

Phylogenetic Classification of Protozoa Based on the Structure of the Linker Domain in the Bifunctional Enzyme, Dihydrofolate Reductase-Thymidylate Synthase*

Received for publication, September 17, 2003, and in revised form, October 6, 2003
Published, JBC Papers in Press, October 9, 2003, DOI 10.1074/jbc.M310328200

Robert H. O'Neil‡, Ryan H. Lilien§, Bruce R. Donald§, Robert M. Stroud¶, and Amy C. Anderson‡||

From the Departments of ‡Chemistry and §Computer Science, Dartmouth College, Hanover, New Hampshire 03755 and the ¶Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, California 94143

We have determined the crystal structure of dihydrofolate reductase-thymidylate synthase (DHFR-TS) from *Cryptosporidium hominis*, revealing a unique linker domain containing an 11-residue α -helix that has extensive interactions with the opposite DHFR-TS monomer of the homodimeric enzyme. Analysis of the structure of DHFR-TS from *C. hominis* and of previously solved structures of DHFR-TS from *Plasmodium falciparum* and *Leishmania major* reveals that the linker domain primarily controls the relative orientation of the DHFR and TS domains. Using the tertiary structure of the linker domains, we have been able to place a number of protozoa in two distinct and dissimilar structural families corresponding to two evolutionary families and provide the first structural evidence validating the use of DHFR-TS as a tool of phylogenetic classification. Furthermore, the structure of *C. hominis* DHFR-TS calls into question surface electrostatic channeling as the universal means of dihydrofolate transport between TS and DHFR in the bifunctional enzyme.

Thymidylate synthase (TS)¹ and dihydrofolate reductase (DHFR) are essential enzymes in the cell cycle of all organisms, since they catalyze the production of dTMP, required for DNA replication. TS converts the substrate, dUMP, to dTMP by reductive methylation using the cofactor, 5,10-methylene tetrahydrofolate, and releases dihydrofolate (1). In the presence of the cofactor NADPH, DHFR reduces dihydrofolate to tetrahydrofolate. The folate cycle is completed by serine hydroxymethyl transferase, which converts tetrahydrofolate back to 5,10-methylene tetrahydrofolate.

Recently, Stechmann and Cavalier-Smith (2) have addressed

* This work was supported by National Institutes of Health (NIH) Grant GM067542 (to A. C. A.), NIH CA63081 (to R. M. S.), and NIH Grant GM65982 (to B. R. D.). Computer equipment for developing the quaternion algorithm was supported by National Science Foundation Grants 98-02068 and IIS-9906790 (to B. R. D.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1QZF) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

|| To whom correspondence should be addressed: Dept. of Chemistry, 6128 Burke Laboratories, Dartmouth College, Hanover, NH 03755. Tel.: 603-646-1154; Fax: 603-646-3946; E-mail: Amy.C.Anderson@Dartmouth.edu.

¹ The abbreviations used are: TS, thymidylate synthase; DHFR, dihydrofolate reductase; ChDHFR-TS, DHFR-TS from *C. hominis*; PfDHFR-TS, DhFR-TS from *P. falciparum*; LmDHFR-TS, DHFR-TS from *L. major*.

the problem of locating the root of the eukaryotic tree, one of the most challenging evolutionary problems. In several protozoa, including Alveolates and Euglenozoa, and in some plants, the genes for DHFR and TS are translated as a single polypeptide, forming a bifunctional enzyme (DHFR-TS), whereas in most animals, fungi, and bacteria, these two enzymes are monofunctional. The monofunctional form of DHFR is a monomer, and that of TS is a dimer. The currently held hypothesis is that the primordial form of DHFR and TS is the monofunctional form and that the genes for DHFR and TS became fused at a single evolutionary point. If the DHFR-TS gene fusion occurred just once, then the fused gene provides an excellent phylogenetic marker, since reversing the fusion would require multiple genetic events. Stechmann and Cavalier-Smith have used the derived gene fusion between DHFR and TS to place the root of the tree below the common ancestor of plants, Alveolates, and Euglenozoa (Fig. 1 shows the overview of the tree; classifications are described in greater detail below).

There has been significant controversy concerning the classification and evolutionary progression of the protozoa (2–5). It is thought that many eukaryotic taxa arose during one explosive event into a cluster that has proved difficult to resolve (6). Sequences of rRNA and four other protein-encoding genes (5) have been used to place the protists into several groups. In this work, we investigated several protozoa with bifunctional DHFR-TS: Alveolata, including the Apicomplexans *Cryptosporidium*, *Toxoplasma*, and *Plasmodium* and Euglenozoa, including the kinetoplastids such as *Leishmania* and the trypanosomes.

To better understand whether bifunctional DHFR-TS is a good tool for the phylogenetic classification of some families of protozoa, we have solved the crystal structure of DHFR-TS from *Cryptosporidium hominis* (ChDHFR-TS), previously called *Cryptosporidium parvum* type 1 (7), to 2.8 Å and performed a structure-function analysis with previously solved structures of DHFR-TS from *Plasmodium falciparum* (PfDHFR-TS) (8) and *Leishmania major* (LmDHFR-TS) (9). The structures differ with respect to the docking of DHFR on TS, the length and interactions of the N terminus, and the structure of the linker domains between the DHFR and TS domains. Major differences exist between the Apicomplexan and the kinetoplastid DHFR-TS structures. The analysis reveals that there are dissimilar associations for DHFR and TS that correspond to two distinct families of protozoa. Furthermore, sequence analysis of additional DHFR-TS genes from various families of protozoa reveals that they, too, fall into these two families.

The structural differences we discovered between the linker domains of the Apicomplexan and kinetoplastid DHFR-TS fam-

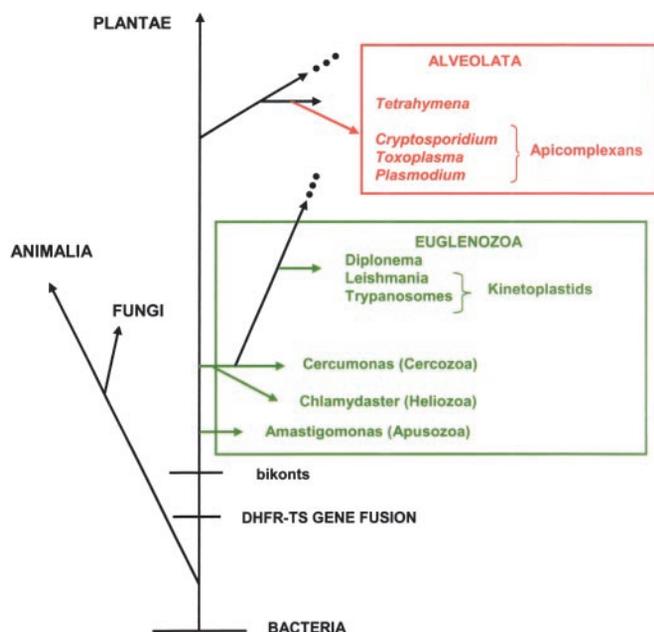


FIG. 1. A portion of the evolutionary tree (as described by Stechmann and Cavalier-Smith (2)). Protist species with available DHFR-TS sequence information are shown in color. Species in green belong to the short linker family, and species in red belong to the long linker family.

