Articles

Structure of a Murine Cytoplasmic Serine Hydroxymethyltransferase Quinonoid Ternary Complex: Evidence for Asymmetric Obligate Dimers†

Doletha M. E. Szebenyi,‡ Xiaowen Liu,§ Irina A. Kriksunov,‡ Patrick J. Stover,*,§ and Daniel J. Thiel‡

Department of Molecular Biology and Genetics and Division of Nutritional Sciences, Cornell University, Ithaca, New York 14853

Received March 20, 2000; Revised Manuscript Received July 28, 2000

ABSTRACT: Serine hydroxymethyltransferase (SHMT) is a pyridoxal phosphate-dependent enzyme that catalyzes the reversible conversion of serine and tetrahydrofolate to glycine and methylenetetrahydrofolate. This reaction generates single carbon units for purine, thymidine, and methionine biosynthesis. The enzyme is a homotetramer comprising two obligate dimers and four pyridoxal phosphate-bound active sites. The mammalian enzyme is present in cells in both catalytically active and inactive forms. The inactive form is a ternary complex that results from the binding of glycine and 5-formyltetrahydrofolate polyglutamate, a slow tight-binding inhibitor. The crystal structure of a close analogue of the inactive form of murine cytoplasmic SHMT (cSHMT), lacking only the polyglutamate tail of the inhibitor, has been determined to 2.9 Å resolution. This first structure of a ligand-bound mammalian SHMT allows identification of amino acid residues involved in substrate binding and catalysis. It also reveals that the two obligate dimers making up a tetramer are not equivalent; one can be described as "tight-binding" and the other as "loose-binding" for folate. Both active sites of the tight-binding dimer are occupied by 5-formyltetrahydrofolate (5-formylTHF), whose N5-formyl carbon is within 4 Å of the glycine α-carbon of the glycine−pyridoxal phosphate complex; the complex appears to be primarily in its quinonoid form. In the loose-binding dimer, 5-formylTHF is present in only one of the active sites, and its N5-formyl carbon is 5 Å from the glycine α-carbon. The pyridoxal phosphates appear to be primarily present as *geminal* diamine complexes, with bonds to both glycine and the active site lysine. This structure suggests that only two of the four catalytic sites on SHMT are catalytically competent and that the cSHMT−glycine−5-formylTHF ternary complex is an intermediate state analogue of the catalytic complex associated with serine and glycine interconversion.

Serine hydroxymethyltransferase (SHMT,[1] EC 2.1.2.1) catalyzes the reversible conversion of serine and tetrahydrofolate (THF) to glycine and 5,10-methyleneTHF. When serine cleavage is catalyzed, this reaction generates single carbon units that are required for purine and thymidine biosynthesis, and for the remethylation of homocysteine to methionine (*1*). The importance of this reaction in DNA synthesis has focused attention on SHMT as a potential target for the development of antineoplastic and antimicrobial agents.

In mammals, SHMT is present in both mitochondria and cytoplasm. The amino acid sequences of human enzymes are 63% identical (*2*), and the enzymes are encoded by two distinct genes that have been cloned and characterized (*3*, *4*). The mSHMT message and activity are present at similar levels in all tissues, and the mSHMT-catalyzed reaction is the primary source of both newly synthesized glycine and one-carbon units for cytoplasmic metabolism (*1*, *3*, *5*). The cSHMT message and protein display a wide range of tissue specific expression. While the cSHMT protein appears to be present in all tissues, the protein and message are highly enriched in kidney, liver, and skeletal muscle (*3*). The human cSHMT hnRNA undergoes tissue specific alternative splicing, resulting in mRNA species that lack either exon 9 or exons 9 and 10 (*3*).

Many aspects of the SHMT catalytic mechanism remain uncertain and unproven as described by Matthews and Drummond (*6*, *7*). To date, most evidence seems to suggest that SHMT catalyzes the retroaldol cleavage of serine to glycine and formaldehyde (Figure 1). This mechanism proposes that the incoming amino acid, either serine or glycine, attacks the Schiff base between the active site lysine and the PLP. Partial displacement of the lysine results in the formation of the *geminal* diamine complex (Figure 1, structure **I**), while full displacement of the lysine results in the formation of a Schiff base between the amino acid and PLP, the external aldimine complex (Figure 1, structure **II**). As proposed by Dunathan (*8*), the amino acid moiety that

† X-ray data were collected at the Cornell High Energy Synchrotron Source (CHESS), which is supported by NSF Grant DMR-9311772, using the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by NIH Grant RR-01646. This study was also supported in part by U.S. Public Health Service Grant DK49621 to P.J.S.

* Corresponding author. Phone: (607) 255-9751. Fax: (607) 255-9751. E-mail: PJS13@cornell.edu.

‡ Department of Molecular Biology and Genetics.

§ Division of Nutritional Sciences.

[1] Abbreviations: THF, (6*S*)-tetrahydrofolate; methyleneTHF, (6*R*)-methylenetetrahydrofolate; 5-formylTHF, (6*S*)-5-formyltetrahydrofolate; PLP, pyridoxal 5′-phosphate; cSHMT, cytoplasmic serine hydroxymethyltransferase; mSHMT, mitochondrial serine hydroxymethyltransferase.

10.1021/bi000635a CCC: $19.00
Published on Web 10/07/2000

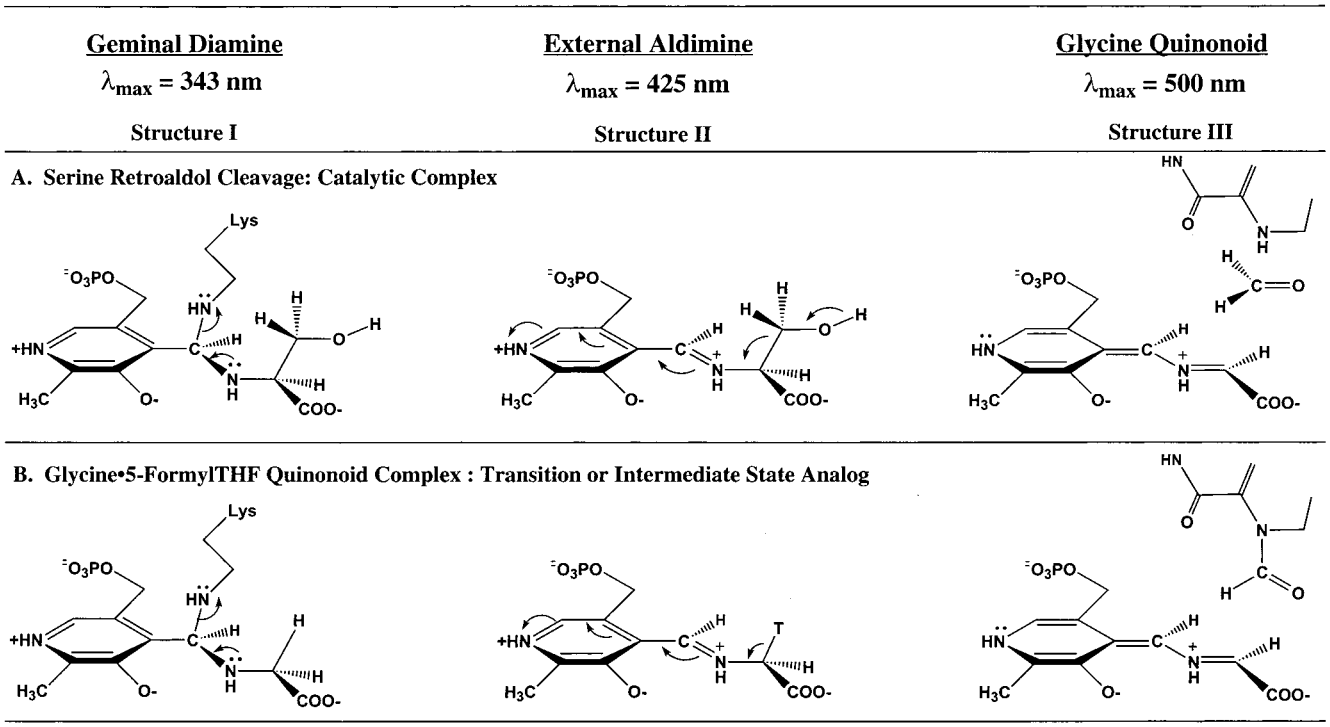

FIGURE 1: Structures associated with (A) the SHMT serine retroaldol cleavage catalytic mechanism and (B) formation of the inhibitory SHMT−glycine−5-formylTHF ternary complex.

resides orthogonal to the plane of the PLP ring is then cleaved (*9*, *10*). In the direction of serine cleavage, the hydroxymethyl group for serine is in this orientation, while for serine synthesis, the pro-2*S* proton of glycine is orthogonal to the PLP ring. Both of these "half" reactions result in the formation of a PLP−glycine quinonoid complex (Figure 1, structure **III**). When catalyzing serine cleavage, formaldehyde is sequestered and removed from the active site by the THF cofactor.

