# Supporting Information

## Koehn et al. 10.1073/pnas.1206077109

### SI Materials and Methods

**Chemicals.** All chemicals were reagent grade and used as purchased without further purification, unless specified. All reagents not specified below were obtained from Sigma. $N^5$,$N^{10}$-methylene-5,6,7,8-tetrahydrofolate ($CH_2H_4$folate) was provided by Eprova. Radiolabeled [2-$^{14}$C]dUMP was purchased from Moravek Biochemicals. Sodium dithionite powder was purchased from J. T. Baker, and Tris(hydroxymethyl)aminomethane was obtained from Research Products International.

**Analytical Methods.** Separations were carried out on a Beckman series HPLC, with UV/Vis diode array detector and 500TR series Packard flow scintillation analyzer (FSA). An analytical reverse-phase Supelco column (Discovery series 250 mm × 4.6 mm) was used with 100 mM $KH_2PO_4$ at pH-6 followed by a gradient of methanol. The concentration of enzyme-bound flavin adenine dinucleotide (FAD) was determined by the 454 nm absorbance ($\varepsilon$ = 11,300 $cm^{-1}{\cdot}M^{-1}$). Single-protein spectra and kinetic measurements were performed on a Hewlett Packard Model 8453 diode array UV-Vis spectrophotometer.

***Protein expression and purification.*** The flavin-dependent thymidylate synthase (FDTS) from *Thermatoga maritima* (TM0449, GenBank accession number NP228259), and its H53A, H53D, E144R, E144A, R174K, and R174A mutants were expressed and purified as previously described (1).

***Crystallization and structure determination.*** The crystals involving the quaternary complexes folate or folate mimics were crystallized at 22 °C in 3–6% PEG 4K (wt/vol), 200 mM NaCl, 100 mM Na/K phosphate (pH 6.58). The protein complex was prepared by first mixing the protein containing FAD (protein purifies in the FAD bound form) in 20 mM Tris (pH 8.0) with 10 M excess of dUMP. This complex was kept at room temperature for 5 min and then treated with 6 mM folate compounds in 20 mM Tris (pH 8.4). These crystals belonged to space group $I4_122$ and contained one monomer per asymmetric unit. Because of the solubility issues, the Raltitrexed complexes were prepared by soaking the drug into FDTS-FAD-dUMP crystals grown in the same crystallization conditions.

Crystals of the E144R and R174K mutants with FAD and with FAD and dUMP were crystallized at 22 °C in 50–60% (wt/vol) PEG 200 and 100 mM Tris buffer, pH 8.0. These crystals belonged to the $P2_12_12_1$ space group with a tetramer in the asymmetric unit. We could not grow crystals of the quaternary complexes with folate or folate mimics from the PEG200 conditions. Therefore, E144R +FAD +Folate crystals were prepared by soaking folate (~10 mM solution) into E144R+FAD crystals grown from the PEG 200 condition. The crystals from the PEG 4K condition were transferred stepwise into a cryoprotective solution containing 24% PEG 4K (wt/vol), 200 mM NaCl, 100 mM Na/K phosphate, and 10% glycerol (vol/vol), and flash cooled. The crystals grown from the PEG200 conditions were flash cooled directly from the drop.

Diffraction data for all of the crystals except E144R+FAD+Folate crystals were collected at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 9-2 using Mar 325 detector (R174K mutant collected using Q315 detector). The diffraction data for the E144R+FAD+Folate crystal were collected at SSRL beamline 12-2 using the Dectris Pilatus 6 M detector. The wavelength used for the folate and folinic acid complexes of the native enzyme was 0.9809Å and the others were collected at 0.9795Å. All data were processed using the XDS package (2). Multiple data sets from other crystals were also collected and refined to reasonable R-factors to confirm the binding mode observed for the folate and folinic acid groups.

Structures of the folate derivative complexes were solved by molecular replacement [MOLREP (3)] using the atomic coordinates of the monomer from the *T. maritima* tetramer (PDB code: 1O26). The E144R and R174K mutants structures were phased by rigid body refinement implemented in the REFMAC (4) program. All of the structures went through several rounds of model building using Coot (5) and refinement using the REFMAC program. The Ramachandran statistics for the refined coordinates showed no outliers. Details of the data collection and refinement statistics are shown in Tables S1–S3.

***Activity determination of FDTS mutants.*** The activities of H53A, H53D, E144R, E144A, R174K, and R174A mutants of *Tm*FDTS were determined relative to the wild-type using the established [2-$^{14}$C]dUMP assay (Table S4), as previously described (6, 7). All experiments were performed anaerobically in 200 mM Tris buffer (exchanged with Ar) pH = 8.0 at 65 °C. Reactions contained 100 μM dUMP (including 0.5 Mdpm [2-$^{14}$C]dUMP), 100 μM $CH_2H_4$folate, 5 mM $CH_2O$ (to stabilize $CH_2H_4$folate) and 5 mM sodium dithionite or 500 μM NADPH. Reactions were initiated by addition of 25 nM to 20 μm (final active-site concentration) of enzyme, quenched with HCl (to a final pH = 1), and stored at −80 °C until analysis. HPLC-FSA analysis was used to determine the conversion of [2-$^{14}$C]dUMP to [2-$^{14}$C]dTMP.

