Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6647–6652, June 1998
Biochemistry

Structural homologies with ATP- and folate-binding enzymes in the crystal structure of folylpolyglutamate synthetase

XIAOLIN SUN*†, ANDREW L. BOGNAR‡, EDWARD N. BAKER*†, AND CLYDE A. SMITH*†§

*Department of Biochemistry, Massey University, Palmerston North, New Zealand; and ‡Department of Medical Genetics and Microbiology, University of Toronto, Toronto, ON Canada M5S 1A8

Communicated by Brian W. Matthews, University of Oregon, Eugene, OR, March 20, 1998 (received for review December 16, 1997)

ABSTRACT **Folylpolyglutamate synthetase, which is responsible for the addition of a polyglutamate tail to folate and folate derivatives, is an ATP-dependent enzyme isolated from eukaryotic and bacterial sources, where it plays a key role in the retention of the intracellular folate pool. Here, we report the 2.4-Å resolution crystal structure of the MgATP complex of the enzyme from *Lactobacillus casei*. The structural analysis reveals that folylpolyglutamate synthetase is a modular protein consisting of two domains, one with a typical mononucleotide-binding fold and the other strikingly similar to the folate-binding enzyme dihydrofolate reductase. We have located the active site of the enzyme in a large interdomain cleft adjacent to an ATP-binding P-loop motif. Opposite this site, in the C domain, a cavity likely to be the folate binding site has been identified, and inspection of this cavity and the surrounding protein structure suggests that the glutamate tail of the substrate may project into the active site. A further feature of the structure is a well defined Ω loop, which contributes both to the active site and to interdomain interactions. The determination of the structure of this enzyme represents the first step toward the elucidation of the molecular mechanism of polyglutamylation of folates and antifolates.**

Both bacteria and eukaryotes require folate coenzymes in a number of important metabolic cycles, where they serve as carriers of one-carbon units, typically in the form of 5,10-methylene-tetrahydrofolate or 10-formyl-tetrahydrofolate. These folate derivatives act as cosubstrates in a host of enzymes involved in one-carbon metabolism (1), including the biosynthesis of methionine and thymidylate and the *de novo* synthesis of purine. Folate antagonists (or antifolates), on the other hand, are potent inhibitors of purine and thymidylate biosynthesis, blocking DNA synthesis and halting cell growth, and are important in cancer chemotherapy. Methotrexate, for example, is a potent inhibitor of dihydrofolate reductase (DHFR) (2) whereas 5-formyl-tetrahydrofolate combined with 5-fluorouracil effectively inhibits thymidylate synthase (3). It has been shown (2) that polyglutamylation of these antifolates (addition of L-glutamate groups at the γ-carboxy of the terminal glutamate) greatly enhances their cytotoxicity and that the enzyme folylpolyglutamate synthetase (FPGS) is responsible for the conversion to the more active form. Consequently, the use of FPGS to produce anticancer drugs with increased cytotoxicity is currently under biochemical and clinical study (2).

Endogenous folates in the cell are predominantly in the form of polyglutamate derivatives, which are the normal substrates for enzymes involved in one-carbon metabolism. It has been shown that many folate-dependent enzymes have a higher affinity for polyglutamate substrates (reviewed in refs. 1 and 4); thymidylate synthase, for example, shows a 60% increase in substrate affinity from monoglutamate 5,10-methylene-tetrahydrofolate to the diglutamate derivative, although further extension has little affect (5). The polyglutamate tail, which is required for folate retention in the cell, can vary from triglutamates in some bacteria to nonaglutamates or greater in eukaryotes. FPGS, which catalyses the MgATP-dependent glutamylation of folate coenzymes (Fig. 1), has been purified from a number of bacterial and mammalian sources (6–8). The genes encoding FPGS have been cloned and sequenced from *Escherichia coli* (7, 9), *Lactobacillus casei* (10), and humans (11), and recently, mutagenesis on the *L. casei* gene (12) identified a region of the enzyme where minor changes to the sequence profoundly affected enzymatic activity. This region of FPGS encompasses a putative glycine-rich P-loop motif (13) similar to that found in a number of nucleotide-binding enzymes, including adenylate kinase (ADK) (14), p21-*ras* (15), and myosin (16). The consensus sequence for the P loop in p21-*ras* and myosin is GxxGxGKT/S whereas ADK has an extra glycine inserted between the lysine and the threonine (GxxGxGKGT). The putative P loop in *L. casei* FPGS has the sequence xxGxxGKGS (10), thus resembling more closely the ADK family.

FPGS forms a quaternary complex with MgATP, a folate derivative, and glutamate (bound in that order), with the ordered release of the products MgADP, the glutamylated folate, and phosphate (6, 17). In the *L. casei* FPGS, 5,10-methylene-tetrahydrofolate is the only effective substrate, and although the enzyme shows a greater affinity for the diglutamate derivative, there is a marked drop-off in activity with the triglutamate until there is no significant interaction with the pentaglutamate (6). Mammalian synthetases prefer tetrahydrofolate derivatives and exhibit a decreasing ratio k_{cat}/K_m as a function of increasing polyglutamate length (1), also indicating a marked decline in their affinity for the longer substrates.

A mechanism for the binding of polyglutamates to FPGS was proposed (1) that sees the pterin moiety binding first, followed by the binding of the polyglutamate tail in a low-affinity glutamate-binding site. The internal polyglutamate residues move through this low-affinity site until the terminal residue is bound, triggering a conformational change resulting in tight binding of the terminal glutamate. Presumably, this conformational change brings the γ-carboxylate near the γ-phosphate of the ATP so that the system then is poised for phosphorylation of the γ-carboxyl and for the subsequent nucleophilic attack by the free amino group of the incoming glutamate on the acyl-phosphate intermediate formed (1, 18).

