CHEMMEDCHEM

DOI: 10.1002/cmdc.200700215

Sequence-Based Identification of Specific Drug Target Regions in the Thymidylate Synthase Enzyme Family

Stefania Ferrari,* Valeria Losasso, and Maria Paola Costi*[a]

Introduction

In the search for antimicrobial agents, the identification of new proteins that are not expressed in humans enables the selective targeting of pathogenic organisms without altering the functions of human cells. Conversely, in the search for receptors that are suitable for drug targeting and that are also present in human cells, it is crucial to identify structural features, or sub-domains, that distinguish these target proteins. The issue of target specificity will be advanced by increasing our knowledge of protein structures and properties. As our understanding of the relationship between protein structure and function improves, the number of potential protein targets will increase. For example, within the families of proteases and kinases in humans and pathogens, the discovery of specific inhibitors has been carried out by matching structural and dynamic specificity with the design of complementary ligands.[1,2] Structural biology of proteins refers to the broader study of protein structure that extends beyond the static description of amino acid sequence and protein folding. Furthermore, conformational analysis incorporates the flexibility and overall plasticity of proteins.[3–9] In the study of protein–ligand interactions, the local motion of protein domains can play a critical role in determining the binding of a ligand, and can ultimately define its efficacy. Structural diversity among proteins within the same family can account for differences in protein flexibility, especially regarding domains with the same folding pattern but that differ in domain length. Conversely, domains with similar folding patterns and length can show the same flexibility and share similar conformational space, even if the amino acid sequences differ.[10] Bearing these considerations in mind, the target specificity of a protein, defined as the capacity of a protein to exclusively bind one ligand over others, can be tracked in terms of sequence structure and flexibility. While the latest structural biology studies require more sophisticated methods, they are firmly grounded in the linear description of sequence/domain similarity.

Members of the thymidylate synthase enzyme family, key proteins in the DNA biosynthetic pathway, have been validated as targets in the anticancer and anti-infective drug discovery areas.[11–13] This family includes two proteins: thymidylate synthase (TS or ThyA; EC 2.1.1.45) and flavin-dependent thymidylate synthase (FDTS, ThyX, or thymidylate synthase complementing protein (TSCP); EC 2.1.1.148). TS and FDTS have the same catalytic functions, but differ in protein structure, fold, and mechanism. They catalyze the reductive methylation of 2′-deoxyuridine-5′-monophosphate (dUMP) into 2′-deoxythymidine-5′-monophosphate (dTMP), assisted by cofactors acting as both carbon unit donors and reducing agents.

TS uses the cofactor 5,10-methylene-5,6,7,8-tetrahydrofolate (MTHF) as both carbon unit donor and reducing agent. TS inhibition in rapidly dividing cells causes cell death through a mechanism referred to as 'thymineless death'.[14–16] We have previously identified TS inhibitors with high species-specific profiles.[17–22] Among them, 3,3-bis-(3-chloro-4-hydroxyphenyl)-1*H*,3*H*-naphtho[1,8-*c*,*d*]pyran-1-one (α156), *N*,*O*-didansyl-L-tyrosine (DDT), and benzyloxycarbonylamino-*N*,*O*-didansyl-L-tyrosine (FD13) (Figure 1) discriminated between human (hTS) and bacterial TS (for example, *Escherichia coli* TS (EcTS) and *Lactobacillus casei* TS (LcTS)). α156 is the leading candidate among the naphthalene derivatives, some of which show a species-

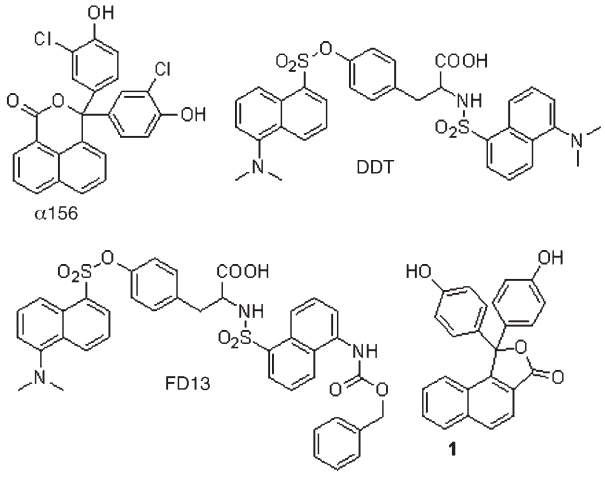

Figure 1. Structures of species-specific inhibitors of bacterial TS.