SHMT catalyzes a second reaction, the glycine-dependent irreversible conversion of 5,10-methenylTHF to 5-formylTHF. The SHMT catalytic mechanism for this reaction shares several similar intermediates with the SHMT-catalyzed interconversion of methyleneTHF and THF (*11*, *12*). Once formed, 5-formylTHF remains bound to the enzyme (*13*). In fact, 5-formylTHF polyglutamates, but not the monoglutamate species, are slow-binding, slow-releasing inhibitors of mammalian SHMT enzymes (*13*). Biophysical studies have suggested that the formation of the cSHMT−glycine−5-formylTHF ternary complex, with both mono- and polyglutamate forms of the folate, is associated with a conformational change in the enzyme, as well as a shift in the equilibrium of the PLP−glycine complex to its quinonoid form (*13*). Previous spectral studies of the rabbit cSHMT enzyme indicate that the PLP quinonoid, external aldimine and *geminal* diamine intermediates exist in equilibrium within the cSHMT−glycine−5-formylTHF ternary complex, with 50% of the active sites present in the quinonoid intermediate, and the other 50% existing as the *geminal* diamine and external aldimine intermediates (*14*). Additionally, titration calorimetric studies suggest that only half of the 5-formylTHF binding sites are occupied within the tetramer of the cSHMT−glycine−5-formylTHF ternary complex (*15*).

The cSHMT−glycine−5-formylTHF quinonoid ternary complex has been suggested to be a transition or intermediate state analogue of the cSHMT−serine−THF catalytic complex (Figure 1, structure **III**) (*13*). The cSHMT−glycine−5-formylTHF ternary complex has been shown to exist in human cultured cells (*16*), suggesting that the cSHMT enzyme exists both in a catalytically active form and as an inactive ternary complex.

Mammalian SHMT is a member of the α class of pyridoxal phosphate-dependent enzymes. The enzyme is a homotetramer with 55 kDa monomeric subunits that has been described as a dimer of obligate dimers (*17*). Amino acid residues from each of the monomeric subunits within the obligate dimers are required to make a single active site (*17*, *18*). In contrast, *Escherichia coli* cSHMT is a dimer in solution and lacks critical amino acid residues at the tetramer interface that are present in mammalian SHMT enzymes (*19*). Crystal structures of unliganded human and rabbit cSHMT have been determined, as well as that of *E. coli* SHMT with bound glycine and 5-formylTHF (*17*−*19*).

In this paper, we describe the structure of the murine tetrameric cSHMT−glycine−5-formylTHF ternary complex at 2.9 Å resolution. We designate the active sites as I−IV; site I includes mostly residues from monomer A but also some from B, and site II includes mostly residues from B but some from A. Likewise, sites III and IV are composed of residues from C and D. This new structure represents a catalytic intermediate state complex and permits the identification of key catalytic residues as well as conformational changes in the mammalian enzyme associated with catalysis.

EXPERIMENTAL PROCEDURES

Preparation and Crystallization of the Enzyme. The murine cSHMT cDNA was generated by RT-PCR using the high-fidelity *Pfu* polymerase. Mouse liver (strain 129/Sv) cDNA was used as a template and generated from total RNA (Purgene, Inc.) and a poly(T) primer. The murine cSHMT

open reading frame was amplified using the forward primer 5′-GCGCCATATGGCAGACAGGGATGCCAC-3′ and the reverse primer 5′-GCGCAAGCTTTCAGAAGTCAGGAAG-GCC-3′. The primers contain *Nde*I and *Hin*dIII sites, respectively. The PCR product was cloned into the pET28a-(+) expression plasmid in frame with a 5′-polyhistidine tag. The nucleotide sequence of three independently derived cDNA clones was confirmed by nucleotide sequencing at the Cornell Biotechnology Center's DNA sequencing facility, and all were found to be identical with the published partial nucleotide sequences (*20*). The cDNA sequence has been deposited in GenBank, accession number AF237702. The selenomethionine-substituted protein was expressed in *E. coli* B834(DE3) (Novagen) following IPTG induction for 4 h. The protein was purified on a metal affinity column (Clontech) following the manufacturer's instructions. The protein was dialyzed overnight at 4 °C against 4 L of 100 mM potassium phosphate (pH 7.5), 50 mM glycine, 5 mM DTT, and 0.1 μM PLP and stored at −80 °C. Throughout this paper, amino acid residues of the mouse cSHMT protein have been numbered on the basis of alignment with the human protein which serves as the reference protein (*2*).

Crystals of both native and selenomethionyl (SeMet) cSHMT were prepared, using the hanging drop vapor diffusion method at room temperature (22 °C). The composition of the final crystallization solution was ∼8% polyethylene glycol (PEG) 8K, 6% ethylene glycol, 0.1 M HEPES (pH 7.7), 200 μM (6*S*)-5-formylTHF, 100 mM glycine, and 4 mg/mL cSHMT protein. Thin rod-shaped crystals appeared overnight and continued to grow for 3−4 days, attaining final diameters of 0.02−0.04 mm and lengths of at least 0.15−0.25 mm. All crystals displayed the characteristic red color associated with the formation of the cSHMT−glycine−5-formylTHF quinonoid ternary complex (*13*).

X-ray Diffraction Data Collection and Processing. Crystals were soaked for a few seconds in a cryoprotectant solution containing 9% PEG 8K, 5% ethylene glycol, 0.08 M HEPES, and 16% glycerol and flash-frozen in liquid nitrogen. Data were collected from both native and SeMet SHMT crystals at the Cornell High Energy Synchrotron Source (CHESS) using the monochromatic station F-1. The highest-resolution data were obtained from SeMet crystals; native crystals consistently diffracted to poorer resolution. From a single cryocooled SeMet crystal, diffraction data were collected to 2.9 Å by the oscillation method, using a Quantum-4 CCD detector from Area Detector Systems Corp. (*21*), and processed with the DPS/Mosflm/CCP4 suite of programs (*22*, *23*). Dps_index (*24*) was used for indexing, MOSFLM (*25*) for refinement and integration, SCALA (*26*) for scaling and merging, and TRUNCATE (*27*) to produce final structure amplitudes. A summary of the data processing statistics is given in Table 1.

The crystals exhibited tetragonal symmetry, in space group $P4_12_12$ or $P4_32_12$, and unit cell dimensions of 142.5 Å × 142.5 Å × 270.9 Å. Diffraction was anisotropic, being stronger along *c* than along *a* and *b*. High redundancy in the data allowed use of the exceedingly weak high-resolution reflections. Anomalous pairs were kept separate during scaling and merging of individual oscillation images, but the anomalous signal was not used for phasing.

Phase Calculations. Phases were determined by the molecular replacement method, using the CCP4 program AMoRe (*28*). The model structure was the rabbit SHMT obligate dimer [Protein Data Bank entry 1CJ0 (*29*)] modified by omission of the external loops comprising residues 269−287 of each monomer (in the numbering system used here) and by conversion to alanine of all residues which differ between the rabbit and mouse proteins (about 10% of the total). Two PLP cofactors per dimer were included. From volume considerations, two dimers per asymmetric unit were expected, and two good solutions were found. The space group was determined to be $P4_32_12$, which gave considerably better statistics than $P4_12_12$. The two top solutions together describe a dimer of dimers arranged very similarly to those in rabbit and human SHMT (*17*, *18*). After rigid-body refinement, the two-dimer solution had a correlation coefficient of 79% and an *R*-factor of 39%. Additional details are given in Table 1.

Table 1: X-ray Data Reduction, Phasing, and Model Refinement Statistics

Data Reduction	
resolution range (Å)[a]	30−2.9
total no. of observations	1046890
no. of rejected[b] observations	7530
no. of rejected[c] reflections	567
no. of final independent reflections	57387
R_{sym}[d]	0.203
$I/\sigma(I)$	3.4
mean *I*/SD (mean *I*)	12.8
completeness	92.8% (100% in the 3.0−2.9 Å range)
multiplicity[e]	10.6 (10.9 in the 3.0−2.9 Å range)
Phasing by Molecular Replacement	
resolution range used (Å)	10−3.0
correlation coefficient/*R*-factor	
dimer 1	0.587/0.496
dimer 2	0.589/0.502
dimers 1 and 2 after rigid-body refinement	0.795/0.387
Model Refinement	
no. of total reflections used	57357
no. of reflections in the test set	5780
no. of total non-hydrogen atoms	14958
overall *B*-factor correction[f]	$B_{11} = B_{22} = -12.8$ Å^2, $B_{33} = 25.6$ Å^2
bulk solvent correction	density level of 0.35 e/Å^3, *B*-factor of 65 Å^2
mean *B* from grouped *B* refinement (Å^2)	85
R/R_{free}	
after stage 1	0.334/0.368
after stage 2	0.294/0.350
after stage 3	0.243/0.289
final	0.228/0.272
rmsd	
bond lengths (Å)	0.008
bond angles (deg)	1.4

[a] Ice rings at 3.96−3.84 and 3.72−3.64 Å were excluded from processing. [b] $E > 10$ for acentric; $E > 12$ for centric reflections, or $|I - I_{ref}| > C\sigma_{ref}$, where I_{ref} is the mean of symmetry-related observations excluding *I*, σ_{ref} is the standard deviation of I_{ref}, and $C = 3$ for sets of more than two observations and $C = 4$ for sets of two observations. [c] $I < -3.7\sigma(I)$, or $I < (\sigma^2/I_{mean} - 4\sigma)$. [d] $R_{sym} = \Sigma|I - I_{ref}|/\Sigma I$, where I_{ref} is the mean of related I^+ or I^- observations. [e] Keeping I^+ and I^- separate. [f] Applied to the F_{obs}.

