Crystal Structure of Human Thymidylate Synthase: A Structural Mechanism for Guiding Substrates into the Active Site†

Celia A. Schiffer,[‡,§,#] Ian J. Clifton,[∥] V. Jo Davisson,[‡,⊥,¶] Daniel V. Santi,[‡,⊥] and Robert M. Stroud*[,‡,§,⊥]

Department of Biochemistry and Biophysics, Graduate Group in Biophysics, and Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0448, and Laboratory of Molecular Biophysics, South Parks Road, Oxford OX1 3QU, England

Received June 29, 1995; Revised Manuscript Received September 5, 1995[⊗]

ABSTRACT: The crystal structure of human thymidylate synthase, a target for anti-cancer drugs, is determined to 3.0 Å resolution and refined to a crystallographic residual of 17.8%. The structure implicates the enzyme in a mechanism for facilitating the docking of substrates into the active site. This mechanism involves a twist of approximately 180° of the active site loop, pivoted around the neighboring residues 184 and 204, and implicates ordering of external, eukaryote specific loops along with the well-characterized closure of the active site upon substrate binding. The highly conserved, but eukaryote-specific insertion of twelve residues 90−101 (h117−128), and of eight residues between 156 and 157 (h146−h153) are known to be α-helical in other eukaryotes, and lie close together on the outside of the protein in regions of disordered electron density in this crystal form. Two cysteines [cys 202 (h199) and 213 (h210)] are close enough to form a disulfide bond within each subunit, and a third cysteine [cys 183 (h180)] is positioned to form a disulfide bond with the active site cysteine [cys 198 (h195)] in its unliganded conformation. The amino terminal 27 residues, unique to human TS, contains 8 proline residues, is also in a region of disordered electron density, and is likely to be flexible prior to substrate binding. The drug resistance mutation, Y6H, confers a 4-fold reduction in FdUMP affinity and an 8-fold reduction in k_{cat} for the dUMP reaction. Though indirectly connected to the active site, the structure suggests a mechanism of resistance that possibly involves a change in structure. This structure offers a unique opportunity for structure-based drug design aimed at the unliganded form of the human enzyme.

Thymidylate synthase (TS) is an essential enzyme in almost all organisms, catalyzing the final step in the sole *de novo* pathway for synthesis of dTMP. It catalyzes the conversion of 2′-deoxyuridylate (dUMP) and 5,10-methylenetetrahydrofolate, to deoxythymidylate (dTMP) and dihydrofolate. Primary sequences show that TS is one of the most highly conserved enzymes (Perry et al., 1990). Crystal structures of TS from several prokaryotic species, *Lactobacillus casei* (Hardy et al., 1987; Finer-Moore et al., 1993) and *Escherichia coli* (Perry et al., 1990); a eukaryote, *Leishmania major* (Knighton et al., 1994); a fungus, *Pneumocystis carinii* (Stroud et al., in preparation); and T4 phage (Finer-Moore et al., 1994) have been determined and indicate that tertiary structure is also very well conserved. The sequence alignment of the species of TS whose three-dimensional structures have been determined is shown in Figure 1A, which can be compared with a ribbon diagram of *E. coli* TS (Figure 1B), whose sequence is shorter than for eukaryotes especially at the N-terminus relative to human, and at two highly conserved inserts. The ribbon diagram illustrates the obligate association of subunits in the TS dimer in a ternary complex (Montfort et al., 1990) with the substrate dUMP, and an anticancer drug, CB3717, that is an analogue of the cofactor methylenetetrahydrofolate.

Crystal structures of several complexes of TS have been determined. Ternary complexes have been solved for TS in *E. coli* Matthews et al., 1990a,b; Montfort et al., 1990; Kamb et al., 1992a,b), *P. carinii* TS (Stroud et al., in preparation), and the bifunctional TS-DHFR *L. major* (Knighton et al., 1994). For *E. coli* TS the structure of a product complex has also been solved (Fauman et al., 1993). These structures provide an understanding of the stereochemistry and mechanism by which the enzyme catalyzes the reaction (Finer-Moore et al., 1990; Stroud & Finer-Moore 1993). Upon binding of the substrate and cofactor to the enzyme, the molecule undergoes a large conformational change, sequestering the ligands from solvent. Since the changes go through several characteristic steps in all species, we regard the structural transitions as steps of a mechanism for orienting the substrates. This mechanism is critically dependent on the carboxy terminal of the protein and is triggered primarily by binding of the cofactor either alone, or in the ordered sequence thought to lead to productive catalysis.

Although bacterial species *E. coli* and *L. casei* TS have already served as a template for several structure-based drug designs (Appelt et al., 1991; Reich et al., 1992; Varney et al., 1992; Shoichet et al., 1993), design of anticancer drugs

† Crystallographic analysis was supported by the National Institutes of Health, Grants RO1-CA41323 to J.F.M. and CA63081 to R.M.S. Protein preparation was supported by CA14394 to D.V.S. V.J.D. was supported by a postdoctoral fellowship from the Damon Runyon−Walter Winchell Cancer Research Fund.

* To whom all correspondence should be addressed.
‡ Department of Biochemistry and Biophysics, UCSF.
§ Graduate Group in Biophysics, UCSF.
∥ Laboratory of Molecular Biophysics.
⊥ Department of Pharmaceutical Chemistry, UCSF.
Present address: Department of Protein Engineering, Genentech Inc., 460 Point San Bruno Blvd., S. San Francisco, CA 94080.
¶ Present address: Department of Medicinal Chemistry and Pharmacology, Purdue University, RHPH, West Lafayette, IN 47907.

⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1995.

