perkin-Elmer Cetus, Boehringer Mannheim, and Promega. The Sequenase 2.0 DNA sequencing kit was purchased from U.S. Biochemical Corporation. Agarose for gel electrophoresis was obtained from Kodak, Ultrafree-MC filter units were from Millipore, and electroporation cuvettes and the gene pulser electroporator were from Bio-Rad. QIA-tip plasmid purification columns were obtained from QIAGEN and Wizard Prep resin and columns from Promega. Noble agar was purchased from Difco. [α-^{35}S]dATP (1000−1500 Ci/mmol) was purchased from New England Nuclear. Synthetic oligonucleotides were obtained from Operon, Midland, or Integrated DNA Technologies. The cofactor 10-formyl-5,8-dideazafolate (fDDF) was purchased from Dr. John Hynes, Medical University of South Carolina. All common buffers and reagents were obtained from Fisher, Sigma, or Aldrich.

Bacterial Strains and Plasmids. The *E. coli* strains used were DH5α (Gibco BRL Life Technologies), BL21(DE3) (Novagen), and TX680 [*ara* Δ(*gpt-pro-lac*) *thi rbs-221 ilvB2102 ilvHI2202 purN′-lacZ*$^+$*Y*$^+$*::*Kan^R *purT*], a generous

† This work was supported by USPHS Grant GM24129 from the National Institutes of Health (S.J.B.).

* To whom correspondence should be addressed. Phone 814-865-2882. FAX 814-865-2973.

⊗ Abstract published in *Advance ACS Abstracts,* June 15, 1996.

[1] Abbreviations: BW1476U89, Burroughs-Wellcome multisubstrate inhibitor; *E. coli*, *Escherichia coli*; fDDF, 10-formyl-5,8-dideazafolate; fGAR, formyl-β-glycinamide ribonucleotide; GAR, β-glycinamide ribonucleotide; GAR transformylase, glycinamide ribonucleotide transformylase; IPTG, isopropyl β-D-thiogalactoside; LB, Luria broth; LD, loop deletion (residues 118-122) mutant; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate.

S0006-2960(95)02871-6 CCC: $12.00

gift from Dr. John M. Smith (R&D Systems, Minneapolis, MN). The λDE3 lysogenization kit from Novagen was used to integrate the T7 polymerase gene into the TX680 host cell chromosome. Two TX680(DE3) colonies were selected and designated MW11 and MW12. These two strains exhibit slightly different properties, with the MW11 cells having slightly faster growth rates, while the MW12 cells generally allow slightly higher levels of induced protein expression. The extent of the differences between these two cell lines was not determined. Bacteria containing plasmids were grown at 37 °C in LB media containing 100 μg/mL ampicillin. Plasmids were purified on a small scale using Promega Wizard Prep columns. Larger scale plasmid purifications were performed with QIA-tip 100 columns, resulting in typical yields of 80 μg of plasmid DNA.

Primers. Synthetic oligonucleotide primers were designed to have a melting temperature of ≥52 °C. Forward and reverse mutagenic primers were 21−27 bases long, with the mutagenic codon in the middle of each primer. For saturation mutagenesis, the NNS codon (N = a mixture of A + C + G +T; S = a mixture of C + G only) was used at the mutational site. The 32-fold degenerate NNS sequence contains the codons for all 20 amino acids as well as the amber stop codon, TAG.

Mutagenesis. In vitro site-directed mutagenesis was performed using the PCR overlap extension method (Ho et al., 1989). One reaction used the mutagenic forward primer and downstream external pUC primer; a second reaction used the mutagenic reverse primer and upstream external pUC primer. PCR was carried out using 1 μg of plasmid template DNA, 0.2 mM each dNTP, 1× PCR buffer, 1.5 mM $MgCl_2$, 2 μM each primer, and 2.5 units of *Taq* polymerase in a total volume of 100 μL. PCR consisted of 30 cycles using 1 min of denaturing at 94 °C, followed by annealing at 52 °C for 2 min and extension at 72 °C for 3 min. The two PCR products were precipitated with ethanol, resuspended in TE (10 mM Tris, 1 mM EDTA, pH 8.0), and loaded onto a 1% agarose gel. Gel fragments containing the desired DNA fragments were excised. DNA was extracted using Millipore Ultrafree-MC filter units, ethanol precipitated, and resuspended in TE. Overlap extension PCR was carried out using the first two PCR products as templates under identical conditions except that the reaction was denatured at 94 °C for 1−5 min prior to addition of enzyme (hot start PCR). The overlap extension product was also gel purified using Millipore Ultrafree-MC filter units. After purification, the final PCR product and vector pJS167 (Inglese et al., 1990a) were digested for 1−10 h with 12 units of *Eco*RI and 20 units of *Pst*I. Digested fragments were run on a 1% agarose gel and purified as described. The mutagenic insert (0.8 kb) and vector fragment (3.0 kb) were ligated using T4 DNA ligase at 16 °C overnight. Ligation mixtures were then transformed into competent DH5α cells either by electroporation or by using $CaCl_2$ transformation procedures (Sambrook et al., 1989). Transformed cells were then plated on LB plates containing 100 μg/mL ampicillin. Plasmid DNA was purified from transformants using Promega Wizard Prep columns or by the alkali lysis method (Sambrook et al., 1989). Mutant candidates were screened by sequencing. Nearly a 100% efficiency of mutagenesis was obtained using these methods.