***Inhibition of enzyme-bound FAD reductive-half reaction.*** All experiments were performed in an anaerobic cuvettete using 200 mM Tris buffer (exchanged with Ar) pH = 8.0 at 37 °C. Reactions contained ~60 μM FDTS (active-site concentration), 100 μM dUMP, with or without 500 μM folinic acid. From the cuvettete sidearm 500 μM NADPH was added to initiate the reaction. The 454 nm absorbance was followed as a function of time (Fig. S6). Reduction rates were approximated by linear fitting corresponding to the first 10% fraction conversion.

***Modeling system set-up.*** The modeling was performed starting from the crystal structures of wild-type *Tm*FDTS complexed with dUMP, FAD, and $CH_2H_4$folate or folinic acid. The crystal structures contain a total of 216 amino acid residues per subunit, 764 ($CH_2H_4$folate structure) and 756 (folinic acid structure) crystallographic water molecules, FAD, dUMP and $CH_2H_4$folate or folinic acid ligands. The structures have missing residues (residues 35, 36, and 37), these residues were not taken into account and fragments ends were constrained. Missing hydrogens were added to the PDB structure using the tleap module of Amber9 package (8). All solvent molecules and ions were removed. Protonation state for all ionizable residues were set corresponding to pH 7. Thus, His residues were modeled as neutral residues with the proton on Nε or Nδ as determined on the basis of possible hydrogen bond interactions deduced from X-ray crystallographic structure. The amber99 force-field (9) parameters were used for all residues.

$CH_2H_4$folate was replaced by the protonated imine form of $CH_2H_4$folate (reactive intermediate, INT). The INT ligand was built starting from the folinic acid structure where the 5-formyl was removed and replaced with a methylene having a double bond to the N5 atom. Minimization of INT was done at the HF/6–31G** level of theory using the gaussian09 package (10). Then, the restrained electrostatic potential (RESP) (11) charge approach was used to derived atomic charges. First, the high-density ESP calculation was performed at HF/6–31G** using Merz–Kollman scheme with the gaussian09 package. Then the charges were derived using Antechamber module of Amber9

package, which performed a two-stage fitting. Finally, Gaff force-field (12) parameters together with RESP charges were used to generate the parameters file for INT. FAD and dUMP RESP charges and parameters were obtained from R.E.DD.B (11).

***Ligand docking and minimizations.*** FDTS is a homotetramer with four active sites each located in the interface of three subunits. Because of its location, the whole tetramer was considered in the minimizations. Minimizations were done using the Sander module of the Amber9 package. $CH_2H_4$folate or folinic acid was removed from all active sites. INT was manually docked into one active site. One short minimization for the entire system was done, and then was followed by a local minimization with restraints. Restraints were applied to all residues outside a radius of 15 Å from N1 atom of dUMP molecule, to keep them fixed at their original positions. Only residues within 15 Å radius from N1 atom of dUMP were allowed to move. The Generalized Born solvation approach was used (13–15).


1. Lesley SA, et al. (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci USA* 99:11664–11669.
2. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66:125–132.
3. Vagin A, Teplyakov A (1997) MOLREP: An automated program for molecular replacement. *J Appl Cryst* 30:1022–1025.
4. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.
5. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
6. Agrawal N, Lesley SA, Kuhn P, Kohen A (2004) Mechanistic studies of a flavin-dependent thymidylate synthase. *Biochemistry* 43:10295–10301.
7. Koehn EM, et al. (2009) An unusual mechanism of thymidylate biosynthesis in organisms containing the thyX gene. *Nature* 458:919–923.
8. Case DA, et al. (2006) *AMBER 9.* (Univ of California, San Francisco, CA).
9. Cheatham TE, 3rd, Cieplak P, Kollman PA (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J Biomol Struct Dyn* 16:845–862.
10. Frisch MJ, et al. (2009) *Gaussian 09, Revision A.02* (Gaussian, Wallingford, CT).
11. Dupradeau F-Y, et al. (2008) R.E.DD.B.: A database for RESP and ESP atomic charges, and force field libraries. *Nucleic Acids Res* 36(Database issue):D360–D367.
12. Wang J, Cieplak P, Kollman PA (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 21:1049–1074.
13. Bashford D, Case DA (2000) Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem* 51:129–152.
14. Simonson T (2001) Macromolecular electrostatics: Continuum models and their growing pains. *Curr Opin Struct Biol* 11:243–252.
15. Tsui V, Case DA (2000-2001) Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers* 56:275–291.