The biochemical and kinetic data accumulated for this class of enzyme raise a number of important questions. To fully understand how this enzyme binds its three substrates, how it determines when the terminal glutamate residue is bound in the

Abbreviations: PPi, pyrophosphate; FPGS, folylpolyglutamate synthetase; DHFR, dihydrofolate reductase; ADK, adenylate kinase; rmsd, rms deviation.

Data deposition: The atomic coordinates have been deposited in the Protein Data Bank, Biology Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1 fgs).

†Present address: School of Biological Sciences, University of Auckland, Auckland, New Zealand.

§To whom reprint requests should be addressed. e-mail: ca.smith@auckland.ac.nz.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

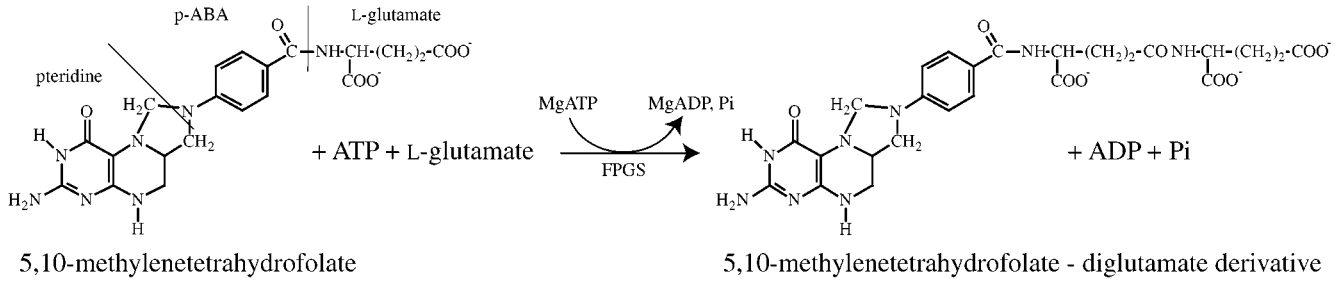

FIG. 1. Reaction catalyzed by folylpolyglutamate synthetase.

low-affinity site, what triggers the resultant conformational change, what the nature of the conformational change is, and how the enzyme knows when the requisite polyglutamate length has been achieved, we have begun an x-ray crystallographic structural analysis of FPGS from *L. casei*, and we report here the structure of the enzyme at 2.4-Å resolution.

MATERIALS AND METHODS

Lactobacillus casei FPGS was purified, and initial apo-FPGS crystals were grown as described (19). By using a variation on these conditions, small, block-like crystals of the MgATP complex of FPGS were grown in hanging drops from 40 mM acetate (pH 5.3), 5 mM MgATP, and 4% polyethylene glycol 4000, with the drop composed of a 1:1 solution of 5 mg/ml FPGS (incubated overnight with 10 mM MgATP) and reservoir solution. These crystals were used to micro-seed drops of protein/polyethylene glycol, yielding crystals with the dimensions $0.5 \times 0.5 \times 0.4$ mm. The crystals reached their optimum size within 2 weeks. Diffraction quality crystals of the apo enzyme at pH 7.5 also were obtained, and this structure will be described elsewhere.

All data were collected at room temperature with a Rigaku RAxis-IIc detector (Rigaku, Tokyo) mounted on a Rigaku RU200 rotating anode generator. The MgATP–FPGS crystals diffracted to a maximum resolution of 2.2 Å and had the following cell dimensions: a = 54.0Å, b = 46.1Å, c = 84.9Å, and $\beta = 107.3°$ in the space group $P2_1$. Based on a molecular weight of 43,000 Da (6), the Matthews coefficient was determined to be 2.62 Da/Å^3, assuming one molecule per asymmetric unit, indicating a solvent content of 53% (20). A complete native data set was collected to 2.4 Å from one crystal and was processed by using DENZO (21) and the CCP4 suite of programs (22). The resulting data set was 100% complete to 2.4 Å with an $R_{merge} = 7.3\%$ (27.2% in the 2.5–2.4 Å shell). These MgATP crystals were used as the native crystals in heavy metal soaking experiments.

Four heavy metal derivatives were used to calculate phases. The first, 20 mM trimethyllead acetate (Pb1) soaked for 2.5 hr, gave a difference Patterson map with one clear site. A second trimethyllead acetate derivative (Pb2, 5 mM, 4 hr) also had the same first site but with higher occupancy. Two more derivatives, ethylmercury chloride (2 mM, 4 hr) and potassium uranyl fluoride ($K_3UO_2F_5$, 2 mM, 6 hr), also were collected, and a first site was found for each in a difference Patterson map. Subsequent cross difference Fourier syntheses were used to locate two additional sites in the lead derivatives, six sites in the uranyl derivative, and three sites in the mercury derivative.

Phases were calculated and refined to 3.0 Å by using the program MLPHARE (22). The overall figure of merit for the resultant multiple isomorphous replacement phases was 0.594. Solvent flattening and histogram matching by using the program DM (23) gave a final figure of merit of 0.70. The multiple isomorphous replacement map was skeletonized, and significant portions of β-strand and α-helical structure then were readily identifiable. A polyalanine model was built into the map by using the molecular graphics program TURBO (Architecture et Fonction des Macromolecules Biologiques and Biographics, Marseilles, France), the phases from this partial model were combined with the multiple isomorphous replacement phases by using SIGMAA (24), and a new map was calculated. Polyalanine addition and phase recombination were repeated twice more, and in the third combined map, alignment with the sequence was attempted. The six tryptophan residues in the enzyme were visible clearly in the electron density, as were many other aromatic residues. In particular, four aromatic residues (YWYF) at positions 129–132 were clearly visible, and the sequence could be traced unambiguously in both directions.

