Crystal Structure of Thymidylate Synthase from T4 Phage: Component of a Deoxynucleoside Triphosphate-Synthesizing Complex[†,‡]

Janet S. Finer-Moore,[*,§] Gladys F. Maley,[||] Frank Maley,[||] William R. Montfort,[§,⊥] and Robert M. Stroud[§]

Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California 94143-0448, and Wadsworth Center, New York State Department of Health, Albany, New York 12201

Received August 12, 1994; Revised Manuscript Received September 30, 1994[⊗]

ABSTRACT: Thymidylate synthase from phage T4 (T4TS) is part of a complex of several enzymes required for coordinate DNA synthesis in infected *Escherichia coli* cells. It has been proposed that similar complexes of enzymes related to DNA synthesis are also functional in eukaryotes [Pardee, A. B. (1989) *Science 246*, 603−608]. To delineate the role of structure in the function of this complex, we have solved the structure of T4TS as a basis for mapping the complex by mutagenesis. The 3.1 Å structure of the unliganded enzyme was determined by molecular replacement and refined to 19.9% for all data. Three inserts and one deletion in the coding region are unique to T4TS, and all sites lie on one side of the enzyme surface, possibly encoding unique T4 specific intermolecular interactions during the infective cycle. The crystal structure is generally in the open, unliganded conformation seen in unliganded *E. coli* TS, as opposed to the closed, ternary complex conformation, except that the critically important C-terminus is inserted into the active site hydrogen bonded to residue Asn85, as seen in functional ternary complex structures. Other differences between *E. coli* TS and T4TS appear to explain the enhanced binding of folyl polyglutamate to the latter.

Thymidylate synthase (TS)[1] provides the sole, essential pathway for synthesis of the DNA base dTMP, catalyzing the reductive methylation of dUMP. TS is found in almost every living species, including viruses such as bacteriophage T4 (Greenberg et al., 1962). In infected cells, T4TS resides in a structured complex of viral enzymes whose activities are kinetically coupled for efficient synthesis of deoxynucleoside triphosphates (Mathews et al., 1988). In particular, T4TS has been shown to associate with deoxycytidylate deaminase (Mathews et al., manuscript in preparation) and with deoxycytidylate hydroxymethylase (Young & Mathews, 1992), which converts dCMP to 5-hydroxymethyl-dCMP (Flaks & Cohen, 1957; Pizer & Cohen, 1962). The hm-dCMP protects viral DNA from degradation by the viral restriction enzymes (Warner & Snustad, 1983). Studies of the T4 phage dNTP-synthesizing enzyme complex have been motivated in part by the intriguing possibility that dNTP synthesis may be physically linked to DNA replication (Mathews et al., 1988), although there is as yet no conclusive evidence that this is the case. However, there is evidence that T4TS interacts with the single-stranded DNA binding protein, T4 gene 32 product, and thus may be directly linked to the phage replication machinery (Wheeler et al., 1992).

Mechanisms for kinetic linkage among dNTP-synthesizing enzymes may be more general than commonly supposed: the conservation of sequence among TS's is higher than for essentially all other enzymes, suggesting additional, as yet unidentified roles for TS. There is evidence for multienzyme complexes that include dNTP-synthesizing enzymes in both prokaryotic and eukaryotic species (Mathews et al., 1988). TS and the next enzyme in the pathway, DHFR, are kinetically linked in protozoan bifunctional enzymes, in which these two enzymes are on the same polypeptide chain (Meek et al., 1985). In this case, products may be tunnelled from one enzyme to the other along a specific electrostatic pathway (Knighton et al., 1994). The interface between DHFR and TS in *Leishmania major* involves mainly nonspecific contacts between residues that are not conserved across species (Knighton et al., 1994). However, the electrostatic surface proposed as the pathway for the products is conserved among even monofunctional TS species (Stroud, 1994), suggesting a common mechanism.

The extremely high conservation of TS across species is perplexing if the enzyme does not also have conserved, nonenzymatic functions. One such function may be autoregulation: TS can autoregulate its own synthesis in both eukaryotic (Chu et al., 1993) and prokaryotic sources (Voeller et al., manuscript in preparation). Site-directed mutagenesis is beginning to explore the roles of the conserved residues.

[†] This work was supported by National Institutes of Health Grant RO1-CA-41323 to J.S.F.-M. and R.M.S., National Cancer Institute Grant (United States Public Health Service) CA-44355 to F.M., and National Science Foundation Grant MCB931621 to G.F.M. Computations at the Pittsburgh Supercomputer Center were supported by National Science Foundation Grant DMB890040P.

[‡] Coordinates for this structure have been deposited in the Brookhaven Protein Data Bank. The access code is 1TIS.

[§] University of California at San Francisco.

[||] New York State Department of Health.

[⊥] Present address: Department of Biochemistry, University of Arizona, Tucson, AZ 85721.

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1994.

[1] Abbreviations: TS, thymidylate synthase; T4TS, thymidylate synthase from phage T4; dUMP, 2′-deoxyuridine 5′-monophosphate; dTMP, thymidine 5′-monophosphate; CH_2H_4folate, 5,10-methylene-5,6,7,8-tetrahydrofolate; hm-dCMP, 5-(hydroxymethyl)deoxycytidine 5′-monophosphate; dNTP, deoxyribonucleoside 5′-triphosphate, PABA, *p*-aminobenzoic acid; P_i, inorganic phosphate; F_o, measured structure factor amplitude; F_c, calculated structure factor amplitude; $(F_o - F_c)\alpha_{calcd}$ map, electron density jmap calculated with coefficients $(|F_o| - |F_c|)$ and phases calculated from the coordinates; $(2F_o - F_c)\alpha_{calcd}$ map, electron density map calculated with coefficients $(2|F_o| - |F_c|)$ and phases calculated from the coordinates; DHFR, dihydrofolate reductase.

0006-2960/94/0433-15459$04.50/0

Another approach to identifying the function of conserved features of TS, and understanding how covariant changes preserve these functions, is to compare the three-dimensional structures of different species of the enzyme. So far, structures of two widely divergent prokaryotic TS species [*E. coli* (Perry et al., 1990; Matthews et al., 1990) and *Lactobacillus casei* (Hardy et al., 1987)] and one eukaryotic species, the bifunctional enzyme from *L major* (Knighton et al., 1994), have been reported. The structure of T4TS represents a fourth species that is widely divergent from the previously reported structures. Although bacteriophage T4 infects *E. coli*, its genome apparently is not derived from its host. Only 48.6% of the *E. coli* TS residues are conserved in T4TS, and respective antibodies to *E. coli* TS and T4TS do not interact with the other protein (Belfort et al., 1983a,c).

Efforts are underway to design drugs that selectively inhibit TS from disease-producing organisms. Although possible in principle, species specific inhibition of TS is complicated by the high conservation of the enzyme. One of the early motivations for studying T4TS was the discovery that a hexaglutamylated cofactor analog selectively inhibits T4TS over *E. coli* TS (Maley et al., 1979b). The poorly conserved polyglutamate binding site is presumably responsible for the selective tight binding of the cofactor analog to T4TS. In order to learn the details of this selective binding, the enzyme was cloned and expressed in *E. coli* (Belfort et al., 1983c) and then crystallized (Maley et al., 1984), and site-directed mutations were made in the polyglutamyl binding region of T4TS (LaPat-Polasko et al., 1990). Our crystal structure analysis was done partly to address the issue of selective binding in TS.