0006-2960/95/0434-16279$09.00/0

```
                 10        20        30        40
               ~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   human MPVAGSELPRRPLPPAAQERDAEPRPPHGELQYLGQIQHILRCGV
L. major                            HEERQYLELIDRIMKTGI
P. carinii                        MVNAEEQQYLNLVQYIINHGE
                                           10
                                  aaaaaaaAaaaaaaaaa
L. casei                          MLEQPYLDLAKKVLDEGH
 E. coli                            MKQYLELMQKVLDEGT
phage T4                            MKQYQDLIKDIFENGY

            50        60        70        80
   human RKDDRTGTGTLSVFGM QARYSLRDE.FPLLTTKRVFWKGVLE
L. major VKEDRTGVGTISLFGA QMRFSLRDNRLPLLTTKRVFWRGVCE
P.carinii DRPDRTGTGTLSVFAPSPLKFSLRNKTFPLLTTKRVFIRGVIE
          20        30        40        50
                 bbbbbbibbbbbb               aaaBaa
L. casei FKPDRTHTGTYSIFGH QMRFDLSKG.FPLLTTKKVPFGLIKS
 E. coli QKNDRTGTGTLSIFGH QMRFNLQDG.FPLVTTKRCHLRSIIH
phage T4 ETDDRTGTGTIALFGS KLRWDLTKG.FPAVTTKKLAWKACIA

              90        100               110       120
               ~~~~~~~~~~~~~~            ~~~~~~~~~~~~~~~~~~~~~~
   human   ELLWFIKGSTNAKELSSKG.......VKIWDANGSRDFLDSLGFSTREE
L. major   ELLWFLRGETSAQLLADKD.......IHIWDGNGSREFLDSRGLTENKE
P. carinii ELLWFIRGETDSLKLREKN.......IHIWDANGSREYLDSIGLTKREE
           60        70               80        90        100
           aaaaaaaa   aaaCaaaa            aaaaDaaaa
L. casei   ELLWFLHGDTNIRFLLQHR.......NHIWDEWAFEKWVKSDEYHGPDM
 E. coli   ELLWFLQGDTNIAYLHENN.......VTIWDEWADEN..........
phage T4   ELIWFLSGSTNVNDLRLIQHDSLIQGKTVWDENYENQAKDLGYHS..

                                                  130       140
                                                  ~~~~       ~~~~
   human .....................................GDLGPVYGFQWRHFGAE
L. major .....................................MDLGPVYGFQWRHFGAD
P. carinii....................................GDLGPIYGFQWRHFGAE
                 110       120       130       140       150
                      aaaaaaaaaEaaaaaaaaa aaaaFaaa    aaaGaaaa
L. casei TDFGHRSQKDPEFAAVYHEEMAKFDDRVLHDDAFAAKYGDLGLVYGSQWRAWHTS
 E. coli .....................................GDLGPVYGKQWRAWPTP
phage T4 .....................................GELGPIYGKQWRDF...

              150       160       170       180       190       200
              ~~~~~~~~~~~~
   human   YRDMESDYSGQGVDQLQRVIDTIKTNPDDRRIIMCAWNPRDLPLMALPPCHALCQFYVVN
L. major   YKGFEANYDGEGVDQIKLIVETIKTNPNDRRLLVTAWNPCALQKMALPPCHLLAQFYVNT
P. carinii YIDCKTNYIGQGVDQLANIIQKIRTSPYDRRLILSAWNPADLEKMALPPCHMFCQFYVHI
                     160       170       180       190       200
                        aaaaaHaaaaaaaa     bbvbb aaIaaa      bbbbivbbbb
L. casei   ........KGDTIDQLGDVIEQIKTHPYSRRLIVSAWNPEDVPTMALPPCHTLYQFYVND
 E. coli   ........DGRHIDQITTVLNQLKNDPDSRRIIVSAWNVGELDKMALAPCHAFFQFYVAD
phage T4   ..........GGVDQIIEVIDRIKKLPNDRRQIVSAWNPAELKYMALPPCHMFYQFNVRN

                     210       220       230       240       250       260
   human   ..SE......LSCQLYQRSGDMGLGVPFNIASYALLTYMIAHITGLKPGDFIHTLGDAHIYLNHIEPL
L. major   DTSE......LSCMLYQRSCDMGLGVPFNIASYALLTILIAKATGLRPGELVHTLGDAHVYRNHVDAL
P. carinii ..PSNNHRPELSCQLYQRSCDMGLGVPFNIASYALLTCMIAHVCDLDPGDFIHVMGDCHIYKDHIEAL
             210               220       230       240       250       260
              b      bbbbiiibbbbaaaaaaaaaaaaaJaaaaaaaaaa     bbbbbiibbbbbb aaaKa
L. casei   ..GK......LSLQLYQRSADIFLGVPFNIASYALLTHLVAHECGLEVGEFIHTFGDAHLYVNHLDQI
 E. coli   ..GK......LSCQLYQRSCDVFLGLPFNIASYALLVHMMAQQCDLEVGDFVWTGGDTHLYSNHMDQT
phage T4   ..GY......LDLQWYQRSVDVFLGLPFNIASYATLVHIVAKMCNLIPGDLIFSGGNTHIYMNHVEQC

              270       280           290               300       310
   human   KIQLQREPRPFPKLRILRKV....EKIDDF........KAEDFQIEGYNPHPTIKMEMAV
L. major   KAQLEVPHAFPTLIFKEER.....QYLEDY........ELTDMEVIDYVPHPAIKMEMAV
P. carinii QQQLTRSPRPFPTLSLNRSI....TDIEDF........TLDDFNIQNYHPYETIKMKMSI
           270       280           290               300       310
           aaaaa       bbbbb                     bbbbb
L. casei   KEQLSRTPRPAPTLQLNPDK....HDIFDF........DMKDIKLLNYDPYPAIKAPVAV
 E. coli   HLQLSREPRPLPKLIIKRKP....ESIFDY........RFEDFEIEGYDPHPGIKAPVAI
phage T4   KEILRREPKELCELVISGLPYKFRYLSTKEQLKYVLKLRPKDFVLNNYVSHPPIKGKMAV
```

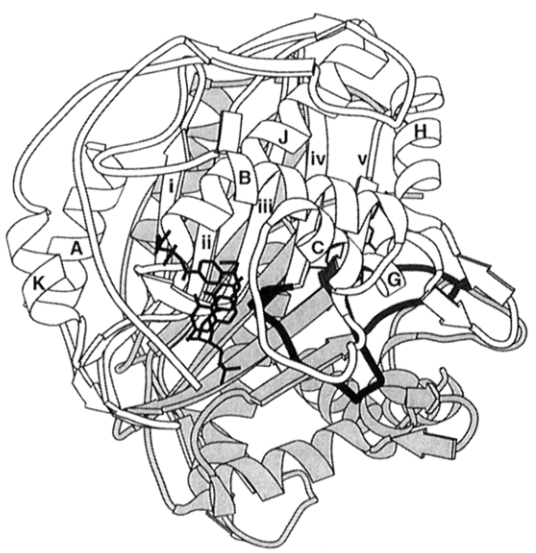

FIGURE 1: (A) Amino acid sequence alignment of the species of TS whose crystal structures have been determined. The number on the top is that of the human sequence. Those residues marked with a (~) are not well ordered in the human crystal structure. The second set of numbers is that of the *L. casei* sequence, used as the primary sequence numbering scheme in the text. The secondary structure is indicated with (*a*) for α-helical and (*b*) for *β*-strands and labeled. The active site loop is highlighted in bold in the human and *L. casei* sequences. (B) Ribbon diagram (Kraulis, 1991) of the crystal structure of the ternary complex of *E. coli* TS (Montfort et al., 1990) with one monomer shown in white and the other shown in gray. The active site loop is highlighted in black and dark gray, and the substrate, dUMP, and cofactor analogue, CB3717, are shown in black in a stick representation.

aimed at human TS would benefit from the structure of the actual target. The structure of the human enzyme can also be used to assist in structure-based approaches to development of species-specific inhibitors with antiproliferative activity against pathogenic TSs versus the human enzyme.

In vitro resistance to fluoro-dUMP, a metabolite of the chemotherapeutic prodrug 5-fluorouracil, has been recently characterized. In a particular human colonic tumor cell line, a single-site mutant of TS was found that encoded a tyrosine to histidine change in the structural gene for TS (Barbour et al., 1990). The context of this residue suggests a mechanism for drug resistance. The primary sequence of the human enzyme has three insertions relative to other species of known structure. Twenty-seven amino acids at the amino terminus are unique to human TS. Two loops of 12 amino acids and 8 amino acids are highly conserved across all eukaryotes, but are not present in any prokaryotic TS sequences, suggesting a eukaryotic-specific function. We report the determination of the crystal structure of unliganded human TS. The unique aspects of the structure are compared with the structures of now five other species of TS.