Refinement of the model was performed by using TNT (25). The initial *R* factor was 0.38 (the free *R* factor using 5% of the reflections was 0.44). Several phases of TNT refinement and manual model building followed in which additional residues were added to the model. Water molecules also were added based on a systematic search of F_o–F_c electron density maps. Each potential water molecule was inspected on the graphics terminal and was included only if it made chemical sense and was not in any region of the molecule that either was defined poorly or had not been built.

The location of the phosphate binding site was established early in the refinement. The two largest peaks in the F_o–F_c maps were consistently in the same place and ≈3.3 Å apart. The shape of the electron density in the $2F_o$–F_c and F_o–F_c maps resembled a pyrophosphate (PPi), with the remainder of the nucleotide not immediately visible. A difference map calculated between the MgATP–FPGS and apo-FPGS data sets also showed unequivocally the location of two of the phosphates of the bound ATP. A third peak near the two phosphate peaks tentatively was assigned

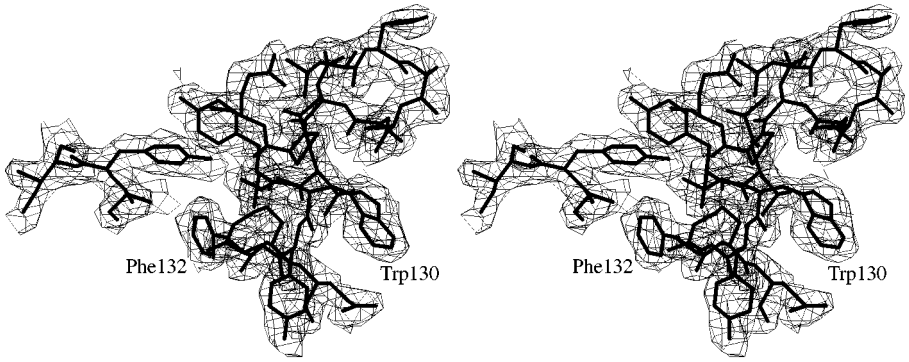

FIG. 2. Stereoview of the final $2F_o$–F_c electron density map for part of the hydrophobic core of the N domain, contoured at 1.0σ. This region contains the four consecutive aromatic residues YWYF (129–132) identified in the initial multiple isomorphous replacements maps.

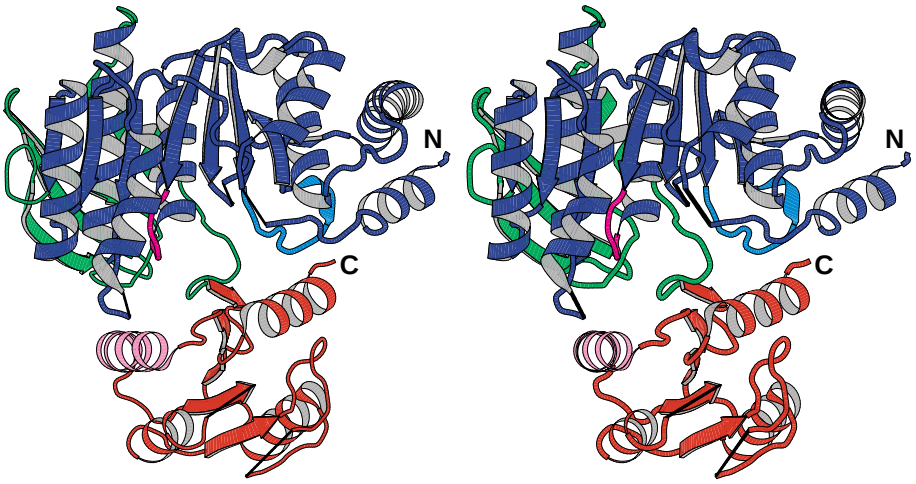

FIG. 3. Stereo ribbon diagram of FPGS. The N domain is composed of a *ras*-like domain (blue) and a three-stranded β-sheet (green) whereas the C domain is shown in red. The N and C termini are indicated, and the breaks in the polypeptide chain are joined by thin black lines. The P loop (46–50) is shown in magenta, the Ω loop (72–81) is shown in cyan, and helix A10 is shown in pink. This figure and Fig. 6 were generated by using MOLSCRIPT (39).

as the third phosphate, but only weak, discontinuous density was apparent for the ribose and adenine groups, and these have not been modeled at this stage. The low occupancy of the nucleotide (estimated from the PPi as 40%) probably resulted from the low pH of crystallization; at this pH (5.3), activity is very low and, even at the optimum pH of ≈10, the affinity of FPGS for MgATP is low (K_m = 5.6 mM) (6). On the opposite side of the PPi, density for a potential metal ion was identified. This has been modeled putatively as Mg^{2+} with an occupancy of 50%.

The final crystallographic *R* factor for MgATP–FPGS is 0.196 for all data (15311 reflections) between 15 and 2.4 Å, with an R_{free} of 0.265 (809 reflections). The final MgATP–FPGS model contains 393 of 428 residues, with five loops missing (residues 18–19, 146–150, 170–176, 342–347, and 373–384), 170 water molecules, one Mg^{2+}, and a PPi. A representative section of the final $2F_o$–F_c density map at 2.4 Å is shown in Fig. 2. The average coordinate error is estimated from a Luzzati plot (26) at 0.20 to 0.25 Å and the overall average temperature factor is 28.5 $Å^2$. A Ramachandran plot (27) of the final model indicates that all of the nonglycine residues lie within the allowed regions of conformational space, with 90.0% lying within the most favorable regions. Statistics on the data collection and the refinement are given in Table 1.