Sequencing. Double-stranded DNA sequencing was performed by the dideoxy chain termination method using the Sequenase 2.0 kit. About 5 μg of plasmid DNA from DH5α cells was denatured in 0.2 M NaOH in a total volume of 20 μL for 5 min. To this were added 8 μL of 3 M NaOAc, pH 5.7, and 80 μL of ethanol simultaneously, with the tube placed into a dry ice/ethanol bath immediately to precipitate the denatured plasmid DNA. After a 15 min precipitation and 15 min centrifugation step, the supernatant was removed and the pellet allowed to dry briefly. Sequencing primer (2 pmol) in 10 μL of 1× Sequenase buffer was added to the denatured plasmid, with the tube being placed immediately into a 55 °C heat block and allowed to cool to <35 °C over 30 min. Synthetic oligonucleotides designed for mutagenesis throughout the gene as well as M13 and pUC primers were utilized as sequencing primers. After annealing, sequencing reactions were performed as described in the Sequenase manual of U.S. Biochemical Corp., except that termination reactions were incubated at 42−48 °C for 5−15 min rather than 37 °C for 5 min, and pyrophosphatase (0.0012 unit) was added in addition to the 3.25 units of Sequenase enzyme. Both strands of the gene were sequenced. Several of the mutations increased the CG content in regions of the gene already quite rich in CG base pairs. Sequencing of these often revealed compressions and other secondary structure artifacts created by the high CG content. We resolved these artifacts using higher than suggested temperatures for the termination reactions (up to 48 °C) and by using dITP and 7-deazaGTP reactions in addition to the dGTP reactions in the Sequenase kit. All mutant clones were sequenced throughout the entire gene encoding region as well as the promoter region to verify the absence of additional mutations.

Screening. Mutant plasmids were transformed into TX680 cells using $CaCl_2$ procedures (Sambrook et al., 1989). Electroporation of TX680 cells resulted in poor transformation efficiencies ($<10^4$/μg of plasmid), while $CaCl_2$ transformations usually resulted in $\geq 10^6$/μg of plasmid. Cells were plated on LB containing 100 μg/mL ampicillin and 25 μg/mL kanamycin to select TX680 transformants. Colonies from these plates were streaked onto minimal selective (purine-deficient) plates containing 30 μg/mL ampicillin. The selective media contained 1× M9 salts, 1 mM $MgSO_4$, 0.2% glucose, 0.06% casamino acids, and 0.00005% thiamine. Difco Noble agar was used, since regular Difco agar was found to contain sufficient levels of purines to allow growth of auxotrophic cells. Hypoxanthine at 50 μg/mL served as a source of purines on positive control plates. Plates were incubated at 30 and 37 °C for varying lengths of time (up to 7 days), and cell growth was assessed visually.

Protein Preparations. Mutant GAR transformylase proteins were prepared using one of two procedures, depending on the expression vector used. Mutant genes in the pJS167 vector were expressed using a temperature induction from the λ_{PL} promoter. These cells were grown in LB media with 100 μg/mL ampicillin at 32 °C up to an OD_{600} of 0.6. Protein production was induced at 42 °C for 10 h, and cells were harvested by centrifugation. Mutant genes cloned into the pMSW2 vector were expressed in MW11 or MW12 cells using an IPTG induction of the T7 promoter. These cells were grown in LB media with 100 μg/mL ampicillin at 37 °C up to an OD_{600} of 0.6. Protein production was induced by the addition of 0.5 mM IPTG, followed by an additional 2−4 h growth at either 30 or 37 °C. Cells were harvested by centrifugation.

Protein Purification. Cells were resuspended in a minimal volume of 50 mM Tris-HCl and 1 mM EDTA, pH 7.5 (total volume not exceeding 100 mL), and lysed by sonication. Lysate was cleared by centrifugation at 40000*g* for 40 min. Supernatant was loaded onto a 5 cm × 120 cm Sephadex G-75 column equilibrated in 50 mM Tris-HCl and 1 mM EDTA, pH 7.5. This column was run at 1 mL/min while collecting 20 mL fractions. Fractions containing GAR transformylase were pooled, and the protein concentration was estimated using the extinction coefficient at 280 nm of 22.1 mM^{-1} cm^{-1} at pH 7.5 for GAR transformylase (Inglese et al., 1990a). Partially purified protein was then loaded in ≤50 mg injections onto a MonoQ HR 10/10 column equilibrated in the same buffer. A KCl gradient of 0−0.5 M in this buffer was applied, with elution peaks monitored at 280 nm. The mutant GAR transformylases eluted between 0.09 and 0.18 M KCl. The purity of the protein was determined by SDS−PAGE. Eluted protein was dialyzed to remove salt. After concentration, glycerol was added to a final concentration of 10%, and aliquots of enzyme (approximately 3 mg/mL) were stored at −70 °C.

Enzyme Assays. GAR transformylase was assayed in 50 mM Tris at pH 7.5. The assay monitors the deformylation of 10-formyl-5,8-dideazafolate ($\Delta\epsilon$ = 18.9 mM^{-1} cm^{-1} at 295 nm) resulting from the transfer of the formyl group to GAR (Smith et al., 1981). Assays were carried out in 1 mL cuvettes thermostated to 25 °C on a Gilford 252 spectrophotometer. Enzyme concentrations varied between 1 and 10 nM, GAR between 5 and 80 μM, and 10-formyl-5,8-dideazafolate between 5 and 100 μM. Reactions were initiated with one of the substrates.

RESULTS AND DISCUSSION

GAR transformylase is a small protein consisting of 212 amino acids encoded by the *E. coli pur*N gene. Sequences from *E. coli* and five additional sources—chicken, human, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, and *Bacillus subtilis*—have revealed that 24 amino acids are strictly conserved in all six known species (Aimi et al., 1990). The recently solved crystal structures of GAR transformylase have made further efforts at elucidating the chemical mechanism possible (Chen et al., 1992; Almassy et al., 1992; Klein et al., 1995). The apoenzyme crystal structure has revealed a proposed binding pocket lined by the side chains of 13 strictly conserved residues (G11, N13, A86, G87, N106, H108, P109, V136, H137, V139, D144, E173, H174) (Chen et al., 1992). The ternary complex with substrate GAR and a folate inhibitor bound has confirmed that this region is indeed the binding pocket and has indicated that three of these residues (N106, H108, D144) are positioned such that they may assist the reaction (Almassy et al., 1992). In order to test the importance of these three amino acids, we have subjected each of them to saturation site-directed mutagenesis as well as conducting limited site-directed mutagenesis on two additional conserved residues (S135, H137) and one nonconserved residue (H119). Figure 1 shows the interactions of these conserved residues with the multisubstrate adduct inhibitor BW1476U89 (Klein et al., 1995).