ilies raised questions concerning the conservation of the electrostatic channeling mechanism between the families. The structure of LmDHFR-TS, the first DHFR-TS structure to be solved, enabled a hypothesis to explain the channeling of dihydrofolate from TS to DHFR (9). Knighton *et al.* reported that a series of positively charged residues lined the surface of the protein, extending from the TS active site to the DHFR active site. These residues were postulated to potentially guide the negatively charged dihydrofolate from TS to DHFR in the sequential reaction. In fact, there is experimental evidence for channeling of dihydrofolate from TS to DHFR (10) in *L. major*, but it is not conclusive that this mechanism is based on electrostatics or extends to all species of DHFR-TS. Stroud (11) has previously postulated an alternative mechanism based on a rapid association of the DHFR and TS active sites by molecular dynamics. Our analysis of the electrostatic potential of ChDHFR-TS shows no pattern of positively charged residues between active sites that would indicate an electrostatic transfer of dihydrofolate.

EXPERIMENTAL PROCEDURES

Crystallization—Pure ChDHFR-TS was obtained from J. Vasquez and R. Nelson (7). The protein was purified on a methotrexate affinity column and eluted with 2 mM dihydrofolate. Fractions containing pure protein were concentrated to 7 mg/ml and incubated with 2 mM ligands (methotrexate, NADPH, CB3717, and dUMP) for approximately 1 h on ice. Using hanging drop vapor diffusion, a promising crystallization condition was refined to 10% polyethylene glycol 6000, 50 mM ammonium sulfate, 150 mM lithium sulfate, and 100 mM Tris, pH 8.0. Crystals grew in 2 weeks. The crystals were soaked in 15% ethylene glycol for 5 min and then in 25% ethylene glycol for another 5 min before being plunged into liquid nitrogen for cryogenic data collection.

Data Collection—Data were collected to 2.8-Å resolution at Brookhaven National Laboratory at beamline X12C on a B4-2k CCD detector. The crystals belong to space group C2 with unit cell edges $a = 214.9$, $b = 116.3$, $c = 219.7$ Å and $\beta = 95.23^\circ$. The data were indexed, integrated, and scaled with Denzo/Scalepack (12) and converted to structure factors with Truncate (13). A random set of reflections (10%) was set aside for the calculation of R_{free} .

Structure Determination—The structure was determined by molecular replacement using a model of thymidylate synthase from *P. carinii* (Protein Data Bank 1F28) (14), from which all ligands were removed, as

a search model (15). A self-rotation map showed a large peak indicating 5-fold noncrystallographic symmetry. Cross-rotation peaks related by 72° were selected by a quaternion algorithm² and subjected to a translation search. Two full dimers of TS were placed in the asymmetric unit, and a third dimer of TS was located across the 2-fold symmetry axis, yielding five monomers per asymmetric unit. The molecules are conjugated in an endless helix with a noncrystallographic 5_1 axis that runs throughout the entire crystal. Five models of DHFR from *P. carinii* (Protein Data Bank 1CD2) (16), without ligands, were placed using the fixed positions of the TS models as initial phase estimates. The starting R-factor was 52%.

Residues for ChDHFR-TS were substituted in the models of *P. carinii* DHFR and TS and refined using noncrystallographic restraints between all monomers of DHFR and TS. Electron density for the ligands and the linker region was visible in the initial maps. The ligands were positioned, the linker region was built, and the entire complex was refined to an R-factor of 24.1% and $R_{free} = 25.8\%$ using CNS. Water molecules were added where adequate hydrogen bonding to a donor or acceptor on the protein was visible and until there was no remaining difference density. The final R-factor was 22.5%, and R_{free} was 24.5%. Data and refinement statistics are found in Table I.

The electron density map for PfDHFR-TS was calculated using CNS and the structure factors for PfDHFR-TS (Protein Data Bank 1J3I) deposited in the Protein Data Bank. The deposited coordinates were used to calculate phase angles.

RESULTS AND DISCUSSION

Description of the Overall Structure—The structure of ChDHFR-TS illustrates that the enzyme is a homodimer based on the canonical dimer interface of TS (Fig. 2). Based on structural comparisons with other DHFR and TS proteins, the DHFR domain is 178 residues, the linker domain is 58 residues, and the TS domain is 284 residues, yielding a 63-kDa monomer. All residues from 3 to 521 are clearly defined in the electron density, allowing the entire protein model to be visualized. The bulk of the DHFR domains do not contact each other, but the linker polypeptide between the DHFR and TS domains crosses from one DHFR monomer (A) to the other DHFR monomer (B), noncovalently connecting the two DHFR domains, and then returns to form the TS monomer of A (see discussion below). In the view shown in Fig. 1, the DHFR active site of the left monomer faces the viewer, and the TS active site is on the side of the enzyme. The DHFR and TS active sites of the same monomer are ~ 74 Å apart along the surface of the enzyme and 45 Å apart if measured directly through the enzyme from the diaminopyrimidine ring of dihydrofolate to the quinazoline ring of CB3717. The molecules in the unit cell are arranged with a noncrystallographic quasi- 5_1 screw axis, forming lines of molecules extending throughout the crystal.

Structure of the *C. hominis* DHFR Domain—The structure of the DHFR domain of ChDHFR-TS generally resembles that of several other DHFR proteins from eukaryotic organisms (16–18). However, DHFR from most other organisms, including *L. major*, has an eight-stranded β -sheet, and *C. hominis* DHFR has a nine-stranded β -sheet, where the last four residues of the linker polypeptide form the ninth strand of the sheet.

The *C. hominis* DHFR active site reveals the conservation of catalytically important residues (17–20). Electron density for the ligands, dihydrofolate and NADPH, was visible in the initial maps (Fig. 3a). In the ChDHFR-TS structure, dihydrofolate is bound in the active site in the same orientation as dihydrofolate in the human DHFR active site (21). NADPH is bound using similar interactions as noted in other structures of eukaryotic species of DHFR (17, 18).

In ChDHFR-TS, the DHFR interface with the TS domain of the same monomer buries a combined surface area of ~ 1300 Å². The interface includes not only residues in the canonical TS

² R. Lilien, C. Bailey-Kellogg, A. Anderson, and B. Donald, manuscript in preparation.