RESULTS AND DISCUSSION

The FPGS molecule is divided into two distinct domains, folded consecutively and joined by a six-residue linker (residues 295–300). The N-terminal domain (residues 1–295) consists of a large, central β-sheet composed of seven strands flanked by seven α-helices. A small, independently folded, three-stranded β-sheet packs against the core of the domain. Two α-helices complete the fold of the N-terminal domain and lead into the C-terminal domain (300–425), which has an α/β structure made up of six β-strands and four flanking helices. The C domain appears to be somewhat more mobile than the N domain (the average main-chain B values are 21.8 $Å^2$ and 35.3 $Å^2$ for the N- and C-terminal domains, respectively). Interactions between the two domains are primarily hydrophobic with a total of only four hydrogen bonds, two of which involve residues in the linker region. The overall structure is shown in Fig. 3.

The N-Terminal Domain. Interrogation of the Protein Data Bank by using DEJAVU (28) showed a broad structural similarity to many nucleotide binding proteins. As pointed out by Schulz (29), nucleotide-binding proteins have different topologies depending on whether they bind dinucleotides or mononucleotides. FPGS falls firmly into the family of mononucleotide-binding proteins, in that the N domain resembles the *ras*-like domain observed in mononucleotide-binding enzymes such as p21-*ras* (15) and ADK (14). The core structure of these enzymes has been reviewed recently (29, 30) and consists primarily of a four-stranded β-sheet plus four invariant α-helices. The majority of the enzymes have extra secondary structure elements and, in some cases, extra domains inserted at various positions within this core, yet they still retain the same structure surrounding the nucleotide-binding site. FPGS possesses a similar core β-sheet consisting of four strands and three helices. A fourth helix that generally is conserved corresponds to a disordered region (residues 145–151) in FPGS, although here, the following residues (152–155) have a conformation that suggests an incipient tendency toward a helical structure for this region. There are two insertions in the core topology (excluding two N-terminal helices): two β-strands and two helices at one end and one β-strand, two helices, and the three-stranded sheet at the other. Three of these additional strands make up the extended seven-stranded β-sheet along with the four core strands. The nomenclature and location of the secondary structure elements in FPGS is given in Fig. 4.

An initial superposition of p21-*ras* (Protein Data Bank code 121P) on to FPGS, based on the β-strand/P-loop/α-helix motif, showed that the core β-sheet and some of the helices matched almost perfectly. A more complete superposition gave an rms deviation (rmsd) of 1.6 Å for five β-strands (A, C, D, E, and F) and the P-loop helix *a*. When all overlapping secondary structure elements were included in the calculation, the rmsd was 2.9 Å for 110 C_α atoms, indicating a remarkable three-dimensional similarity despite the lack of sequence identity (<15%). Structural comparisons with elongation factor Tu (33) and guanine nucleotide-binding protein $G_{i\alpha}$ (34) gave similar results of 3.1 Å (112 C_α) and 3.2 Å (107 C_α), respectively. The ADK family of enzymes matched the core of FPGS even more closely. Superposition of five strands (A, B, C, D, and E) and three helices (*a*, *d*, and *h*) of ADK (14) onto FPGS gave an rmsd of 1.9 Å for 72 core C_α atoms

Table 1. Summary of crystallographic data collection and refinement

	native	Pb1	Pb2	Hg	U
Resolution, Å	2.4	2.5	3.0	3.0	3.0
Measured reflections	39117	17080	9264	12415	11019
Unique reflections	15725	9145	6290	6156	7456
Completeness, %	100	64.5	78.6	75.1	93.1
R_{merge}, %*	7.3	9.1	9.9	7.5	8.3
R_{iso}, %†		23.3	21.1	15.0	30.6
Phasing power‡		1.49	1.17	1.20	1.70
Resolution range, Å	15.0–2.4				
R factor/R_{free}, %	19.6/26.5				
rms deviations	bond distances 0.008 Å			angles 1.46°	

*$R_{merge} = \Sigma_i | I_i - \bar{I}_i | / \Sigma \bar{I}_i$ where $\bar{I}_i$ is the mean intensity.
†$R_{iso} = \Sigma | | F_{PH} | - | F_P | | / | F_P |$ where $| F_{PH} |$ and $| F_P |$ are the observed derivative and native structure factor amplitudes.
‡Phasing power = *rms* F_H/*rms* lack-of-closure summed over all reflections.