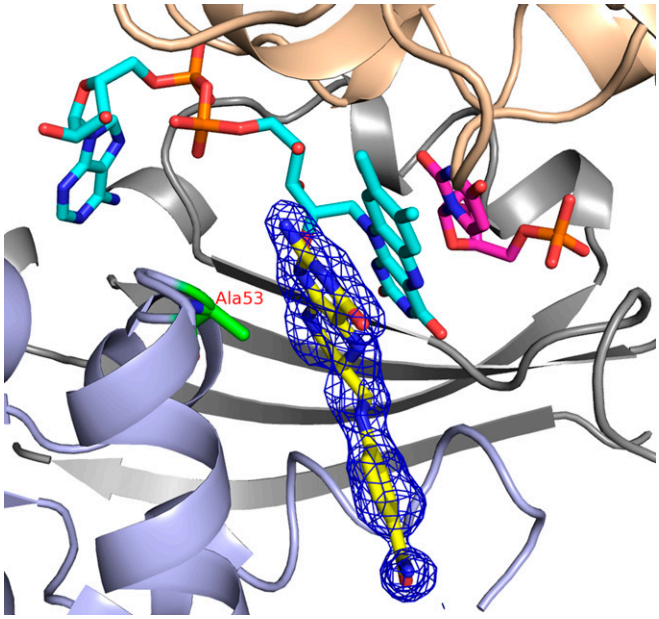


**Fig. S1.** The crystal structure of H53A *Tm*FDTS in complex with FAD, dUMP, and folate. A view of the omit map contoured at 3σ for the folate. The mutated alanine residue is shown by stick representation in green and the other colors are as in Fig. 3. PDB code: 4GTF.