MATERIALS AND METHODS

Purification and Crystallization of Phage T4-Encoded TS. T4TS was isolated from a cloned gene (Belfort et al., 1983c) and purified as previously described (LaPat-Polaska et al., 1990). Crystals were grown at pH 7.5 by vapor diffusion against a 25−30% ammonium sulfate solution from 4.35 mg/mL T4TS in 0.2 M phosphate buffer. Large, square-pyramidal crystals formed, which diffracted to 2.8 Å on still photographs. One crystal was characterized by precession photography. A second crystal, about 250 μm on an edge, was used for data collection.

Data Collection. Data (3.1 Å) were collected at room temperature on a Siemens IPC area detector with a three-circle goniometer, using a 200 μm Cu Kα focal spot source from a Rigaku rotating target tube with a graphite monochrometer. Oscillation widths were 12 min in ω, and counting times were 35 min per degree of rotation. Data frames were processed by a modified version of the software of Blum et al. (1987). Data were reduced to intensities using the programs of Howard et al. (1987).

Molecular Replacement and Refinement. The structure was solved by molecular replacement using the Crowther rotation function (Crowther, 1972), with a radius of integration from 4.0 to 23.5 Å and data between 10 and 4 Å resolution, followed by a 0.3 Å grid *R*-factor translation search using a program written by E. Dodson and P. Evans. Rotation and translation parameters were then optimized using rigid body rotation and a translation grid search (SIXDREF, V. Ramalingam, personal communication).

The structure was refined using computer graphics (FRODO) (Jones, 1985), in conjunction with molecular dynamics coupled with *R*-factor and energy minimization (XPLOR) (Brunger et al., 1987). All recorded data were used for the computation of $(F_o - F_c)\alpha_{calcd}$ and $(2F_o - F_c)\alpha_{calcd}$ maps and for refinement. Bulk solvent contributions to the calculated structure factors were included. Individual *B*-factors were refined, but were restrained to be correlated to the *B*-factors of neighboring atoms. Restrained *B*-factors were determined to be an appropriate model of thermal motion, on the basis of a 1% decrease in the free *R*-factor (Brunger, 1992), in going from an overall *B*-factor to refined, restrained atomic *B*-factors.

Table 1: Data Statistics for T4TS

resolution range (Å)	no. of collected reflections	% of possible reflections collected	no. of total observations	av I/σ (I)	R_{sym} (%)
∞−4.9	1627	94	11681	47.2	6.7
4.9−3.9	1498	93	8758	25.2	10.0
3.9−3.4	1431	91	6416	11.3	15.2
3.4−3.1	1057	68	2983	5.1	21.5
3.1−2.9	381	24	453	2.3	24.2
2.9−2.7	192	12	227	2.0	25.7
total	6186	65	30518	22.2	6.8

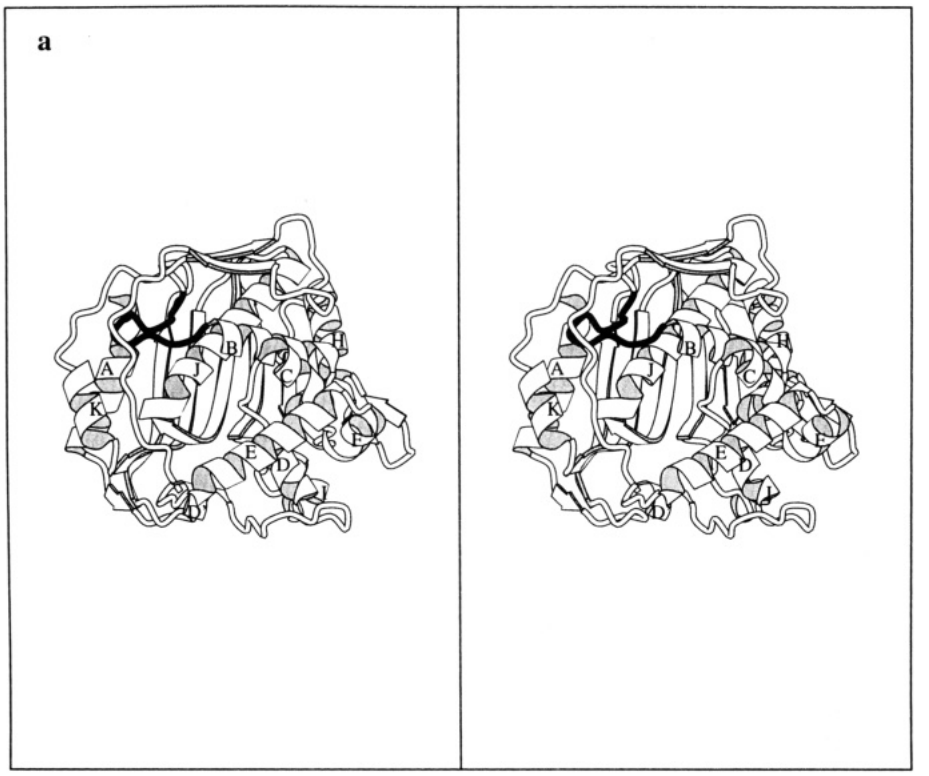
Structure Analysis. Homologous TS structures were overlapped as described previously (Montfort et al., 1990). The significance of the differences between atoms in the overlapped structures was estimated as a function of the atomic *B*-factors (B) of the compared atoms using an equation of the form $\sigma_{(x,y,z)}(B) = (aB^2 + bB + c)$, where a, b, and c are empirically determined constants (Chambers & Stroud, 1979; Perry at al., 1990; Fauman & Stroud, manuscript in preparation). Solvent accessible surface areas, as defined by F. M. Richards (1977), were calculated using the program DMS, written by Conrad Huang at the University of California at San Francisco.

RESULTS AND DISCUSSION

Structure Solution

Crystals of T4TS are tetragonal, with a monomer in the asymmetric unit and unit cell dimensions of $a = b = 68.02$ Å and $c = 140.8$ Å. Every species of TS studied so far exists as an obligate 2-fold symmetric dimer. The crystallographic 2-folds in the T4TS crystals therefore necessarily coincide with the dimer 2-fold. On the basis of systematic absences, the space group is either $P4_32_12$ or $P4_12_12$; molecular replacement identifies the correct space group as $P4_32_12$. Table 1 summarizes the statistics for the X-ray data.

Molecular replacement was done with a dimer search model derived from the refined crystal structure of *L. casei* TS (Finer-Moore et al., 1993). Regions where there are inserts in or deletions from *L. casei* TS relative to T4TS were deleted from the search model (Figure 1), and all nonconserved, non-glycine residues were changed to alanines. The truncated *L. casei* TS dimer was oriented with the 2-fold axis relating the protomers parallel to the Cartesian y-axis. The best-scoring rotation solution orients the dimer 2-fold along the $(x,x,0)$ crystallographic 2-fold. A translation search in space group $P4_32_12$ using a 0.27 Å grid gives a significant peak, $R = 51.7\%$, slightly displaced from the crystallographic 2-fold, at $x = 2.4$ Å and $y = 1.0$ Å; the next highest solution has an *R*-factor of 54% in a background