Saturation site-directed mutagenesis of N106, H108, and D144 was performed using the PCR overlap method (Ho et al., 1989). The 32-fold degenerate NNS codon was used at the mutagenic site in synthetic oligonucleotides. This codon allows all 20 amino acids as well as the amber termination codon to be placed at the target site. Approximately 50 transformants from each mutagenesis were sequenced. The sequencing revealed that 14−17 different mutants of each residue had been obtained. It was not practical to identify all 20 possible mutants at each position using this strategy, as this would require sequencing of approximately 160 clones to obtain a 95% confidence level of obtaining a complete set (Climie et al., 1990). Mutagenesis efficiencies obtained were close to 100% for each experiment. For each mutant constructed, the entire gene was sequenced to verify the absence of additional mutations. This proved to be a prudent precaution, as we found 59 secondary mutations, yielding an error frequency of 6.3×10^{-4}/nucleotide or approximately 1 error for each 1600 nucleotides polymerized.

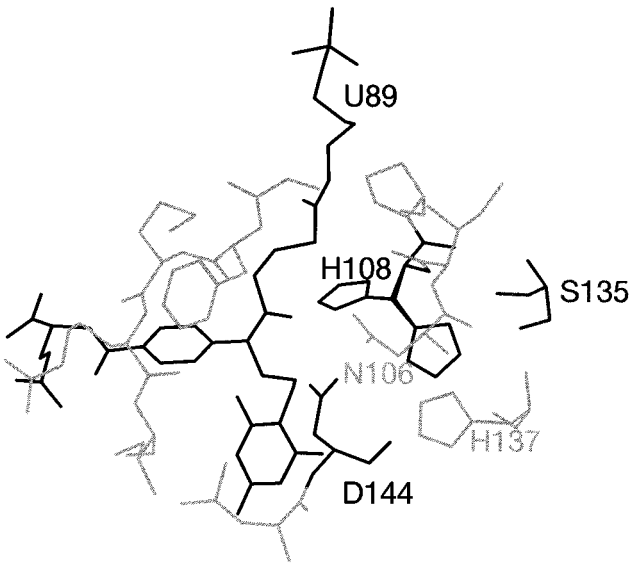

FIGURE 1: Active site of GAR transformylase. The enzyme was crystallized in a binary complex with the multisubstrate adduct inhibitor BW1476U89, labeled U89, from Klein et al. (1995). Residues shown are those within an 8 Å sphere around the formyl oxygen.

Table 1: Mutagenesis of Aspartic Acid 144

codon	frequency	codon	frequency
AAC	2	AAG	1
ACC	2	ACG	4
AGC	2	AGG	2
ATC	0	ATG	0
CAC	3	CAG	3
CCC	2	CCG	4
CGC	2	CGG	1
CTC	0	CTG	1
GAC	5	GAG	1
GCC	0	GCG	5
GGC	0	GGG	0
GTC	0	GTG	3
TAC	2	TAG	1
TCC	1	TCG	1
TGC	1	TGG	1
TTC	1	TTG	1

Data for the mutagenesis of D144 are presented in Table 1. Since the mutagenesis used a mixture of primers, a potential bias toward those containing the wild-type codon or only a single nucleotide substitution is possible during the annealing reaction. However, sequencing information revealed little preference for wild-type or single base changes. The distribution of codons obtained at the mutagenic site

Table 2: Activities of PurN Mutants

	N106	H108	H119	S135	H137	D144
A	+	−	+++	+		++
C						−
D	+					+++
E	−	+				+
F					+++	−
G	+	−				
H	−	+++	+++		+++	−
I	−	−				
K	−					−
L	−	+		+		−
M		−				
N	+++	−				−
P	−	+				−
Q		−			+++	−
R	−	−				−
S	+	+		+++		+
T	−	+				−
V						−
W	−	−				−
Y	−					+

shows a near random distribution of nucleotides at each of the first two positions and a mixture of only C and G at the third position. There is a slight bias against T at each of the first two positions, probably due to a smaller amount of this nucleotide being used in the synthesis of the oligonucleotide rather than a systematic preference for the other nucleotides during the PCR mutagenesis.

Specific site-directed mutagenesis was also utilized on three additional residues. H119 was mutated to alanine, S135 to alanine and leucine, and H137 to glutamine and phenylalanine. Sequencing of transformants from these experiments also revealed that a 100% mutagenesis efficiency had been obtained.

One additional and rather dramatic mutation constructed in GAR transformylase was the removal of a five-residue loop that folds down over the binding pocket in the presence of folate. Residues 118−122 were deleted and replaced by two serine residues. This diserine substitution was chosen in order to maintain the hydrophilicity of this loop and to bridge the gap between the loop termini adequately. This loop deletion mutant will be designated "LD" for simplicity.

One extremely efficient means of screening multiple mutants is a positive growth screen, such as functional complementation of auxotrophic cells. By transforming mutant plasmids into TX680 auxotrophic cells, only those mutants capable of some level of GAR transformylase activity will allow growth under conditions in which the bacteria must use *de novo* purine biosynthesis in order to replicate. Results from the functional complementation screen are presented in Table 2. Growth of the mutants falls into four categories: identical to wild type with growth observed in <18 h (+++); somewhat slower than wild type with growth observed in 18−36 h (++); substantially slower than wild type with 2−6 days required for observable growth (+); and no growth (−). Blank entries in Table 2 refer to mutants not obtained during sequencing. Aliquots of protein lysates from each mutant in Table 2 were analyzed by SDS−PAGE in order to verify that all mutants were producing protein. These gels showed that all mutant genes except the LD mutation were producing roughly equivalent amounts of protein (data not shown). Protein production from the LD mutant was minimal and will be discussed later.