TABLE I
Crystallographic statistics

Parameter	Value
Space group	C2
Unit cell edges (Å)	214.9 × 116.3 × 219.7 × 90 × 95.23 × 90
Resolution (Å)	2.8
No. of reflections used	111,000
Completeness (%)	87.9
R_{merge}	10.5
R_{cryst} (%)	22.5
R_{free} (%)	24.5
Total no. of atoms	22,242
No. of water molecules	397
Root mean square deviation bonds (Å)	0.009
Root mean square deviation angles (degrees)	1.604
Average B-factor (Å ²)	33.0

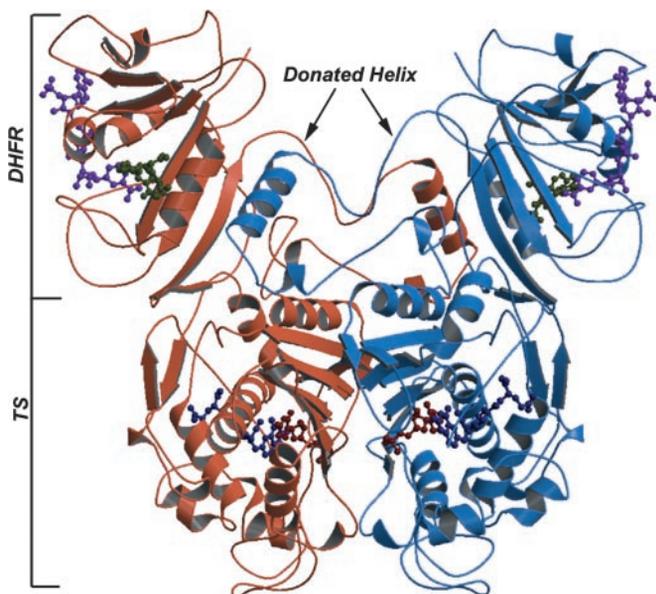


FIG. 2. DHFR-TS from *C. hominis*. Chain A is colored red, and chain B is colored blue. The ligands, NADPH (purple) and dihydrofolate (green) in the DHFR active site and dUMP (red) and CB3717 (blue) in the TS active site, are shown in ball-and-stick models. The helix in the linker domain is noted.

and DHFR domains but also includes several residues that belong to the linker between the domains. In general, the TS residues in the interface are widely conserved across most species; the DHFR residues are not generally conserved. The interface between *C. hominis* DHFR and *C. hominis* TS is primarily hydrophobic and includes a number of van der Waals' interactions between hydrophobic side chains. There are two hydrogen bonds (Glu¹³⁹-Tyr⁵¹⁰ and Ser¹⁶⁹-Gln⁴⁸⁶) between residues belonging to the canonical DHFR and TS domains. There are an additional three hydrogen bonds (Arg²¹⁰-Glu²⁷⁶, Arg²¹⁵-Arg²⁷⁵, and Arg²³³-Asp²⁴²) stabilizing the interface between residues in the linker and residues in TS. Overall, there are relatively few interactions between the residues in the canonical DHFR and TS domains and the interactions that are present involve weak van der Waals' forces.

Structure of the *C. hominis* TS Domain—The overall structure of the TS domain of ChDHFR-TS strongly resembles that of human TS (22, 23) and TS from other eukaryotic sources (24, 25). The substrate, dUMP, and the cofactor mimic, CB3717, are bound to *C. hominis* TS in both active sites. The nucleophilic cysteine residue, Cys⁴⁰², is covalently attached to the C-6 position of dUMP. CB3717 is bound in a similar manner as seen in other structures of monofunctional TS (24, 26) and Lm-DHFR-TS (9).

Structure of the ChDHFR-TS Linker Domain—Several

DHFR-TS proteins, notably from Apicomplexan protozoa, include a long linker between the DHFR and TS domains. The linker polypeptide, even within the Apicomplexans, varies significantly in length; the entire linker in *C. hominis* is 58 amino acids, the linker in *Toxoplasma gondii* is 72 amino acids, and the linker in *P. falciparum* is 89 amino acids.

The linker in ChDHFR-TS plays an important role in the structure of the enzyme. Compared with the structure of DHFR-TS from PfDHFR-TS (8) in which major portions of the linker domain are not built into the model, the entire linker domain of ChDHFR-TS is clearly visible in the electron density (Fig. 3b). The chain leaves DHFR of monomer A at residue 178, crosses to DHFR of monomer B, forms an 11-residue helix, termed the donated helix, that packs against the DHFR active site residues of monomer B, crosses back to DHFR of monomer A to form the last strand of the 9-stranded β -sheet of the DHFR domain, and then finally joins the TS domain of monomer A. The crossover is repeated for both dimers, resulting in four chains crossing the cleft between the DHFR domains: one leaving the A monomer and going to the donated helix (tether 1), one returning to A after forming the donated helix (tether 2), and the 2-fold related chains leaving and returning to B. There is one hydrogen bond between the linker polypeptides of the A and B monomers (Gln^{192A} with Gln^{192B}). There are several additional hydrogen bonds between tethers 1 and 2 within one monomer: Arg^{230A} with Gln^{184A}, Asp^{201A} with Arg^{190A}, and Asn^{185A} with Asn^{187A}. The donated helix (residues 196–207) packs tightly against the opposite monomer on the back side of its DHFR active site using mainly hydrophobic interactions (Table II). Clearly, all of the interactions of the donated helix are important in domain stabilization.

The donated helix may be implicated in the mechanism of the bifunctional enzyme. Genetic analysis, kinetic measurements, and molecular dynamics simulations identify a network of coupled motions that occur during DHFR catalysis (27). Molecular dynamics simulations suggest that Phe³¹ (*Escherichia coli* numbering) approaches the bound dihydrofolate, directing the conformational changes that bring the substrate to the transition state (27, 28). Phe³¹ is strictly conserved, and mutations of this residue strongly decrease the rate of hydride transfer (29).