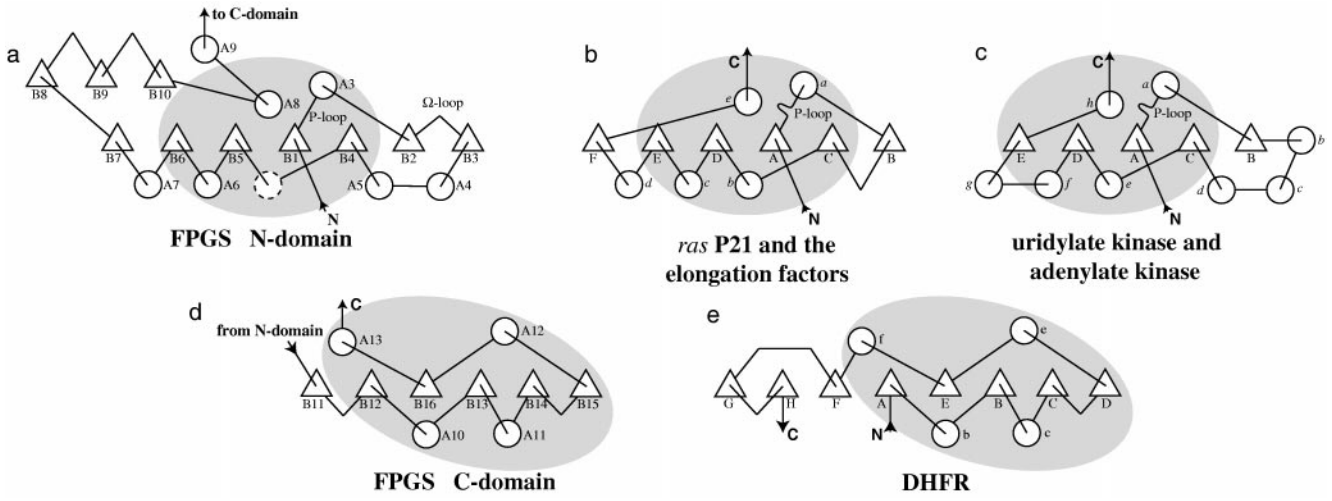

FIG. 4. Topology diagrams of (*a*) the N domain of FPGS, (*b*) p21-*ras* and G proteins, (*c*) ADK and uridylate kinase, (*d*) the C domain of FPGS, and (*e*) DHFR. The regions of structural similarity between the N domain of FPGS, the G proteins, and the kinases are indicated by the shaded areas in *a*–*c*. Likewise, structurally similar regions in the C domain of FPGS and DHFR are shaded in *d* and *e*. Secondary structure assignments for FPGS are: N domain; A1, 2–11; A2, 22–38; A3, 49–63; A4, 91–113; A5, 119–135; A6, 177–189; A7, 201–216; A8, 260–280; A9, 284–294; B1, 40–45; B2, 67–72; B3, 83–86; B4, 139–144; B5, 160–164; B6, 194–198; B7, 218–222; B8, 227–234; B9, 237–244; B10, 247–254; C domain; A10, 320–332; A11, 351–359; A12, 390–402; A13, 412–425; B11, 302–305; B12, 309–313; B13, 336–341; B14, 361–365; B15, 386–388; B16, 406–410.

and 3.0 Å for 125 C_α atoms in all matching elements. Uridylate kinase (35) showed an even greater structural similarity: 1.8 Å (72 C_α) and 2.8 Å (130 C_α).

Previous comparisons (30, 36) have noted the remarkable similarity in nucleotide binding domains that includes not only the GTPases and kinases but also the motor proteins myosin, kinesin, and ncd. FPGS now joins this growing superfamily. Given the remarkable similarity in three-dimensional structure between the G proteins, the kinases, FPGS, the motor proteins, and other nucleotide-binding enzymes containing a *ras*-like domain, it is conceivable that all of these enzymes may have diverged from an ancient nucleotide-binding enzyme.

The Nucleotide-Binding Site. The putative nucleotide-binding P loop, residues 46–52, joins strand B1 to helix A3. Although its sequence (GTNGKGS) resembles the consensus P-loop sequence found in adenylate and uridylate kinase (GxxGxGKGT), it is two residues shorter. It also differs somewhat in its conformation. Typically, the P loop folds to form a small "ledge" perpendicular to the plane of the central β-sheet (30); the triphosphate group sits on this ledge and interacts with amide nitrogens from the P loop and the N terminus of the helix, and the nucleoside projects over the side of the ledge. In FPGS, on the other hand, the shorter loop does not project in the same way (Fig. 5).

Despite the difference in P-loop conformation, it appears that, in FPGS, the ATP still binds to part of this loop and to the N terminus of helix A3. Assuming that the PPi ion modeled from difference electron density maps represents the α- and β-phosphates of the nucleotide, with the γ-phosphate occupying the weaker third site (Fig. 5), the remainder of the nucleotide then projects into the cleft between the N and C domains, coincident with weak electron density. In this orientation, the PPi is in a position to hydrogen bond with amide nitrogens from the helix A3 N terminus and the P loop, and the putative γ-phosphate is adjacent to the conserved lysine of the P loop (Lys50). This residue has the same conformation as in the p21-*ras*-GTP complex (15) and myosin (31) (Fig. 5). The putative γ-phosphate position also would make it accessible to the terminal carboxylate residue of the folate (see below).

A special feature of the FPGS active site is an Ω loop (32) of 10 amino acids (Ser73–Arg82) that links strands B2 and B3. Based on a sequence alignment of 17 known FPGS sequences (data not shown), this loop appears to be conserved in both length and sequence. The Ser73–Pro74 dipeptide (positions 1 and 2 in the Ω loop) is conserved in all known FPGS sequences. After this is a region typically comprising a large hydrophobic residue (Tyr, Phe, or His) followed by two small hydrophobic residues (Ile, Leu, or Val). In position 9, there is a conserved acidic residue (Glu in all sequences except mouse and *Streptococcus pneumoniae*, which both have an Asp), and it is this residue that stabilizes the loop by making intraloop interactions across to the mainchain nitrogen atoms at positions 4–6. In the current structure, Pro74 has a *cis*-conformation, and it is reasonable to assume, given the high level of sequence identity in this region, that this *cis* peptide will be a conserved feature in the other FPGS molecules.

Of most significance, the presence of this *cis*-proline causes the carbonyl oxygen of the preceding residue (Ser73) to be oriented into the active site in bonding distance of the putative cation binding site (Fig. 5). Other ligands to this site include phosphate oxygens of the PPi and the sidechain of Glu143, which recently