In order to provide a more quantitative estimate as to the significance of the growth times observed for the various mutant proteins, the Michaelis−Menten parameters for select proteins in each category have been determined and are reported in Table 3. The range in k_{cat}/K_m (fDDF) for the (+++) to (−) classification is a factor of 12 000-fold. Similarly, the (+++) to (++) and (+) categories represent decreases in k_{cat}/K_m (fDDF) of 12-fold and 2000-fold. Hence, the cutoff for successful complementation of the auxotroph appears to be a difference in k_{cat}/K_m (fDDF) of somewhere between 2000- and 12 000-fold.

One conclusion that can be immediately drawn from the complementation data is that none of these residues are absolutely required for catalytic function. Each residue can be replaced by a range of substitutions leading to an active enzyme. These substitutions are not limited to conservative changes but rather encompass alterations in both size and polarity. The malleability of the GAR transformylase active site in accommodating mutations is not unprecedented among folate-requiring enzymes. Extensive mutagenesis of active site residues in thymidylate synthase and dihydrofolate reductase has shown that most active site mutants in these enzymes retain at least some catalytic activity, although some of these mutants use alternative mechanisms (Howell et al., 1986) or change the substrate specificity (Hardy & Nalivaika, 1992; Liu & Santi, 1992, 1993). It remains to be determined whether mutations at the GAR transformylase active site have had similar effects.

The required chemical steps in the GAR transformylase reaction are the transfer of a formyl group from 10-formyltetrahydrofolate to the amino group of GAR, removal of one proton from the amino group of GAR, and protonation on N^{10} of the product tetrahydrofolate. It is unknown whether these steps occur concertedly or how the enzymic groups participate in the mechanism. In order to glean mechanistic insights, mutagenic results from each position will be examined separately and then summarized within the current mechanistic hypothesis.

H137 is strictly conserved in all six known GAR transformylase sequences. Although H137 is in the substrate binding pocket, it is too far from the formyl group to directly participate in the formyl transfer. However, the crystal

Table 3: Comparison of Growth Times to Kinetic Parameters for GAR Transformylase Mutant Proteins[a]

growth category/enzyme	k_{cat} (s^{-1})	K_m (GAR)	K_m (fDDF)	k_{cat}/K_m (fDDF)
(+++)/wt	16 ± 1	19 ± 4	17 ± 3	0.94 ± 0.17
(++)/D144A	2.0 ± 0.2	29 ± 2	25 ± 3	0.08 ± 0.01
(+)/N106S	6 ± 1			
(+)/N106G	0.36 ± 0.05		900 ± 200	0.0004 ± 0.0001
(−)/H108A	0.00079 ± 0.00007	54 ± 5	10 ± 1	0.00008 ± 0.00001
(−)/H108Q	0.000031 ± 0.000003	18 ± 3	24 ± 5	0.0000013 ± 0.0000003

[a] K_m values are given in micromolar.

structures indicate that H137 could participate in the active site hydrogen-bonding network. If the chemical mechanism required the transfer of a proton to bulk solvent, H137 could be a member of a proton transfer chain leading to the protein surface. Complementation results showed that the H137F and H137Q mutants both possessed substantial catalytic activity, indistinguishable from that of the wild-type enzyme in the complementation assay. These results indicate that H137 likely does not participate in catalysis and is not required for proton transfer.

S135 is represented by the conservative substitution of threonine in most GAR transformylase sequences. The crystal structure indicates that S135 could also participate in the active site hydrogen bond network. S135A and S135L mutants both demonstrated catalytic efficiencies reduced approximately 3 orders of magnitude as compared to the wild-type enzyme in the complementation assay (Table 3). These results infer that S135 probably does contribute to the hydrogen-bonding network, most likely through assisting in the orientation of H108.

H119 is one of three histidines in proximity to the formylfolate cofactor. Crystallographic results have shown that H119 is located on a flexible loop consisting of amino acids 118−122 (Klein et al., 1995). This loop folds down over the folate, presumably securing it in place in the binding pocket. Although this residue is not conserved in other species, it has previously been implicated as possibly being involved in the catalytic mechanism. Affinity labeling experiments with N^{10}-(bromoacetyl)-5,8-dideazafolate on a D144N mutant showed a stoichiometric labeling of H119, indicating its proximity to and possible involvement in the active site (Inglese et al., 1990b). However, the H119A mutant demonstrates full activity in the complementation assay, indicating that the imidazole of H119 is not participating as a general acid−base catalyst.

In order to test the importance of the flexible loop containing H119, a loop deletion mutation (LD) was constructed in which residues 118−122 were deleted and replaced by a bridge of two serines. The LD mutant was not capable of complementing TX680 auxotrophs. This mutant did not express well in TX680, MW11, or MW12 auxotrophic cells. High levels of expression were possible only when the gene was cloned into the T7 expression plasmid pMSW2 and expressed in BL21(DE3) cells. Protein expression in these cells confirms that the LD mutant can be expressed and folded into a stable conformation. The inability of the LD mutant to complement can thus be ascribed to either minimal protein levels in the auxotrophic cells or to a lack of activity, presumably due to weakened or abolished folate binding. Indeed, preliminary experiments have shown that folate binding is substantially weaker in the LD mutant as compared to wild type.

N106 is strictly conserved in all known sequences, and its side chain is very close to the catalytic center of the enzyme (Figure 1). The backbone amide of N106 forms two hydrogen bonds with that of H137, while the side chain amide nitrogen makes hydrogen bonds with the side chain of H137 and the oxygen of the cofactor formyl group (Klein et al., 1995). Results of saturation mutagenesis at this position indicate that no mutant proteins show full activity and that only a few residues can substitute for N106 to give partially active protein. The N106D and N106S proteins can probably maintain at least some of the hydrogen bonds normally present; however, both demonstrate substantially decreased activity (ca. 2000-fold) from wild-type levels. The nearly isosteric substitutions N106L and N106I would eliminate side chain hydrogen bonds, and as expected, these proteins demonstrate no activity in our complementation assay. The N106T protein could potentially restore one hydrogen bond as compared to the N106L and N106I proteins; however, N106T is not active. Although the N106A and N106G proteins lack side chains capable of hydrogen bonding with the cofactor, both exhibit activity comparable to N106D and N106S. We believe the most likely explanation for the activity of these mutants is that one or two water molecules can fill the gap created by the loss of the N106 side chain. The active site already contains several water molecules, and hydrogen bonding may be maintained either by movement of those already present or by importing additional water molecules, as noted for dihydrofolate reductase (Howell et al., 1986). Confirmation of this rationale will have to await a crystal structure of either N106G or N106A.