[a] *Dr. S. Ferrari, Dr. V. Losasso,[+] Prof. M. P. Costi*
Dipartimento di Scienze Farmaceutiche
Università degli Studi di Modena e Reggio Emilia
Via Campi 183, 41100 Modena (Italy)
Fax: (+39) 059-205-5131
E-mail: stefania.ferrari@unimore.it
mariapaola.costi@unimore.it

[+] *current address:*
Statistical and Biological Physics Group
Scuola Internazionale Superiore di Studi Avanzati (SISSA)
via Beirut 2–4, 34014 Trieste (Italy)

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

specificity index (SI) of up to 10^3 ($SI = K_{i(hTS)}/K_{i(pathogen\ TS)}$).[18] α156 is known to inhibit LcTS with a K_i value 43-fold lower than that against hTS.[18] X-ray crystallography studies revealed that this compound binds in a non-conserved region of the enzyme active site and interacts with the *small domain* (SD) region that is structurally different between LcTS and hTS.[18] Based on the crystallographic analysis of the LcTS–α156 complex, an initial hypothesis was formulated to explain the activity data of α156: variations in the SD region account for differences in inhibition activity. Further investigation of the inhibition activity profile of α156 surprisingly revealed that this compound shows the same affinity toward EcTS and LcTS even though the SD region of EcTS is also structurally different from that of LcTS (and that of hTS).[19] To understand how variations in sequences and structures, as deduced from crystal structure comparison, are able to affect the binding and thus the inhibition activity of α156, it was necessary to consider not only static structural differences between TSs of various species, but also how these structural differences can influence protein dynamics and flexibility, and how these enzyme features can affect ligand binding.[19] Docking calculations and crystal-structure-based estimates of the essential dynamics of TSs derived from five different species (*Pneumocystis carini* TS (PcTS), *Leishmania major* TS (LmTS), LcTS, EcTS, and hTS) showed that for LcTS and EcTS, differences in the protein structures and thus in the protein dynamics make the active site more accessible to α156 than in the other TSs, thereby enhancing the specificity of α156.[18,19] FD13, a didansyltyrosine derivative, is able to discriminate among LcTS, EcTS, and hTS with SI values of 100 due to interaction with specific residues present only in the SD of LcTS.[21] New 1,2-naphthalene derivatives were recently synthesized and tested against a TS-based bio-library of mostly pathogen TS enzymes, and then compared with hTS and human dihydrofolate reductase (hDHFR).[22] Within this series, the inhibitor with greatest species specificity was 1,1-bis-(4-hydroxyphenyl)-1*H*-naphtho[1,2-*c*]furan-3-one (compound **1**, Figure 1), with a K_i value against *Cryptococcus neoformans* TS (CnTS) of 0.4 μM and SI = 330. These studies demonstrate that it is possible to selectively target a specific region within highly conserved proteins to achieve specific inhibition at the molecular level.

In 1989, Dynes and Firtel discovered a gene in *Dictyostelium discoideum* that codes for a protein with thymidylate synthase activity.[23] This protein has no sequence homology to TS and it is not present in the human genome.[24] Several organisms, some of which are pathogens of great interest for pharmaceutical research, possess this protein with thymidylate synthase function, called FDTS.[24–30] Thus, FDTS is an excellent target for the design of selective microbial growth inhibitors. Whereas in most organisms the presence of FDTS excludes that of TS and vice versa, in *Mycobacterium tuberculosis* as well as some other organisms, both enzymes are present.[26,28,31] Recent studies have shown that FDTS adopts MTHF as a carbon unit donor and uses $FADH_2$ as a reducing agent.[24,26,28,31–35] The differences in primary structure and catalytic mechanism between TS and FDTS suggest independent origins for the two enzymes.[31] However, no inhibitors have been discovered yet for FDTS.

To further the discovery of structure-based specific inhibitors, we have identified putative drug target regions on thymidylate synthases from pathogenic organisms. To this end, we collected all the available TS and FDTS sequences from pathogenic organisms of pharmaceutical interest. We grouped the TS proteins in three bio-libraries based on featured structural regions throughout the comparative analyses of TS sequences with those of EcTS, LcTS, and hTS. The three families, EcTS-like, LcTS-like, and hTS-like, should identify proteins with the same flexibility properties. Using these three proteins as structural templates, we will provide a useful tool to achieve specific inhibition by ad hoc designed molecules by exploiting the structural and dynamic variation of these TSs.[18,19,21,22] Starting from the information provided herein, further flexibility studies among similar proteins can be undertaken. This more sophisticated part of the work is beyond the scope of the present review. With this analysis, we aimed to introduce an additional tool for examining the specificity potential of structure-based selected inhibitors.

The web search for protein sequences was conducted through the Protein database at the National Center for Biotechnology Information (NCBI) through the Entrez interface (http://www.ncbi.nlm.nih.gov/). A text-based query was carried out to search and collect all TS protein sequences that have been identified. The results (updated May 31, 2007) consisted of the TS sequences of 154 pathogenic organisms, grouped in 71 different genera. Among them, 45 organisms have the recently discovered FDTS protein (Table 1 and Table 2) and 11 organisms have both FDTS and TS proteins (Table 1). There are 14 organisms that express the bifunctional DHFR–TS protein. *Aspergillus nidulans* has a bifunctional enzyme, different from DHFR–TS, in which the C-terminal domain corresponds to TS (Supporting Information). ClustalW, provided by the Network Protein Sequence Analysis server, was used for the sequence alignment.[36,37] Each alignment was verified using the available information and published data, and, if necessary, corrected using GeneDoc software.[38] The TS sequences were then grouped based on two featured regions (the SD and the *interface loop* (IL)). The collected FDTS sequences were compared with the FDTS sequences of *Thermotoga maritima* (TmFDTS), *Mycobacterium tuberculosis* (MtFDTS) and *Paramecium bursaria Chlorella virus 1* (Pbcv1FDTS). To date, these are the only FDTS sequences with a reported 3D crystallographic structure.[26,28,34,39–41] We regularly update the TS/FDTS sequence collection, which is available at http://cdm.unimo.it/home/dipfarm/costi.mariapaola/TS-FDTS.htm.