b

```
                                  1       A     16 -i      27     i   38
                                  |-------------| |^|      |^^^^^^^^^^|
L. casei                          MLEQPYLDLAKKVLDEGHFKPDRTHTGTYSIFGHQMRFDLSK

T4 Phage                            MKQYQDLIKDIFENGYETDDRTGTGTIALFGSKLRWDLTK
E. coli                             MKQYLELMQKVLDEGTQKNDRTGTGTLSIFGHQMRFNLQD
Human     MPVAGSELPRRPLPPAAQERDAEPRPPHGELQYLGQIQHILRCGVRKDDRTGTGTLSVFGMQARYSLRD

                     54     B     67  71  C  78          83   D  91           104
                      |------------|   |------|           |-------|             |
L. casei  GFPLLTTKKVPFGLIKSELLWFLHGDTNIRFLLQHR        NHIWDEWAFEKWVKSDEYHGPDMTDF

T4 Phage  GFPAVTTKKLAWKACIAELIWFLSGSTNVNDLRLIQHDSLIQGKTVWDENYENQAKDLGYHS
E. coli   GFPLVTTKRCHLRSIIHELLWFLQGDTNIAYLHENN       VTIWDEWADEN
Human     EFPLLTTKRVFWKGVLEELLWFIKGSTNAKELSSKG       VKIWDANGSRDFLDSLGFSTREE

          D'     113        E       131   F  140  145  G 152              162
          ----|   |-----------------| |------|    |------||^|           |^||----
L. casei  GHRSQKDPEFAAVYHEEMAKFDDRVLHDDAFAAKYGDLGLVYGSQWRAWHTS         KGDTIDQLGD

T4 Phage                                GELGPIYGKQWRDFG              GVDQIIE
E. coli                                 GDLGPVYGKQWRAWPTP          DGRHIDQITT
Human                                   GDLGPVYGFQWRHFGAEYRDMESDYSGQGVDQLQR

            H   174    180 v    I 191    198  iv  207    210  iii 222          J
          -------|     |^^^| |----|      |^^^^^^^^|     |^^^^^^^^|--------------
L. casei  VIEQIKTHPYSRRLIVSAWNPEDVPTMALPPCHTLYQFYVNDGKLSLQLYQRSADIFLGVPFNIASYALL

T4 Phage  VIDRIKKLPNDRRQIVSAWNPAELKYMALPPCHMFYQFNVRNGYLDLQWYQRSVDVFLGLPFNIASYATL
E. coli   VLNQLKNDPDSRRIIVSAWNVGELDKMALAPCHAFFQFYVADGKLSCQLYQRSCDVFLGLPFNIASYALL
Human     VIDTIKTNPDDRRIIMCAWNPRDLPLMALPPCHALCQFYVVNSELSCQLYQRSGDMGLGVPFNIASYALL

               244   250       ii   263   K  272        IIa
          -------|     |^^^^^^^^^^||--------|        |^^^|
L. casei  THLVAHECGLEVGEFIHTFGDAHLYVNHLDQIKEQLSRTPRPAPTLQLNPD             KHDIFDF

T4 Phage  VHIVAKMCNLIPGDLIFSGGNTHIYMNHVEQCKEILRREPKELCELVISGLPYKFRYLSTKEQLKYVLKL
E. coli   VHMMAQQCDLEVGDFVWTGGDTHLYSNHMDQTHLQLSREPRPLPKLIIKRK              PESIFDY
Human     TYMIAHITGLKPGDFIHTLGDAHIYLNHIEPLKIQLQREPRPFPKLRILRK              VEKIDDF

                IIb
               |^^^|
L. casei  DMKDIKLLNYDPYPAIKAPVAV
T4 Phage  RPKDFVLNNYVSHPPIKGKMAV
E. coli   RFEDFEIEGYDPHPGIKAPVAI
Human     KAEDFQIEGYNPHPTIKMEMAV
```

FIGURE 1: (a) Ribbon drawing of *L. casei* TS, which was used as the model for solving the T4TS crystal structure by molecular replacement. The so-called folate binding loop, which includes residues involved in binding the polyglutamate tail of the cofactor (Maley et al., 1982; Kamb et al., 1992a), is shaded black. All plots in this paper are convergent ("crossed-eyes") stereoviews generated with MOLSCRIPT (Kraulis et al., 1991). (b) The aligned *L. casei* TS (Maley et al., 1979a), *E. coli* TS (Belfort et al., 1983b), human TS (Takeishi et al., 1985), and T4TS (Chu et al., 1984) sequences, with the *L. casei* TS residues included in the search model underlined. Above the *L. casei* sequence, the secondary structure for the *L. casei* structure (Finer-Moore et al., 1993) is marked: |- - -|, helix; ^^^, β-strand.

Table 2: Refinement Statistics for T4TS

Parameters		
no. of atoms	2362	
no. of waters	20	
no. of discretely disordered side chains	1	
B-factor model	restrained, isotropic	
Diffraction Agreement		
resolution (Å)	50.0−2.7	7.0−2.7
no. of reflections	6160	5537
σ cutoff	0.0	0
R_{crys} (%)	19.9	17.8
Stereochemical Ideality		
bond lengths (Å)	0.016	
bond angles (deg)	3.8	
torsions (deg)	25.6	
av *B* (Å^2)	28	

of $R = 56\%$. After rigid body refinement of the molecular replacement search model, the *R*-factor for 4.5 Å data is 48.5%.

A $(2F_o - F_c)\alpha_{calcd}$ map derived from the molecular replacement solution shows density for several T4TS side chains not present in *L. casei* TS, confirming that the molecular replacement solution is correct. Subsequent maps show well-defined density for the inserts unique to T4TS. When all residues had been built into convincing density [there were no breaks in the backbone density, except for the most disordered stretch of residues, 272(230)−275(233),[2] and all side chains except that of Lys311(281) were fit to density], the structure was refined against 7−2.7 Å data, omitting the solvent mask, to give a final *R*-factor of 17.7% (19.9% for all data). Table 2 lists refinement statistics.

In the final structure, conserved parts of the sequence are also structurally conserved with respect to the *L. casei* TS and *E. coli* TS crystal structures, as expected. The unique parts of the T4TS sequence are clearly delineated in well-defined density; (F_o-F_c) omit maps, in which these inserts were left out of the F_c calculations, show their structures unambiguously (Figure 2). The average *B*-factors for the small domain, residues 71(69)−97(102), and the C-terminal insert, residues 287a(246)−287l(257), are 28.5 and 24 Å^2, respectively, which is about the same as the average *B*-factor for the entire protein, 28.2 Å^2. The only (ϕ,ψ) violation in the T4TS inserts is for Ser97(102), where $\phi = 15°$ and $\psi = 77°$. There are three other (ϕ,ψ) violations in the structure, two of which fall in disordered regions of the structure. In all cases, the residues can be fit to density in allowable conformations, but they move slightly away from the allowed conformations during energy minimization and must be put back into allowable conformations.

Component of the dNTP-Synthesizing Complex

Inserts Alter the Surface Structure, Providing Unique Interfaces for Protein/Protein Interactions. In T4 phage-infected cells, T4TS and other phage-encoded enzymes that synthesize the deoxyribonucleotide triphosphates for phage DNA assemble as a complex (Mathews et al., 1988). Thus, the enzyme may have evolved a unique surface for the purpose of binding other T4 enzymes. Recently, T4TS has been shown by means of anti-idiotypic antibodies to bind specifically to another member of the dNTP synthetase aggregate, dCMP hydroxymethylase (Young & Mathews, 1992).

As shown in Figure 1, the major domain of thymidylate synthase, which is highly conserved among all species, consists of three layers: first, the six-stranded β-sheet, which makes up most of the dimer interface; second, the hydrophobic helix J, which makes up the hydrophobic core of the protomer, and amphipathic helices A, on one edge of the protein, and G and H on the opposite edge; and the third layer, consisting of helices B and K and the long C-terminal loop.