EXPERIMENTAL PROCEDURES

Protein Preparation and Crystallization. Human TS was cloned and expressed in *E. coli* in a vector that utilized a 115-base 5′-untranslated leader sequence from the *L. casei* TS gene. The enzyme was purified and shown to have native-like activity as described by Davisson et al. (1989). This sample was later shown to be heterogeneous, with a large percentage of the protein having 13 additional residues which came from the leader sequence attached to the N-terminus (Davisson et al., 1994). Nevertheless, the protein had wild-type activity (Davisson et al., 1994). Crystals of human TS were grown at room temperature, in hanging drops, from Tris-HCl buffer, in the absence of phosphate or other ligands using 1.5 M ammonium sulfate as a precipitating agent with 20 mM β-mercaptoethanol between pH 6.5 and 8. Dissolved crystal retained enzymatic activity. The protein crystallized in the space group $P3_121$ with unit cell dimensions, determined by diffractometry as $a = b = 96.7$ Å, $c = 84.1$ Å (Schiffer et al., 1991). There is one monomer in the asymmetric unit. The crystal used for the final data collection was irregular in shape with the average dimension of approximately 0.4 mm and came from a drop with a pH of 7.0.

Data Collection. The data used in refinement of the structure were collected at room temperature at the synchrotron source in Daresbury, U. K. Intensities were recorded using a wavelength of 0.9 Å at station 9.5, SRS Daresbury, on a Hendrix-Lentfer image plate scanner system made by Mar Research. The crystal to detector distance was 179 mm. Forty 75-second images were collected with a 1.5-degree oscillation covering a total of 60 degrees on one crystal. The data were processed with the programs OSCGEN and MOSFLM (Leslie et al., 1988).

Evaluation of the Data. Table 1 shows the statistics for the processed diffraction data. The cumulative Rsymm was 9.8% to 2.8 Å resolution. Beyond a resolution of 3.0 Å, Rsymm became significantly worse, in the highest resolution shell becoming 35.3.% This is due to the radiation sensitivity of the crystals which can be seen as the value of $\langle I\rangle/\langle \mathrm{sig}I\rangle$ that falls below 3.0 Å resolution (see the three highest resolution bins in Table 1). Thus only the data to 3.0 Å ($R_{symm} = 8.5\%$) was used in the subsequent refinement. This data is 94.9% complete to this resolution, and 93.8% complete in the highest resolution shell (3.12 − 3.0 Å).

Table 1: Crystallographic Data Statistics from Monochromatic Synchrotron Data

highest resolution	$\langle I\rangle/\langle \mathrm{sig}I\rangle$	no. of independent reflections	local R_{symm}	cumulative R_{symm}
11.6	8.98	209	0.058	0.058
8.53	10.6	835	0.051	0.052
7.06	10.6	1136	0.056	0.053
6.15	10.7	1326	0.062	0.055
5.53	9.2	1491	0.069	0.058
5.06	9.5	1645	0.064	0.059
4.70	10.4	1795	0.060	0.059
4.40	9.5	1848	0.062	0.060
4.15	9.8	2020	0.064	0.061
3.94	7.8	2211	0.078	0.063
3.76	7.1	2215	0.093	0.065
3.60	6.6	2386	0.103	0.068
3.46	6.5	2422	0.109	0.071
3.34	5.1	2661	0.142	0.074
3.23	4.5	2633	0.163	0.078
3.13	3.9	2707	0.198	0.081
3.03	3.3	2925	0.236	0.085
2.95	2.8	2953	0.287	0.089
2.87	2.4	3036	0.316	0.094
2.80	2.1	3077	0.353	0.098

[a] $R_{symm} = ([\Sigma hkl\Sigma_{i=1,N}\,\omega_i(I_{avg} - I_i)^2]/[\Sigma hkl\Sigma_{i=1,N}\,\omega_i(I_i)^2])^{1/2}$, where $\omega_i = 1/\sigma_i^2$ and $I_{avg} = 1/N(\Sigma_{i=1,N}\,I_i)$.

Molecular Replacement. The structure was solved by molecular replacement by use of a truncated version of the highly refined unliganded *L. casei* TS structure (Finer-Moore et al., 1993) as a search model. To make the truncated model, the variable domain (residues 90−139),[1] a 52-residue insert in *L. casei* TS relative to human and most other species (Hardy et al., 1987; Finer-Moore et al., 1993), was omitted. Those residues of the *L. casei* TS model that were not identical to the human sequence (Takeishi et al., 1985) were replaced by alanine. The search model did not include 27 residues at the amino terminus, and two insertions of 12 and 8 amino acids residues that are highly conserved in eukaryotic species of TS at residues 90 (h117) to 101 (h128), and inserted between 156 and 157 (h146−h152) respectively (Perry et al., 1990).

An orientation was obtained from the data to 3.0 Å resolution using the Crowther rotation function (Crowther, 1972), in which the molecular 2-fold axis corresponded to a crystallographic 2-fold axis, as dictated by the density and symmetry of the crystals. The peak height of the solution was +72 with the lowest background peak height at −34 and an average value of +17. The translation search was carried out along the crystallographic 2-fold axis. The minimum in the R-factor grid search for translation was 50% within a background of 58%−62%. After 6 cycles of least squares refinement (Hendrickson & Konnert, 1978; Furey, 1984), the *R*-factor dropped to 36%. Several side-chains unique to human TS were visible in the resulting difference density map calculated using phases from the model, and

[1] For ease of comparison, the numbering system used for amino acid residues of TS refers to the sequence of the reference *L. casei* enzyme sequence with the homologous numbering of the human (h) TS sequence in parentheses. All numbering in the figures and tables and the discussion of hydrogen bonds refers, for simplicity to the corresponding residue numbers of the *L. casei* sequence.

terms ($F_o - F_c$). This confirmed that the molecular replacement solution was correct.

Structure Refinement. The structure was refined using all data to 3.0 Å resolution. Side-chains of human TS were substituted into the molecular replacement solution model, without fitting the density, using the graphics program INSIGHT (Dayringer, 1986). Energy minimization served to eliminate several close contacts in the model. This model was subjected to rigid body refinement of the entire molecule. Refinement continued with several rounds of "slow cool" molecular dynamics runs, energy minimization, and individual temperature-factor refinement within the X-PLOR package (Brunger, 1990) followed by manual rebuilding in FRODO and difference Fourier refinement (Chambers & Stroud, 1977) of the regions where either residues did not fit the $2F_o - F_c$ maps or where negative peaks appeared in an ($F_o - F_c$) map. A solvent mask was used within X-PLOR to take into account the contribution of bulk solvent to the diffraction intensities, thus allowing the low-resolution data to be included in the refinement. The starting temperature for the "slow cool" molecular dynamics was either 2000 or 3000 K depending on the run, using 0.5 femtosecond time steps, and the temperature was reduced every 50 steps by 25 degrees until the temperature was 300 K. Structures were then energy minimized and the coupled temperature factors were refined until the energy function converged. Finally the structure was examined using difference Fourier syntheses to confirm placement of the residues (Chambers & Stroud, 1977).