Electron Density Maps and Model Refinement. A $2F_o - F_c$ electron density map was calculated from the rotated and translated models using the program CNS, version 0.9a (*30*), and performing density modification to reduce model bias; a solvent mask was automatically generated, followed by density truncation in protein regions and solvent flipping in

solvent regions. When examined with O, version 6.22 (*31*), the initial map suggested that mouse liganded cSHMT is similar in overall conformation to the unliganded rabbit and human proteins, but that significant differences exist in several regions, particularly in the C-terminal domains of each monomer.

Structure refinement was initiated using a model including only residues 15−272 and 285−346 of each monomer, and no PLP. Residues that differ in rabbit and mouse proteins were mutated from Ala to the proper mouse amino acid only where clear side chain density was visible in the initial map. Subsequent steps were carried out using CNS (for refinement and map generation) and O (for examination of maps and manual rebuilding) in several stages.

In stage 1, monomer A was manually adjusted to better agree with the current electron density maps, and was then used to generate monomers B−D using noncrystallographic symmetry (NCS) operations. The tetramer was then refined using 4-fold NCS restraints throughout the molecule. Anisotropic overall *B*-factor and bulk solvent corrections were applied, and simulated annealing molecular dynamics were used to refine the atomic coordinates against diffraction data from 30 to 3 Å. New $2F_o - F_c$ and $F_o - F_c$ maps were calculated using density modification, and the cycle was repeated.

In stage 2, monomer A alone continued to be modified and used to generate monomers B−D. Refinement of grouped *B*-factors and atomic coordinates was carried out by energy minimization without simulated annealing, with the same 4-fold NCS restraints as before, and maps were calculated without density modification.

Next, monomers A−D were each manually modified to fit the new maps (stage 3). NCS restraints were removed in regions where the maps suggested variations among the monomers, and were relaxed for all side chain atoms, and refinement proceeded using the data out to 2.9 Å. Data from 3.0 to 2.9 Å were very weak, but their inclusion lowered R_{free} as well as *R*.

In the final step (stage 4), simulated annealing omit maps were calculated for each portion of the molecule to be rebuilt, and the region covered by NCS restraints was further reduced, to about 75% of the molecule in the last cycles. After the final rebuilding of the model, all remaining missing side chains were added in stereochemically preferred conformations where electron density was not clear.

Residues were inserted, and side chains added to residues that had initially been modeled as alanine, as they became visible in the maps, a process that continued through all four stages. During stage 1, strong features in the electron density maps were visible for the PLP cofactor and in the expected positions of bound glycine and 5-formylTHF (in site I, which was the only one being examined at this stage), and these species were added to the model. During stage 3, 5-formylTHF was removed from site III, which clearly lacked density for folate. Four-fold NCS restraints were applied to the PLP−glycine complex and 5-formylTHF in stages 1 and 2, but not thereafter.

Initial coordinates and topology parameters for the PLP−glycine complex were generated from those for the PLP−serine complex in tryptophan synthase (PDB entry 1BEU), as stored in the HIC-Up database (*32*). Similarly, coordinates and restraints for THF as found in thymidylate synthase (PDB entry 2TDD) were obtained from the HIC-Up database.

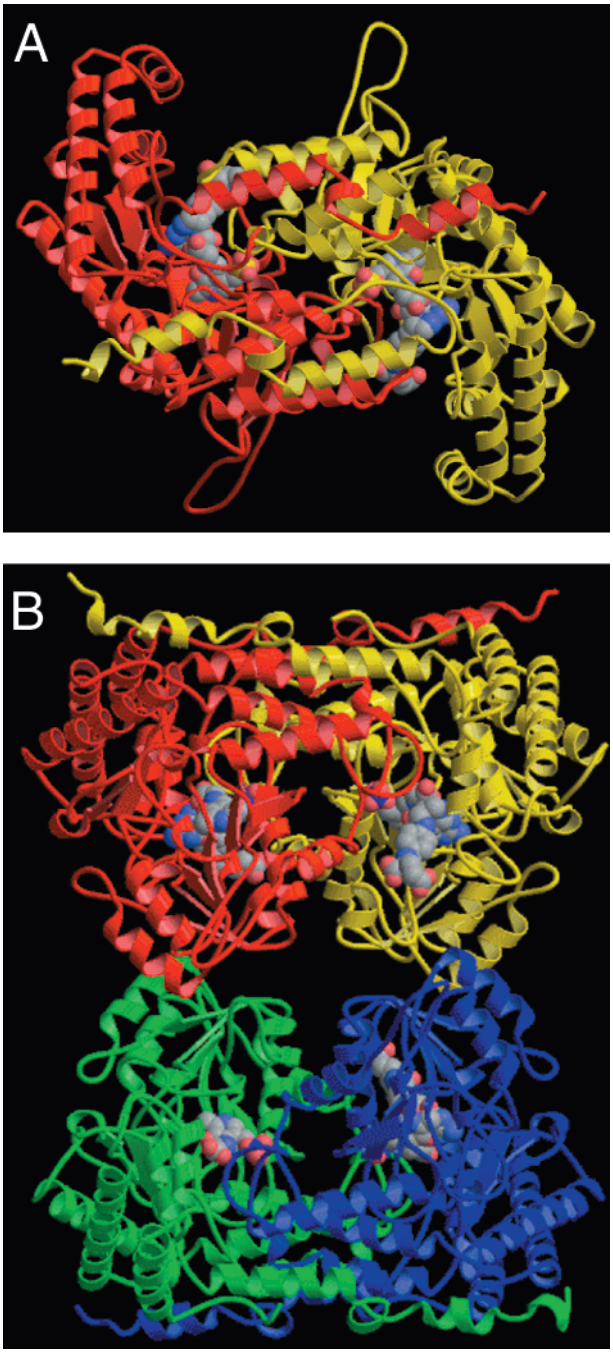

FIGURE 2: Overall structure of murine cSHMT with glycine and 5-formylTHF bound. (A) The tight-binding obligate dimer, viewed down its approximate 2-fold axis. Monomer A is in red, and monomer B is in yellow. PLP, bound glycine, and 5-formylTHF in sites I (on the left) and II (on the right) are shown in space-filling form. The loose-binding dimer appears to be very similar, except that only one folate site is occupied. (B) The complete tetramer, viewed along one of the approximate 2-fold axes relating the two obligate dimers; this view is perpendicular to that in panel A. The tight-binding dimer, containing chains A (red) and B (yellow) and active sites I (on the left) and II (on the right), is at the top, and the loose-binding dimer, containing chains C (green) and D (blue) and active sites III (on the left) and IV (on the right), is at the bottom. PLP, bound glycine, and 5-formylTHF are shown in space-filling form. This figure was drawn using MOLSCRIPT (*37*) and Raster3D (*38*).