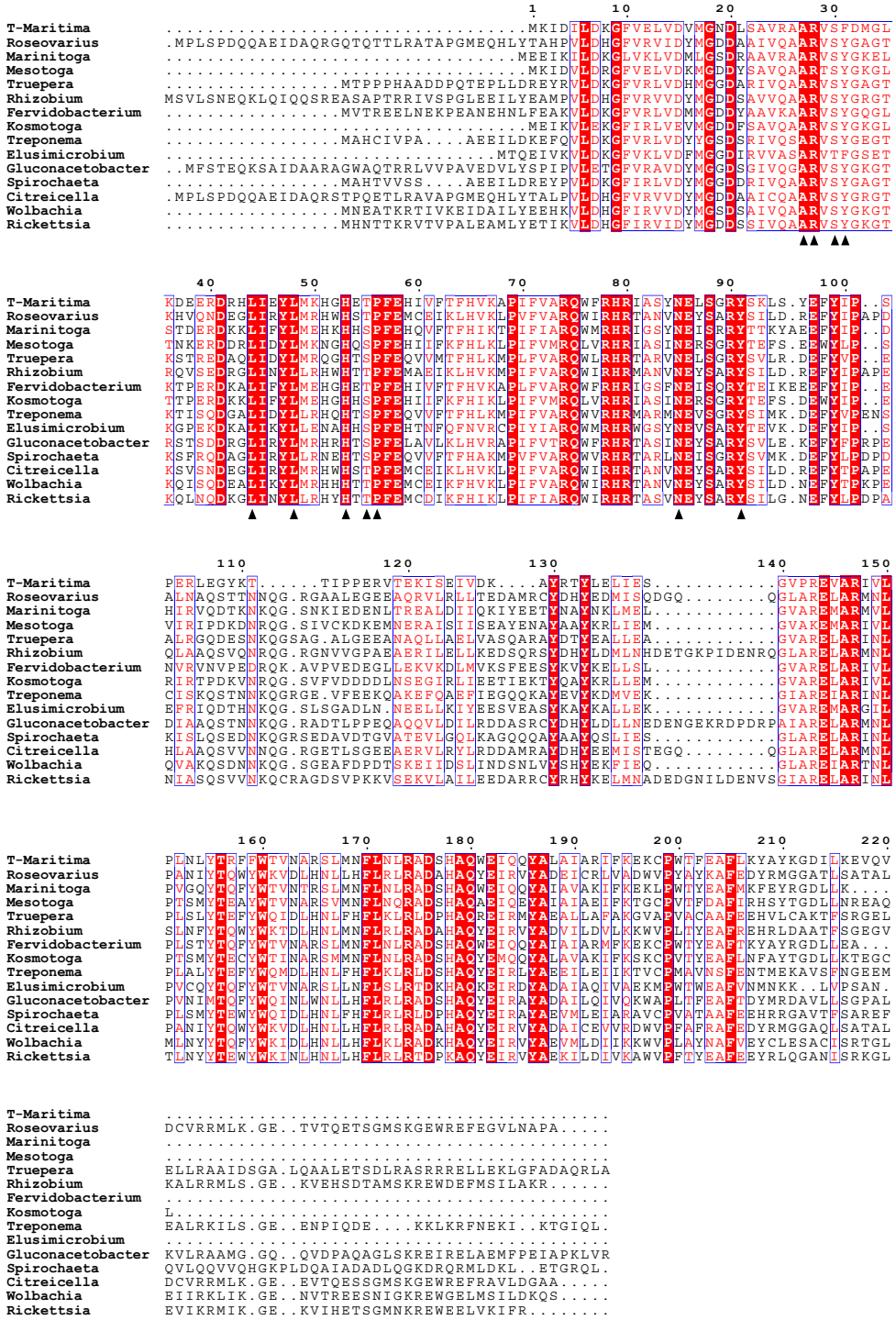


**Fig. S2.** Alignment chart of FDTS sequences related to the *T. maritime* FDTS. The sequences were aligned using the ClustalW program (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The strictly conserved residues are highlighted with red background. Residues that show similar properties are shown with a blue box around them. The amino acids near the folate binding site are marked by black triangles at the bottom of each column. Aligned sequences are: *T. maritima* (NP_228259.1), Roseovarius (ZP_01038219.1), *Marinitoga piezophila* (YP_005097630.1), *Mesotoga prima* (YP_006345493.1), *Truepera radiovictrix* (YP_003703722.1), *Rhizobium leguminosarum* (EIW44808.1), *Fervidobacterium pennivorans* (YP_005471930.1), *Kosmotoga olearia* (YP_002939761.1), *Treponema succinifaciens* (YP_004366481.1), *Elusimicrobium minutum* (YP_001874966.1), *Gluconacetobacter oboediens* (ZP_08897936.1), *Spirochaeta Africana* (YP_005476544.1), Citreicella (ZP_05781443.1), Wolbachia endosymbiont of *Drosophila melanogaster* (NP_966910.1), and *Rickettsia prowazekii* (NP_220685.1). The sequence alignment figure was generated using ESPript (1).

1. Gouet P, Courcelle E, Stuart DI, Métoz F (1999) ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics* 15:305–308.

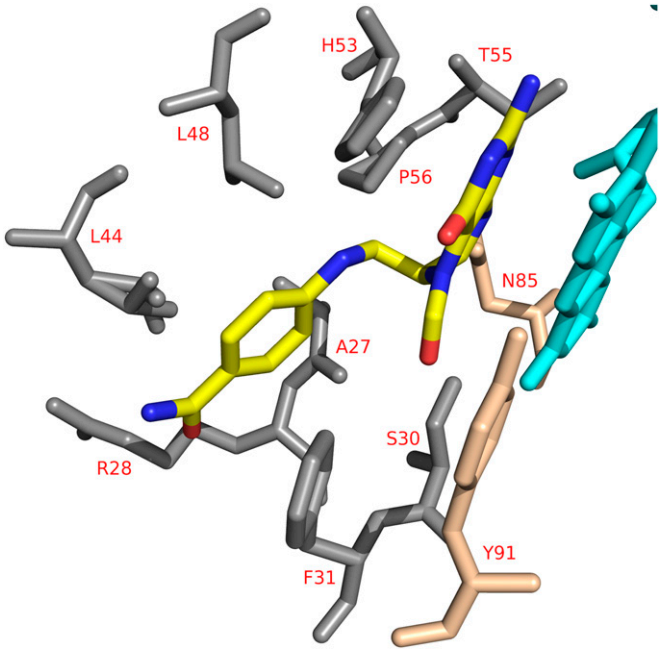


**Fig. S3.** A view from the structure presented in Fig. 3*A* of *Tm*FDTS in complex with FAD, dUMP, and $CH_2H_4$folate (PDB code: 4GT9). This view emphasizes the folate binding site and specifies residues that interact with the folate ligand.

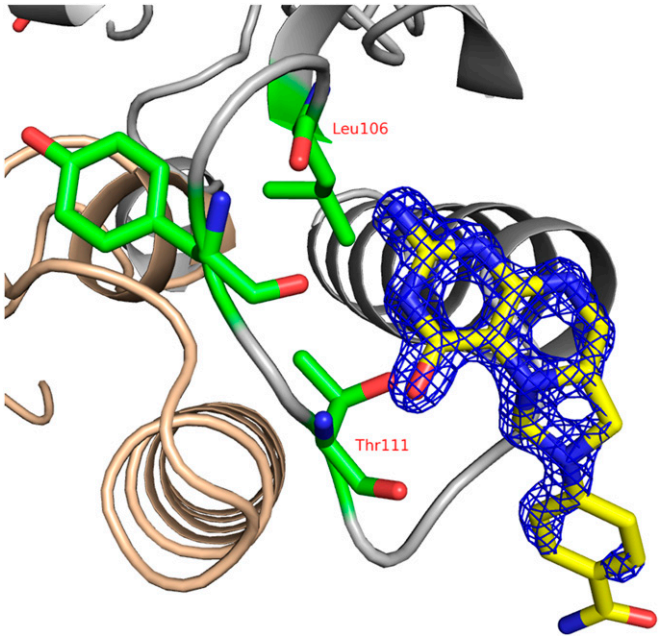


**Fig. S4.** Remote folate binding site. A view of the 2Fo-Fc electron density for the second folate contoured at 1.5σ. $CH_2H_4$folate is shown in yellow and the hydrogen bonding residues are shown as sticks. PDB code: 4GT9.