The donated helix in ChDHFR-TS contacts residues in *C. hominis* DHFR that are equivalent to those identified as being involved in the coupled motions of catalysis (see Fig. 4). Phe³⁶ in ChDHFR-TS is equivalent to Phe³¹ in *E. coli*, and the helix containing Phe³⁶ and Phe³⁵ in the ChDHFR-TS active site packs against the donated helix from the linker of the opposite monomer. The closely fitting interaction between the donated helix and the catalytic residues implies that the donated helix may be affected by the conformational changes occurring during catalysis and may represent a signaling mechanism be-

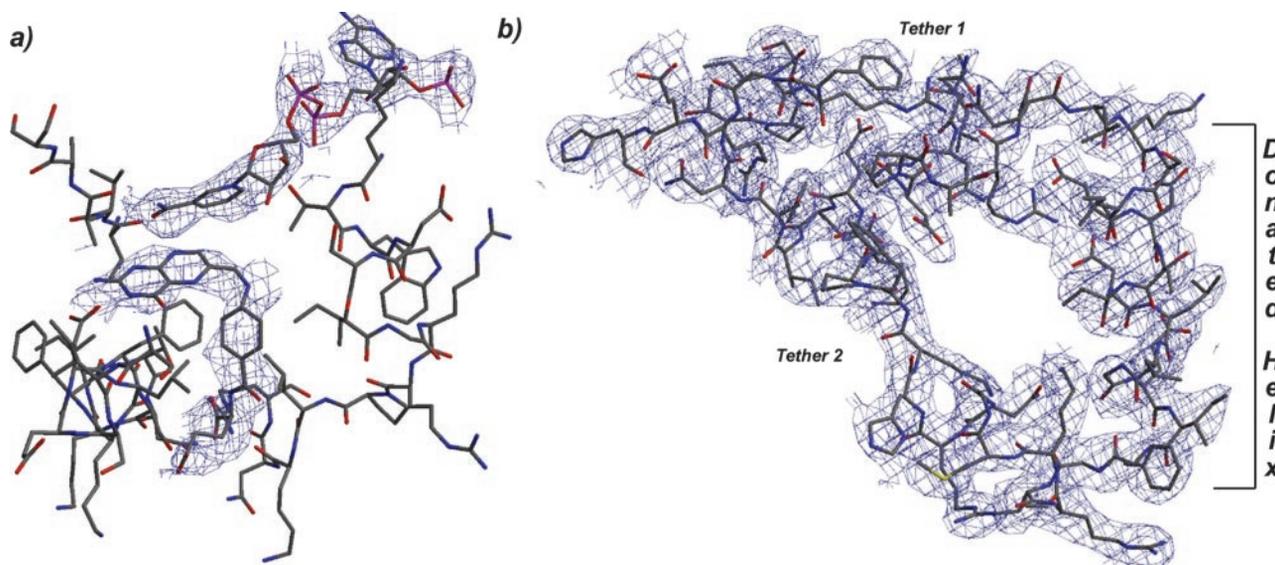


FIG. 3. **Initial and final electron density maps overlaid on the final model.** *a*, initial density ($2F_o - F_c$, contoured at 1σ) for dihydrofolate and NADPH in the DHFR active site. The positioned molecular replacement search models, which did not include ligands, were used to calculate phases for the initial map. *b*, final density map ($2F_o - F_c$, 1σ) for the linker region of monomer A. The donated helix and tethers are noted.

TABLE II
Interactions between the donated helix and the opposite monomer

Residue in donated helix	Residue in opposite monomer
Hydrogen bonds	
Ser ¹⁹⁵	Glu ¹⁷⁶
Asp ¹⁹⁸	Asn ⁴²
Arg ²¹⁰	Glu ³¹
Arg ²¹⁰	Cys ¹⁶⁴ (backbone oxygen)
Arg ²¹⁰	Glu ²⁷⁶
Van der Waals' interactions	
Ile ¹⁹⁶	Val ¹⁵⁷ , Leu ¹⁹³
Val ²⁰⁰	Tyr ¹⁵⁹
Leu ²⁰²	Lys ³⁸ , Ile ³⁹ , Tyr ¹³²
Leu ²⁰³	Phe ³⁵ , Phe ¹⁶³ , Phe ¹⁷²
Ile ²⁰⁶	Phe ³⁵
Phe ²⁰⁷	Phe ¹⁶³ , Phe ¹⁷²

tween the two monomers of the enzyme. The tethers crossing back to the originating monomer could possibly carry a signal revealing the catalytic status of the opposite monomer.

The human monofunctional DHFR positions several residues at the same location as the donated helix. Residues Tyr¹⁶², Pro¹⁶³, Gly¹⁶⁴, and Val¹⁶⁵ of human DHFR form a loop that extends away from the body of DHFR and makes very similar interactions with the back side of the active site. However, if the donated helix is crucial to a mechanism in the bifunctional enzymes, it must be significantly different from any mechanism in the monofunctional enzymes if, in fact, the monofunctional enzymes associate in the cell.

Comparison of All Bifunctional DHFR-TS Structures—The docking surface of *C. hominis* DHFR and *C. hominis* TS consists of mainly hydrophobic residues and does not contribute to the docking of DHFR on TS in a lock-and-key fashion. The docking area for DHFR on TS appears to provide only an attractive and not an orienting force. The fact that the contacts between DHFR and TS are nonspecific was previously recognized in an analysis of the LmDHFR-TS structure by Knighton *et al.* (9). In fact, the three structures of bifunctional DHFR-TS show different docking surfaces for DHFR on TS despite the conservation of TS residues in that area, giving further evidence that the TS surface provides no common orienting forces. A structural alignment of ChDHFR-TS, PfdDHFR-TS, and LmDHFR-TS reveals that the linker polypeptides 58, 93, and 2 residues, respectively, play a crucial role in the orientation of

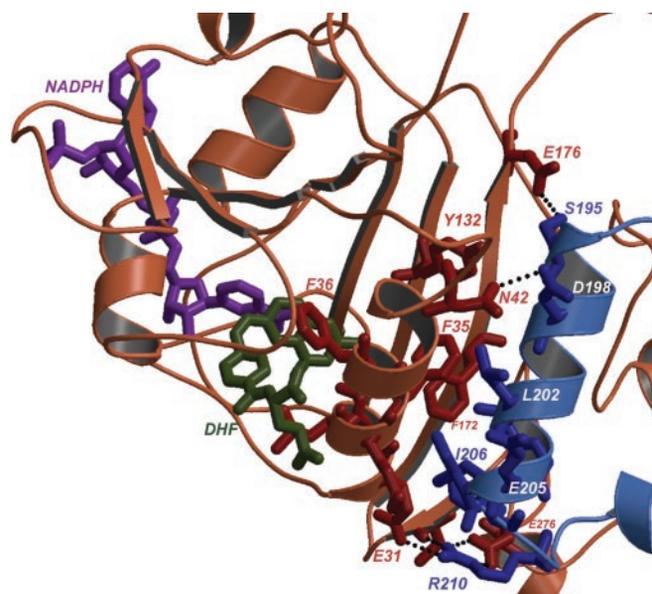


FIG. 4. **The donated helix (blue) interacts with the residues of the DHFR active site of the opposite monomer (red).** NADPH is shown in purple, and dihydrofolate (DHF) is shown in green. Hydrogen bonds are shown as dashed lines. Key residues in the catalytic mechanism of DHFR (Phe³⁶) and in the interactions of the donated helix with the opposite monomer are noted in one-letter codes.