1. TS Sequence Analysis

Featured structural regions have been identified from previous studies: 1) SD (residues 96–131 in hTS, 67–90 in EcTS, 69–142 in LcTS), which varies in length and sequence in EcTS (24 residues), hTS (36 residues), and LcTS (74 residues); 2) IL (residues 143–158 in hTS, 102–109 in EcTS, 154–161 in LcTS), in which eukaryotic organisms show a specific eight-residue insertion that is missing in LcTS and EcTS (Figures 2a and 3a).[18,19,42–45] These variations make the enzymes structurally and dynamical-

Table 1. Eleven pathogenic organisms for which both thymidylate synthases, TS and FDTS, have been identified.[a]

Organism	Identity [%][b]	Comparison with TS model sequences
Bacillus anthracis	TS: 43 (hTS), 50 (EcTS), 68 (LcTS) FDTS: 16 (TmFDTS), 25(MtFDTS), 12 (Pbcv1FDTS)	SD and HIL as LcTS (identity: 65 and 50%, respectively)
Bacillus cereus	TS: 50 (hTS), 43 (EcTS), 69 (LcTS) FDTS: 16 (TmFDTS), 24(MtFDTS), 12 (Pbcv1FDTS)	SD and HIL as LcTS (identity: 66 and 50%, respectively)
Bacillus weihenstephanensis	TS: 43 (hTS), 50 (EcTS), 68 (LcTS) FDTS: 17 (TmFDTS), 24(MtFDTS), 12 (Pbcv1FDTS)	SD and HIL as LcTS (identity: 65 and 50%, respectively)
Clostridium difficile	TS: 27 (hTS), 34 (EcTS), 28 (LcTS) FDTS: 20 (TmFDTS), 24(MtFDTS), 17 (Pbcv1FDTS)	SD 2 aa residues shorter than EcTS, HIL as EcTS (identity: 0%) insertions: 9 aa (res. 58–59 of hTS), 1 aa (res. 307–308 of hTS) deletions: res. 98–99 of hTS
Corynebacterium diphtheriae	TS: 47 (hTS), 61 (EcTS), 47 (LcTS) FDTS: 20 (TmFDTS), 53(MtFDTS), 16 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 50 and 75%, respectively) insertion: 1 aa (res. 282–284 of hTS)
Corynebacterium jeikeium	TS: 44 (hTS), 60 (EcTS), 47 (LcTS) FDTS: 20 (TmFDTS), 60(MtFDTS), 16 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 63 and 75%, respectively) insertion: 1 aa (res. 282–284 of hTS)
Mycobacterium avium	TS: 44 (hTS), 65 (EcTS), 49(LcTS) FDTS: 18 (TmFDTS), 18 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 58 and 75%, respectively) deletion: 1 aa (res. 279–286)
Mycobacterium bovis	TS: 44 (hTS), 66 (EcTS), 50(LcTS) FDTS: 19 (TmFDTS), 18 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 63 and 63%, respectively) deletion: 1 aa (res. 279–286)
Mycobacterium leprae	TS: 44 (hTS), 64 (EcTS), 49(LcTS) FDTS: 18 (TmFDTS), 16 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 67 and 50%, respectively) deletion: 1 aa (res. 279–286)
Mycobacterium tuberculosis	TS: 44 (hTS), 65 (EcTS), 50 (LcTS) FDTS: 19 (TmFDTS), 17 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 63 and 63%, respectively) deletion: 1 aa (res. 279–286)
Mycobacterium ulcerans	TS: 45 (hTS), 64 (EcTS), 50 (LcTS) FDTS: 18 (TmFDTS), 18 (Pbcv1FDTS)	SD and HIL as EcTS (identity: 58 and 75%, respectively) deletion: 1 aa (res. 279–286 of hTS)

[a] As of May 31, 2007. [b] Percent identities relative to the models' sequence as indicated.