D144 is another strictly conserved amino acid involved in hydrogen bonding with the formyl group of the cofactor. The crystallographic data show that the carboxyl oxygens of D144 show favorable polar interactions with N^{10} and the formyl oxygen of the cofactor as well as hydrogen bonding to the H108 side chain and a conserved water molecule (Almassy et al., 1992; Klein et al., 1995). Previous mutagenesis of this residue to asparagine resulted in an inactive enzyme (Inglese et al., 1990b), which we have confirmed using functional complementation. The conservative D144E mutation, moreover, does yield active enzyme, although its activity is reduced by approximately 3 orders of magnitude. The decreased activity of the D144E enzyme likely indicates that the carboxylic acid functionality has been somewhat disrupted from its optimal geometry. D144S and D144A are also of equivalent activity levels, likely through recruitment of water molecules. Surprisingly, the D144Y mutation is also active, despite a ΔpK_a of ca. 6 units between the phenolic side chain of tyrosine and the carboxyl of aspartate (Dawson et al., 1969). While the active site properties may alter these pK_a values from their "normal" levels, it is probable that the D144Y substitution may change the rate-limiting step in the kinetic sequence, if D144 is involved in proton transfer.

Saturation mutagenesis of the remaining strictly conserved active site residue, H108, also revealed several amino acid substitutions that lead to active proteins, although their activities are reduced by 3 orders of magnitude. Crystallographic data suggest that the H108 imidazole resides in two different conformations, as shown in Figure 1 (Klein et al., 1995). In one position, the imidazole ring nitrogens form hydrogen bonds with the side chains of D144 and S135. In the other H108 conformation, the imidazole ring nitrogens form hydrogen bonds with the side chain of D144 and the formyl oxygen of the cofactor (Figure 2). The orientation and position of the H108 side chain in this second conformation would allow it to stabilize a negative charge that develops on the formyl oxygen during the transfer process. The active proteins H108E, H108S, and H108T could potentially preserve at least some of the hydrogen-bonding interactions of the wild type. However, the H108P and H108L proteins should abolish any hydrogen-bonding interactions from this position, suggesting their activity may arise from a different active site conformation, giving rise

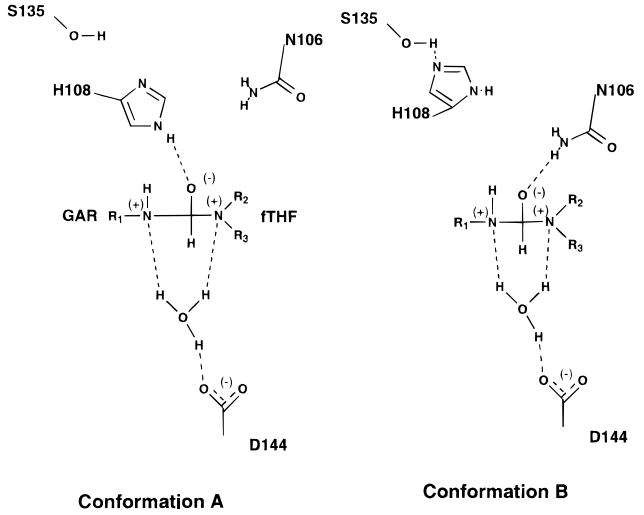

FIGURE 2: Hydrogen-bonding schemes. Two hydrogen-bonding schemes for the proposed transition state intermediate are shown corresponding to the two distinct conformations present in the BW1476U89 multisubstrate adduct inhibitor complex (Klein et al., 1995).

to a less efficient kinetic sequence that might not involve hydrogen bonding for this residue.

The overall results from the mutagenesis indicate that there are no irreplaceable residues involved in the catalytic mechanism of GAR transformylase. In every case studied, there are a range of substitutions allowed at each position. While most of the acceptable substitutions are relatively conservative changes, some are quite different in size, polarity, and potential function. These unexpected results demonstrate the importance of making multiple mutations at each position before attempting to assign a structure−function role to each residue. If we had constructed only alanine substitutions of the active site residues, or only investigated the D144N mutation, we might have falsely concluded that H108 and D144 were absolutely essential for catalysis.

Since no single residue making hydrogen bonds with the substrates is irreplaceable, it is likely that the multiple proton transfers required for catalysis may occur alternatively via water molecules rather than involving enzymic residues directly. Alternatively, some fraction of them may be eliminated without total loss of activity. We propose that these proton transfers between ligands and a water molecule are optimally assisted by residues N106, H108, and D144 and expand upon a previously proposed mechanism involving D144 alone (Inglese et al., 1990b).

In order to understand how the enzyme may facilitate catalysis, each protonation state of the putative tetrahedral intermediate will now be examined. The progression of the catalytic mechanism is presented in Figure 3. Nucleophilic attack by the amino group of GAR upon the formyl carbon of 10-formyltetrahydrofolate would lead to the initial formation of the tetrahedral species **1** that may be protonated to give **2**. The required proton switch from the amino group of GAR to the N^{10} of the cofactor probably proceeds through **3** and **4** to yield the zwitterionic **5**, which collapses, with cleavage of the formyl carbon−N^{10} bond, yielding the products formyl-GAR and tetrahydrofolate.