The small domain is variable in length and sequence among different species. It spans residues 71−139 and includes helix C and the residues following up to helix G. The small domains of eukaryotic TS species are conserved with one another in length (31 amino acids) and sequence. *L. casei* TS has a much larger small domain (69 amino acids), which is similar only to the small domain of TS from *Staphylococcus aureus* transposon Tn*4003* (Rouch et al., 1989). The small domain of T4TS is 34 amino acids long and poorly conserved with all other TS species. A cleft separates the large and small domains, and the only contacts between the domains are between residues on helix B and Leu74 on helix C.

Inserts in T4TS relative to *E. coli* TS are in the small domain and in the third layer of the major domain. There is a five-residue deletion in the eukaryotic loop between helices G and H, which protrudes from one side of the protein and helps form one edge of the dimer interface. The inserts and the deletion are clustered on one side of the protein (Figure 3). Taken together, they provide a flatter surface on this side of the protein than in other TS species.

The surface defined by the inserts contains a number of residues that are moderately hydrophobic (such as tyrosine), or whose side chains contain a large hydrocarbon surface (such as lysine or arginine). Horton and Lewis (1992) find that the free energy of association of proteins correlates well with the amount of surface area buried in the protein/protein interface when no conformational changes are required for association. The interaction energy driving the association often includes sizable contributions from hydrogen-bonding and ionic interactions between polar groups, but is usually dominated by the entropic contribution from burying hydrophobic surface area, as well as nonspecific van der Waals interactions between apolar atoms. Thus, the relatively hydrophobic surface formed by the T4TS inserts may interact favorably with hydrophobic surfaces of other proteins.

One of the crystal contact interfaces in T4TS involves these unique insert residues and may reflect the packing of TS molecules with themselves or with other proteins, such as dCMP deaminase or dCMP hydroxymethylase, in the infected *E. coli* cells. At this interface, the C-terminal loop insert is close-packed against a portion of the small domain insert and Tyr192(150), without an intermediate layer of water, and makes several nonspecific (nonhydrogen-bonding) protein/protein contacts (Figures 4 and 5). Two solvent-exposed hydrophobic residues, Leu78d(80) and Ile78e(81), in the insert after the C helix are also partially buried at this

[2] Residue numbers refer to the *L. casei* sequence and are followed by the T4TS numbering in parentheses. Residues that are part of an insert in T4TS relative to *L. casei* TS are numbered with the *L. casei* sequence number at the position of the insert, followed by a letter designating the position of the T4TS residue in the insert (a is the first residue, etc.). Residues from the second protomer are indicated with a prime (′).

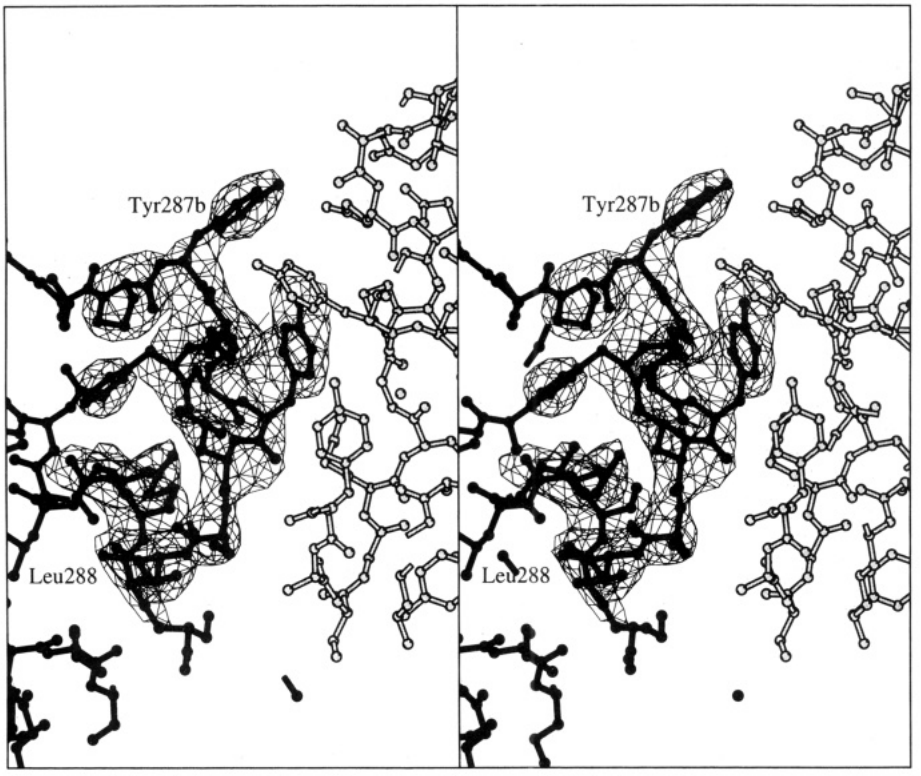


FIGURE 2: $(F_o - F_c)\alpha_{calcd}$ omit map, showing a C-terminal insert unique to T4TS, residues 287a(246)−287l(257), which was excluded from the calculation of structure factors.

crystal contact by proximity to Leu173(131) in the symmetry-related protein molecule. Residue 173 is hydrophilic in all other TS species. It is unusual for it to be substituted by a hydrophobic leucine in T4TS since it is completely exposed to solvent in the dimer; therefore, it may also contribute to a binding surface. A total of 512 Å^2 of hydrophobic surface area is buried at this interface, of which 172 Å^2 is from the unique T4TS inserts.

The protein surface whose features are altered by inserts and a deletion in T4TS is distinct from, but adjacent to, the DHFR/TS interface seen in the bifunctional enzyme from *L. major*. In the bifunctional enzyme, DHFR lies along the extended loop leading to the C-terminus (Knighton et al., 1994), on top of the molecule as viewed in Figure 1a. The most extensive contacts, most of them nonspecific van der Waals interactions, are between an N-terminal extension of the *L. major* DHFR/TS and a shallow channel on the TS surface running between the second and third layers of the protein. The N-terminal extension is in contact with one end of the C-terminal loop, where T4TS has a twelve-residue insert, and the eukaryotic loop, where T4TS has a five-residue deletion and eukaryotic species have an insert. The N-terminal extension is not present in DHFR from T4 phage. DHFR is part of the T4TS dNTP-synthesizing complex, but specific contacts between T4TS and DHFR have not yet been demonstrated.

The role of the T4TS inserts in binding other proteins is being tested by mutational analysis. Thus, mutations of residues on the putative binding surfaces might preclude proper incorporation of T4TS into the dNTP synthetase aggregate and destroy coupling of the dNTP synthetase reactions. Residues to test for possible involvement in protein/protein interfaces include residues in the T4TS inserts, as well as residues along the extended loop leading to the C-terminus, where DHFR contacts TS in *L. major* DHFR/TS. The crystal structure of a cocrystal of T4TS and dCMP hydroxymethylase might provide a direct view of the contacts between these proteins in the dNTP synthetase aggregate.

Differences in the Highly Conserved Active Site Are Minor. The structure of the ligand binding cavity is extremely well-conserved between T4TS and *E. coli* and *L. casei* TS's. Therefore, the same residues most likely contact dUMP and folate in the ternary complex of T4TS as in *E. coli* TS. The roles of TS residues in ligand binding and catalysis that were deduced from the *E. coli* TS ternary complex structure (Stroud & Finer-Moore, 1993) clearly seem to hold for T4TS as well.