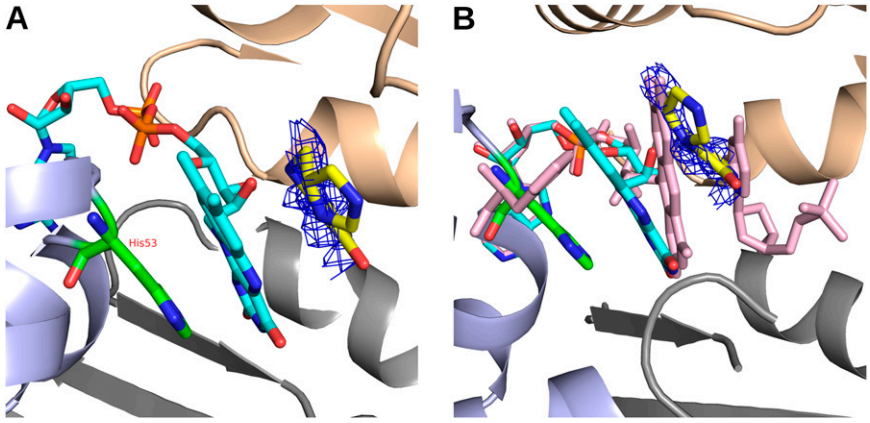


**Fig. S5.** A view of the binding mode in the E144R+FAD+folate structure (PDB code: 4GTE). (*A*) The electron density omit map contoured at 2.5σ showing the approximate location of the folate. (*B*) Comparison of the active sites of the E144R+FAD+folate and wt+FAD+dUMP (PDB code: 1O26, in pink) structures.

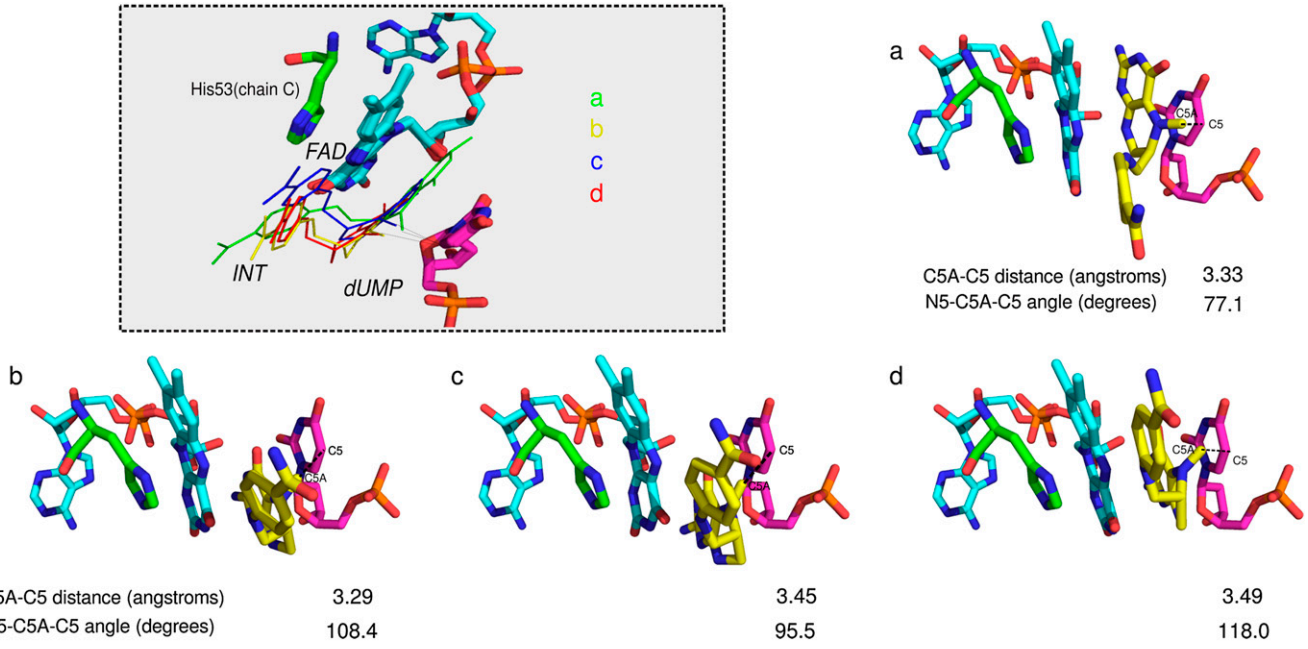


**Fig. S6.** Alternative modeled structures for the FDTS-FAD-dUMP-iminium $CH_2H_4$folate complex where the methylene in close to C5 of dUMP (same color code as in the main text figures). Orientations in *A*, *B*, *C*, and *D* are not identical to Fig. 5, although the docking procedures are the same as described for Fig. 5. (*Inset*) The superimposed intermediate docked orientations in different colors: *A* (green), *B* (yellow), *C* (blue), and *D* (red).