Table 2. Thirty-four pathogenic organisms with known FDTS (for which no TS enzyme has been identified).[a]

Organism	Identity [%][b]	Organism	Identity [%][b]
Anaplasma phagocytophilum	31 (TmFDTS), 13 (MtFDTS), 17 (Pbcv1FDTS)	*Ehrlichia chaffeensis*	30 (TmFDTS), 14 (MtFDTS), 19 (Pbcv1FDTS)
Borrelia afzelii	27 (TmFDTS), 13 (MtFDTS), 17 (Pbcv1FDTS)	*Granulibacter bethesdensis*	25 (TmFDTS), 16 (MtFDTS), 16 (Pbcv1FDTS)
Borrelia burgdorferi	27 (TmFDTS), 13 (MtFDTS), 17 (Pbcv1FDTS)	*Helicobacter pylori*	19 (TmFDTS), 15 (MtFDTS), 15 (Pbcv1FDTS)
Borrelia hermsii	30 (TmFDTS), 14 (MtFDTS), 18 (Pbcv1FDTS)	*Neorickettsia sennetsu*	30 (TmFDTS), 15 (MtFDTS), 17 (Pbcv1FDTS)
Campylobacter coli	17 (TmFDTS), 25 (MtFDTS), 14 (Pbcv1FDTS)	*Orientia tsutsugamushi*	31 (TmFDTS), 17 (MtFDTS), 17 (Pbcv1FDTS)
Campylobacter concisus	16 (TmFDTS), 22 (MtFDTS), 12 (Pbcv1FDTS)	*Rickettsia akari*	30 (TmFDTS), 16 (MtFDTS), 17 (Pbcv1FDTS)
Campylobacter curvus	16 (TmFDTS), 20 (MtFDTS), 11 (Pbcv1FDTS)	*Rickettsia bellii*	29 (TmFDTS), 16 (MtFDTS), 16 (Pbcv1FDTS)
Campylobacter fetus	18 (TmFDTS), 22 (MtFDTS), 14 (Pbcv1FDTS)	*Rickettsia canadensis*	29 (TmFDTS), 16 (MtFDTS), 17 (Pbcv1FDTS)
Campylobacter jejuni	16 (TmFDTS), 23 (MtFDTS), 14 (Pbcv1FDTS)	*Rickettsia conorii*	29 (TmFDTS), 17 (MtFDTS), 17 (Pbcv1FDTS)
Campylobacter lari	19 (TmFDTS), 23 (MtFDTS), 15 (Pbcv1FDTS)	*Rickettsia felis*	30 (TmFDTS), 16 (MtFDTS), 17 (Pbcv1FDTS)
Campylobacter upsaliensis	17 (TmFDTS), 24 (MtFDTS), 14 (Pbcv1FDTS)	*Rickettsia prowazekii*	30 (TmFDTS), 16 (MtFDTS), 17 (Pbcv1FDTS)
Chlamydia trachomatis	10 (TmFDTS), 5 (MtFDTS), 4 (Pbcv1FDTS)	*Rickettsia rickettsii*	29 (TmFDTS), 17 (MtFDTS), 17 (Pbcv1FDTS)
Chlamydophila pneumoniae	7 (TmFDTS), 4 (MtFDTS), 3 (Pbcv1FDTS)	*Rickettsia sibirica*	29 (TmFDTS), 17 (MtFDTS), 17 (Pbcv1FDTS)
Clostridium botulinum	20 (TmFDTS), 24 (MtFDTS), 15 (Pbcv1FDTS)	*Rickettsia typhi*	30 (TmFDTS), 16 (MtFDTS), 17 (Pbcv1FDTS)
Clostridium novyi	22 (TmFDTS), 24 (MtFDTS), 16 (Pbcv1FDTS)	*Treponema denticola*	30 (TmFDTS), 17 (MtFDTS), 22 (Pbcv1FDTS)
Clostridium perfringens	21 (TmFDTS), 24 (MtFDTS), 15 (Pbcv1FDTS)	*Treponema pallidum*	26 (TmFDTS), 16 (MtFDTS), 16 (Pbcv1FDTS)
Clostridium tetani	22 (TmFDTS), 23 (MtFDTS), 15 (Pbcv1FDTS)	*Tropheryma whipplei*	33 (TmFDTS), 18 (MtFDTS), 21 (Pbcv1FDTS)

[a] As of May 31, 2007. [b] Percent identities relative to the models' sequence as indicated; for bifunctional enzymes only the FDTS domain was considered.

ly different. The SD and IL are flexible regions essential for TS activity. Previous functional studies indicate that the SD is located near the active site, and through extension of the SD, the active site is enlarged and becomes more accessible to ligand binding. The movement of the SD, as substrates and inhibitors bind in the active site, has been reported.[19,46,47] The