The one exception to structural conservation of active site residues is invariant Arg178′(136′), which is one of four arginines hydrogen-bonded to P_i or to the phosphate moiety of dUMP in *E. coli* and *L. casei* TS structures (Stroud & Finer-Moore, 1993). In the unliganded T4TS structure, the Arg178′(136′) side chain is oriented away from the active site cavity. Arg178′(136′) is hydrogen-bonded to Asn257-(215) in the β-sheet, and there is an ordered water between the guanidinium group and the inorganic phosphate. However, rotation about one bond in the side chain is all that would be required to swing the guanidinium group of the Arg into the structurally conserved orientation. Since mutagenesis of T4TS indicates that Arg178′-(136′) has a significant role in ligand binding (J. Pedersen-Lane and F. Maley, unpublished results), Arg178′(136′) likely is in the structurally conserved orientation in T4TS ternary complexes.

Residue 85 (*L. casei* numbering) forms part of one wall of the binding site cavity, against which the pterin ring of the cofactor packs. It is a partially conserved ligand binding residue: it is either a tryptophan or an asparagine in all but one of 28 available TS sequences. In the *E. coli* TS−dUMP−CB3717 complex, the C-terminus has moved into

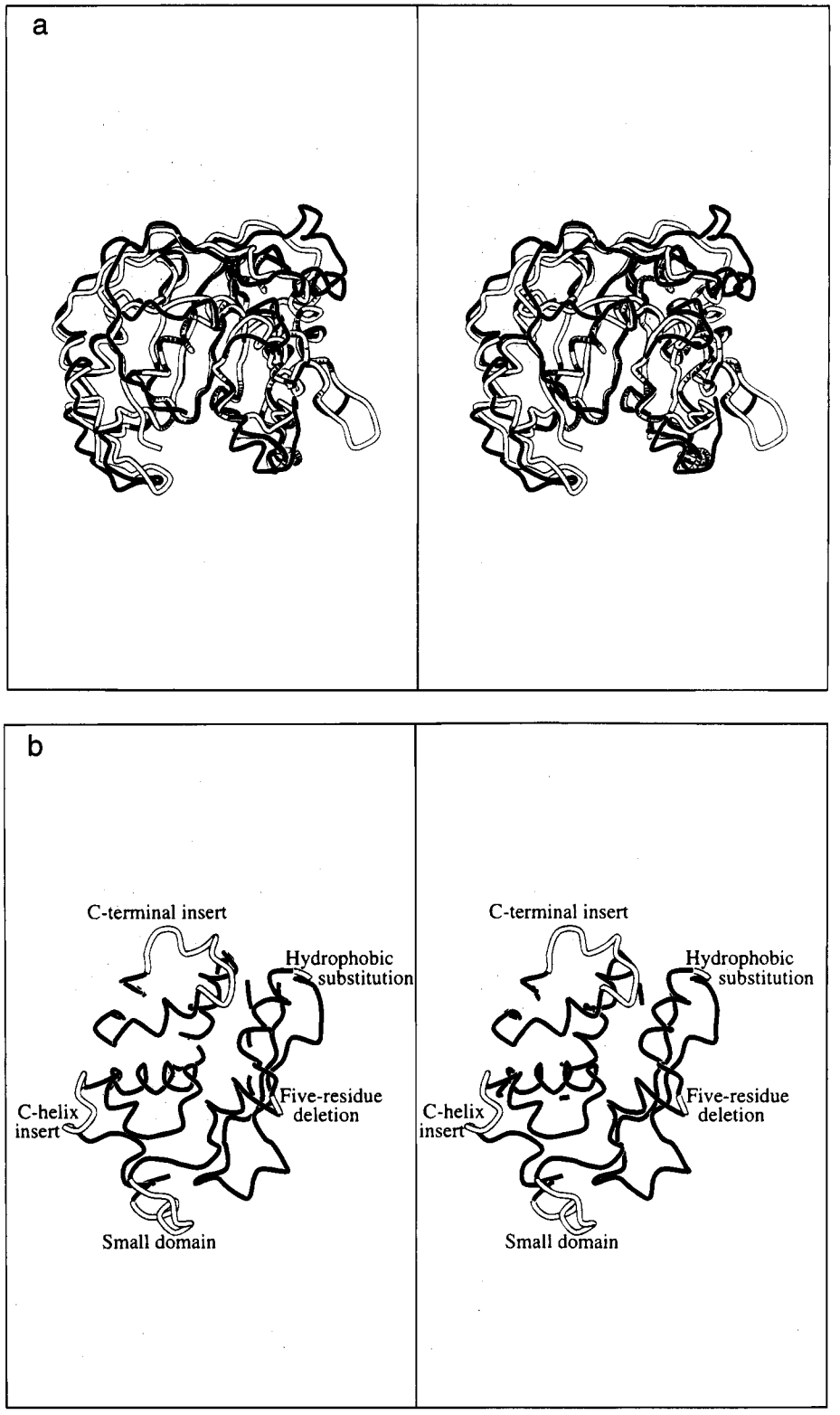


FIGURE 3: (a) α-carbon tracing of T4TS (in black) superimposed on the α-carbon tracing of the smaller *E. coli* TS structure, shown with white coils. (b) View of the T4TS α-carbon tracing made by a 90° clockwise rotation of the orientation of the molecule in part a, showing clustering of the two inserts, the deletion, and a hydrophobic substitution on one face of the protein. The inserts, hydrophobic substitution, and residues bordering the deletion are indicated with white coils.

the active site to form part of the cofactor binding site, and its new conformation is stabilized by Trp85 (83 in *E. coli* TS), which hydrogen bonds to the C-terminus through its side chain nitrogen (Montfort et al., 1990). When T4TS is overlapped with *E. coli* TS, the nitrogen of the asparagine side chain overlaps the hydrogen-bonding nitrogen of the Trp85 side chain in the *E. coli* structure, as would be expected if the hydrogen-bonding function of this side chain is conserved.

The inserts in and deletion from T4TS do not alter the fold of the remainder of the protein. Secondary structure elements, including the β-ribbon bordering the insert in the C-terminal loop, are preserved. The β-ribbon in the C-terminal loop and the C helix, which also borders an insert,

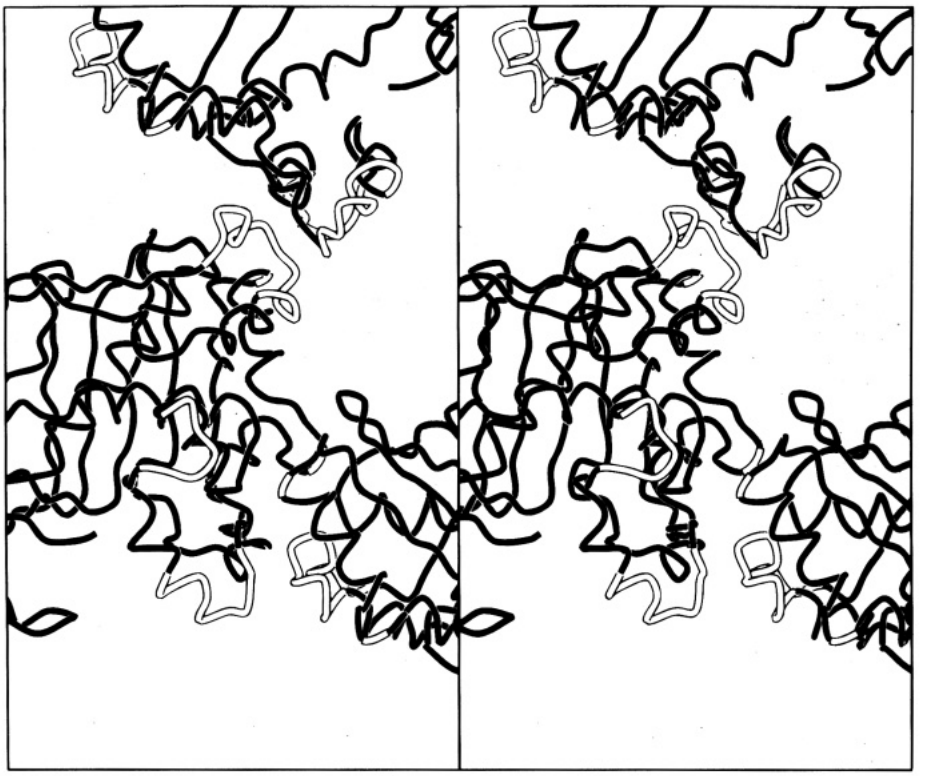


FIGURE 4: Tracing of T4TS molecules in the crystal. Inserts, a hydrophobic substitution, and residues bordering the deletion are shown in white. The inserts are involved in some of the crystal contacts.