In stage 4, the parameters for the PLP−glycine complex were altered to reflect the different conformation of the species in the I and II versus III and IV sites. The final refinement

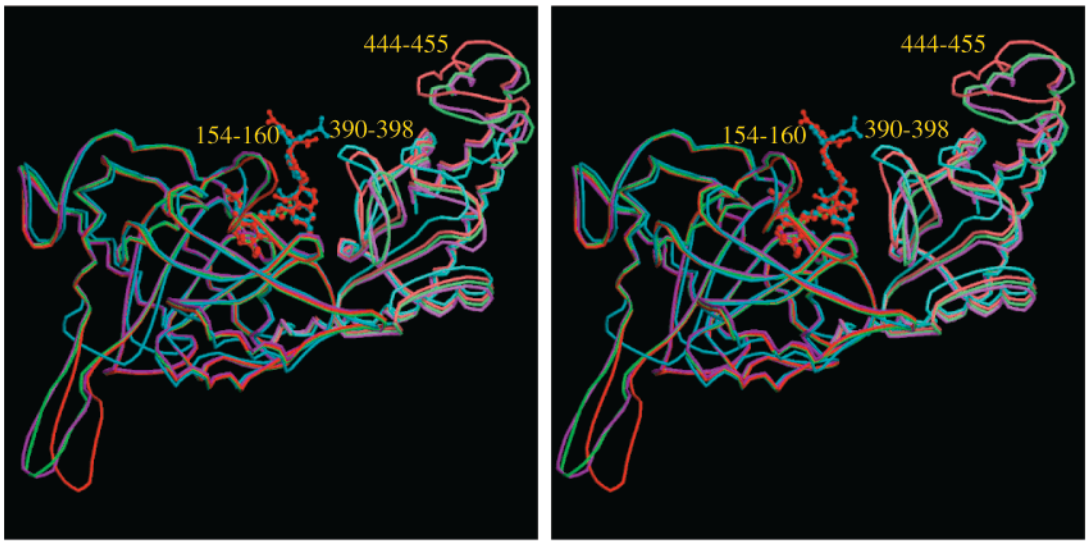

FIGURE 3: Comparison of large and small domains in tight-binding monomer A (red) and loose-binding monomer C (green) with those in human cSHMT (magenta) and *E. coli* SHMT (turquoise). Large domains (on the left, in darker colors) were aligned on the basis of residues in the central β-sheet. Small domains are on the right, in lighter colors. PLP molecules are included in ball-and-stick form, as are bound glycine and 5-formylTHF, when present. For clarity, the polypeptide chains have been truncated at both ends; residues corresponding to mouse SHMT residues 60−475 are shown. The importance of the labeled loops is discussed in the text. The extended loop at lower left, comprising residues 270−285, makes no contacts with ligands or other monomers and appears to be unimportant to SHMT structure or function. This figure was drawn using MOLSCRIPT (*37*) and Raster3D (*38*).

statistics are summarized in Table 1. Coordinates have been deposited in the Protein Data Bank (*29*), as entry 1EJI.

RESULTS

Properties of the Murine cSHMT Structure. The structures of unliganded human and rabbit cSHMT enzymes have already been reported and compared to the structures of other members of the α class of PLP-dependent enzymes (*17*, *18*), and the structure of a liganded SHMT from *E. coli* has recently been determined (*19*). The amino acid sequence of murine cSHMT is 91, 90, and 50% identical to the human, rabbit, and *E. coli* proteins, respectively. This work will focus on the structure of the murine cSHMT−glycine−5-formylTHF ternary complex and analyses of the active sites, in comparison with those of the previously known structures.

The final structure includes all protein residues (7−484) plus one PLP−glycine complex in each monomer and a 5-formylTHF molecule in three of the four active sites. Because of the limited resolution of the data, no solvent molecules have been included. The final *R*-factor was 0.23, and R_{free} was 0.27. A final ΔF map contoured at 4σ shows a few small isolated features, some of which probably represent water molecules, with the remainder being noise. Evaluation of the stereochemistry by PROCHECK (*33*) gave statistics that were better than average for a structure to 2.9 Å resolution. More than 97% of the main chain torsion angles fall in preferred or allowed regions of the Ramachandran plot, and only 1% in the "disallowed" region. As in the rabbit and human proteins, there is one *cis*-proline at residue 298, and the conformation at residues 133 and 257 is strained. Residues 356 and 456 have unfavorable ϕ and ψ angles in at least three of the monomers; these occupy tight turns between helices. Figure 2 depicts the overall structure of murine cSHMT.

Conformational Changes in the Enzyme Associated with Substrate Binding and Evidence for Nonequivalent Obligate Dimers. Several studies have suggested that SHMT undergoes a conformational change upon binding substrates (*11*, *15*, *34*). Substrate binding results in increases in the thermostability of the enzyme and the enthalpy of denaturation and a decrease in the exchange rates of backbone amide protons. The greatest conformational changes in SHMT occur when the one-carbon site is occupied, by either the hydroxymethyl group of serine or the methyl or formyl group of N5-substituted folates. Comparison of the *E. coli* SHMT structure with bound 5-formylTHF and glycine to the unliganded human and rabbit protein structure revealed a difference of about 5° in the hinge angle between the large (residues 54−321 in the human numbering) and small (residues 322−484) domains (*17*−*19*). It was unclear whether this difference was due to substrate binding or to the several deletions in the bacterial protein relative to the mammalian enzymes. The structure of the murine cSHMT−glycine−5-formylTHF ternary complex, described here, resolves this question. It indicates that substrate-induced conformational changes in mammalian SHMTs involve only small displacements, unlike the situation in the related protein aspartate aminotransferase (*35*). The overall conformation of each mouse cSHMT monomer is very similar to the human and rabbit monomers, with only a slight change in the interdomain angles (Figure 3). There are some local differences in loop regions, as discussed below. In the human and rabbit crystal structures, the two obligate dimers that comprise each tetramer are identical, and are related by an exact 2-fold axis. In the liganded mouse protein structure, this is not true. Although the tetramer appears to be very similar overall to the human and mouse forms, its symmetry is broken in small but significant ways.

The most notable feature of the murine cSHMT ternary complex is this lack of structural symmetry and the corresponding difference in substrate binding properties between the obligate dimers. One of the dimers binds folate tightly (tight-binding A−B dimer, containing active sites I and II), and the other dimer binds the folate either more loosely or

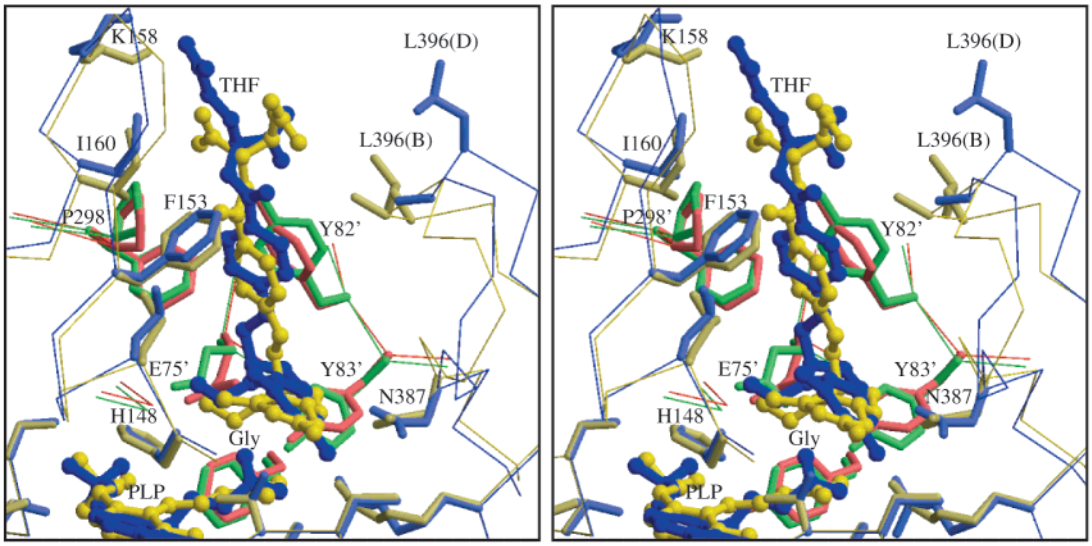

FIGURE 4: Comparison of folate binding sites II and IV. 5-FormylTHF, PLP, glycine, and some important surrounding protein side chains in site II (tight-binding) are shown in yellow, except for residues from monomer A, which are in red. Site IV (loose-binding) is in blue, except that residues from monomer C are in green. The view is similar to that of Figure 3, with the solvent-exposed surface of the molecule at the top. This figure was drawn using MOLSCRIPT (*37*) and Raster3D (*38*).