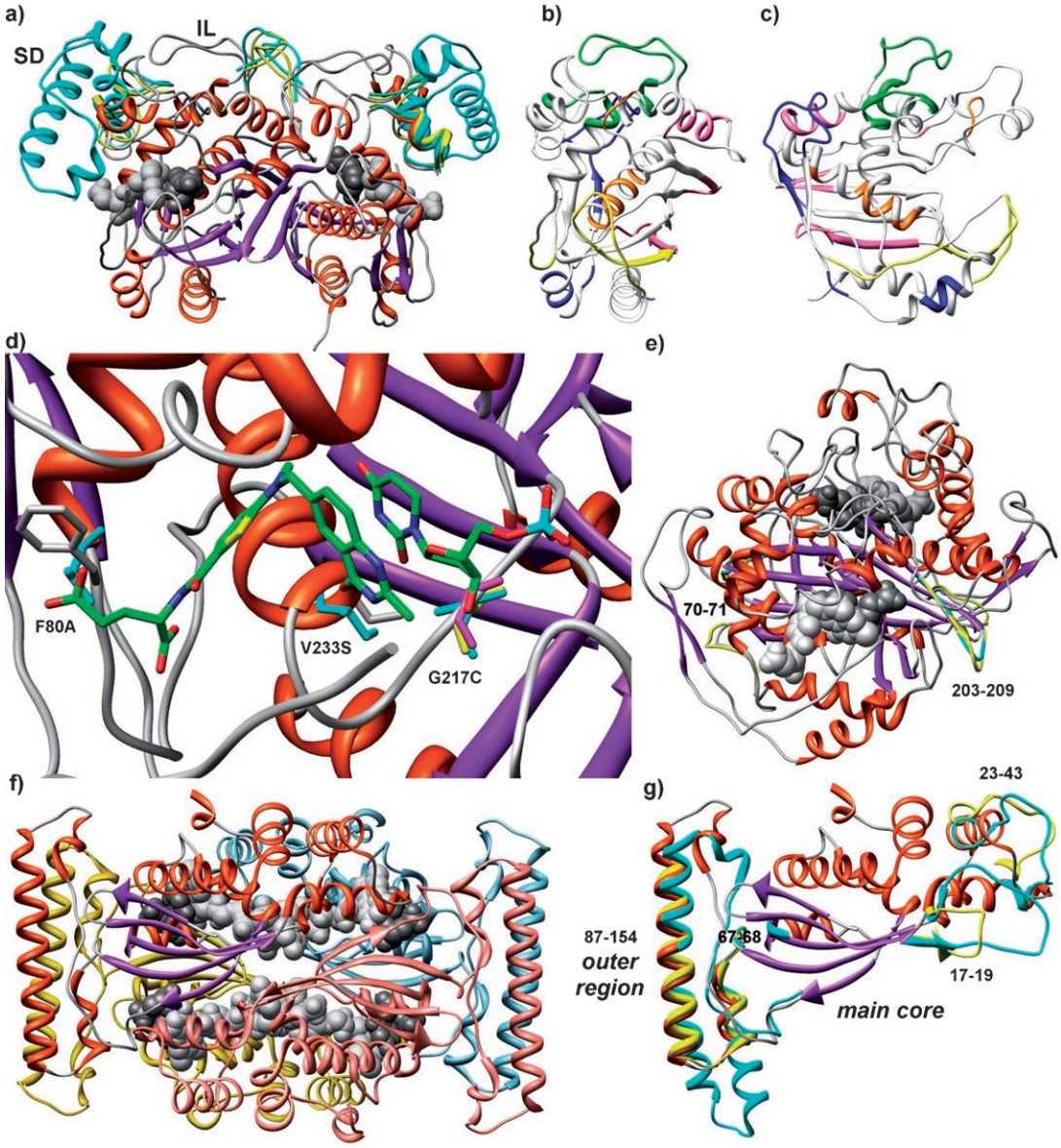

Figure 2. 3D protein structures of thymidylate synthase: a) Comparison of hTS (PDB code 1HVY), EcTS (yellow, PDB code 3TMS), and LcTS (cyan, PDB code 1TSL). The ribbon of hTS is colored by secondary structure elements: α helices in red, β strands in violet, other regions in gray. For clarity, only featured regions (SD and IL) of EcTS and LcTS are shown. Substrate (dUMP, dark gray) and a cofactor analogue (ZD1694, light gray) are bound in the active site. b),c) Regions (SD and IL in green, monomer–monomer interface regions in pink, regions near or in the active site in yellow, regions on the surface in blue) of TS in which insertions and/or deletions with respect to hTS have been found in TS sequences from pathogens. For clarity, only one monomer of the hTS (PDB code 1HVY) is shown. In panel c), the structure is rotated ~90° around the vertical axis with respect to that in panel b). d) Differences in active site residues of human (gray, PDB code 1HVY), *Cryptosporidium hominis* (cyan, PDB code 1SEJ), *Leishmania major* (yellow, PDB file from D. A. Matthews) and *Plasmodium falciparum* (magenta, PDE code 1J3I) TS; only three non-conserved residues are indicated. Substrate (dUMP) and a cofactor analogue (ZD1694) are bound in the active site, colored by atom type (C in green, O in red, N in blue, S in yellow, P in cyan). The ribbon of hTS is colored by secondary structure elements: α helices in red, β strands in violet, other regions in gray. e) Species-specific insertions in *Cryptosporidium hominis* (cyan, PDB code 1SEJ) and *Leishmania major* (yellow, PDB file from D. A. Matthews) TS with respect to hTS (PDB code 1HVY); hTS sequence numbering. Substrate (dUMP, dark gray) and a cofactor analogue (ZD1694, light gray) are bound in the active site. The ribbon of hTS is colored by secondary structure elements: α helices in red, β strands in violet, other regions in gray. f) TmFDTS crystallographic structures (PDB code 1O26): the four monomers of the homotetramer are colored differently (by secondary structure element (α helices in red, β-strands in violet, other regions in gray, pink, yellow, and light blue). Substrate (dUMP, dark gray) and an oxidized cofactor (FAD, light gray) are bound in the active site. g) Species-specific insertions in *Mycobacterium tuberculosis* (cyan, PDB code 2AF6) and *PBCV-1* (yellow, PDB code 2CFA) FDTS with respect to TmFDTS (PDB code 1O26); TmFDTS sequence numbering.