the DHFR domain relative to TS. In the structures of all three enzymes, there are examples of dimers where the monomers are not related by crystallographic symmetry, providing evidence that the orientations of the monomers and the linker domains are determined within the molecule rather than by crystal packing forces. The short linker in LmDHFR-TS creates a taut tether, restricting the range of DHFR orientations relative to TS. In ChDHFR-TS, the linker creates an orienting force, composed of van der Waals' interactions and hydrogen bonds, that positions the DHFR monomers in relation to the TS domain (Fig. 5). In PfdDHFR-TS, the donated helix is strongly negatively charged (Glu²⁸⁵, Asp²⁸⁴, and Asp²⁸⁸), complementing the positively charged groove on DHFR into which it fits (Lys⁵⁶, Lys¹⁸⁷, Lys¹⁸⁰, Lys¹⁶⁰, Lys²²⁷, and Lys²³²) (Fig. 5). The electrostatic interactions of the donated helix in PfdDHFR-TS are the primary determinants of the PfdDHFR-TS orientation

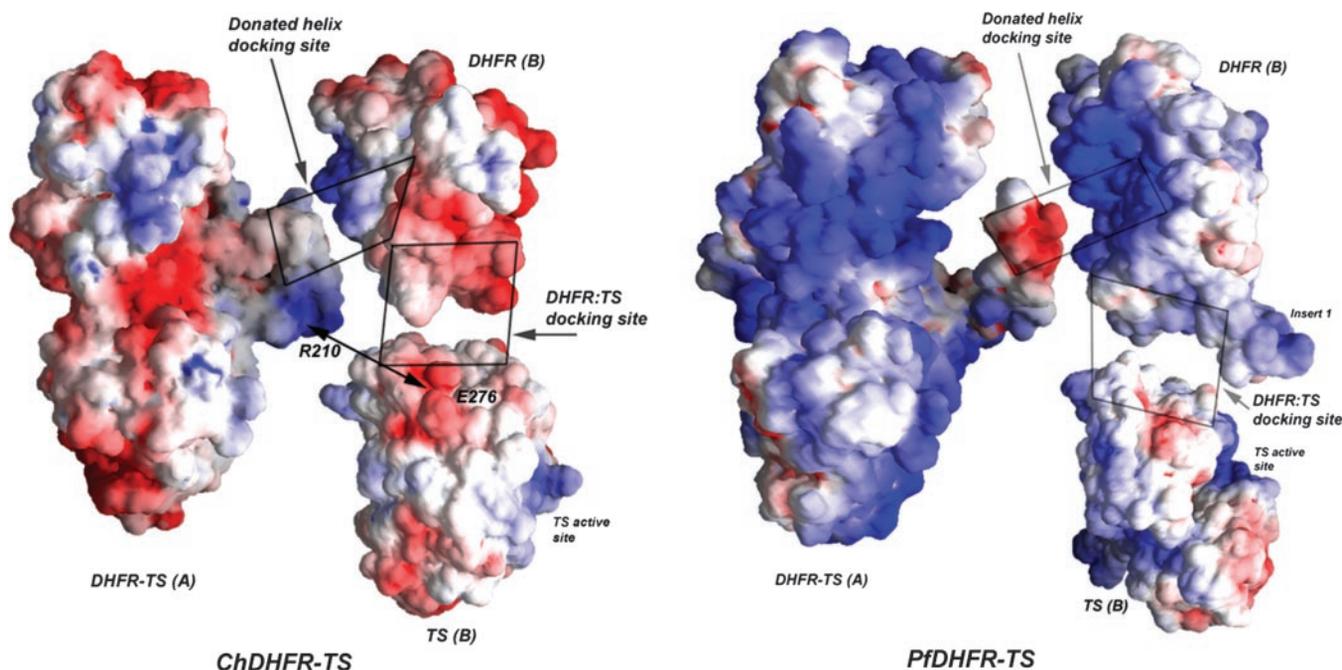


FIG. 5. **Electrostatic potential surfaces for ChDHFR-TS and PfDHFR-TS.** All surfaces are contoured from -5 kiloteslas (red) to $+5$ kiloteslas (blue). The A monomer is kept as a single surface, and the B monomer is divided between the DHFR domain and the TS domain (the linker is omitted for clarity in monomer B). The interface area between DHFR (B) and TS (B) is opened slightly to show details of the docking region. In ChDHFR-TS, Arg^{210A} and the area surrounding Glu^{276B} (including Glu^{311B} and the backbone oxygen of Cys^{164B}; these residues not shown for clarity) form hydrogen bonds. *Inset 1* in PfDHFR-TS is defined as residues 20–36.

and are supplemented by the nonspecific van der Waals' interactions of the *P. falciparum* DHFR to *P. falciparum* TS interface.

A comparison of ChDHFR-TS and PfDHFR-TS reveals that the two linker domains are structurally similar despite the fact that *C. hominis* DHFR and *P. falciparum* DHFR do not contact TS in the same orientation. When the TS domains of ChDHFR-TS and PfDHFR-TS are aligned, *P. falciparum* DHFR is rotated 22.6° relative to *C. hominis* DHFR. The active site of DHFR, specifically the phosphate moiety near the adenine ring of NADPH, forms the pivot point of this rotation. The placement of the donated helix, at the back side of the opposite DHFR active site, is the same in ChDHFR-TS and PfDHFR-TS. If the orientation differences are taken into account and the structure of ChDHFR-TS is overlaid on PfDHFR-TS and if the electron density map for PfDHFR-TS is visualized using the structure factors deposited with the coordinates in the Protein Data Bank, then electron density corresponding to tether 1 of PfDHFR-TS is evident in the PfDHFR-TS electron density map, although its connectivity is poor. Overall, the location of the donated helix and the observation of electron density corresponding to tether 1 leads to the conclusion that the linker in PfDHFR-TS may also cross from one DHFR domain to the opposite and then cross back, similarly to ChDHFR-TS.

Surprisingly, the distances between all active sites, measured directly through the enzyme, are conserved within 1 \AA in both ChDHFR-TS and PfDHFR-TS (Table III). The impact of this conservation of active site distances is potentially important. If we hypothesize that a third conserved enzyme, perhaps serine hydroxymethyl transferase, is interacting in a complex composed of all of the folate cycle enzymes in the cell, the overall geometry of the interaction would be conserved across a wide variety of Apicomplexan species of the bifunctional enzyme.

The orientation of *L. major* DHFR relative to *L. major* TS is drastically different from the orientation of DHFR to TS in ChDHFR-TS or PfDHFR-TS. LmDHFR-TS has a two-residue

TABLE III
Distances between active sites in DHFR-TS (\AA)
Distances between the phosphate near the adenine of NADPH and the phosphate of dUMP were measured.