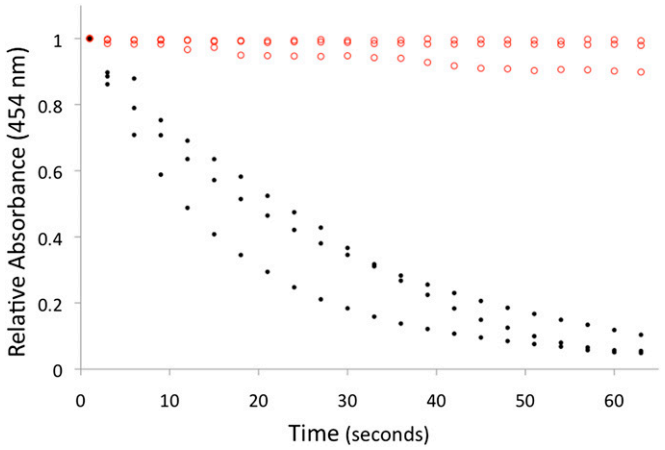


**Fig. S7.** Folinic acid inhibition of FAD reduction by NADPH (single reductive-half reaction). The 454-nm absorbance of FAD decreases with reduction to $FADH_2$. A marked decrease in the rate of reduction is observed in the presence of folinic acid (red open circles) versus the control without folinc acid (black solid circles).

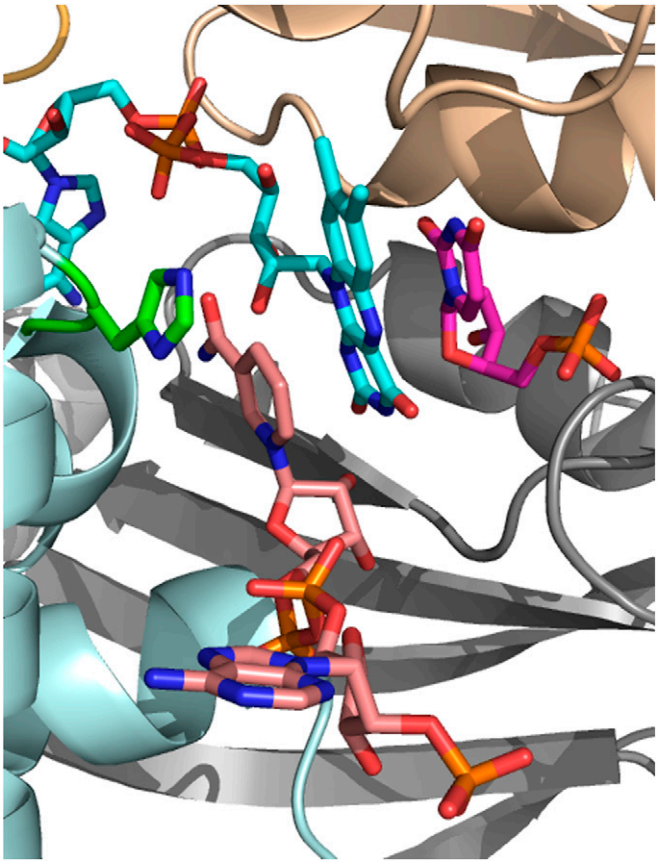

**Fig. S8.** $NADP^+$ (in pink) docked into the binding site of folate derivatives found in the crystal structures presented in Fig. 2. The binding cavity is wide and could easily accommodate $NAD^+$, NADPH, NADH, (both substrates of this enzyme) or other reducing agents.