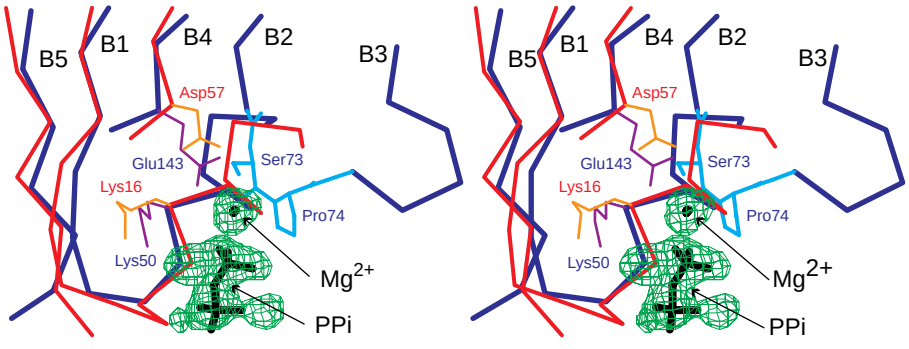

FIG. 5. Stereoview of the P loop in FPGS (blue bonds) overlaid with the P loop from p21-*ras* (red bonds). The location of the pyrophosphate and the Mg^{2+} are given in black, superimposed on F_o–F_c omit density for the PPi and the Mg^{2+}. The proposed location of the γ-phosphate is the lobe of F_o–F_c density between the PPi and the sidechain of Lys50. The conserved lysine residue in the P loop is indicated for FPGS (Lys50, magenta) and p21-*ras* (Lys16, orange), along with a conserved acidic residue adjacent to the active site, which is Glu143 in FPGS and is Asp57 in p21-*ras*. The first two residues of the Ω loop (Ser73 and Pro74) are indicated in cyan.

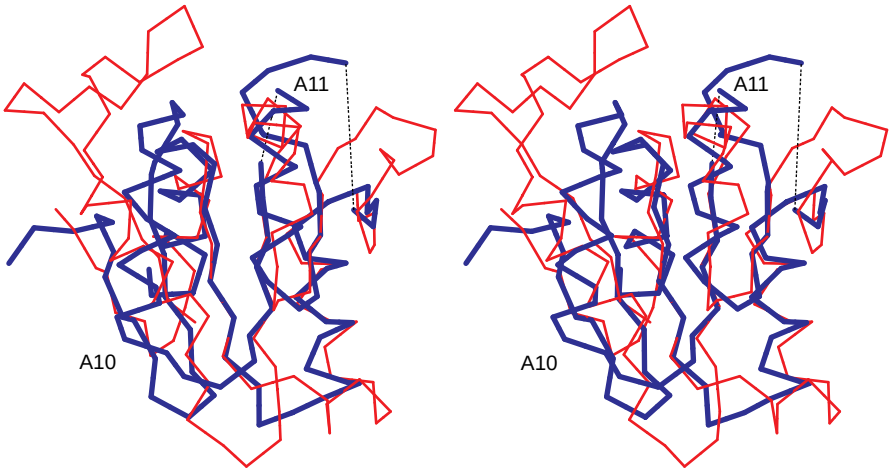

FIG. 6. Superpositions of FPGS (blue, thick bonds) onto human dihydrofolate reductase (1DHF, red, 1.6 Å, 52 C_α). Note the different orientations of the two helices A10 and A11, which brings them closer together in FPGS. The program LSQKAB (22) was used for all superpositions. This figure was generated by using MOLSCRIPT (39).

has been implicated in enzyme activity (Y. Sheng and A.L.B., unpublished observations). Interestingly, the G proteins have an aspartate in the same position as Glu143 (Fig. 5) interacting with the essential cation via a water molecule. This aspartate (Asp57 in p21-*ras*) is part of the B motif (or switch II region) of the G proteins, which exhibits a conformational change on nucleotide hydrolysis.

The location of the Ω loop also suggests a wider role in the activity of FPGS. Not only does it contribute to the active site, through the *cis*-proline, but it also mediates interactions with the C-terminal domain via hydrophobic interactions and two hydrogen bonds between the carbonyl oxygen atoms of Ile76 and Met77 and the sidechain of Gln421. Both the complexity of the enzymatic reaction and the domain structure of the enzyme suggest the importance of relative domain movements. The location of the Ω loop thus suggests that it could play a key role in any such movement.

The C-Terminal Domain. A search of the Protein Data Bank by using DEJAVU showed that the C-terminal domain of FPGS bears an unexpected and striking similarity to dihydrofolate reductase, strongly implicating this domain as the site of folate binding. DHFR from several species consistently gave rmsds substantially lower than any other molecule. The rmsds for the human (Protein Data Bank code 1drf) and the chicken (Protein Data Bank code 8dfr) structures were 1.1 Å and 1.3 Å, respectively, for six secondary structure elements. The next best hits were uridine diphosphogalactose 4-epimerase and 3-α,20-β-hydroxysteroid dehydrogenase with rmsds of 3.4 Å and 4.0 Å, respectively. Visual inspection of the resulting superpositions showed that, although the N-terminal 130 residues of both structures showed a superficially similar topology to the C domain of FPGS and to DHFR (in that the secondary structure elements were in similar locations), the connectivity was different. Thus, the C domain of FPGS bears by far the strongest resemblance to DHFR, possibly indicative of a functional similarity.

Despite minimal sequence identity (≈15%), five β-strands and two helices (B12–B16 and A12 and A13) in the C domain of FPGS can be superimposed onto equivalent secondary structural elements in DHFR with an rmsd of only 1.6 Å for 52 C_α positions (Fig. 6). The only secondary structural differences between the two structures are at the N- and C-terminal ends of the molecules and in the position of the helix connecting strands B13 and B14. The C domain of FPGS begins with the two antiparallel strands B11 and B12 whereas DHFR begins with strand A, the equivalent of B12. The loop connecting strand B12 with helix A10 in FPGS is significantly shorter (4 residues) than the corresponding region in DHFR (17 residues), and, consequently, the orientation of helix A10 is different, rotated in toward the plane of the β-sheet