	DHFR(A)-DHFR(B)	DHFR(A)-TS(A)	DHFR(A)-TS(B)
ChDHFR-TS	86.5	66.7	69.8
PfDHFR-TS	87.4	67.1	69.8

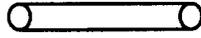
linker between the DHFR and TS domains. *L. major* DHFR is rotated upside down relative to *C. hominis* DHFR, causing the body of *L. major* DHFR to rest on the shoulder of the *L. major* TS domain. The DHFR and TS active sites are exposed on the same exterior side of the molecule in LmDHFR-TS. In contrast, in ChDHFR-TS, the DHFR and TS active sites are on orthogonal faces of the enzyme.

In addition to the different DHFR and TS orientations, ChDHFR-TS and LmDHFR-TS differ in the placement and function of the N-terminal amino acids of DHFR. ChDHFR-TS and PfDHFR-TS do not have an N-terminal extension, relative to the canonical fold of DHFR, which interacts with the TS domain. In LmDHFR-TS, the 22-residue N-terminal extension on DHFR fits into a groove on the exterior of TS and stabilizes the interaction between *L. major* DHFR and *L. major* TS. In contrast, the N terminus of ChDHFR-TS and PfDHFR-TS points upward away from TS and remains close to DHFR.

Sequence Analysis of DHFR-TS within Multiple Protist Families—Using the three structures analyzed above, we created a functional analysis of all available complete sequences in the Euglenozoa and Alveolata families. This functional analysis aligns the residues corresponding to the canonical folds for DHFR and TS (Fig. 6). The beginning and the end of the DHFR and TS domains were determined by structural alignment for ChDHFR-TS, PfDHFR-TS, and LmDHFR-TS. The beginning and the end of the DHFR and TS domains of *T. gondii* (Fig. 6, *TgDHFR-TS*), *Trypanosoma cruzi*, and *Trypanosoma brucei* DHFR-TS were determined by sequence alignments, since the

LONG LINKER FAMILY

Pf	MMEQVCDV	9	231	KMLNEQNCIKGEEKNDMPLKNDKDTCHMKKLTIFYKNV
Ch	MSE	4	DHFR 178	KKTLQNCDPARGQLKS-----
Tg	MQK	4	322	KTDDAATAEPSNAMSSLTSTRETTPVHGLQAPSS-----



Pf	DKYKINYENDDDEEEDDFVYFNFNKEKEEKNKNSIHPNDFQIYNSLKYKY-	323	608
Ch	-----IDDTVDLLGEIFGIRKMGNRHKFPKKEIYNTPSIRFGRE	236	TS 521
Tg	-----AAAIAPVLAWMDEEDRKKREQKELIRAVPHVHFRG---	323	610

SHORT LINKER FAMILY

Lm	MSRAAARFKIIMPETKADFAPP	23	239	RN	242	520
Tc	MSLFKIRMPETVAEGTRL----	19	DHFR 232	RN	235	TS 521
Tb	MLSLTRILRKKIPVHELAKGIS	23	238	RN	241	527

FIG. 6. Functional alignment of several protozoal DHFR-TS sequences. The canonical DHFR and TS folds are enclosed in boxes with the start and end residues noted. The species are *C. hominis* (Ch), *P. falciparum* (Pf), *T. gondii* (Tg), *L. major* (Lm), *T. cruzi* (Tc), and *T. brucei* (Tb). The donated helix is noted with a cylinder. Only fully sequenced genes are shown.

structures of these proteins have not yet been determined. DHFR-TS from Euglenozoa such as *Leishmania*, *T. cruzi*, and *T. brucei* all have an N-terminal extension relative to the start of the canonical DHFR fold and short, two-residue (Arg-Asn) linkers. The Apicomplexans *Cryptosporidium* and *Toxoplasma* do not have N-terminal extensions and generally have longer linkers. Secondary structure prediction suggests a helix within the linker of *T. gondii* DHFR-TS at positions 288–297. This helix may be, in fact, longer than nine residues, but has been predicted conservatively. *Plasmodium falciparum*, an Apicomplexan, has a long linker and a partial (eight-residue) N-terminal extension, although this extension points away from TS and does not form an interaction with TS (8) similar to that seen in LmDHFR-TS. The functional analysis leads to the conclusion that there are two families of DHFR-TS structures: a long linker family that resembles ChDHFR-TS and PfDHFR-TS and includes a donated helix (Fig. 7, a and b) and a short linker family that resembles LmDHFR-TS and uses the N-terminal extension that appears to stabilize the DHFR-TS interaction (Fig. 7, c and d).

Several partial sequences of DHFR-TS from other protist families, three of which are distantly related to the Alveolates or Euglenozoa, were analyzed and found to fall into one of the two structural families. The sequence of DHFR-TS from *Tetrahymena pyriformis* (all partial sequences, discussed in Ref. 2, include a portion of the DHFR domain, the entire linker domain, and a portion of the TS domain), a member of the Alveolates (see Fig. 1), but not an Apicomplexan, also has a long, 58-residue, linker domain between DHFR and TS with helical content corresponding to the same region as ChDHFR-TS. The sequence of the DHFR-TS gene from *Diplonema ambulator* (2), another protozoa in the Euglenozoa family, but not a kinetoplastid like *Leishmania* or the trypanosomes, confirms that it has a short, two-residue linker (Val-Asn) between the DHFR and TS domains. Sequences of DHFR-TS from *Chlamydomonas stermi* (from the Heliozoa phylum), *Cercomonas longicauda* (from the Cercozoa phylum), and *Amastigomonas debruynei* (from the Apusozoa phylum) also have short (2–6-residue) linker domains. Apusozoa is on the first branch after the DHFR-TS fusion event. Heliozoa and Cercozoa are on the second branch, the same branch off which the Euglenozoa originate, but are only distantly related to the Euglenozoa. Based on the length of these various linker domains, we hypothesize that