IL, depending on its extension, can affect protein flexibility and thus ligand binding.[19] The EcTS and LcTS short ILs move together with the interface region, remaining distant from the enzyme active site entrance. On the other hand, the hTS, LmTS, and PcTS long ILs move together with the surface regions, such as SDs, that modify the entrance of the active site, thus affecting the accessibility to bind the enzyme. After an in-depth analysis of the TS sequences, it became evident that the

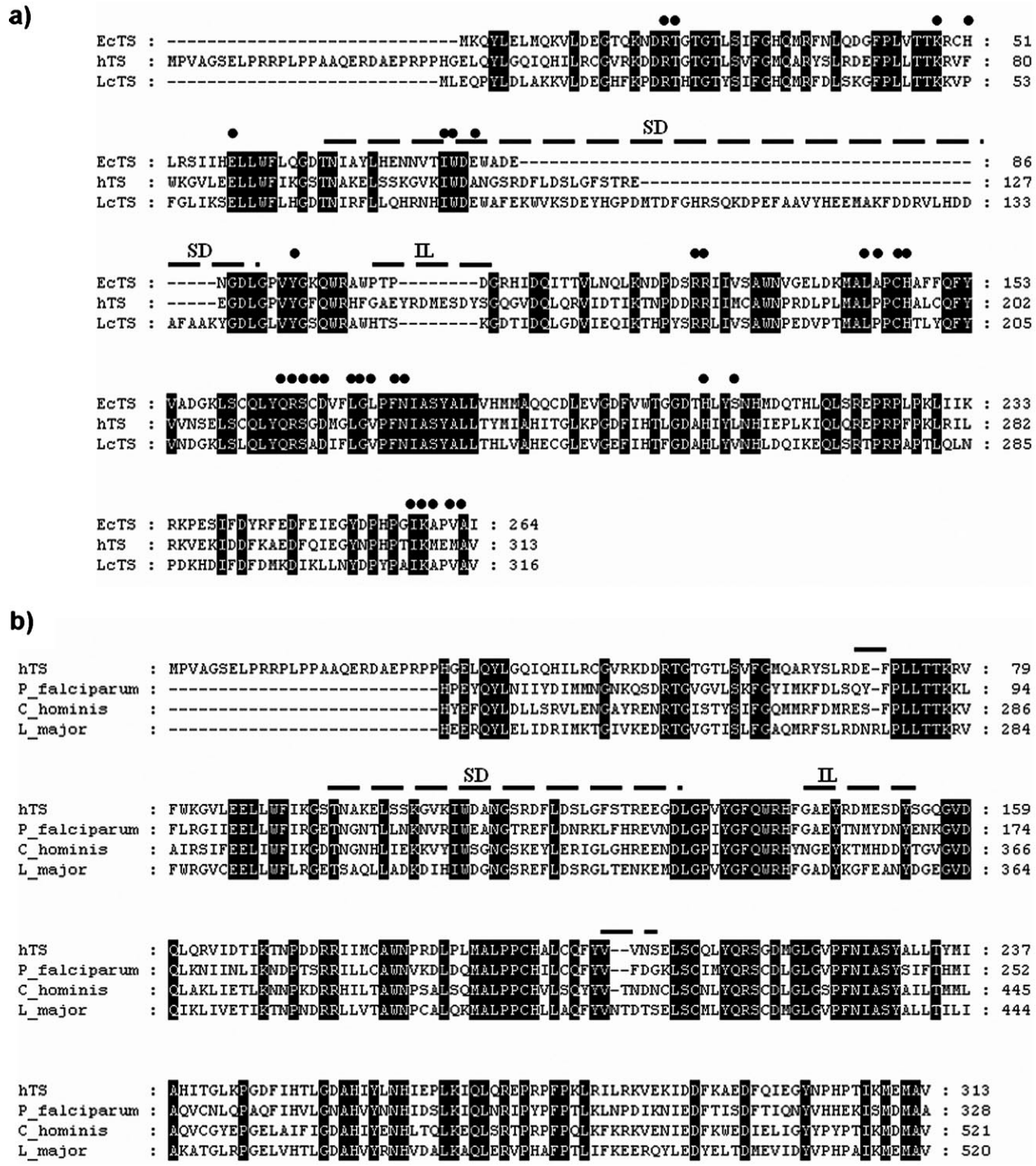

Figure 3. TS sequence alignments: identical residues among the aligned sequences are highlighted in dark boxes; SD and IL and other species-specific insertions are indicated with dashed lines. a) hTS, EcTS, LcTS sequence alignment; active site residues are indicated (●). b) Human, *Cryptosporidium hominis*, *Leishmania major*, and *Plasmodium falciparum* TS sequence alignment.

majority of the collected sequences could be grouped based on their SD and IL sequences using hTS, LcTS, and EcTS as templates (Table 3).