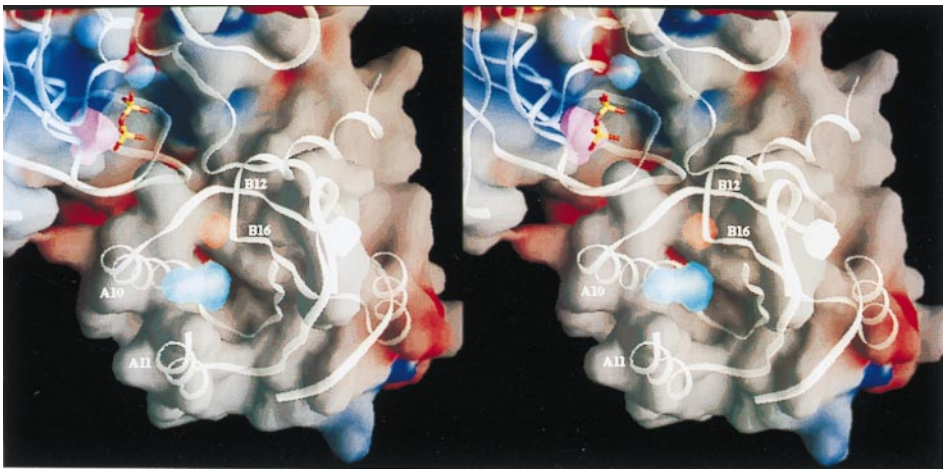

FIG. 7. Stereoview of the molecular surface (gray, semitransparent) calculated for the N domain (at top of figure) and the C domain (lower) of FPGS, colored according to electrostatic charge (red = negative, blue = positive). The cavities within the molecule are shown in various colors. A backbone trace of the molecule is shown in white, with the location of the pyrophosphate indicated in red and yellow sticks (center), lying in a deep interdomain cleft. The putative folate binding site is indicated by the cyan and light red cavities between helices A10 and A11 and strands B12 and B16. This figure was generated by using GRASP (40).

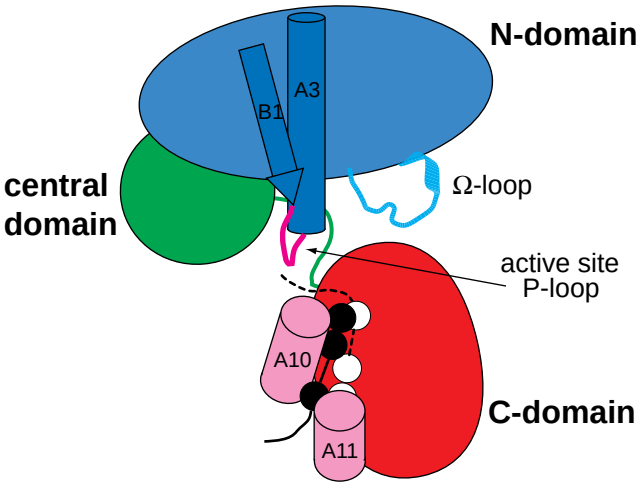

FIG. 8. Cartoon representation of FPGS showing the two possible binding modes of folate [given as two fused circles (pteridine), a single circle (p-aminobenzoic acid), and a long line (the polyglutamate tail)] adjacent to helices A10 and A11. Binding mode *i* (see text) as observed in DHFR is represented by a black folate, and mode *ii*, proposed for FPGS, is given as a white folate.

by ≈11°. At the C terminus of DHFR after helix f, there are three or four extra strands [depending on the source of the enzyme (37, 38)], whereas strand B11 of FPGS is in the same position as DHFR strand F but runs in the opposite direction.

Folate Binding. In human DHFR (37, 38), the folate substrate is bound in a pocket formed from strands A and E, helix b, and the loops after helix c. The pteridine group of the folate is buried against the back surface of the binding pocket (strands A and E) whereas the hydrophobic p-aminobenzoic acid moiety packs against hydrophobic residues from the inner side of the covering helix b and the loops at the C terminus of helix c. The glutamate group extends out into solution between helices b and c.

In FPGS, a similar hydrophobic cavity exists (Fig. 7), formed between strands B12 and B16 and helix A10 (equivalent to strands A and E and helix b in DHFR); we presume this cavity to be the site of folate binding. Compared with DHFR, the B12–A10 loop is shorter, however, and helix A10 is orientated somewhat differently, so that the folate binding mode cannot be exactly the same. We can identify two alternative possibilities: (*i*) Folate binds as in DHFR with the glutamate tail projecting out between helices A10 and A11, or (*ii*) the pteridine and p-aminobenzoic acid groups bind the other way around with the glutamate tail projecting into the active site from under the B12–A10 loop (Fig. 8). Binding mode *i* appears to be disallowed by the different orientation of helix A10, which does not allow space between A10 and A11 for the polyglutamate tail. It also would have as a disadvantage the fact that a very large domain reorientation would be needed to bring the glutamate tail close to the active site for addition of further Glu residues. Binding mode *ii*, on the other hand, would place the glutamate tail into the active site, between the domains and close to the ATP. Given that FPGS and DHFR have different substrate specificities [*L. casei* FPGS utilizes 5, 10-methylene-tetrahydrofolate (6) whereas the preferred substrate for DHFR is dihydrofolate], it is not unlikely that there will be a difference in folate substrate binding between the two enzymes. Moreover, the elongated shape of the cyan cavity (Fig. 7) suggests that the fused ring system of the pteridine group would be bound more readily in this site than in the smaller site indicated by the red cavity. Hence, we favor binding mode *ii*, although it has yet to be proven.