RESULTS

The Final Statistics. In the final structure the root mean square (r.m.s.) deviation in bond lengths is 0.019 Å, the bond angles are 3.5 degrees, and the dihedral angles are 25.2 degrees. The average temperature factor was 18 Å^2 and the *R*-factor was 17.8% for all data from 7.0 to 3.0 Å. No water molecules or solvent mask were included.

The Active Site. Although 17 of the 25 residues (180–204, highlighted in bold in Figure 1a) in the loop containing the catalytic cysteine are identical between the human and *L. casei* enzymes, the electron density for these residues was discontinuous when phases were computed based on the conformation seen in the *L. casei* enzyme structure. The electron density was broken along both the side chains and the main chain, indicating a unique and different structure of the human active site in the apoenzyme. Because of this difference, these 25 residues were rebuilt into an ($F_o - F_c$)α_{calc} "omit map," where the active site loop was excluded from calculation of the phases (Chambers & Stroud, 1977). Once the structure was rebuilt for this region, and refined, both the main chain and the side chains fit the electron density without ambiguity (Figure 2A). This is in direct contrast to a model based on the *L. casei* conformation, and refined. A map phased on this alternate tested structure shows broken and discontinuous density, indicating that the *L. casei* conformation does not apply to the human enzyme at the active site (Figure 2B). As definitive evidence for the accuracy of this key alteration in structure, the rebuilt region of structure (as seen in Figure 2A) was placed in the map phased on the *L. casei* conformation (the map seen in Figure 2B) that was therefore biased in favor of the *L. casei* conformation (Figure 2B). The configuration seen in the human structure fits the electron density better than the *L. casei* conformation as seen in Figure 2B without any main chain discontinuities. Additional support for this unique configuration in human TS is that the average temperature factor for these twenty-five residues dropped from 30 Å^2

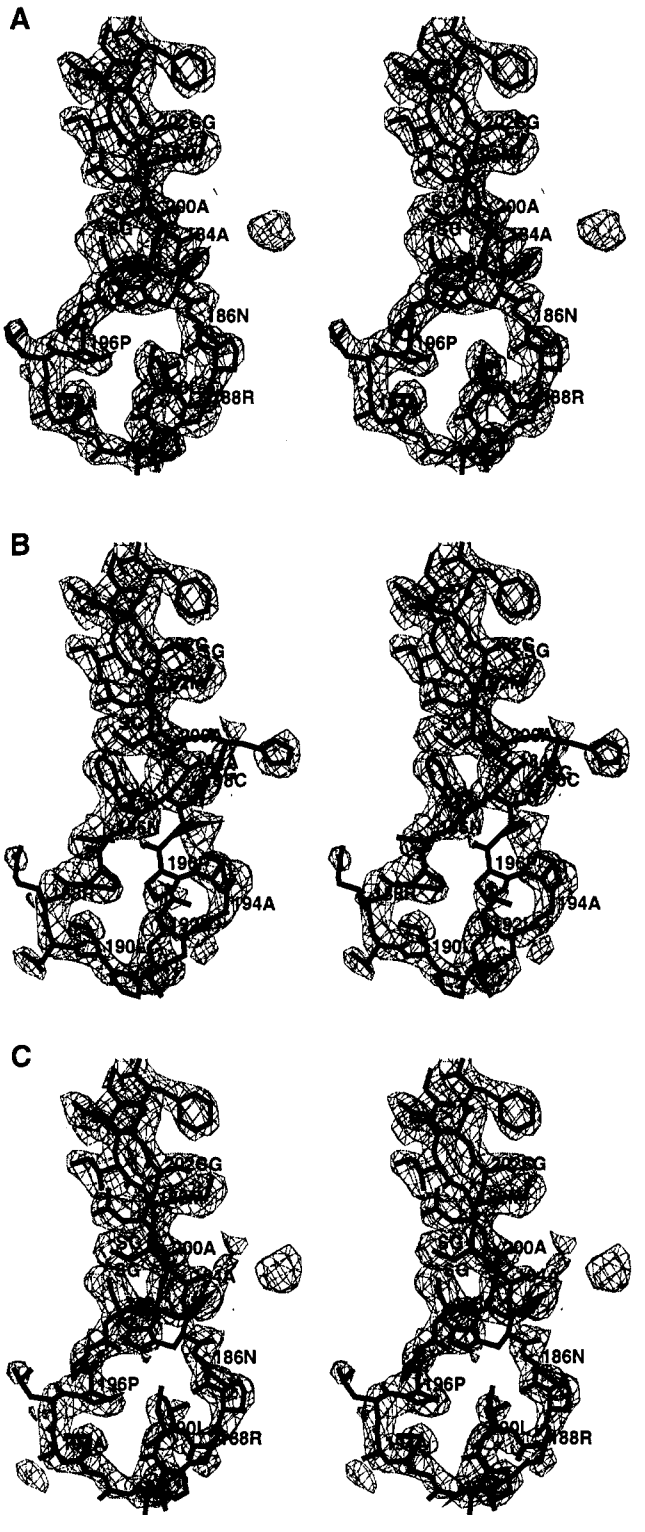

FIGURE 2: (A) $2F_o - F_c$ electron density (Sack, 1988) contoured at one sigma around the active site loop. Structure for the active-site loop, 180–204, and final density map. (B) *L. casei* configuration of the loop, and density map phased on the *L. casei* configuration showing discontinuities along the main chain and side chains. (C) Final structure, superimposed on the density map phased on the *L. casei* configuration. (Numbering corresponds to the *L. casei* sequence.)

for the *L. casei* conformation, to 17 Å^2 for the correct structure; concomitantly the crystallographic *R*-factor was reduced from 21.3% to a final value of 17.8% using all data from 7.0 to 3.0 Å. The major variation lies in a 180° twist of the loop between residues 184 and 200 that forms part of the wall of the active site cavity. This critical region contains the catalytic cysteine, Cys 198 (h195), and His 199 (h196), and must reorient along with other previously characterized changes that constitute closure of the active site, a structural step in the mechanism of substrate binding for productive catalysis (Stroud & Finer-Moore, 1993).

The Amino Terminal Sequence, and the Two Eukaryote-Specific Insertions. One region of the human sequence is an extension of the amino-terminus with an additional 27 amino acid residues that is unique to human TS. It includes eight proline residues and seven charged residues; most of it lies in weak density, indicating that it is poorly ordered in the human TS structure. Prediction of secondary structure, carried out using the program PREDICT (written by Stroud) (Finer-Moore & Stroud 1984; Finer-Moore et al., 1989), and incorporating amphipathic Fourier analysis alongside statistical determinants, predicts a predominantly random coil structure for this segment of the sequence.

There are two insertions within the sequence that are highly conserved in length, sequence, and character in all eukaryotes as a class, versus all prokaryotes, suggesting eukaryote-specific functions for them (Hardy et al., 1987; Perry et al., 1990). The first insertion introduces twelve residues into all eukaryotes, between residues 90 and 101 (h117-h128), although nonhomologous inserted sequences do occur in prokaryotes at this site. At this location in *L. casei* TS the structure of a much larger 52 amino acid residue insertion has been determined (Finer-Moore et al., 1994). The second insert places eight residues in all eukaryotes between 156 and 157 (h146−h153) versus prokaryotic species which generally have no inserts at all at this site. However, these regions plus some additional flanking regions are poorly defined in the electron density, indicating that they are flexible in the crystal structure. There are however, clues as to where these regions are expected, since the conformation of this region including both eukaryotic inserts has been recently determined in the structure of TS from the eukaryote *P. carinii* (Stroud et al., in preparation).