TS is a highly conserved enzyme, with 79% of the analyzed TS sequences showing at least 40% identity to one of the model sequences. Only 25 sequences (21% of the TS sequences analyzed) have an identity percentage equal to or less than 35% (Figure 4 and Supporting Information); these are: *Bacillus licheniformis*, *Bordetella* spp. (*bronchiseptica*, *parapertussis*, *pertussis*), *Burkholderia* spp. (*ambifaria*, *dolosa*, *mallei*, *multivorans*, *pseudomallei*, *vietnamiensis*), *Clostridium difficile*, *Fusobacterium nucleatum*, *Haemophilus* spp. (*ducreyi*, *influenzae*), *Streptococcus* spp. (*agalactiae*, *mutants*, *pneumoniae*, *pyogenes*, *sanguinis*), and *Vibrio* spp. (*alginolyticus*, *cholerae*, *parahaemolyticus*, *vulni-*

Table 3. 120 pathogenic organisms with known TS sequences.[a]

TS sequence with SD and IL as those of hTS
Ascaris suum, *Aspergillus* spp. (*clavatus*, *fumigatus*, *nidulans*), *Candida* spp. (*albicans*, *glabrata*), *Chaetomium globosum*, *Coccidioides immitis*, *Cryptococcus neoformans*, *Cryptosporidium* spp. (*hominis*, *parvum*), *Debaryomyces hansenii*, *Filobasidiella neoformans*, *Human herpesvirus* (*3*, *8*), *Leishmania* spp. (*amazonensis*, *braziliensis*, *donovani*, *infantum*, *major*, *tropica*), *Neosartorya fischeri*, *Plasmodium* spp. (*falciparum*, *malariae*, *vivax*), *Pneumocystis carinii*, *Schistosoma japonicum*, *Toxoplasma gondii*, *Trichinella spiralis*, *Trypanosoma* spp. (*cruzi*, *brucei*).
TS sequence with SD and IL as those of EcTS
Aeromonas hydrophila, *Bacteroides* spp. (*fragilis*, *thetaiotaomicron*), *Bartonella* spp. (*bacilliformis*, *henselae*, *quintana*), *Brucella* spp. (*abortus*, *melitensis*, *suis*), *Chromobacterium violaceum*, *Corynebacterium* spp. (*diphtheriae*, *jeikeium*), *Coxiella burnetti*, *Enterobacter* sp., *Legionella pneumophila*, *Mycobacterium* spp. (*avium*, *bovis*, *leprae*, *tuberculossis*, *ulcerans*), *Neisseria* spp. (*gonorrhoeae*, *meningitidis*), *Nocardia farcinica*, *Porphyromonas gingivalis*, *Pseudomonas* spp. (*aeruginosa*, *mendocina*), *Salmonella* spp. (*enterica*, *paratyphi*, *typhi*, *typhimurium*), *Serratia proteamaculans*, *Shewanella putrefaciens*, *Shigella* spp. (*boydii*, *dysenteriae*, *flexneri*, *sonnei*), *Stenotrophomonas maltophilia*, *Yersinia* spp. (*bercovieri*, *enterocolitica*, *frederiksenii*, *intermedia*, *mollaretii*, *pestis*, *pseudotuberculosis*).
TS sequence with SD and IL as those of LcTS
Bacillus spp. (*anthracis*, *cereus*, *weihenstephanensis*), *Enterococcus* spp. (*faecalis*, *faecium*), *Listeria monocytogenes*, *Staphylococcus* spp. (*aureus*, *epidermidis*, *haemolyticus*, *saprophyticus*).
TS sequence with SD and/or IL different from the three models
Acinetobacter baumannii, *Bacillus* spp. (*amyloliquefaciens*, *licheniformis*), *Bordetella* spp. (*brochiseptica*, *parapertussis*, *pertussis*), *Burkholderia* spp. (*ambifaria*, *dolosa*, *mallei*, *multivorans*, *pseudomallei*, *vietnamiensis*), *Clostridium difficile*, *Comamonas testosteroni*, *Encephalitozoon cuniculi*, *Francisella tularensis*, *Fusobacterium nucleatum*, *Haemophilyus* spp. (*ducreyi*, *influenzae*), *Mycoplasma* spp. (*genitalium*, *penetrans*, *pneumoniae*), *Novosphingobium aromaticivorans*, *Pasteurella multocida*, *Propionibacterium acnes*, *Streptococcus* spp. (*agalactiae*, *mutants*, *pneumoniae*, *pyogenes*, *sanguinis*), *Vibrio* spp. (*alginolyticus*, *cholerea*, *parahaemolyticus*, *vulnificus*).

[a] As of May 31, 2007; each sequence has been classified in one of the four groups based on the sequence and length of the two featured regions: the small domain (SD) and the loop at the interface (IL).