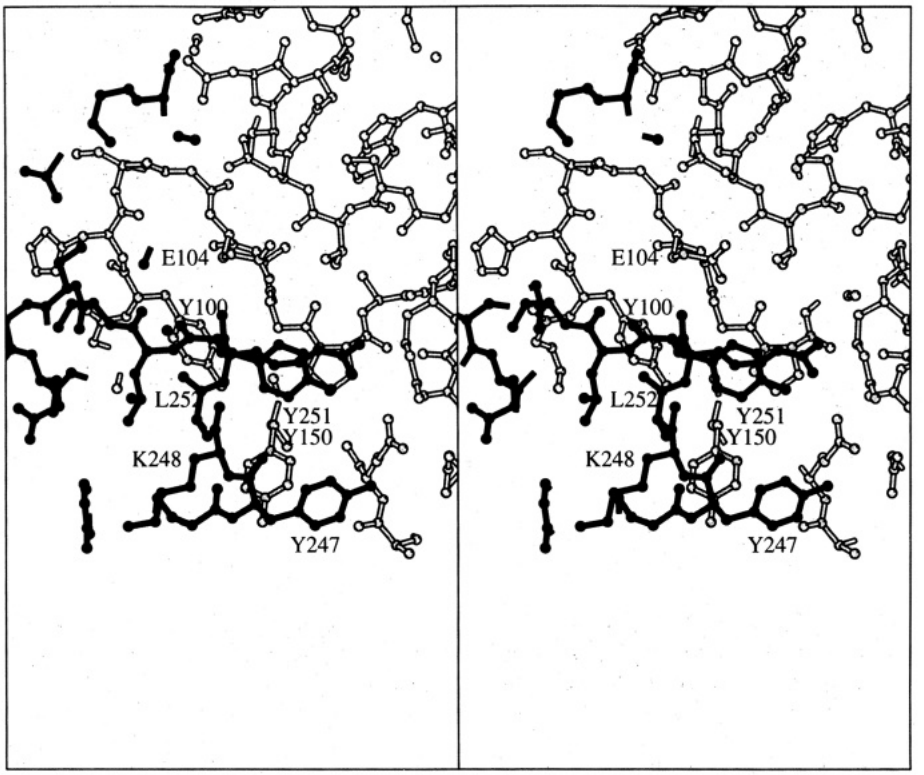


FIGURE 5: Closer view of one of the interfaces shown in Figure 4 detailing the crystal contacts. The two symmetry-related molecules forming the contact are in different shades of gray. The symmetry-related molecules contain many atoms within 4 Å of each other, but there are few specific hydrogen-bonding contacts between them.

are shifted significantly with respect to the native *E. coli* TS structure (Figure 3a). For the residues bordering the inserts, these shifts are 7 or 8 times the random error in position differences calculated as a function of the *B*-factor (Chambers & Stroud, 1979; Perry et al., 1990; Fauman & Stroud, manuscript in preparation). This plastic accommodation of sequence inserts is similar to that described in the comparison of *L. casei* TS to *E. coli* TS (Perry et al., 1990). *L. casei* TS has a 30 amino acid insert in the small domain relative to *E. coli* TS. As a consequence, helices B, C, G, and I in the vicinity of the *L. casei* TS small domain are shifted relative to the *E. coli* TS structure.

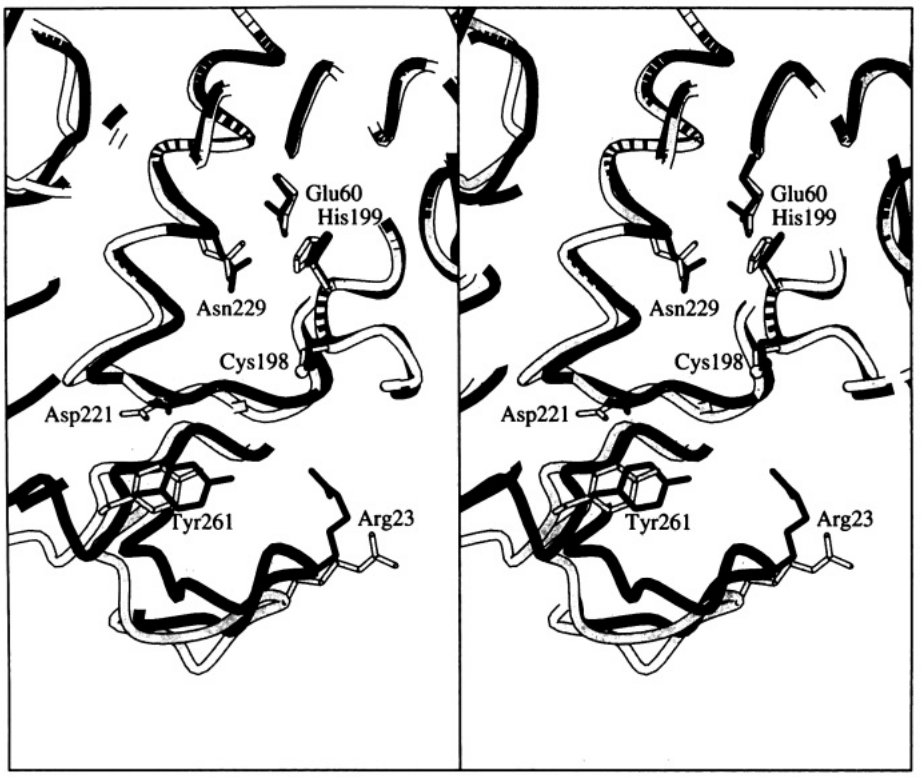


FIGURE 6: T4TS α-carbon tracing (gray) superimposed on the *E. coli* TS ternary complex (black) (ligands not shown). Side chains of ligand binding residues are plotted. The shifts of protein segments toward the active site when the ternary complex is formed have not taken place in T4TS, even though the C-terminus is moved into the active site and is hydrogen-bonded to Asn85(90), as in the *E. coli* TS ternary complex.

Structural Features Related to Ligand-Induced Conformational Change

The Folate Binding C-Terminal Residues Move Independently of the Ternary Conformational Change in TS. When *E. coli* TS binds its cofactor, it undergoes a major conformational change in which segments of the protein shift toward the active site. Most of the segmental shifts are about 1 Å or less, but the C-terminal residues move 5 Å into the active site where they form part of the cofactor binding site (Montfort et al., 1990; Matthews et al., 1990; Kamb et al., 1992b). The *L. casei* enzyme undergoes a similar conformational change upon ternary complex formation (unpublished results). The crystallographic *B*-factors for the C-terminus in unliganded *E. coli*, *L. casei*, and human TS (unpublished) crystal structures are high, indicating that the C-terminus is relatively unconstrained, but the conformation of the C-terminal residues in all three structures is the same. Thus, the conformational change upon binding ligands may be a conserved requirement for these three species.