The pK_a estimates of these intermediates have been made using the methods of Fox and Jencks (1974). These authors have estimated acidities of alcohols and aliphatic ammonium ions using $\rho\sigma$ structure−reactivity relationships (Ritchie &

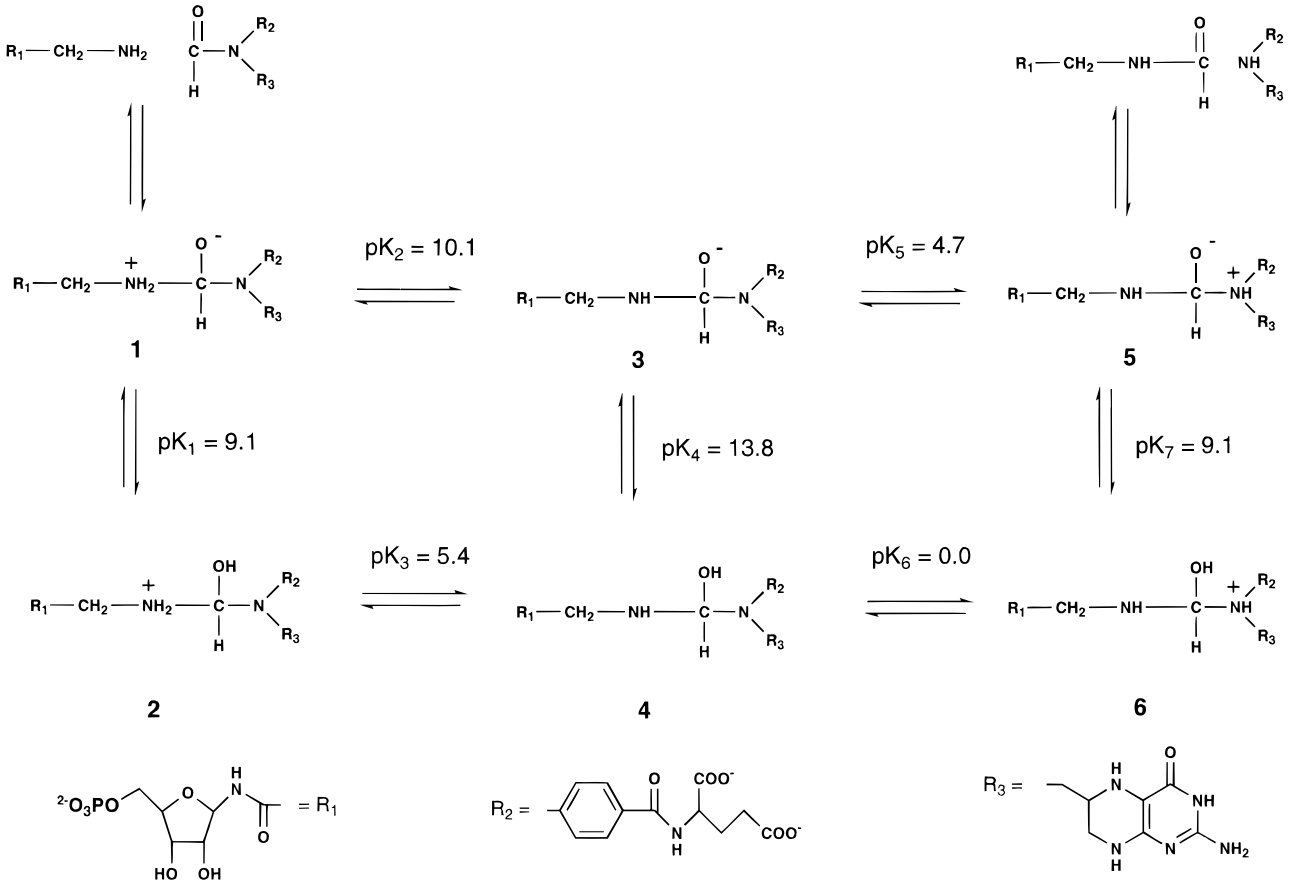

FIGURE 3: Progression of the catalytic mechanism. Estimates of pK_a values were made using the methods of Fox and Jencks (1974).

Sager, 1964). Model compounds with known pK_a values are used for base values, and the alterations caused by the substituent groups are calculated. For the dissociation of alcohols and ammonium ions, $\rho_I = -8.4 \pm 1.0$ (Fox & Jencks, 1974). The values of σ_I are 0.25 for hydroxyl groups, 0.10 for amine, and 0.25 for amide substituents, with an intermediate value of 0.18 for methoxyamine groups (Fox & Jencks, 1974). With these values available, the individual pK_a values in Figure 3 can be calculated to an accuracy of within approximately 1 pK unit.

The value of pK_1 was estimated from the reference compound $CH_3N^+H_2CH_2OH$, with a pK_a of 9.98 for the hydroxyl group (Hine et al., 1970). Correcting this value for the effect of the N^{10} substituent ($\rho_I \times \sigma_I = -8.4 \times 0.10 = -0.84$) yields p$K_1 = 9.1$. The value of p$K_3$ was estimated as follows. Addition of a hydroxyl group to the methyl group of a secondary ammonium ion decreases the pK_a by approximately 1.88 units (Hine et al., 1970; Kallen & Jencks, 1966), while the pK_a alteration caused by the addition of N^{10} to the methyl group is $-8.4 \times 0.10 = -0.84$ unit. On the basis of the pK_a of 8.15 for the amino group of GAR (Dawson et al., 1969), pK_3 is $8.15 - 1.88 - 0.84 = 5.4$. In order to obtain the value of pK_4, we used methanol with p$K_a = 15.5$ (Ege, 1984). Upon correction for both the folate N^{10}-amino group ($-8.4 \times 0.10 = -0.84$) and the amino group of GAR ($-8.4 \times 0.10 = -0.84$), the calculated pK_4 is 13.8. The value of pK_2 is then pK_3 + pK_4 − pK_1 = 10.1.