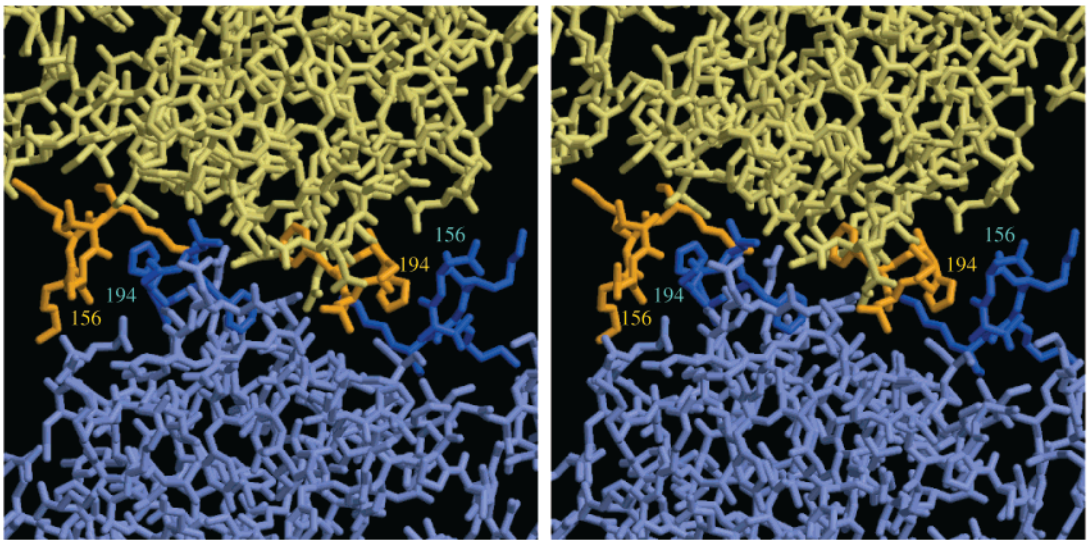

FIGURE 5: Interface between monomers B (yellow) and D (blue). Residues 155−159 and 191−194 in each monomer are highlighted in darker colors; contacts between these residues are probably critical in mediating the conformational change of the CD dimer when folate binds to the A−B dimer. The view is from the right side of Figure 2, looking toward the center of the tetramer. RasMol (*39*) was used to generate this figure, along with MOLSCRIPT and Raster3D.

not at all (loose-binding dimer CD, containing active sites III and IV). These differences in substrate binding properties of the two dimers are not associated with any major structural differences, but result from subtle shifts in conformation (Figure 3). In the crystal structure, even the two sites within each dimer are not identical; in the loose-binding dimer, only active site IV contains a bound 5-formylTHF, and there are less obvious but significant differences between tight-binding sites I and II.

Figure 4 shows differences in protein conformation around active sites II (tight-binding) and IV (loose-binding). Site I is similar to site II, and site III is similar to site IV. The most obvious difference between sites II and IV is the position of the loop of residues 390−398 (Figures 3 and 4), and one might hypothesize that the loop must be positioned as in site II for tight folate binding to occur. However, there is no obvious intra- or intermolecular interaction among SHMT chains that would cause the loops in sites III and IV to be positioned differently with respect to those in sites I and II. Instead, we suggest that the presence of a tightly bound folate induces the shift in the loop of residues 390−398. This shift results in the motion of the helix of residues 445−455 seen in Figure 3; the helix is in contact with the loop both directly and through the intervening helix of residues 369−379.

A more promising candidate for the cause of differential binding of 5-formylTHF is the shift in the loop-strand comprising residues 150−165. Hydrophobic residues from this stretch of chain form one side of the folate binding site and may be critical in promoting tight binding. Moreover, a means exists for the conformation of residues 150−165 in site IV to be altered by binding of folate to site II. Figure 5 depicts the contact region between monomers B and D, which provide most of the residues in sites II and IV,

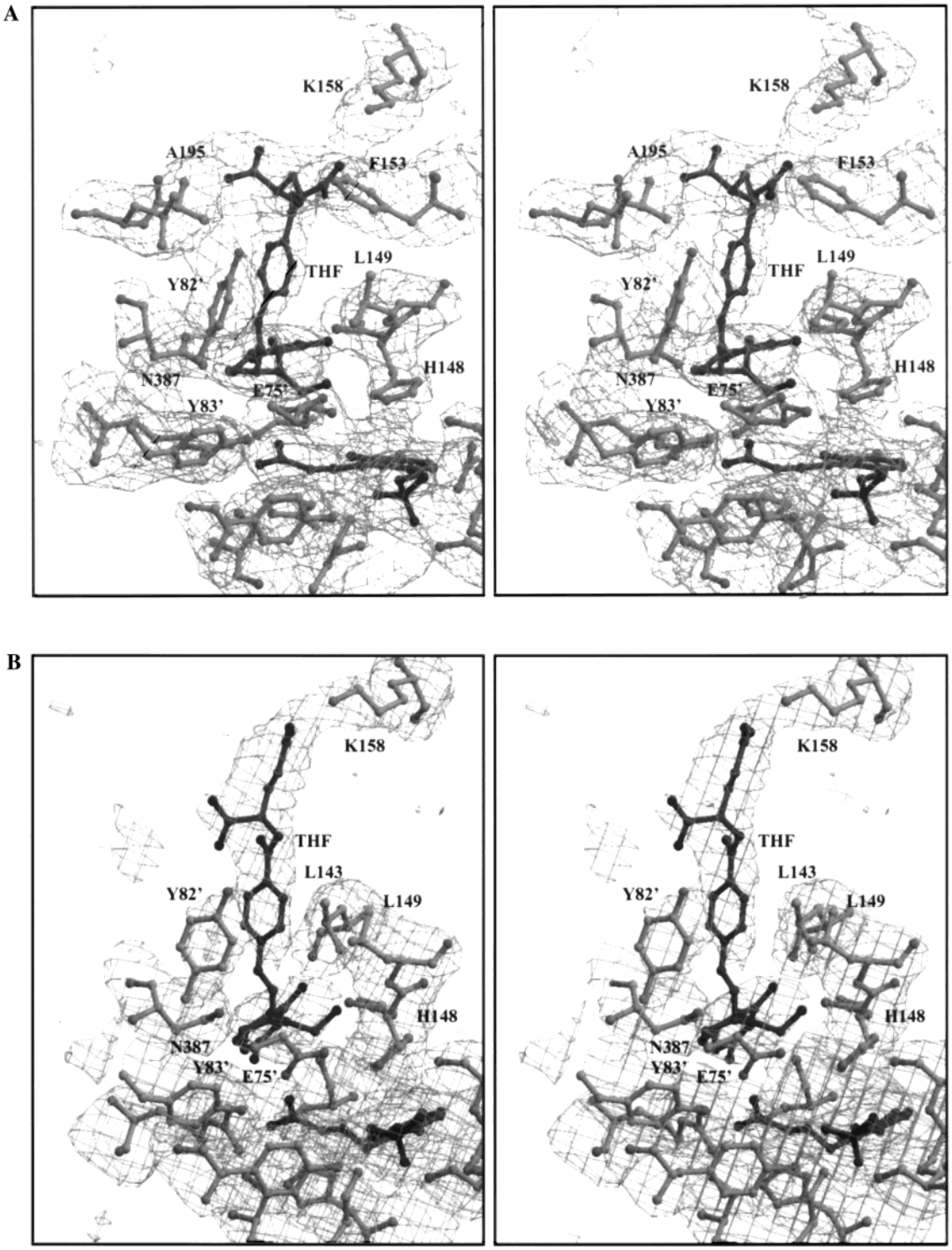

FIGURE 6: Electron density and corresponding molecular structure for 5-formylTHF and environs. ΔF omit maps were calculated for each site from a model with PLP, bound glycine, 5-formylTHF, and a small surrounding region omitted, and the remaining structure was refined by simulated annealing. The contour level is 2.5σ. 5-FormylTHF and the PLP−glycine complex are in dark gray, and the protein is in medium gray: (A) site I and (B) site IV. This figure was drawn using MOLSCRIPT (*37*) and Raster3D (*38*).

respectively. The tetramer contact surface is small, involving only residues 135−137, 154−157, 168−171, and 189−194 of each monomer. This region and a similar one between monomers A and C are the only contacts between obligate dimers. We propose that tight binding of 5-formylTHF to site II results in a shift of the loop of residues B154−B160. This shift is transmitted to D via His194, leading to an overall displacement of monomer D. However, the interaction of residues D154−D160 with residue B194 restricts the movement of the loop of residues D154−D160, thereby altering its position with respect to the rest of monomer D, with a corresponding alteration in the folate binding properties of site IV. Similarly, binding of 5-formylTHF to site I alters the binding properties of site III.

Site IV binds 5-formylTHF in a different conformation compared to sites I and II (Figure 6), with significant alteration in the protein−ligand contacts. Site III, however, despite its similarity to IV, is empty of ordered 5-formylTHF.