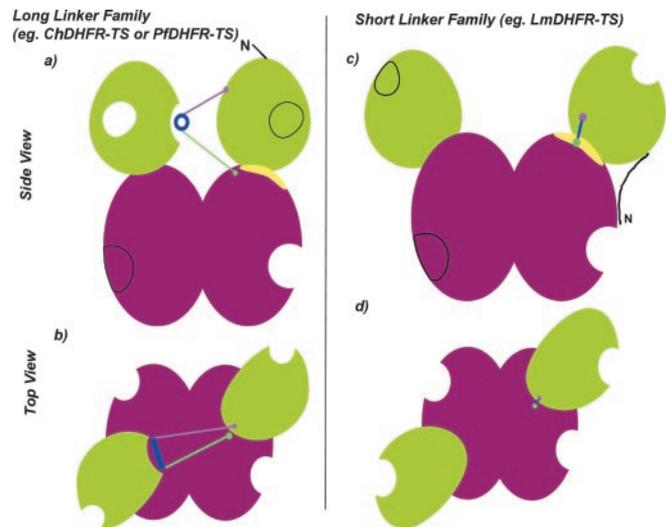


FIG. 7. Schematic diagram of the long linker and short linker families of DHFR-TS structures. TS is shown in maroon, DHFR is shown in green, and the N terminus is noted with the letter N. Active sites are shown as cavities. The tether 1 attachment point is shown in purple, and the tether 2 attachment point is shown in bright green. The entire docking site between TS and DHFR for all species is shown in yellow. Features of the linker are shown only in one monomer for clarity. a, side view of the long linker family. The donated helix is shown in blue, tether 1 is purple, and tether 2 is bright green. The active site of the right side DHFR monomer is on the back side, with its location noted with an outlined circle. b, top view of the long linker family. c, side view of the short linker family. d, top view of the short linker family.

the structure of *Tetrahymena* more closely resembles the Apicomplexan DHFR-TS and that the structures of DHFR-TS from Apusozoa, Heliozoa, and Cercozoa more closely resemble the structure of LmDHFR-TS.

The long linker family includes the Apicomplexans and may extend to the Alveolates (shown in red in Fig. 1); the short linker family includes the kinetoplastids and may extend to the Euglenozoa, Cercozoa, Heliozoa, and Apusozoa (shown in green in Fig. 1). The two classes appear to be evolutionarily distinct and appear to have evolved separately after the DHFR-TS gene fusion event. The two families of DHFR-TS structures confirm the separate branches shown in the evolutionary tree derived by Stechmann and Cavalier-Smith (2). Since Apusozoa repre-

sent the first branch after the DHFR-TS fusion event and appear to belong to the short linker family, the short linker family may represent the primitive condition for DHFR-TS.

Electrostatic Channeling—Electrostatic channeling, using several positively charged residues aligned between the TS and DHFR active sites, has been postulated to guide dihydrofolate from the TS active site to the DHFR active site in the LmDHFR-TS and PfDHFR-TS enzymes. In *L. major* TS, three basic side chains (Lys²⁸², Arg²⁸³, and Arg²⁸⁷) are proposed to bind the polyglutamylated tail of the folate. These basic residues are adjacent to a highly basic insertion in *L. major* DHFR (Lys⁶⁶, Lys⁶⁷, Lys⁷², and Lys⁷³), forming a putative electrostatic “highway” (11) in LmDHFR-TS. In PfDHFR-TS, there are two positively charged grooves lined with residues conserved between plasmodial DHFR-TS enzymes and hypothesized to be involved in electrostatic channeling (8). Biochemical experiments show that the rates of substrate transfer in *T. gondii* DHFR-TS are altered in the presence of varying amounts of salt, lending evidence to the electrostatic channeling hypothesis (30). Rapid chemical quench assays (10) verify that, for bifunctional LmDHFR-TS enzymes, there is no lag in tetrahydrofolate production or accumulation of dihydrofolate, unlike the case for monofunctional enzymes. Brownian dynamics simulation experiments determined that almost all substrate molecules with charge -2 leaving the TS active site reached the DHFR active site in LmDHFR-TS (31, 32). However, in recent experiments (33), charge reversal or charge neutralization mutants in the putative channeling region of LmDHFR-TS were shown not to interfere with the kinetic channeling of substrate.

There is currently no evidence supporting or denying substrate channeling in ChDHFR-TS. However, the structure of ChDHFR-TS does not show a pattern of positively charged residues like the one seen in LmDHFR-TS, giving no structural support to the theory of electrostatic channeling in ChDHFR-TS. Three of the basic residues implicated in the electrostatic highway, Lys²⁸², Arg²⁸³, and Arg²⁸⁷ in *L. major* TS, are conserved and located near the *C. hominis* TS active site (Lys²⁸⁴, Lys²⁸⁵, and Arg²⁸⁹ in ChDHFR-TS). One of the four residues in the basic insertion found in *L. major* DHFR is conserved in *C. hominis* DHFR (Lys⁴⁸), but its location, 46 Å from the positively charged residues near the TS active site, is drastically different from that of the equivalent residue in *L. major* DHFR, due to the significantly different docking of DHFR on TS in ChDHFR-TS relative to that in LmDHFR-TS. An electrostatic potential map of ChDHFR-TS does not reveal any other pattern of positively charged residues that would create an alternate electrostatic highway if, in fact, ChDHFR-TS does channel substrate (Fig. 8). Therefore, a pattern of positively charged residues between the TS and DHFR active sites is not a common trait between *all* of the bifunctional DHFR-TS enzymes.

A previously proposed, alternate version of substrate channeling proposes that dynamical motion can bring the two active sites close together (11). Whereas the ChDHFR-TS structure cannot rule out the possibility that the two active sites approach each other, it does seem very unlikely, since the DHFR and TS active sites are on orthogonal faces of the enzyme and are cross-tied by the linker between the domains. In fact, the opening of the DHFR active site points toward the 2-fold axis between the two monomers and the opening of the TS active site points in the opposite direction, away from the 2-fold axis. The enzyme would need to undergo a major rearrangement in order to bring the two sites into proximity.

Conclusion—The crystal structure of DHFR-TS from *C. hominis* reveals a long linker and a new interaction, the donated

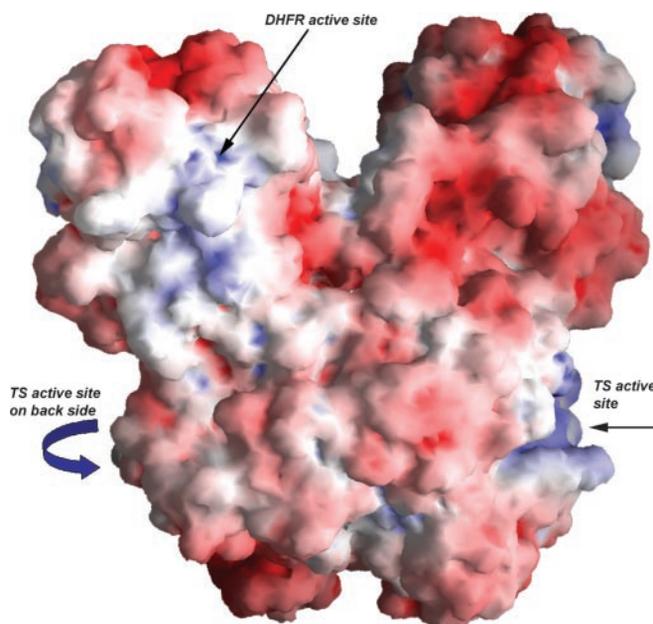


FIG. 8. An electrostatic potential map for ChDHFR-TS. A gradient from -10 /kiloteslas (red) to 10 /kiloteslas (blue) reveals no pattern of positively charged residues between the TS and DHFR active sites. The TS and DHFR active sites are noted; the second DHFR active site is not noted but is on the back side of the enzyme.