Several different approaches were pursued to resolve the structure in these regions. These methods included placement of dummy atoms into the poorly defined density (Finer-Moore et al., 1994) and use of an automated density skeletonization routine, PRISM (Baker et al., 1993; Bystroff et al., 1993; Wilson & Agard, 1993). Finally, since the sequence for these eukaryotic inserts is highly conserved between the human and *P. carinii* enzymes (Figure 1A), these predominantly helical segments of the *P. carinii* structure were built into the corresponding region of the model of the human enzyme. However, the electron density in (2Fo-Fc)αcalc maps was more poorly defined after including these regions in the refinement, indicating that they are different from the structure seen in the fully liganded structure of *P. carinii* TS. One possible interpretation is that in the unliganded eukaryotic enzymes, there are multiple conformations for these regions. This flexibility may be functionally important for associations between TS and other enzymes in the cell.

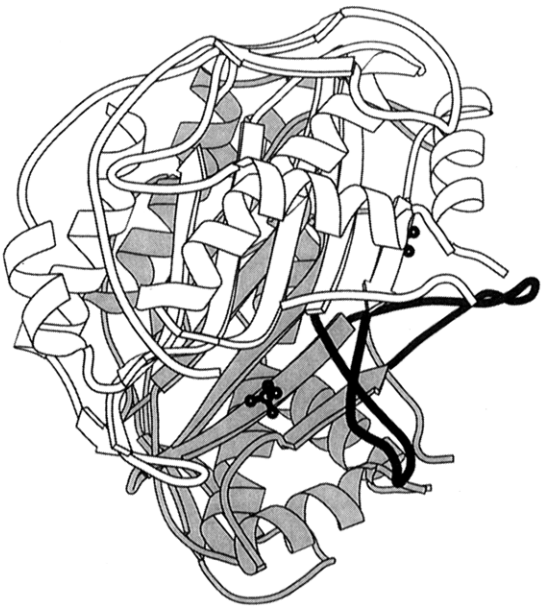

FIGURE 3: Ribbon diagram (Kraulis, 1991) of the crystal structure of the dimer of human TS, with one monomer shown in white and the other shown in gray. The active site loop is highlighted in black and dark gray, and the sulfate ion is shown in a "ball-and-stick" representation.

DISCUSSION

Mechanism of Substrate Binding in Human TS Differs from That in Other Species. We have previously shown that substrate binding in TS involves a large conformational rearrangement that sequesters and orients the two substrates for catalysis (Perry et al., 1990; Kamb et al., 1992; Stroud & Finer-Moore, 1993). These mechanisms appear to be conserved across the six species of TS whose structures have been determined (Figure 1A). The tertiary structure of human apo-TS follows the same general fold as that found for the structures of TS from *L. casei* and *E. coli* (Perry et al., 1990). TS is an obligate dimer, with the dimer interface being a tight interaction between two six-stranded β-sheets, one from each subunit. The surfaces opposite the dimerization interface of the β-sheets form the back walls of the two active-site cavities. The catalytic cysteine, which is essential for enzymatic activity, resides at the base of the fourth β-strand. Michael addition of this cysteine to C-6 of the substrate pyrimidine ring activates the substrate for condensation with cofactor, 5,10-methylene-5,6,7,8-tetrahydrofolate. In the core of the monomer is the J-helix, a 22-residue-long hydrophobic helix, which contains several of the active-site residues that interact directly with the substrate and cofactor. Surrounding this central helix and contributing to much of the solvent-exposed surface area of the molecule is a network of shorter α-helices.

The major difference in the tertiary structure of human TS compared with that of other species is in the conformation of the conserved active site loop (180−204, highlighted in bold in Figure 1A). In the *L. casei* TS structure this region includes the end of β-strand V, a short α-helix, the I-helix, and the beginning of β-strand IV. The structure of the dimer of human TS, with the active site loop highlighted in black in one monomer and gray in the other, is shown in Figure 3. This loop has locally high primary sequence homology between the *L. casei* and the human sequence. However, the conformation of the entire active-site loop of the human

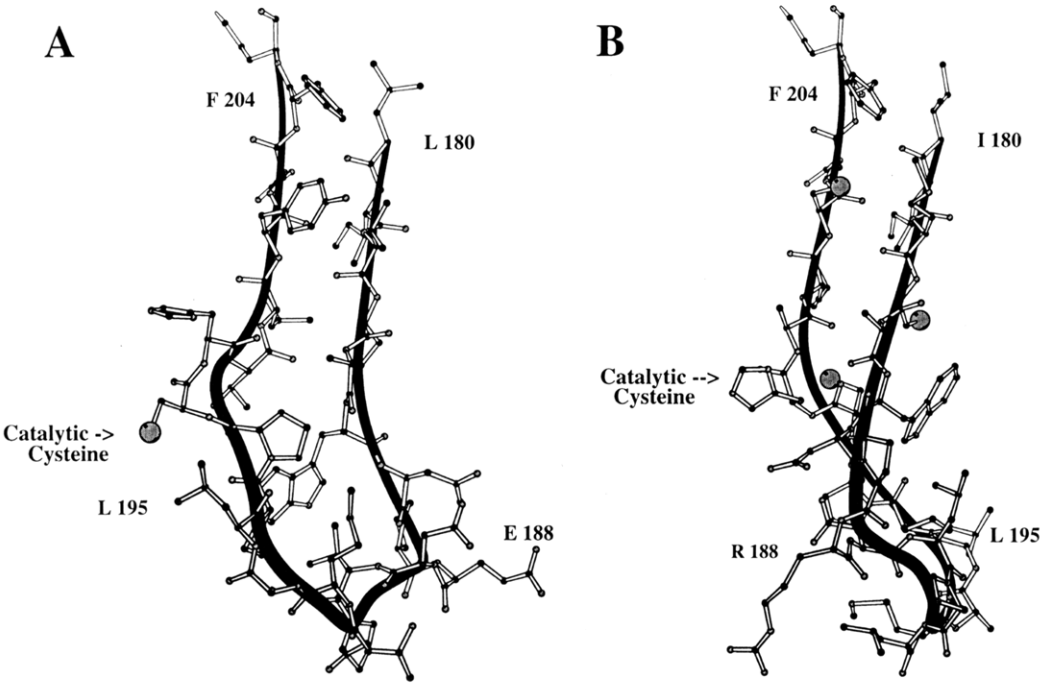

FIGURE 4: Ribbon diagram (Kraulis, 1991) of different topologies of the active site loop (A) from *L. casei* and (B) from human. The radii of the sulfurs in the Cys side-chains are expanded for better visualization. (Numbering corresponds to the *L. casei* sequence.)

TS structure is twisted by approximately 180 degrees relative to the conformation in the *L. casei* enzyme structure (Figure 4). The points of rotation are at neighboring residues 184 and 200 on two adjacent β-strands. Since the catalytic cysteine is buried in this structure, the protein must undergo a major conformational change upon binding ligands, to catalyze the enzymatic reaction, converting dUMP into dTMP.