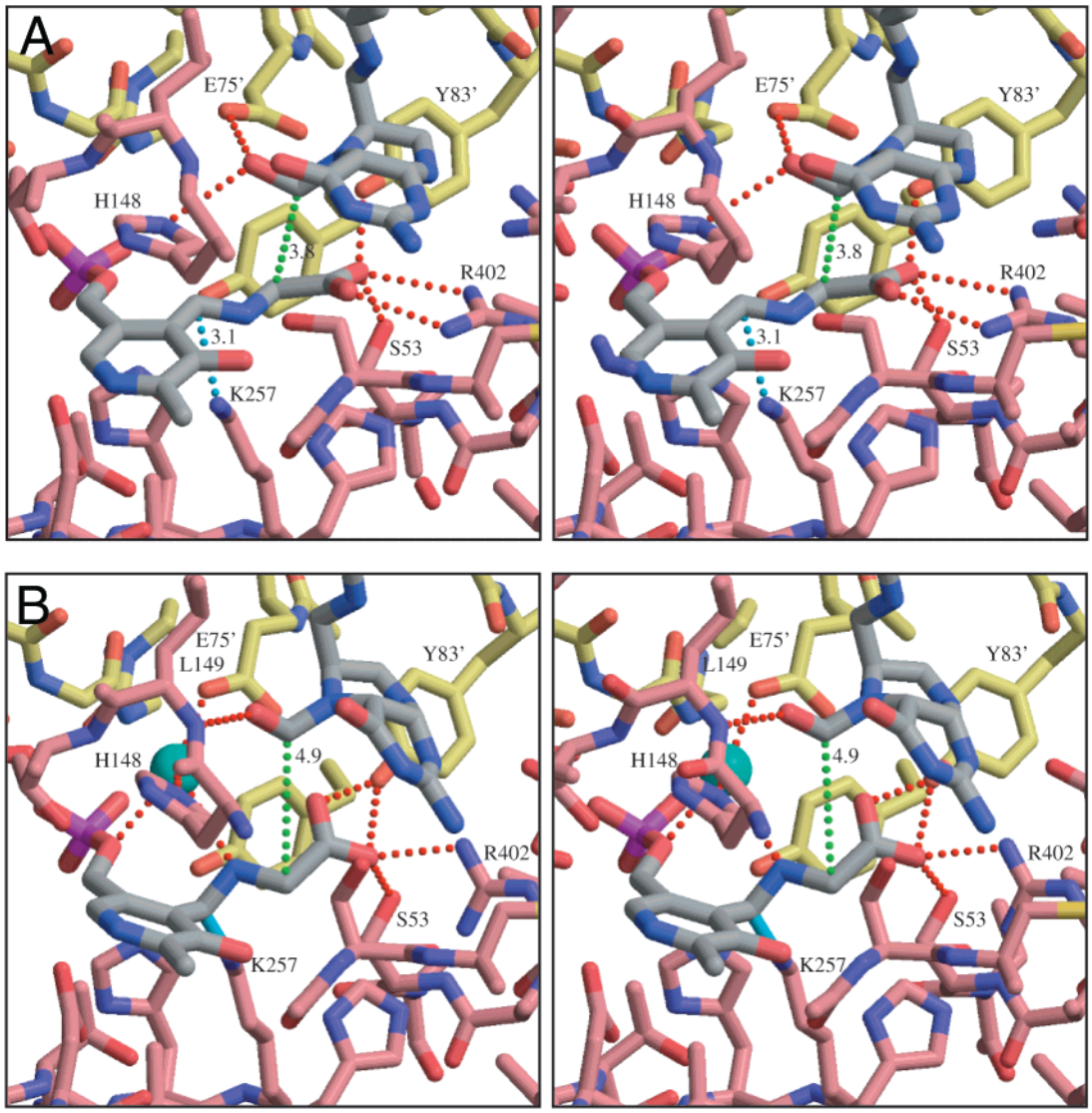

FIGURE 7: Detail of active sites: (A) site I and (B) site IV. The two protein chains involved are distinguished by pink and yellow colors of the carbon atoms; 5-formylTHF, PLP, and bound glycine are in standard CPK colors. Hydrogen bonds involving the glycine and the 5-formyl group of THF are shown in red. The distance between C_α atoms of glycine and C atoms of the 5-formyl group is indicated in green. K257 N_ζ is bonded to C4A of PLP in site IV; in site I, it is 3.1 Å away, as shown in blue. This figure was drawn using MOLSCRIPT (*37*) and Raster3D (*38*).

The reason for the difference between sites III and IV is unclear. Interaction between sites III and IV is clearly possible, as both involve the same polypeptide chains (monomers C and D). Hence, binding of folate to site IV may have a subtle effect on the conformation of site III, such that its affinity is reduced. Alternatively, the different crystal contacts involving the two monomers may be responsible. The same effects result in small but clearly visible differences between sites I and II.

Analysis of the Active Site. Figures 7 and 8 depict the architecture of the cSHMT active site for both the tight-binding and loose-binding dimer active sites. For each site, residues from the monomer having fewer contacts with that site are designated by primed sequence numbers. The contacts with the phosphate moiety of the PLP cofactor are largely similar among the four active sites. The side chains of Ser119, Ser121, and Tyr73′ donate hydrogen bonds, and the main chain amide nitrogens of Gly120, Ser121, and Gly303′ also make contacts with the phosphate. The side chain of His148 is stacked with the PLP. In the tight-binding dimer active sites, side chains of His231, Ser203, and Lys257 are within hydrogen bonding distance of O3 of the PLP ring. In the loose-binding dimer, the contacts with O3 of the PLP ring are similar, with the exception that Ser203 is not within hydrogen bonding range. The PLP environment, including most of the hydrogen-bonding amino acid residues and the ring-stacking histidine, is very similar to that found in previously determined SHMT structures (*17*−*19*).

The catalytic environment surrounding the glycine substrate differs among the active sites. At 2.9 Å resolution, fine details cannot be distinguished, but careful refinement and examination of electron density maps suggest the following; in tight-binding sites I and II, the N_ζ of Lys257 (the lysine that forms a Schiff base with the C4A of PLP in the unliganded enzyme) is about 3 Å from the C4A of PLP and is probably hydrogen bonded to PLP O3. A mixture of *geminal* diamine and external aldimine/quinonoid states (Figure 1) is present, with site I having mostly the latter and site II a more equal mixture. The reddish color of the crystals suggests the presence of substantial amounts of the quinonoid

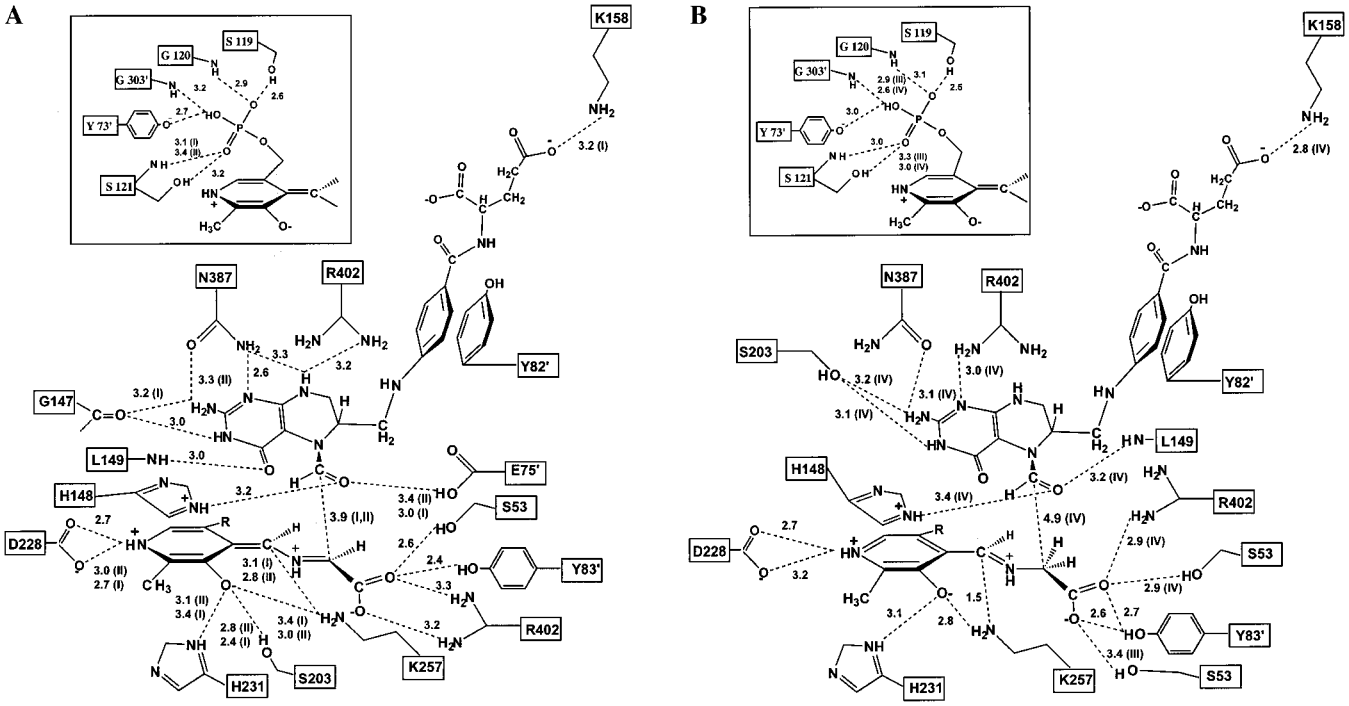

FIGURE 8: Schematic representation of the cSHMT active site in the tight-binding (A) and loose-binding (B) dimers. All contact distances are reported in angstroms and differ by less than 0.2 Å between the active sites within the dimer, unless otherwise noted. A cutoff of 3.5 Å was used to determine possible hydrogen bonds. Amino acid residues that contribute to the active site are shown in boxes. Residues that contribute to an active site that originate from the opposite monomer within the obligate dimer are indicated with a prime.

form. In the loose-binding sites, the bonding between Lys257 N_ζ and PLP C4A is considerably stronger, and most complexes are in the *geminal* diamine state; some fraction may lack glycine and contain the lysine−PLP Schiff base of the unliganded form. Residues Arg402, Tyr83′, and Ser53 anchor the glycine carboxyl group in a similar but not identical manner in the tight-binding and loose-binding dimers. Probable hydrogen bonds in the active sites are shown in Figures 7 and 8.