The value of pK_6 was estimated using *p*-aminobenzoic acid as a model compound, with the amine p$K_a = 2.59$ (Dawson et al., 1969). Correcting this for a methyl substituent, which increases the pK_a by 0.3 unit (Fox & Jencks, 1974), for the hydroxyl group, which will alter the pK_a by $-8.4 \times 0.25 = -2.1$ units, and for the GAR amino group, which will change the pK_a by $-8.4 \times 0.10 = -0.84$ units, yields p$K_6 = 0.0$. In order to obtain pK_7, $CH_3N^+H_2CH_2OH$ was again used. Correcting this hydroxyl pK_a of 9.98 for the effects of N^{10} ($-8.4 \times 0.10 = -0.84$) gives p$K_7 = 9.1$. Finally, pK_5 is then pK_4 + pK_6 − pK_7 = 4.7.

Although the overall transfer reaction is thermodynamically favorable, the relative levels of the key protonated intermediates (**5**/**6**) are unfavorable. The above analysis suggests that the proton switch from **1** would strongly favor the formation of **4**. At neutral pH, the levels of **3** and **6** relative to **4** are both disfavored by approximately a factor of 10^7. Therefore, after formation of **4**, very little of the desired intermediate **5** would be expected to exist under equilibrium conditions. In effect, **4** acts as a sort of "trap" in this scheme, impeding completion of the transfer process.

In order to serve as an efficient catalyst, the enzyme most likely alters the pK_a values in this scheme. One would expect the enzyme to stabilize the oxyanion with a positively charged residue. This residue should decrease pK_1, pK_4, and pK_7. Additionally, either hydrophobic or acidic groups placed near the amino group of GAR would lower pK_2, helping to shift the equilibrium toward **3** from **1**. Finally, an anionic group near N^{10} of tetrahydrofolate should raise pK_5, enhancing the formation of **5** from **3**.

On the basis of structure−reactivity considerations, the intrinsic reactivity of **1** relative to **5** may be as great as ca. 10^5 (the difference in pK_as) so that the level of **5** relative to **1** may differ by as much as 10^{-3} with the net flux favoring the formyl transfer process by a hundredfold. Any reduction in this intrinsic reactivity (i.e., $\alpha = 0.5$) would require a corresponding decrease in the ratio of the level of **5** to **1**, so that the levels of **5** and **1** are nearly balanced.

The GAR transformylase crystal structures corroborate this mechanism. The structure with the highest resolution available (1.96 Å) is the BW1476U89 multisubstrate adduct inhibitor complex (Klein et al., 1995), which approximates the transition state by having a tetrahedral carbon and hydroxyl group in the position of the formyl group. This structure is composed of two enzyme dimers, with each monomeric unit showing a slightly different active site conformation, primarily involving alterations in the positions of the inhibitor and the side chain of H108 (Klein et al., 1995). These two conformations are believed to represent protonated and nonprotonated forms of H108. Since either or perhaps even both of these conformations may be important in the catalytic mechanism, each will be examined separately.

In conformation "A" from Klein et al. (1995), the putatively charged imidazolium of H108 makes a strong hydrogen bond (2.65 Å) with the formyl oxygen (Figure 2). This interaction would be expected to stabilize the oxyanion, substantially decreasing pK_4, as well as pK_1 and pK_7. The proximity of the imidazolium to the GAR amine would also be expected to lower pK_2, although the decrease in pK_2 would be less than the decrease in pK_4 since the imidazolium does not hydrogen bond to the GAR amino group. In fact, no enzymic side chains form hydrogen bonds directly with the GAR amino group in the one crystal structure available with bound GAR (Almassy et al., 1992). Additionally, the carboxylate side chain of D144 interacts with N^{10} of folate through a bridging water molecule. This interaction would be expected to raise pK_5, stabilizing the protonated form of N^{10}. Overall, if H108 decreases pK_2 slightly while substantially decreasing pK_4, and D144 increases pK_5, the net equilibrium in Figure 3 may actually shift in favor of **5**. The only necessary chemical transformations required would be the switch of a proton from the GAR amino group to the N^{10} of tetrahydrofolate, which should easily be accomplished by solvent molecules in the active site.

One substantial problem with this potential mechanism is that there is nothing to prevent the protonated imidazolium of H108 from simply transferring a proton to the oxyanion. If the H108 pK_a is near the normal value of 6−7 for histidine residues, one should expect the imidazolium to rapidly catalyze the formation of **4** from **3**. After proton transfer, the uncharged imidazole should still hydrogen bond to the hydroxyl group of **4**, further stabilizing this species. However, the species of H108 hydrogen bonding to the oxyanion might not be the imidazolium (p$K_a \sim 6$) but rather the neutral imidazole, with a pK_a of ∼14.5 (Ege, 1984). In this case, a neutral H108 hydrogen bonding to the oxyanion would favor **3** over **4**.

Although an uncharged histidine serving as an acid is unusual, it is not unprecedented. In triosephosphate isomerase, the imidazole side chain of H95 acts as an acid, transferring a proton to an oxyanion intermediate (Lodi & Knowles, 1991). In this enzyme, H95 is positioned at the amino end of an α-helix which acts to decrease the imidazole pK_a (Hol, 1985). No similar α-helix is present in GAR transformylase. The main chain amide of Q97 also hydrogen bonds to the imidazole of H95 in triosephosphate isomerase, perhaps providing supplemental stabilization of the anionic imida-

zolate (Lodi & Knowles, 1991). In the catalytic site of GAR transformylase, it is possible that N106 may assist by orienting the imidazole of H108.

Conformation "B" from Klein et al. (1995) presents another possible side chain array to facilitate increased levels of species **5**/**6** (Figure 2). In this conformation, the imidazole of H108 does not hydrogen bond with the formyl oxygen and is 3.8 Å away from this atom. Instead, the amide hydrogens of the N106 side chain are within 2.8 Å of the formyl oxygen. N106 (pK_a ~15) would be expected to stabilize the oxyanion analogously to the uncharged imidazole of H108 in conformation A. Asparagine side chains have been previously reported to stabilize oxyanions in subtilisin Carlsberg (McPhalen & James, 1988) and in subtilisin BPN′ (Takeuchi et al., 1991). The proximity of N106 and H108 may lower pK_2 through dipole effects, although neither one of them forms hydrogen bonds with the GAR amine. As in conformation A, pK_5 could be elevated by the carboxylate side chain of D144 through a bridging water molecule. The required proton transfer steps would again occur via water molecules in the active site.