In the structure of the human enzyme, the active-site cysteine 198 (h195) is deeply buried in the β-sheet dimer interface of TS. The rearranged conformation of the active-site loop of human TS puts the catalytic cysteine close to cysteine 183 (h180). The human TS structure is the first TS three-dimensional structure to be determined with a cysteine at this position. The distance between the beta carbons of these two cysteine residues is 4 Å. The electron density clearly indicates, however, that the two cysteines are not linked by a disulfide bond (Figure 2A). Although TS from other species have cysteine at this second position in their primary sequence, all the tertiary structures of TS thus far determined have a serine at this second site. Also two other cysteine residues occur within the human TS monomer and are close enough to form a disulfide bond. These residues are on adjacent β-strands at positions 202 (h199) and 213 (h210), but the sulfhydryl groups are pointed away from each other and there is no bridging electron density. The fact that neither potential disulfide bond is seen in the structure is not surprising since the protein was crystallized in the presence of a reducing agent. Furthermore, the protein was expressed from the naturally reducing environment of the cytoplasm in *E. coli* cells (Davisson et al., 1989, 1994). Nevertheless, the proximity and orientation of two independent pairs of buried sulfhydryls seems unlikely to have occurred randomly and suggests the possibility of crosslinking under certain physiological conditions as yet undefined.

The structure of the active site loop in human TS is well ordered, having an average temperature factor of 17 Å^2. In both structures, human and *L. casei*, this region is stabilized by an extensive network of hydrogen bonds. In human TS stabilization of this region involves 26 hydrogen bonds, and in *L. casei* TS 29 hydrogen bonds. In Table 2, this hydrogen bonding network is compared between the two proteins. For ease of comparison, the hydrogen bonds are divided into

Table 2. Hydrogen Bonds in the Active Site Loop

human	*L. casei*
conserved	
Q**163**OE1 − S**183**N	Q**163**OE1 − S**183**N
Q**163** NE2 − S**183**O	Q**163** NE2 − S**183**O
R**179**O − I**181**N	R**179**O − I**181**N
L**180**O − F**204**N	L**180**O − F**204**N
M**182**N − C**202**O	V**182**N − Y**202**O
A**184**N − A**200**O	A**184**N − A**200**O
N**186**OD1 − R**188**N	N**186**OD1 − E**188**N
N**186**O − D**189**N	N**186**O − D**189**N
D**189**O − L**192**N	D**189**O − L**192**N
C**198**O − A**200**N	C**198**O − A**200**N
L**201**O − Y**216**N	L**201**O − Y**216**N
swapped	
R**179**NH1 − D**189′**OD1	R**179**NH1 − A**194′**O
R**166**NH1 − A**194′**O	
W**185**NE1 − P**196**O	
N**186**ND2 − **SO_4^{2-}**	N**186**ND2 − D**189**OD2
D**189**OD2 − R**218**NH1	P**196**O − C**198**N
D**189**OD2 − R**218**NH2	P**196**O − R**218**NH1
P**197**O − H**199**N	P**197**O − T**200**OG1
P**197**O − R**218**NH2	P**197**O − R**218**NE
H**199**O − R**218**N	H**199**O − L**201**N
L**201**N − Y**216**O	C**198**O − R**218**N
A**200**N − A**184**O	T**200**N − Y**216**O
varied	
P**187**O − D**189**N	W**185**N − T**200**O
P**187**O − L**190**N	N**186**O − E**188**N
L**190**O − L**192**N	N**186**O − V**190**N
	E**188**OE1 − S**156'**N
	D**189**O − M**193**N
	D**189**O − T**192**OG1
	C**198**O − T**200**OG1
	H**199**NE1 − N**229**OD1

three categories: conserved, swapped, and varied. When a hydrogen bond is between the same atoms of the same residues in both structures then it is defined as "conserved." An example of this is the hydrogen bond between **182**N and **202**O in both the human and the *L. casei* structure. "Swapped" hydrogen bonds belong to a set of atoms that were hydrogen bonded in a particular network in one structure and then rearranged to find other hydrogen bonding mates in the second structure. For example in the structure of human TS, D**189**OD2 hydrogen bonds to R**218**NH1; but in the structure of *L. casei* TS, D**189**OD2 hydrogen bonds to N**186**ND2, and R**218**NH1 hydrogen bonds to P**196**O. The "varied" category of hydrogen bonds are ones that do not have an analogous bond in the other structure. These differences are due to either a difference in solvent exposure from the conformational rearrangement of the active site loop, or a change in amino acid sequence between the two species. The extensive rearrangement of the hydrogen bonding network in this region shows that although the conformation in the human TS structure may not correspond to an enzymatically active conformation, it represents a stable, low-energy conformation of the molecule.

TS normally provides four conserved arginine side chains to coordinate the phosphate of dUMP. Phosphate anion also binds at the same site as phosphate in dUMP. Human TS was crystallized in the absence of phosphate; however, a large positive peak in the electron density map exists at the phosphate binding site which has been characterized in the apo-structures of *E. coli* and *L. casei* TS. This peak is likely to be a sulfate ion, since the protein was crystallized with ammonium sulfate (1.5M) as the precipitating agent. However, these crystals are unable to accommodate phosphate, as addition of the ion causes the crystals to crack. The density peak is, however, surrounded by the four arginines Arg 23 (h50), Arg 215 (h218), Arg 178″ (h175′), and Arg 179′ (h176′), which bind the phosphate of dUMP in the Michaelis complex, or phosphate in unliganded *E. coli* and *L. casei* TS. Arg 188 (h185) is also nearby, although this arginine is not conserved, and is near the ion binding site because of rearrangement of the active site. When crystals were grown in the presence of FdUMP and CB3717, which leads to formation of a stable ternary complex, they appeared to have the same morphology as the unliganded crystals, and it remains to be established whether under the crystallization conditions these ligands bind to the protein.

Flexibility in the Amino Terminal, and Conserved Eukaryotic Insertions. The unique amino terminal sequence is 27 residues longer than *L. casei* TS and contains eight proline residues and seven charged residues. This amino terminal extension is longer than any other species of TS for which there are currently at least 25 determined sequences. This extension, which is 29 residues longer than the shortest of the TS sequences, suggests that it may have a function unique to the human enzyme. Perhaps it interacts in forming complexes with other proteins in the cell some of which are involved in DNA replication. The prediction of a random coil structure, and the fact that the electron density is weak in this region, is consistent with multiple conformations.

The two other insertions in the human TS sequence are present and highly conserved in all the currently sequenced eukaryotic TSs, suggesting some as yet undefined role. Located at positions 90−101 (h117−h128), and inserted between 156 and 157 (h146−h153) in the human TS sequence, they are conserved in length and homologous in sequence to all other eukaryotes. However in the present unliganded structure the electron density where the regions should be found is disordered, indicating that these insertions may be are flexible. Yet in the recently determined fully liganded *P. carinii* TS crystal structure (Perry et al., in preparation), the only other eukaryotic TS structure known at this time, both eukaryotic insertions are helical, and lie in adjoining regions of the structure. Because of the high conservation of sequence we might expect that the structure of the human enzyme is likely to fold in an analogous manner. In the apo-human TS crystal form, however, the density is disordered. The *R*-factor is satisfactory for a well-refined structure at 17.8% to 3.0 Å resolution, and thus these disordered regions contribute little to the X-ray data. Therefore a probable scenario is that the helical segments become ordered as part of the mechanism of substrate binding. The two inserts may also be involved in signaling between, and assembly of multienzyme replitase complex inside the cell (veer Reddy et al., 1983; Pucinski et al., 1990).