The folate binding sites are clearly distinct between the tight-binding and loose-binding dimer active sites (Figures 4 and 8), and it appears that it is binding of folate to one dimer that reduces the affinity of the second dimer for 5-formylTHF. This observation of negative cooperativity explains previous titration calorimetric studies that suggested only half of the 5-formylTHF binding sites were occupied within the cSHMT tetramer (*15*). When these sites are compared with the corresponding site in *E. coli* SHMT (*19*), there is a general overall resemblance but numerous differences in detail. The *E. coli* folate-binding site is the tightest; N5 of 5-formylTHF is 4.0 Å from C_α of the bound glycine, and the surrounding loops enclose the ligand tightly. In sites I and II, the distance between N5 of 5-formylTHF and C_α of glycine is 4.3 Å and the loop of residues 390−398 is somewhat farther from the ligand than the corresponding *E. coli* loop. In site IV, the distance is increased further, to 4.8 Å, and the loop of residues 390−398 is swung away to make a more open site. The *p*-aminobenzyl ring of 5-formylTHF is stacked with Tyr82′ in all of the occupied active sites. In sites I and II, the long axis of the quinazoline ring of the folate is oriented approximately 90° to the *p*-aminobenzylglutamate (PABA) moiety; the ring is relatively flat and has a propeller twist of about 25° (relative to a plane perpendicular to the PABA). This is similar to the conformation of the 5-formylTHF in the *E. coli* protein (*19*). In site IV, the long axis of the ring is similarly oriented, but the ring is much more nonplanar, with a propeller twist varying from ∼40 to 70°. The γ-carboxylate of the folate glutamate forms a hydrogen bond to the side chain of Lys158 in active sites I and IV. The orientation of the 5-formylTHF glutamate moiety and its (few) contacts with the protein are different in each of the three active sites (Figures 4 and 6).

The quinazoline environment is clearly different in the tight-binding and loose-binding dimers; there are almost no common interactions with protein. In tight-binding dimer active sites I and II, the quinazoline is hydrogen bonded to the main chain carbonyl of Gly147 at N3 and (in site I) at the C2 amino group. The Asn387 side chain is hydrogen bonded with N1, N8, and (in site II) the C2 amino group. Arg402 also forms a hydrogen bond with N8. The Leu149 main chain nitrogen forms a hydrogen bond with the C4 carbonyl of THF. In the loose-binding dimer, Ser203 replaces the contacts made by the main chain carbonyl of Gly147, and Arg402 interacts with N1 of THF. In active site IV, there are no apparent contacts with the C4 carbonyl or N8 of THF.

The THF N5 formyl group is also anchored differently in the two dimers. In tight-binding dimer sites I and II, the formyl carbonyl group is within hydrogen binding distance of the side chains of His148 and Glu75′. In site IV, His148 is also involved, but the main chain amino group of Leu149 replaces the Glu75′ side chain. In the *E. coli* SHMT structure, the formylquinazoline group of 5-formylTHF makes contacts similar to those in site I (*19*).

The X-ray structure active site of thymidylate synthase (TS) with bound 5-fluorouracil and methyleneTHF reveals the presence of several ordered water molecules. Mutational studies have shown that none of the active site residues are completely essential for TS activity, and it has been suggested that these water molecules are the only essential acid and base catalysts (*36*). In contrast, in cSHMT there is insufficient

space available to position multiple water molecules in the active site. The final ΔF map for murine SHMT has peaks corresponding to one possible H_2O site each in active sites II and IV (none visible in site I or III). In site II, the water molecule is coordinated by the side chains of His148 (2.8 Å) and Glu75′ (2.4 Å), and by O3 of the THF (2.7 Å). This corresponds to the location of a bound water molecule in the *E. coli* SHMT structure. In site IV, the coordinating moieties include the side chains of His148 (2.8 Å) and Glu75′ (2.4 Å), the O4 of the PLP cofactor (3.1 Å), and the glycine nitrogen (3.2 Å). These water molecules may help to stabilize the precise conformation of the ligands in sites II and IV, but it is unlikely that ordered water participates as a catalyst in the SHMT reaction mechanism.

Is the SHMT−Glycine−5-FormylTHF Ternary Complex a Transition State Analogue of the Catalytic Complex? If the SHMT−glycine−5-formylTHF ternary complex is a structural analogue of the catalytic complex, then the N5 formyl group should occupy the one-carbon transfer site between the glycine anion and N5 of THF. Additionally, the geometry of the active site should be appropriate for catalysis. In active sites I and II, the N5 formyl carbon and C_α of the glycine anion are separated by 3.9 Å. Using this structure as a model, the angle for nucleophilic attack by the N5 iminium cation of methyleneTHF on the glycine C_α, as determined by the angle formed by the glycine C_α, the formyl carbon, and N5 of THF, is 135°. This angle is appropriate for S_N2 nucleophilic attack, but the distance between the formyl carbon and glycine anion is not optimal. In site IV, the C_α resides 4.9 Å from the formyl carbon, and nucleophilic attack by the N5 iminium cation at the glycine C_α would occur with an angle of 110°. If the formyl group occupies the same relative position as liberated formaldehyde, active sites I and II appear to be competent for catalysis, and the formaldehyde is oriented appropriately. In contrast, both the distance and attack angle suggest that site IV is not competent for catalysis.

DISCUSSION

Relevance of Structure to SHMT Catalysis. The elucidation of the mouse cSHMT ternary complex structure provides both insights into the enzyme's mechanism and a structural rationale for previous biophysical and mechanistic studies of the enzyme mechanism. This structure enables examination of the SHMT active site in both the *geminal* diamine (active sites III and IV) and the (mostly) quinonoid ternary complex (active sites I and II).

Previous spectral studies suggested that the addition of saturating levels of glycine and 5-formylTHF to cSHMT resulted in 50% of the enzyme adopting the quinonoid complex, while 50% was in the *geminal* diamine and external aldimine complex (*13*). These results were interpreted to mean that all three PLP−glycine complexes (Figure 1) exist in equilibrium within all of the active sites of the enzyme. The X-ray structure now demonstrates that only the active sites in the tight-binding dimer have any substantial fraction of the SHMT−glycine−5-formylTHF quinonoid ternary complex; this complex is most likely in equilibrium with the external aldimine complex. The active sites in the loose-binding dimer clearly exist in the *geminal* diamine form.

These studies also suggest that only the active sites in the tight-binding dimer of cSHMT are competent for catalysis. The inability of 5-formylTHF to effectively bind to the loose-binding dimers suggests that these sites are not capable of catalyzing serine cleavage, a THF-dependent reaction. Additionally, we may infer that they are not competent for catalyzing the folate-independent cleavage of allothreonine to glycine and acetaldehyde. Previously, 5-methylTHF and 5-formylTHF, and their polyglutamate derivatives, were found to be partial competitive inhibitors of allothreonine cleavage (*13*). At saturating levels of an N5-substituted inhibitory folate, the rate of allothreonine cleavage is equal to and limited by the rate of release (off rate) of the product, glycine, from the quinonoid ternary complex (*13*). This off rate is dependent upon both the nature of the N5 substitution and the length of the polyglutamate. Since release of the product glycine from the enzyme is both limiting for allothreonine cleavage and dependent upon the nature of the folate bound, it is likely that only SHMT active sites that bind 5-formylTHF tightly are catalyzing the cleavage of allothreonine. Therefore, the active sites in the loose-binding dimer, which do not bind folate tightly, are not catalytically competent for folate-independent allothreonine cleavage.

The mammalian cSHMT enzymes also catalyze the folate-independent exchange of the pro-2*S* proton of glycine with a k_{cat} of 0.02 s^{-1}. This rate can be increased by the addition of THF, but is decreased to undetectable levels by the addition of 5-formylTHF (*13*). The latter observation implies that only sites which bind 5-formylTHF can be involved, and is further evidence that the active sites in the loose-binding dimer are not competent to exchange the pro-2*S* proton of glycine with solvent and are therefore catalytically inactive.