helix, in the bifunctional DHFR-TS protein. The structures of the Apicomplexan DHFR-TS proteins, ChDHFR-TS and PfDHFR-TS, show the donated helix, and the structure of the kinetoplastid LmDHFR-TS does not. A functional alignment, partitioning several sequences of DHFR-TS proteins into DHFR, linker, and TS domains, suggests that there are two structural families of DHFR-TS proteins that evolved separately after the gene fusion event. The observation of these two families coincides with other genetic evidence that separates the Alveolates, of which the Apicomplexans are a member, and the Euglenozoa, of which *Leishmania* is a member. Furthermore, if the donated helix, at a conserved location with respect to TS and the opposite DHFR active site, confers an important structural or functional role to the Apicomplexans, it would confirm the advanced evolutionary placement of the Alveolates over the Euglenozoa in the phylogenetic tree (2) (see Fig. 1). Finally, the structure of ChDHFR-TS does not show a pattern of positively charged residues between the DHFR and TS active sites, suggesting that it belongs to a different class of enzymes that may not use electrostatic channeling in the transfer of dihydrofolate from the TS active site to the DHFR active site.

Acknowledgments—We thank John Vasquez and Rick Nelson of San Francisco General Hospital for providing pure ChDHFR-TS as well as the staff at beamline X12C at Brookhaven National Laboratory for providing synchrotron access. Chris Bailey-Kellogg contributed to the development of algorithms to determine relevant peaks in the cross-rotation function.

REFERENCES

- Carreras, C., and Santi, D. (1995) *Annu. Rev. Biochem.* **64**, 721–762
- Stechmann, A., and Cavalier-Smith, T. (2002) *Science* **297**, 89–91
- Van de Peer, Y., and Wachter, R. (1997) *J. Mol. Evol.* **45**, 619–630
- Sogin, M. (1997) *Curr. Opin. Genet. Dev.* **7**, 792–799
- Baldauf, S., Roger, A., Wenk-Seifert, I., and Doolittle, W. (2000) *Science* **290**, 972–977
- Knoll, A. (1992) *Science* **256**, 622–627
- Vasquez, J., Gooze, L., Kim, K., Gut, J., Petersen, C., and Nelson, R. (1996) *Mol. Biochem. Parasitol.* **79**, 153–165
- Yuvaniyama, J., Chitnumsub, P., Kamechongwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M., and Yuthavong, Y. (2003) *Nat. Struct. Biol.* **10**, 357–365
- Knighton, D., Kan, C., Howland, E., Janson, C., Hostomska, Z., Welsh, K., and Matthews, D. (1994) *Nat. Struct. Biol.* **1**, 186–194
- Liang, P., and Anderson, K. (1998) *Biochemistry* **37**, 12195–12205

11. Stroud, R. (1994) *Nat. Struct. Biol.* **1**, 131–134
12. Otwinowski, Z. (1993) Denzo/Scalepack, SERC Daresbury Laboratory, Warrington, UK
13. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D* **50**, 760–763
14. Anderson, A., O'Neil, R., Surti, T., and Stroud, R. (2001) *Chem. Biol.* **8**, 445–457
15. Brunger, A. (1998) *Acta Crystallogr. D* **54**, 905–921
16. Cody, V., Galitsky, N., Rak, D., Luft, J., Pangborn, W., and Queener, S. (1999) *Biochemistry* **38**, 4303–4312
17. Cody, V., Luft, J., Ciszak, E., Kalman, T., and Freisheim, J. (1992) *Anti-Cancer Drug Des.* **7**, 483–491
18. Champness, J., Achari, A., Ballantine, S., Bryant, P., Delves, C., and Stammers, D. (1994) *Structure* **2**, 915–924
19. Li, R., Sirawaraporn, R., Chitnumsub, P., Sirawaraporn, W., Wooden, J., Athappilly, F., Turley, S., and Hol, W. (2000) *J. Mol. Biol.* **295**, 307–323
20. Bolin, J., Filman, D., Matthews, D., Hamlin, R., and Kraut, J. (1982) *J. Biol. Chem.* **257**, 13650–13662
21. Oefner, C., D'Arcy, A., and Winkler, F. (1988) *Eur. J. Biochem.* **174**, 377–385
22. Sayre, P., Finer-Moore, J., Fritz, T., Biermann, D., Gates, S., MacKellar, W., Patel, V., and Stroud, R. (2001) *J. Mol. Biol.* **313**, 813–829
23. Almgren, R., Waddling, C., Maley, F., Maley, G., and Van Roey, P. (2001) *Protein Sci.* **10**, 988–996
24. Anderson, A., Perry, K., Freymann, D., and Stroud, R. (2000) *J. Mol. Biol.* **297**, 645–657
25. Sotelo-Mundo, R., Siesla, J., Dzik, J., Rode, W., Maley, F., Maley, G., Hardy, L., and Montfort, W. (1999) *Biochemistry* **38**, 1087–1094
26. Montfort, W., Perry, K., Fauman, E., Finer-Moore, J., Maley, G., Hardy, L., Maley, F., and Stroud, R. (1990) *Biochemistry* **29**, 6964–6977
27. Agarwal, P., Billeter, S., Rajagopalan, P., Benkovic, S., and Hammes-Schiffer, S. (2002) *Proc. Natl. Acad. Sci.* **99**, 2794–2799
28. Watney, J., Agarwal, P., and Hammes-Schiffer, S. (2003) *J. Am. Chem. Soc.* **125**, 3745–3750
29. Chen, J.-T., Taira, K., Tu, C., and Benkovic, S. (1987) *Biochemistry* **26**, 4093–4100
30. Trujillo, M., Donald, R., Roos, D., Greene, P., and Santi, D. (1996) *Biochemistry* **35**, 6366–6374
31. Elcock, A., Potter, M., Matthews, D., Knighton, D., and McCammon, A. (1996) *J. Mol. Biol.* **262**, 370–374
32. Elcock, A., Huber, G., and McCammon, A. (1997) *Biochemistry* **36**, 16049–16058
33. Atreya, C., Johnson, E., Williamson, J., Chang, S., Liang, P., and Anderson, K. (2003) *J. Biol. Chem.* **278**, 28901–28911