In contrast, the three C-terminal residues of unliganded T4TS are already at the bottom of the active site cavity (Figure 3a) with the C-terminus hydrogen-bonded to Asn85, as in the ternary complex of *E. coli* TS with CB3717 and dUMP. However, the protein is not in the productive, closed conformation characteristic of the ternary complexes (Figure 6), and the final four residues are not in a conformation that would allow productive cofactor binding analogous to that seen in other TS species. The nonconserved C-terminal conformation may result from crystal packing, since Cβ of Ala315(285) is 3.5 Å from the side chain carboxyl group of Glu275(233) from a symmetry-related molecule. If the C-terminus were in the conserved conformation, it would make energetically unfavorable contacts with residues Glu275(233)−Lys277(235) in the symmetry mate.

Hydrogen Bonding Surrounding the Folate Binding Loop Is Different. We have postulated that a cluster of conserved hydrogen bonds involving invariant residues preserves structural integrity during the conformational change (Finer-Moore et al., 1993). Invariant residues Thr48 and Thr49 are key residues in the hydrogen-bonding network that form four hydrogen bonds to the side chain of Glu3 in *L. casei* TS (Figure 7). In most other TS species, residue 3 is a Glu or an Asp, but *E. coli* TS and T4TS have a two-residue deletion at the N-terminus; thus, the corresponding residue in these two species is the N-terminal methionine. In *E. coli* TS, the N-terminus becomes carbamylated, and the carbamyl group mimics the glutamate side chain of L. casei TS, forming all four hydrogen bonds to Thr48 and Thr49 (Fauman et al., 1994). Thus, Thr48 and Thr49 have a key role in structure and, we speculate, in the cofactor-induced conformational change.

Curiously, there is no density for a carbamate group at the N-terminus of T4TS. The N-terminal methionine and Thr48(46) and Thr49(47) have moved apart relative to their positions in *E. coli* TS. The movement of the folate binding loop also positions it too far from the C-terminal segments to hydrogen bond to Arg274(231) and Tyr304(274). The same segments that are well ordered and linked by hydrogen bonds in *E. coli* and *L. casei* TS's have high *B*-factors in T4TS: 2−4 times higher than in *L. casei* TS, which has an overall *B*-factor similar to that of T4TS. This suggests that the hydrogen-bonding network identified in the bacterial TS's, which is absent from the T4TS crystals, makes the

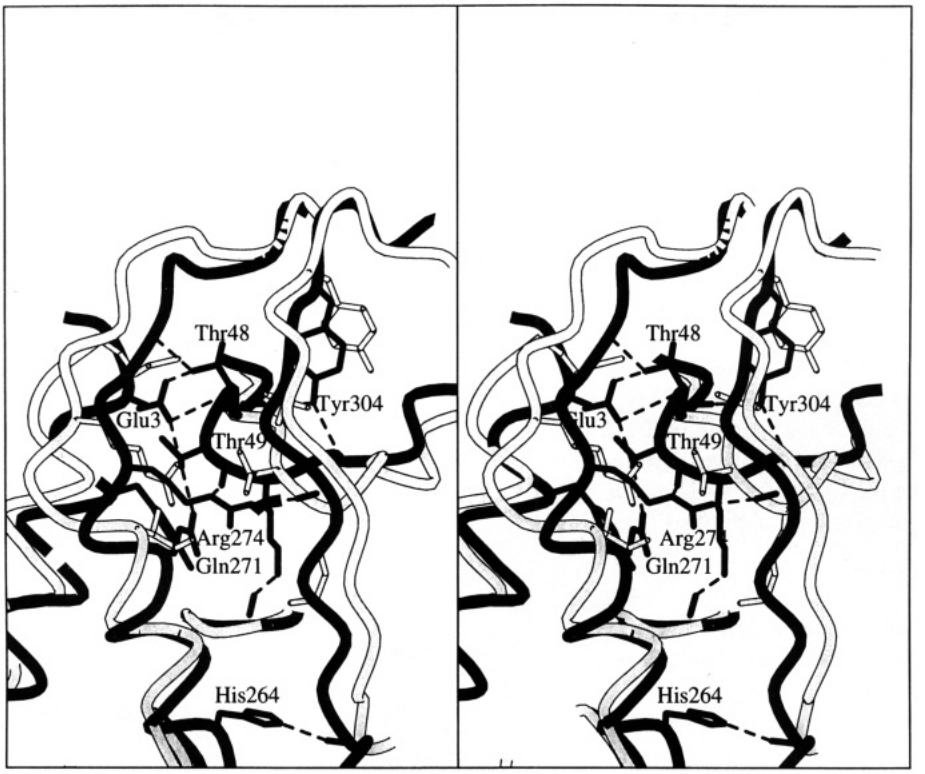


FIGURE 7: Illustration of the hydrogen bonding of the folate binding loop (see Figure 1) to helices A and K and segments of the long C-terminal loop in *L. casei* TS. Side chains and carbonyl groups involved in the hydrogen-bonding network are shown, along with the black α-carbon tracing. The gray α-carbon tracing is for T4TS superimposed on *L. casei* TS. The hydrogen-bonding scheme is not conserved in the T4TS crystal structure.

protein more rigid. A prerequisite of such a hydrogen-bonding scheme in *E. coli* TS or T4TS is carbamylation of the N-terminus, a modification that, although not seen in T4TS, would be expected *in vivo*. If the conserved hydrogen-bonding network were absent from T4TS *in vivo*, we predict an effect on the conformational change that occurs upon cofactor binding. Lack of carbamylation in the T4TS crystals may result from the altered conformation of the C-terminal residues and therefore may be another artifact of crystal packing.

The Response of T4TS and E. coli TS to Polyglutamyl Folates Is Dramatically Different

In the presence of Mg^{2+}, inhibition of *E. coli* TS by polyglutamylated folates progressively increases as the number of glutamate residues in the polyglutamyl moiety increases, while inhibition of T4TS by polyglutamylated folates is unaffected by the number of glutamate residues in the polyglutamyl tail (Kisliuk et al., 1979). However, under low-salt conditions, T4TS is strongly inhibited by pteroyl hexaglutamate in concentrations at which *E. coli* TS shows no inhibition (Maley et al., 1979b). Addition of Mg^{2+} to the assay mixture diminishes binding of the hexaglutamyl inhibitor to T4TS while increasing its binding to *E. coli* TS.

In vivo, folates are modified by as many as seven glutamate residues. The crystal structure of the ternary complex of *E. coli* TS with dUMP and a tetraglutamylated folate has revealed the binding site for a portion of the polyglutamate tail of folates (Kamb et al., 1992a). The tetraglutamyl binding site is a shallow cleft on the protein surface lined with hydrophilic, poorly conserved side chains. Variability in the sequences forming the cleft makes it an attractive target for species specific drugs.