Mechanism of Drug Resistance. The mechanisms by which cells become drug resistant, and novel stratagems for maintaining efficacious drug therapy are of increasing importance as resistance to present day drugs develops. Human TS is the physiological target for the anticancer drug 5-fluorouracil, and remains a good target for development of antiproliferative and anticancer drug development. In a particular human colonic tumor cell line, drug resistance to FdUMP, the TS-active metabolite of 5-fluorouracil, was found to be accompanied by a single site mutant located in the structural gene for TS. The structure of human TS makes it possible to ask how resistance is evoked and to strategize alternatives. The single mutation, Y6H, from a highly conserved tyrosine at residue 6 (h33) to a histidine (Barbour et al., 1990) occurs in the A-helix and confers approximately a 3- to 4-fold reduction in FdUMP affinity (increase in K_d) for human TS (Hughey et al., 1993), constituting resistance to the inhibitor. The activity of the enzyme is likewise compromised by the Y6H mutation, showing an 8-fold drop in k_{cat} for the dUMP reaction. Although the Tyr 6 side-chain is not in the active site of the molecule, the hydroxyl oxygen of Tyr 6 forms a hydrogen bond to the backbone carbonyl oxygen of residue 222 (h219) at the first turn of the central hydrophobic J-helix, very close to Asp 221 whose side chain does participate in the reaction chemistry (Figure 5). The reason the Y6−O222 hydrogen bond might be functionally important is that the first turn of the highly conserved J helix (residues 222−245) is not quite a true α-helix, but is more open with an extra residue in the first turn. However when Tyr 6 is changed to phenylalanine, in the Y6F mutation, the enzymatic properties are very similar to that of the wild type (Hughey et al., 1993). Invariant Pro 227 (h224) also participates in maintaining this unusual first turn of the J helix since the amide is unavailable for the $i - (i + 4)$ hydrogen bond with the carbonyl of 224 (h221) to make a classical α-helix. The extension of this first turn is necessary to properly orient key components of the active site, and is a conserved structural feature of all TS species whose structures have been determined to date (Stroud and Finer-Moore, 1993). The residue most obviously influenced by the conformation of this first loop of the helix is the invariant Asp 221 (h218), which just precedes the hydrogen

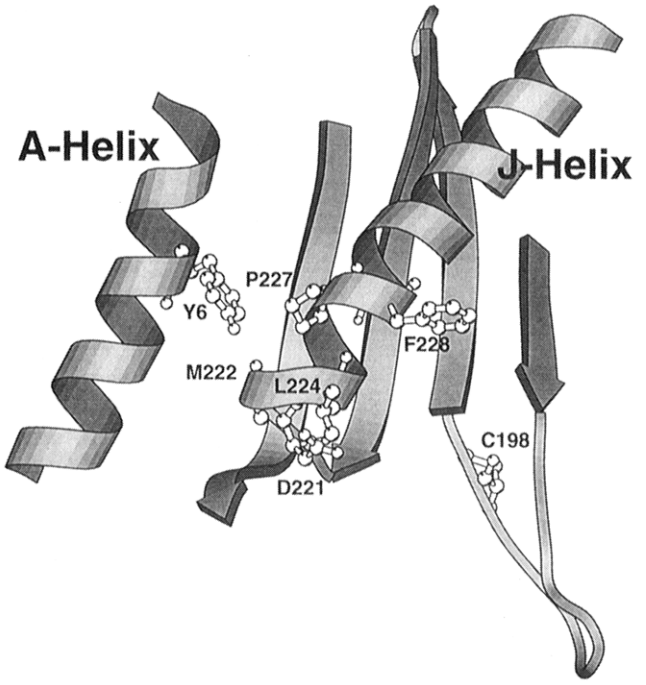

FIGURE 5: Tyr 6 (h33), near the end of the A-helix hydrogen bonds to the backbone oxygen of residue 222 (h219) at the first turn of the J-helix. Also shown are Asp 221 (h218), Cys 198 (h195), Leu 224 (h221), Pro 227 (h223) and Phe 228 (h225), which are residues known to be important to the coordination of the substrate and cofactor.

bond formed to Tyr 6 (h33). Furthermore, the Y6H mutation involves a change in charge. The positive charge of the His 6 could possibly be attracting the negatively charged side chain of Asp 221, causing its orientation to be no longer optimal for ligand binding. In the ternary complex the main-chain amide of Asp 221 hydrogen-bonds to the pyrimidine of dUMP, while the side chain hydrogen-bonds to the cofactor; thus Asp 221 may be the ultimate effector of changes in activity. Leu 224 (h221) and Phe 228 (h225), also in the J helix, form a hydrophobic pocket for the PABA ring of the cofactor, and their orientation could also be altered if this initial extended turn of the J-helix was perturbed. Mutation of Tyr 6 (h33) to a His changes the size of the internal side chain, while it maintains hydrogen bonding potential, although the donor sites are very differently oriented and it adds a positive charge. These two factors can then serve to alter the conformation of the active site and the electrostatic environment, resulting in diminished activity, and in drug resistance.

CONCLUSIONS

The human TS structure provides the first three-dimensional view of a mammalian TS and represent one state of a therapeutically important drug target useful for structure-based drug development. The structure indicates a large degree of similarity in the overall fold to the bacterial enzymes, which is consistent with the high degree of sequence identity. The major exception is the active site loop. This loop is arranged such that the catalytic cysteine is buried, and thus must reorient on binding of substrate.

Major conformational changes occur during substrate binding and orientation in all forms of TS. Changes in enzyme conformation are implicated in correctly orienting substrate and cofactor for the multistep reaction, and in appropriately sequestering the substrate or cofactor from the external solvent environment. The conformational change serves as a way of docking, or guiding the substrates into the active site. In human TS an additional possible function for the orientation of the active site may be to protect the key catalytic cysteine from being modified in the cell. Thus to bind and orient substrates, for catalysis of the reaction of dUMP to dTMP, human TS undergoes additional conformational changes. The molecule must reorient the active site residues as well as undergoing the conformational changes, well characterized in prokaryotic TSs, which lead to closure of the domains around the substrate and cofactor (Finer-Moore et al., 1990; Stroud & Finer-Moore, 1993). Based on the comparison with the structure of liganded TS from *P. carinii*, we hypothesize that these changes may play a role in substrate binding, or in signaling and assembly within the replitase complex.

A drug-resistant mutation, Y6H, can be interpreted in terms of induced change in the J-helix, eventually mediated through residues in the active site.

This unliganded structure of human TS suggests a novel course for future structure-based drug design of chemotherapeutic agents targeting human TS. Inhibitors designed to prevent the structure from rearranging the active site loop may be added to existing drug design schemes.