Mutagenesis results do not indicate which residue is most responsible for stabilizing the oxyanion, N106 or H108. No mutants of either residue demonstrate substantial activity, although a few mutants of both N106 and H108 do show some activity in our complementation assay (Table 2). We propose that the neutral imidazole of H108 and the amide side chain of N106 are both involved in stabilizing the oxyanion of **3**. It is unclear whether only one of these residues hydrogen bonds directly to the oxyanion, with the other providing supplemental stabilization, or whether both hydrogen bond directly to the oxyanion. Since mutations of either residue can lead to a partially active enzyme, we believe that both side chains are directly involved in stabilization of the oxyanion through hydrogen bonds to lone pair electrons. Because these residues show no consistent specificity in their replacement, other factors such as the introduction of water molecules or alternate side chain conformations must be important. Two different active site conformations observed in the structure of the enzyme inhibitor complex are in accord with this rationale.

D144 is involved in the protonation of N^{10} through a water molecule. Mutants of D144 would likely not be able to elevate pK_5, leading to the decreased activity levels shown in Table 2. In addition, D144 mutant proteins would lack the carboxylate that may be required to stabilize and mediate proton transfer to N^{10}.

In summary, we have used site-directed mutagenesis to alter each polar residue within 6 Å of the catalytic center of GAR transformylase. Screening of these mutants by functional complementation has shown that none of these residues are absolutely essential for catalytic activity, although several of them (N106, H108, S135, and D144) are important for full activity. We believe that the side chains and water molecules within the active site are configured to perturb the pK_as of the attacking and departing nitrogens, thus biasing the proton switch acting on the pool of intermediates toward the species required for transfer.

ACKNOWLEDGMENT

We gratefully acknowledge John M. Smith for providing us with the TX680 cell line, pJS167 plasmid, and technical advice, Brian J. Kearney for technical assistance, and Jon Stewart and R. Ilene Kaufman for helpful discussions, and we especially thank Jairo Arevalo and Ping Chen for providing us with crystallographic coordinates.

REFERENCES

Aimi, J., Qiu, H., Williams, J., Zalkin, H., & Dixon, J. E. (1990) *Nucleic Acids Res. 18*, 6665−6672.
Almassy, R. J., Janson, C. A., Kan, C. C., & Hostomska, Z. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6114−6118.
Chen, P., Schulze-Gahmen, U., Stura, E. A., Inglese, J., Johnson, D. L., Marolewski, A., Benkovic, S. J., & Wilson, I. A. (1992) *J. Mol. Biol. 227*, 283−292.
Climie, S., Ruiz-Perez, L., Gonzales-Pacanowska, D., Prapunwattana, P., Cho, S. W., Stroud, R., & Santi, D. V. (1990) *J. Biol. Chem. 265*, 18776−18779.
Cunningham, B. C., & Wells, J. A. (1989) *Science 244*, 1081−1085.
Dawson, R. M. C., Elliot, D. C., Elliot, W. H., & Jones, K. M. (1969) *Data for Biochemical Research*, 2nd ed., Oxford University Press, Oxford.
Ege, S. (1984) *Organic Chemistry*, D. C. Heath and Co., Lexington, MA.
Fox, J. P., & Jencks, W. P. (1974) *J. Am. Chem. Soc. 96*, 1436−1449.
Hardy, L. W., & Nalivaika, E. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9725−9729.
Hine, J., Via, F. A., Gotkis, J. K., & Craig, J. C. (1970) *J. Am. Chem. Soc. 92*, 5186−5193.
Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989) *Gene 77*, 51−59.
Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol. 45*, 149−195.
Howell, E. E., Villafranca, J. E., Warren, M. S., Oatley, S. J., & Kraut, J. (1986) *Science 231*, 1123−1128.
Inglese, J., Johnson, D. L., Shiau, A., Smith, J. M., & Benkovic, S. J. (1990a) *Biochemistry 29*, 1436−1443.
Inglese, J., Smith, J. M., & Benkovic, S. J. (1990b) *Biochemistry 29*, 6678−6687.
Kallen, R. G., & Jencks, W. P. (1966) *J. Biol. Chem. 241*, 5864−5878.
Klein, C., Chen, P., Arevalo, J. H., Stura, E. A., Marolewski, A., Warren, M. S., Benkovic, S. J., & Wilson, I. A. (1995) *J. Mol. Biol. 249*, 153−175.
Liu, L., & Santi, D. V. (1992) *Biochemistry 31*, 5100−5104.
Liu, L., & Santi, D. V. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8604−8608.
Lodi, P. J., & Knowles, J. R. (1991) *Biochemistry 30*, 6948−6956.
Marolewski, A., Smith, J. M., & Benkovic, S. J. (1994) *Biochemistry 33*, 2531−2537.
McPhalen, C. A., & James, M. N. G. (1988) *Biochemistry 27*, 6582−6598.
Nygaard, P., & Smith, J. M. (1993) *J. Bacteriol. 175*, 3591−3597.
Ritchie, C. D., & Sager, W. F. (1964) *Prog. Phys. Org. Chem. 2*, 323−397.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, New York.
Santi, D. V., Pinter, K., Kealey, J., & Davisson, V. J. (1990) *J. Biol. Chem. 265*, 6770−6775.
Smith, G. K., Mueller, W. T., Benkovic, P. A., & Benkovic, S. J. (1981) *Biochemistry 20*, 1241−1245.
Takeuchi, Y., Satow, Y., Nakamura, K. T., & Mitsui, Y. (1991) *J. Mol. Biol. 221*, 309−325.

BI9528715