By examining the active sites in the loose-binding and tight-binding dimers, we are able to identify key amino acid residues that are responsible for catalyzing the *geminal* diamine−quinonoid interconversion. The key mechanistic step in the conversion of the external aldimine to the glycine quinonoid is the abstraction of the pro-2*S* proton of glycine. Lys257 is the best candidate base for pro-2*S* proton abstraction. Its N_ζ atom resides 3.6 Å from the α-carbon of glycine in tight-binding dimer active sites I and II, and 2.6 Å from the α-carbon of glycine in loose-binding dimer active sites III and IV. His231 is a less likely candidate base; its $N_{\delta 1}$ atom is about 4.9 Å away from the glycine C_α in sites I and II and about 4.4 Å away in sites III and IV. The difference in distances between the two dimers is not primarily due to movement of His231 and Lys257, but rather to differences in the PLP−glycine conformation. The structure also suggests that Glu75′, which forms a hydrogen bond with the formyl carbonyl moiety, is important in orienting the hydroxymethyl group of serine in an orientation that is appropriate for dehydration of the serine aldimine, and likely participates in catalysis. This contact is not present in the *geminal* diamine complexes.

At least three mechanisms have been proposed for the reversible transfer of formaldehyde from serine to N5 of THF. Two of these mechanisms predict the formation of free formaldehyde as an intermediate, while the third suggests direct attack of the THF on the serine aldimine (the direct displacement mechanism) (*6*). The retroaldol cleavage mechanism predicts the generation of free formaldehyde from the serine aldimine in a THF-independent step, followed by nucleophilic attack on the free formaldehyde by N5 of THF. An alternative mechanism suggests that formaldehyde is

trapped as a thiolhemiacetal, and then transferred to N5 of folate. The structure of the cSHMT−glycine−5-formylTHF ternary complex indicates that a direct displacement mechanism, requiring nucleophilic attack by the glycine anion at the N5 iminium cation (shown in our structure as the N5 formyl group), cannot be ruled out. The relative geometry of active sites I and II is optimal for S_N2 nucleophilic attack, although the distance between the formyl carbon and the glycine anion (3.9 Å) is rather long for this mechanism. Similar limitations exist for the retroaldol cleavage mechanism, as free formaldehyde would be expected to occupy the same site as the N5 formyl carbon, resulting in optimal geometry for attack, but the attack distance is again not optimal. The closest cysteine to the active site, Cys204, appears to be too far away to permit formation of a thiolhemiacetal, thus ruling out the third possible mechanism.

This structure also enables us to predict the function of the alternatively spliced forms of the human cSHMT enzyme (*16*). Alternative splicing resulting from loss of exons 9 and 10 in the human mRNA would result in the loss of residues 273−351. None of these residues contribute to the catalytic active site or subunit interfaces, and therefore, these spliced variants would be expected to have catalytic activity. However, deletion of these residues would probably affect the positions of residues Asn387 and Arg402, which aid in binding the quinazoline ring of THF, and some changes in THF contacts would be expected. Nonetheless, alternative splicing would not be expected to significantly alter the integrity of the active site.

In the crystal structure of *E. coli* SHMT with bound glycine and 5-formylTHF (*19*), all active sites of the two obligate dimers in the asymmetric unit were found to bind glycine and 5-formylTHF in a very similar manner. However, the *E. coli* enzyme is not a tetramer in solution; critical amino acid residues found at the tetramer interface of mammalian SHMT enzymes are lacking (*19*). Hence, the *E. coli* protein is not expected to exhibit negative cooperativity associated with 5-formylTHF binding; that property is unique to tetrameric SHMT enzymes. The *E. coli* SHMT structure is most likely complexed as an external aldimine (Figure 1, structure **II**), as the crystals do not have the characteristic red color associated with formation of the quinonoid ternary complex (Figure 1, structure **III**) (*19*). The active sites of the *E. coli* external aldimine ternary complex most closely resemble active sites I and II in the tight-binding dimer of the mouse ternary complex, with some differences in the contacts with the quinazoline moiety of 5-formylTHF. Structures of SHMT active sites are now available for each of the reaction intermediates associated with SHMT catalysis (Figure 1): the initial state in the unliganded human and rabbit enzymes (*17*, *18*), the *geminal* diamine complex in mouse enzyme sites III and IV, the external aldimine complex in the *E. coli* enzyme (*19*), and the quinonoid complex in mouse enzyme sites I and II.

ACKNOWLEDGMENT

We thank Tadhg Begley for helpful discussions.

REFERENCES

1. Appling, D. R. (1991) *FASEB J. 5*, 2645.
2. Garrow, T. A., Brenner, A. A., Whitehead, V. M., Chen, X., Duncan, R. G., Korenberg, J. R., and Shane, B. (1993) *J. Biol. Chem. 268*, 11910.
3. Girgis, S., Nasrallah, I., Suh, J. R., Oppenheim, E., and Stover, P. (1998) *Gene 210*, 315.
4. Stover, P., Chen, L., Suh, J. R., Stover, D., Keyomarsi, K., and Shane, B. (1997) *J. Biol. Chem. 272*, 1842.
5. Pfender, W., and Pizer, L. I. (1980) *Arch. Biochem. Biophys. 200*, 503.
6. Matthews, R. G., and Drummond, J. T. (1990) *Chem. Rev. 90*, 1275.
7. Matthews, R. G., Drummond, J. T., and Webb, H. K. (1998) *Adv. Enzyme Regul. 38*, 377.
8. Dunathan, H. C. (1966) *Proc. Natl. Acad. Sci. U.S.A. 55*, 712.
9. Besmer, P., and Arigoni, D. (1968) *Chima 22*, 494.
10. Jordan, P. M., and Akhtar, M. (1970) *Biochem. J. 116*, 277.
11. Stover, P., and Schirch, V. (1992) *Biochemistry 31*, 2155.
12. Stover, P., and Schirch, V. (1990) *J. Biol. Chem. 265*, 14227.
13. Stover, P., and Schirch, V. (1991) *J. Biol. Chem. 266*, 1543.
14. Schirch, L. (1984) in *Chemistry and Biochemistry of Folates* (Blakley, R. L., and Benkovic, S. J., Eds.) Vol. 1, pp 399, John Wiley & Sons, New York.
15. Huang, T., Wang, C., Maras, B., Barra, D., and Schirch, V. (1998) *Biochemistry 37*, 13536.
16. Girgis, S., Suh, J. R., Jolivet, J., and Stover, P. J. (1997) *J. Biol. Chem. 272*, 4729.
17. Renwick, S. B., Snell, S., and Baumann, U. (1998) *Structure 6*, 1105.
18. Scarasdale, J. N., Kazanina, G., Radaev, S., Schirch, V., and Wright, H. T. (1999) *Biochemistry 38*, 8347.
19. Scarsdale, J. N., Radaev, S., Kazanina, G., Schirch, V., and Wright, H. T. (2000) *J. Mol. Biol. 296*, 155.
20. Nakshatri, H., Bouillet, P., Bhat-Nakshatri, P., and Chambon, P. (1996) *Gene 174*, 79.
21. Szebenyi, D. M. E., Arvai, A., Ealick, S., LaIuppa, J. M., and Nielsen, C. (1997*) J. Synchrotron Radiat. 4*, 128.
22. Nielsen, C., Arvai, A., Szebenyi, D. M. E., Deacon, A., Thiel, D. J., Bolotovsky, R., Van Zandt, K. C., and Rossmann, M. (1998) ACA Annual Meeting 1998, Abstract 11.06.06.
23. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760.
24. Steller, I., Bolotovsky, R., and Rossmann, M. G. (1997) *J. Appl. Crystallogr. 30*, 1036.
25. Leslie, A. G. W. (1990) in *Crystallographic Computing*, Oxford University Press, Oxford.
26. Evans, P. R. (1993) in *Proceedings of the CCP4 Study Weekend on Data Collection and Processing*, p 114, SERC Daresbury Laboratory, Warrington, U.K.
27. French, G. S., and Wilson, K. S. (1978) *Acta Crystallogr. A34*, 517.
28. Navaza, J. (1994) *Acta Crystallogr. A50*, 157.
29. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res. 28*, 235.
30. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewsi, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905.
31. Jones, T. A. (1982) in *Computational Crystallography* (Sayre, D., Ed.) pp 303, Clarendon Press, Oxford.
32. Kleywegt, G. J., and Jones, T. A. (1998) *Acta Crystallogr. D54*, 1119.
33. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283.
34. Schirch, V., Shostak, K., Zamora, M., and Gautam-Basak, M. (1991) *J. Biol. Chem. 266*, 759.
35. McPhalen, C. A., Vincent, M. G., and Jansonius, J. N. (1992) *J. Mol. Biol. 225*, 495.
36. Carreras, C. W., and Santi, D. V. (1995) *Annu. Rev. Biochem. 64*, 721.
37. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946.
38. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol. 277*, 505.
39. RasMol Version 2.7.1, original program by Roger Sayle, modifications by Herbert Bernstein. (http://www.bernstein-plus-sons.com/software/rasmol/).

BI000635A