ACKNOWLEDGMENT

We thank Dr. Janet Finer-Moore, Dr. A. Kossiakoff, Dr. J. Hajdu, and Dr. P. Greene for their assistance. The structure has been deposited in the Brookhaven Protein Data Bank, 1TDS.

REFERENCES

Anderson, C. M., Stenkamp, R. E., McDonald, R. C., & Steitz T. A. (1978) *J. Mol. Biol. 123*, 207−219.
Appelt, K., Bacquet, R. J., Bartlett, C. A., Booth, C. L. J., Freer, S. T., et al., (1991) *J. Med. Chem. 34*, 1925−1934.
Baker, D., Bystroff, C., Fletterick, R. J., & Agard, D. A. (1993) *Acta Crystallogr. D49*, 429−439.
Barbour, K. W., Berger, S. H., & Berger, F. G. (1990) *Mol. Pharmacol. 37*, 515−518.
Blum, M., Metcalf, P., & Harrison, S. C., & Wiley, D. C. (1987) *J. Appl. Crystallogr. 20*, 235−242.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
Brunger, A. T. (1990) *X-PLOR, version 2.1*, Yale University, New Haven, CT.
Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.
Bystroff, C., Baker, D., Fletterick, R. J., & Agard, D. A. (1993) *Acta Crystallogr. D49*, 440−448.
Chambers, J. L., & Stroud, R. M. (1977) *Acta Crystallogr. B33*, 1824−1837.
Crowther, R. A. (1972) in *The Molecular Replacement Method* (Rossmann, M. G., Ed.) pp 173−185, Gordon & Breach, New York.
Davis, S. T., & Berger, S. H. (1993) *Mol. Pharmacol. 43*, 702−708.
Davisson, V. J., Siraworaporn,W., & Santi, D. V. (1989) *J. Biol. Chem. 264*, 9145−9148.
Davisson, V. J., Siraworaporn,W., & Santi, D. V. (1994) *J. Biol. Chem. 269*, 30740.
Dayringer, H. E., Tramontano, A., Sprang, S. R., & Fletterick, R. J. (1986) (INSIGHT 2.5) *J. Mol. Graphics 4*, 82−87.
Fauman, E. B., Rutenber, E. E., Maley, G. F., Maley, F., & Stroud, R. M. (1994) *Biochemistry 33*, 1502−1511.
Finer-Moore, J., & Stroud, R. M. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 155−159.

Finer-Moore, J., Bazan, F., Rubin, J., & Stroud, R. M. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G., Ed.) pp 719−759, Plenum Press, New York.

Finer-Moore, J., Fauman, E. B., Foster, P. G., Perry, K. M., Santi, D. V., & Stroud, R. M. (1993) *J. Mol. Biol. 232*, 1101−1116.

Finer-Moore, J. S., Montfort, W. R, & Stroud, R. M. (1990) *Biochemistry 29*, 6964−6977.

Finer-Moore, J. S., Maley, G. F., Maley, F., Montfort, W. R., & Stroud, R. M. (1994) *Biochemistry 33*, 15459−15468.

Furey, W. J. (1984) in *Methods and Applications in Crystallographic Computing* (Hall, S. R., & Ashinda, T., Eds.) pp 353−371, Clarendon, Oxford, U.K.

Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V., & Stroud, R. M. (1987) *Science 235*, 448−455.

Hendrickson, W. A., & Konnert, J. H. (1978) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., Subramanian, E., & Yathindra, N., Eds.) pp 43−57, Pergamon, Oxford.

Howard, A. J., Nielsen, C., & Xuong, Ng H. (1985) *Methods Enzymol. 114*, 454−472.

Hughey, C. T., Barbour, K. W., Berger, F. G., & Berger, S. H. (1993) *Mol. Pharmacol. 44*, 316−323.

Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., & Nagai, K. (1991) *EMBO J. 10*, 3447−3456.

Jones, T. A. (1985) *Methods Enzymol. 115*, 157−171.

Kamb, A., Finer-Moore, J. S., & Stroud, R. M. (1992a) *Biochemistry 31*, 9883−9890.

Kamb, A., Finer-Moore, J. S., & Stroud, R. M. (1992b) *Biochemistry 31*, 10315−10321.

Knighton, E. R., Kan, C.-C., Howland, E., Janson, C. A., Hostomska, Z., Welsh, K. M., & Matthews, D. A. (1994) *Nature Struct. Biol. 1*, 186−194.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 23*, 946−950.

Leslie, A. G. W., Brick, P., & Wonacott, A. J. (1988) *An Improved Program Package for the Measurement of Oscillation Photographs*, Imperial College, London.

Matthews, D. A., Applet, K., Oatley, S. J., & Xuong, N. H. (1990a) *J. Mol. Biol. 214*, 923−936.

Matthews, D. A., Villafranca, J. E., Janson, C. A., Smith, W. W., Welsh, K., & Freer, S. (1990b) *J. Mol. Biol. 214*, 937−948.

Montfort, W. R., Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Maley, G. F., Maley, F., & Stroud, R. M. (1990) *Biochemistry 29*, 6977−6986.

Perry, K., Fauman, E., Finer-Moore, J. S., Montfort, W. R., Maley, C. F., Maley, F., & Stroud, R. M. (1990) *Proteins 8*, 315−333.

Perryman, S. M., Rossana, C., Deng, T., Vanin, E. F., & Johnson, L. F. (1986) *Mol. Biol. Evol. 3*, 313−326.

Plucinski, T. M., Fager, R. S., & Reddy, G. P., (1990) *Mol. Pharmacol. 38*, 114−120.

Reich, S. H., Fuhry, A. M., Nguyen, D., Pino, M. J., et al. (1992) *J. Med. Chem. 35*, 847−858.

Sack, J. S. (1988) *J. Mol. Graphics 6*, 224−225.

Santi, D. V., & Danenberg, P. V. (1984) *Folates and Pterins, Vol. 1: Chemistry and Biochemistry of Folates* (Blakley, R. L., & Benkovic, S. J., Eds.) pp 345−398, John Wiley and Sons, New York.

Schiffer, C. A., Davisson, V. J., Santi, D. V., & Stroud, R. M. (1991) *J. Mol. Biol. 219*, 161−163.

Shoichet, B. K., Stroud, R. M., Santi, D. V., Kuntz, I. D., & Perry, K. M. (1993) *Science 259*, 1445−1450.

Stroud, R. M., & Finer-Moore, J. S. (1993) *FASEB J. 7*, 671−677.

Stroud, R. M., Freymann, D., Finer-Moore, J.-F., Govinda-Rajan, R., Santi, D. V., & Perry, K. M., in preparation.

Takeishi, K., Saneda, S., Ayusawa, D., Shimizu, K., Gotoh, O., & Seno, T. (1985) *Nucleic Acids Res. 13*, 2035−2043.

Varney, M. D., Marzoni, G. P., Palmer, C. L., Deal, J. G., et al. (1992) *J. Med. Chem. 35*, 663−676.

veer Reddy, G. P., & Pardee, A. B. (1983) *Nature 304*, 86−88.

Wilson, C., & Agard, D. A. (1993) *Acta Crystallogr. A 49*, 97−104.

BI951476B