CONCLUSIONS

FPGS catalyzes a complex reaction involving three substrates (ATP, folate, and glutamic acid) and a series of steps in which glutamic acid residues are added one at a time to a growing chain. The binding site for glutamate is not yet known, but the structure of FPGS reveals an intriguing modular construction, comprising an ATP-binding domain homologous with other ATP- and GTP-dependent enzymes and a domain homologous with DHFR, presumed to be the site of folate binding. Given that the two domains are connected only by a single piece of polypeptide and that the area of interdomain contact is small and predominantly hydrophobic in nature, it seems reasonable to assume that the catalytic reaction of FPGS is coupled to a conformational change involving some rearrangement of the two domains. Further structural analyses will be aimed at defining the specificity of folate binding and its relationship with that of antifolate drugs and the nature of the polyglutamylation reaction.

This research was supported by the Health Research Council of New Zealand, the Medical Research Council of Canada (Grant MT9822 to A.L.B.), and Massey University (Doctorate Scholarship to X.S.). E.N.B. also received research support as an International Research Scholar of the Howard Hughes Medical Institute.

1. Shane, B. (1989) *Vitam. Horm.* **45,** 263–335.
2. Synold, T. W., Willits, E. M. & Barredo, J. C. (1996) *Leukemia and Lymphoma* **21,** 9–15.
3. Keyomarsi, K. & Moran, R. G. (1988) *J. Biol. Chem.* **263,** 14402–14409.
4. McGuire, J. J. & Coward, J. K. (1984) in *Folates and Pterins*, eds. Blakley, R. L. & Benkovic, S. J. (Wiley, New York), Vol. 1, pp. 135–190.
5. Lu, Y.-Z., Aiello, P. D. & Matthews, R. G. (1984) *Biochemistry* **23,** 6870–6878.
6. Bognar, A. L. & Shane, B. (1983) *J. Biol. Chem.* **258,** 12574–12581.
7. Bognar, A. L., Osborne, C., Shane, B., Singer, S. C. & Ferone, R. (1985) *J. Biol. Chem.* **260,** 5625–5630.
8. Cichowicz, D. J. & Shane, B. (1987) *Biochemistry* **26,** 504–512.
9. Bognar, A. L., Osborne, C. & Shane, B. (1987) *J. Biol. Chem.* **262,** 12337–12343.
10. Toy, J. & Bognar, A. L. (1990) *J. Biol. Chem.* **265,** 2492–2499.
11. Garrow, T. A., Admon, A. & Shane, B. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 9151–9155.
12. Toy, J. & Bognar, A. L. (1994) *Arch. Biochem. Biophys.* **314,** 344–350.
13. Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990) *Trends Biochem. Sci.* **15,** 430–434.
14. Muller, C. W. & Schulz, G. E. (1992) *J. Mol. Biol.* **224,** 159–177.
15. Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W. & Wittinghofer, A. (1990) *EMBO J.* **9,** 2351–2359.
16. Rayment, I., Rypniewski, W. R., Schmidt-Bäse, K., Smith, R., Tomchick, D. R., Benning, M. M., Winkelmann, D. A., Wesenberg, G. & Holden, H. M. (1993) *Science* **261,** 50–58.
17. Cichowicz, D. J. & Shane, B. (1987) *Biochemistry* **26,** 513–521.
18. Bannerjee, R., Shane, B., McGuire, J. J. & Coward, J. K. (1988) *Biochemistry* **27,** 9062–9070.
19. Cody, V., Luft, J. R., Pangborn, W., Toy, J. & Bognar, A. L. (1992) *J. Mol. Biol.* **224,** 1179–1180.
20. Matthews, B. W. (1968) *J. Mol. Biol.* **33,** 491–497.
21. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, eds. Sawyer, L., Isaacs, N. & Bailey, S. (SERC Daresbury Laboratory, Daresbury, England), pp. 56–62.
22. Collaborative Computing Project 4 (1994) *Acta Crystallogr. D* **50,** 760–763.
23. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter in Protein Crystallography* **31,** 34–38.
24. Read, R. J. (1986) *Acta Crystallogr. A* **42,** 140–149.
25. Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987) *Acta Crystallogr. A* **43,** 489–501.
26. Luzzati, V. (1952) *Acta Crystallogr.* **5,** 802–810.
27. Ramakrishnan, C. & Ramachandran, G. N. (1965) *Biophys. J.* **5,** 909–933.
28. Kleywegt, G. J. & Jones, T. A. (1997) *Methods Enzymol.* **277,** 525–545.
29. Schulz, G. E. (1993) *Curr. Opin. Struct. Biol.* **2,** 61–67.
30. Smith, C. A. & Rayment, I. (1996) *Biophys. J.* **70,** 1590–1602.
31. Fisher, A. J., Smith, C. A., Thoden, J., Smith, R., Sutoh, K., Holden, H. M. & Rayment, I. (1995) *Biochemistry* **34,** 8960–8972.
32. Fetrow, J. S. (1995) *FASEB J.* **9,** 708–717.
33. Kjeldgaard, M., Nissen, P., Thirup, S. & Nyborg, J. (1993) *Structure* **1,** 35–50.
34. Coleman, D. E., Berghuis, A. M., Lee, E., Linder, M. E., Gilman, A. G. & Sprang, S. R. (1994) *Science* **265,** 1405–1412.
35. Muller-Dieckmann, H.-J. & Schulz, G. E. (1994) *J. Mol. Biol.* **236,** 361–367.
36. Kull, J. F., Sablin, E. P., Lau, R., Fletterick, R. J. & Vale, R. D. (1996) *Nature (London)* **380,** 550–555.
37. Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C. & Kraut, J. (1982) *J. Biol. Chem.* **257,** 13650–13662.
38. Davies, J. F., Delcamp, T. J., Prendergast, N. J., Ashford, V. A., Freisheim, J. H. & Kraut, J. (1990) *Biochemistry* **29,** 9467–9479.
39. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**.
40. Nicholls, A., Sharp, K. A. & Honig, B. (1991) *Proteins* **11,** 281–296.