The polyglutamyl binding clefts of T4TS and *E. coli* TS are significantly different. The polyglutamyl tail rests on the face of the imidazolidine ring of His51 in the *E. coli* TS complex. In T4TS, residue 51 is an alanine; thus, the floor of the binding site formed by the histidine is absent, and the polyglutamyl tail can bind deeper into the cleft. T4TS has an extra lysine near the C-terminus: Lys313(283), whose side chain is in the polyglutamyl binding cleft in the T4TS structure. In *E. coli* TS this residue is a proline. There is also a lysine just after the first T4TS insert (Val80 in *E. coli* TS), whose side chain is in the polyglutamyl binding cleft. Finally, the T4TS sequence connecting the ends of the two paired β-strands in the C-terminal loop, which contains one of the T4TS inserts, is especially rich in lysines and arginines. There are five lysines and two arginines in this sequence, compared to two arginines and a lysine in *E. coli* TS. The side chains of some of these residues may be close enough to the polyglutamyl binding cleft to interact with a negatively charged polyglutamyl tail. Side chains of Arg295(265) and Lys297(267) in particular are close to the projected path of the polyglutamyl tail beyond the third glutamyl residue (Kamb et al., 1992a). Arg295(265) aligns with an arginine in *E. coli* TS, but in *E. coli* TS this arginine is in a different position, farther from the polyglutamyl binding cleft. The extra positively charged residues in T4TS compared to *E. coli* TS may contribute to polyglutamyl binding via electrostatic or hydrogen-bonding interactions. The effects of Mg^{2+} on folate binding in *E. coli* TS and T4TS cannot be fully explained with the available data. Mg^{2+} may shield the polyglutamyl tail from interaction with charged residues or influence the conformation of the folate binding site.

CONCLUSION

The protein surface of T4TS is modified by inserts and deletions relative to the normal host *E. coli* enzyme. The structure of the interior of the enzyme is conserved, particularly the active site cavity. Side chains of residues that have been noted as important in the enzymatic reaction are structurally conserved, validating their assigned roles in catalysis or ligand binding. Critical differences in the polyglutamyl binding site contribute to the differential binding of polyglutamylated folate inhibitors to T4TS and *E. coli* TS.

Differences between T4TS and the bacterial structures provide new interfaces for the quaternary interactions made specifically by T4TS and for the binding of T4TS selective inhibitors. Most relevant are the insertions that may be involved in promoting the binding of T4TS to the T4 phage deoxynucleoside triphosphate multienzyme complex.

ACKNOWLEDGMENT

We are very grateful to Nancy Helmers for crystallizing T4TS in the Stroud laboratory. We thank Ralph Reed for helpful discussions.

REFERENCES

Belfort, M., Maley, G. F., & Maley, F. (1983a) *Proc. Natl. Acad. Sci. U.S.A. 80*, 1858−1861.
Belfort, M., Maley, G., Pederson-Lane, J., & Maley, F. (1983b) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4914−4918.
Belfort, M., Moelleken, A., Maley, G. F., & Maley, F. (1983c) *J. Biol. Chem. 258*, 2045−2051.
Blum, M., Metcalf, P., Harrison, S. C., & Wiley, D. C. (1987) *J. Appl. Crystallogr. 20*, 235−242.
Brunger, A. T. (1992) *Nature 355*, 472−475.
Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458−460.
Chambers, J. L., & Stroud, R. M. (1979) *Acta Crystallogr. B35*, 1861−1874.
Chu, F. K., Maley, G. F., Maley, F., & Belfort, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3049−3053.
Chu, E., Voeller, D., Koeller, D. M., Drake, J. C., Takimoto, C. H., Maley, G. F., Maley, F., & Allegra, C. J. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 517−521.
Crowther, R. A. (1972) in *The Molecular Replacement Method* (Rossman, M. G., Ed.) pp 173−178, Gordon and Breach, New York.
Fauman, E. B., Rutenber, E. E., Maley, G. F., Maley, F., & Stroud, R. M. (1994) *Biochemistry 33*, 1502−1511.
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.
Flaks, J. G., & Cohen, S. S. (1957) *Biochim. Biophys. Acta 25*, 667−668.
Greenberg, G. R., Somerville, R. L., & DeWolfe, S. (1962) *Proc. Natl. Acad. Sci. U.S.A. 48*, 242−247.
Hardy, L. W., Finer-Moore, J. S., Montfort, W. R., Jones, M. O., Santi, D. V., & Stroud, R. M. (1987) *Science 235*, 448−455.
Horton, N., & Lewis, M. (1992) *Protein Sci. 1*, 169−181.
Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 20*, 383−387.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157−171.
Kamb, A., Finer-Moore, J., Calvert, A. H., & Stroud, R. M. (1992a) *Biochemistry 31*, 9883−9890.
Kamb, A., Finer-Moore, J. S., & Stroud, R. M. (1992b) *Biochemistry 31*, 12876−12884.
Kisliuk, R. L., Gaumont, Y., Baugh, C. M., Galivan, J. H., Maley, G. F., & Maley, F. (1979) in *Developments in Biochemistry* (Kisliuk, R. L., & Brown, G. M., Eds.) Vol. 4, pp 431−435, Elsevier/North Holland, Amsterdam.
Knighton, D. 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. 24*, 946−950.
LaPat-Polaska, L., Maley, G., & Maley, F. (1990) *Biochemistry 29*, 9561−9572.
Maley, G. F., Bellisario, R. L., Guarino, D. U., & Maley, F. (1979a) *J. Biol. Chem. 254*, 1301−1304.
Maley, G. F., Maley, F., & Baugh, C. M. (1979b) *J. Biol. Chem. 254*, 7485−7487.
Maley, G. F., Maley, F., & Baugh, C. M. (1982) *Arch. Biochem. Biophys. 216*, 551−558.
Maley, F., Belfort, M., & Maley, G. (1984) *Adv. Enzyme Regul. 22*, 413−430.
Mathews, C. K., Moen, L. K., Wang, Y., & Sargent, R. G. (1988) *Trends Biochem. Sci. 13*, 394−397.
Matthews, D. A., Villafranca, J. E., Janson, C. A., Smith, W. W., Welsh, K., & Freer, S. (1990) *J. Mol. Biol. 214*, 937−948.
Meek, T. D., Garvey, E. P., & Santi, D. V. (1985) *Biochemistry 24*, 678−686.
Montfort, W. R., Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Maley, G. F., Hardy, L., Maley, F., & Stroud, R. M. (1990) *Biochemistry 29*, 6964−6977.
Pardee, A. B. (1989) *Science 246*, 603−608.
Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F., & Stroud, R. M. (1990) *Proteins: Struct., Funct., Genet. 8*, 315−333.
Pizer, L. I., & Cohen, S. S. (1962) *J. Biol. Chem. 237*, 1251−1259.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Rouch, D. A., Messerotti, L. J., Loo, L. S. L., Jackson, C. A., & Skurry, R. A. (1989) *Mol. Microbiol. 3*, 161−175.
Stroud, R. M. (1994) *Nature Struct. Biol. 1*, 131−134.
Stroud, R. M., & Finer-Moore, J. S. (1993) *FASEB J. 7*, 671−677.
Takeishi, K., Kaneda, S., Ayusawa, D., Shimizu, K., Gotoh, O., & Seno, T. (1985) *Nucleic Acids Res. 13*, 2035−2043.
Warner, H. R., & Snustad, D. P. (1983) in *Bacteriophage T4* (Mathews, C. K., Kutter, E. M., Mosig, G., & Berget, P. B., Eds.) pp 103−109, American Society for Microbiology, Washington D.C.
Wheeler, L., Wang, Y., & Mathews, C. K. (1992) *J. Biol. Chem. 267*, 7664−7670.
Young, J. P., & Mathews, C. K. (1992) *J. Biol. Chem. 267*, 10786−10790.