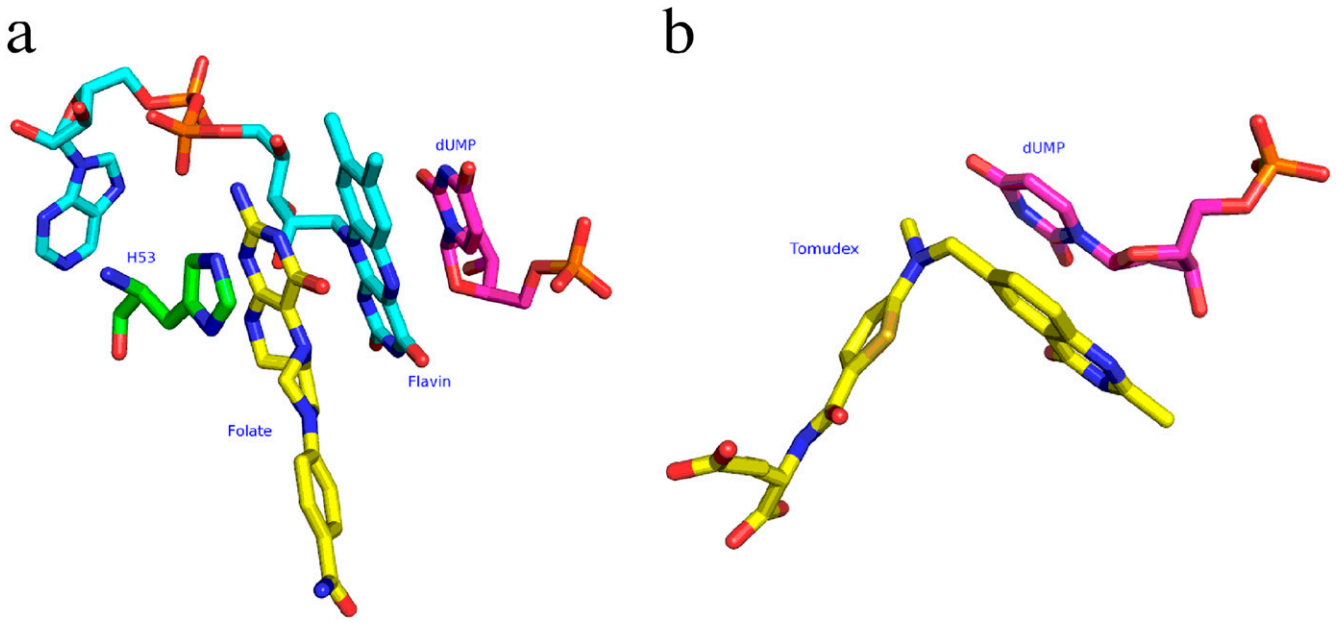


**Fig. S9.** Comparison of folate binding modes in FDTS and TSase. (*A*) A view of the stacking of the pterin with FAD and His53 in the *Tm*FDTS enzyme. (*B*) Stacking of the pterin and dUMP in the classic *Ec*TSase. Compounds are shown with stick representation and labeled.

**Table S1. Data collection and refinement statistics: Wild type complexes**

| Data and refinement | FAD+dUMP+Folate | FAD+dUMP+Folinic Acid | FAD+dUMP+ Raltitrexed |
|---|---|---|---|
| Data collection | | | |
| Space group | $I4_122$ | $I4_122$ | $I4_122$ |
| Cell dimensions | | | |
| a, b, c (Å) | 110.3, 110.3, 120.6 | 110.3, 110.3, 121.0 | 110.6, 110.6, 121.7 |
| α, β, γ (°) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| Resolution (Å) | 25.6–1.39 (1.43–1.39) | 40.0–1.50 (1.54–1.50) | 40.0–1.70 (1.74–1.70) |
| $R_{sym}$ or $R_{merge}$* | 4.3 (87.6) | 4.7 (95.3) | 5.0 (83.1) |
| I/σI | 27.7 (2.7) | 24.3 (2.4) | 21.0 (2.1) |
| Completeness (%) | 99.6 (99.9) | 99.8 (100.0) | 99.0 (99.9) |
| Redundancy | 8.4 (8.4) | 8.4 (8.5) | 6.6 (5.3) |
| Refinement | | | |
| Resolution (Å) | 27.6–1.39 | 39.0–1.50 | 39.1–1.70 |
| No. reflections | 70,453 | 56,486 | 39,174 |
| $R_{work}/R_{free}$ † | 15.2/16.4 | 16.4/17.9 | 17.1/18.3 |
| No. atoms | | | |
| Protein | 1845 | 1920 | 1830 |
| Ligand/ion | 121/1 | 98/1 | 88/1 |
| Water/buffer atoms | 209/4 | 189 | 126 |
| B-factors | | | |
| Protein | 22.3 | 30.0 | 36.2 |
| Ligand/ion | 23.1/25.3 | 22.9/29.2 | 30.5 |
| Water/buffer atoms | 34.4/46.5 | 39.5 | 42.6/32.0 |
| R.m.s deviations | | | |
| Bond lengths (Å) | 0.015 | 0.018 | 0.017 |
| Bond angles (°) | 1.756 | 1.796 | 1.740 |

*$R_{sym} = \sum |Iavg - Ii|/\sum Ii$
†Rfactor = $\sum |Fp - Fpcalc|/\sum Fp$, where Fp and Fpcalc are observed and calculated structure factors; $R_{free}$ is calculated with 5% of the data.

**Table S2. Data collection and refinement statistics: E144R complexes**

| Data and refinement | E144R+FAD | E144R+FAD+dUMP | E144R+FAD+Folate |
|---|---|---|---|
| Data collection | | | |
| Space group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Cell dimensions | | | |
| a, b, c (Å) | 54.1, 117.1, 142.1 | 54.6, 117.4, 142.6 | 54.7, 117.1, 141.9 |
| α, β, γ (º) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| Resolution (Å) | 30.0–1.97 (2.02–1.97) | 30.0–1.76 (1.81–1.76) | 40.0–1.89 (1.94–1.89) |
| $R_{sym}$ or $R_{merge}$* | 6.0 (88.6) | 6.1 (98.8) | 8.2 (97.0) |
| I/σI | 16.4 (1.6) | 21.9 (2.0) | 14.5 (1.7) |
| Completeness (%) | 98.6 (99.6) | 99.8 (99.9) | 99.2 (99.4) |
| Redundancy | 4.2 (4.3) | 6.6 (6.4) | 5.4 (3.7) |
| Refinement | | | |
| Resolution (Å) | 29.3–1.97 | 29.4–1.76 | 38.5–1.89 |
| No. reflections | 60,636 | 86,827 | 69,504 |
| $R_{work}/R_{free}$ † | 19.3/23.6 | 17.9 (21.5) | 18.4 (21.0) |
| No. atoms | | | |
| Protein | 7,263 | 7,174 | 7,246 |
| Ligand/ion | 194 | 292 | 206 |
| Water | 265 | 373 | 167 |
| B-factors | | | |
| Protein | 41.7 | 40.3 | 39.7 |
| Ligand/ion | 48.2 | 38.6 | 49.1 |
| Water | 44.0 | 42.8 | 41.7 |
| R.m.s deviations | | | |
| Bond lengths (Å) | 0.016 | 0.017 | 0.018 |
| Bond angles (º) | 1.638 | 1.684 | 1.666 |

*$R_{sym} = \sum |Iavg - Ii|/\sum Ii$
†Rfactor = $\sum |Fp - Fpcalc|/\sum Fp$, where Fp and Fpcalc are observed and calculated structure factors; $R_{free}$ is calculated with 5% of the data.

**Table S3. Data collection and refinement statistics: H53A and R174K complexes**

| Data and refinement | H53A+FAD+dUMP+Folate | R174K+FAD |
|---|---|---|
| Data collection | | |
| Space group | $I4_122$ | $P2_12_12_1$ |
| Cell dimensions | | |
| a, b, c (Å) | 109.8, 109.8, 122.3 | 53.8, 116.5, 141.2 |
| α, β, γ (º) | 90.0, 90.0, 90.0 | 90.0, 90.0, 90.0 |
| Resolution (Å) | 40.0–1.77 (1.82–1.77) | 40.0–2.17 (2.23–2.17) |
| $R_{sym}$ or $R_{merge}$* | 6.0 (96.9) | 6.5 (85.9) |
| I/σI | 23.3 (2.7) | 18.1 (1.9) |
| Completeness (%) | 99.9 (100.0) | 99.8 (99.5) |
| Redundancy | 9.1 (9.2) | 5.0 (4.6) |
| Refinement | | |
| Resolution (Å) | 38.8–1.77 | 38.1–2.17 |
| No. reflections | 34,763 | 45,284 |
| $R_{work}/R_{free}$ † | 16.3/18.3 | 18.9/24.0 |
| No. atoms | | |
| Protein | 1,864 | 7,233 |
| Ligand/ion | 96/1 | 141 |
| Water | 124 | 81 |
| B-factors | | |
| Protein | 34.3 | 34.2 |
| Ligand/ion | 27.9/38.3 | 59.4 |
| Water | 41.3 | 46.7 |
| R.m.s deviations | | |
| Bond lengths (Å) | 0.016 | 0.015 |
| Bond angles (º) | 1.699 | 1.697 |

*$R_{sym} = \sum |Iavg - Ii|/\sum Ii$
†Rfactor = $\sum |Fp - Fpcalc|/\sum Fp$, where Fp and Fpcalc are observed and calculated structure factors; $R_{free}$ is calculated with 5% of the data.

**Table S4. Activity of *Tm*FDTS mutants. Reactions contained 100 μM dUMP (including 0.5 Mdpm [2-$^{14}$C]dUMP), 100 μM $CH_2H_4$folate, and 5 mM sodium dithionite or 500 μM NADPH**

| FDTS enzyme | [$^{14}$C]dTMP formation | Rate ($min^{-1}$) ($10^3$) | % Relative activity |
|---|---|---|---|
| Wild-type | Yes | 2,135 ± 16 | 100.0 ± 0.8 |
| E144A | Yes | 2.42 ± 0.04 | 0.113 ± 0.002 |
| E144R | Yes | 0.34 ± 0.03 | 0.016 ± 0.002 |
| R174A | Yes | 0.017 ± 0.003 | 0.0008 ± 0.0002 |
| R174K | No | ND | ND |
| H53A | Yes | 29.6 ± 0.4 | 1.39 ± 0.02 |
| H53D | Yes | ND | ND |

FDTS activity was assessed by quantifying [2-$^{14}$C]dTMP formation. Average rates and their SDs were determined from three independent measurements. ND, denotes that the average rates were not determined.