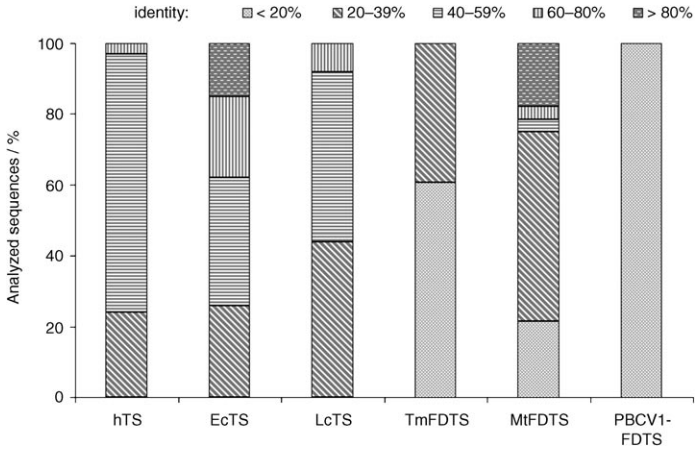

Figure 4. Distribution of percent identity (indicated) with respect to model sequences.

ficus). Because TSs from these pathogens show limited identity to hTS, the prospects are good for designing drugs that specifically inhibit these enzymes.

Eight of the 32 residues involved in substrate and cofactor binding are invariant in all the analyzed TS sequences; another 11 residues are invariant in $<90\%$ of the sequences (Table 4). 26% of the analyzed sequences have SDs and ILs as long as those in hTS; 37% of the analyzed sequences have SDs and ILs as long as those in EcTS; 8% of the analyzed sequences have SDs and ILs as long as those in LcTS. 29% of the analyzed sequences have insertions or deletions within the SD and IL (Table 3). In particular, the SD in *Acinetobacter baumannii* is shorter by three residues, and the IL is shorter by one residue with respect to hTS. *Bacillus amyloliquefaciens* has an SD and IL one residue longer than those of EcTS, while *Bacillus licheniformis* has an IL and SD that are respectively one and two residues longer than those of EcTS. The *Bordetella* spp. (*bronchiseptica*, *parapertussis*, *pertussis*) and the *Burkholderia* spp. (*ambifaria*, *dolosa*, *mallei*, *multivorans*, *pseudomallei*, *vietnamiensis*) have an IL region 26 residues longer than that of hTS. The SD of *Clostridium difficile* is two residues shorter than in EcTS. The SD in *Comamonas testosteroni* is one residue longer than the analogous region of EcTS, while *Encephalitozoon cuniculi* is one residue shorter in the SD region with respect to hTS. *Francisella tularensis* and *Fusobacterium nucleatum* show IL region deletions of 11 residues with respect to hTS. *Haemophilus* spp. (*ducreyi*, *influenzae*), *Pasteurella multocida*, and *Vibrio* spp. (*alginolyticus*, *cholerae*, *parahaemolyticus*, *vulnificus*) show an SD as long as that in hTS, but an IL region as long as those in EcTS and LcTS. *Mycoplasma* spp. (*genitalium*, *penetrans*, *pneumoniae*) show a 17-residue insertion in the SD region while lacking 13 residues in the IL region with respect to hTS. *Novosphingobium aromaticivorans* shows SD and IL regions that are 16 and two residues longer, respectively, than the corresponding sites in hTS. *Propionibacterium acnes* has an IL one residue shorter with respect to EcTS. *Streptococcus* spp. (*agalactiae*, *mutans*, *pneumoniae*, *pyogenes*, *sanguinis*) are characterized by an SD region three residues longer and an IL region one residue shorter than EcTS. These variations in the SDs and ILs featured structural regions present opportunities to selectively inhibit the TS activity in these pathogens over hTS.

In 40% of the analyzed sequences, neither insertions nor deletions are present with respect to the model sequences, while

Table 4. TS residues involved in substrate and cofactor binding, and their conservation within pathogen TS sequences.[a]

Residue[b]		Residue[b]	
R50	R (94%), D, E, K	R215	R (100%)
T51	T (92%), Y (5%), G, V	S216	S (100%)
K77	K (82%), R (13%), L (5%)	G217	C (42%), A (31%), G (15%), N (12%)
F80	H (36%), F (32%), A (9%), P (8%), Y (6%), L (5%), N, S	D218	D (100%)
E87	E (95%), K (5%)	L221	L (94%), V (6%)
I108	I (78%), V (11%), T (18%), F, S	G222	G (93%), A (6%), W
W109	W (95%), H, R, T	V223	V (57%), L (46%), I, T (6%), H (5%), N, S
N112	N (64%), W (36%), Y	F225	F (92%), I (5%), A, W
Y135	Y (99%), F	N226	N (100%)
R175	R (100%)	H256	H (100%)
R176	R (90%), G (10%)	Y258	Y (100%)
L192	L (99%), V	I307	I (84%), L (14%), V
P193	P (66%), A (13%), R (10%), L (6%), M, T, Y	K308	K (63%), S (11%), T (10%), Q (6%), A, D, H, L, P, R, Y
C195	C (100%)	M309	A (47%), M (29%), Y (11%), F, G (6%), I
H196	H (84%), M (9%), A (6%), L, V	M311	V (35%), M (34%), I (15%), F (10%), L (6%)
Q214	Q (93%), I (6%), A	A312	A (74%), S (25%), P

[a] As of May 31, 2007. [b] hTS sequence numbering.

50% of them contain insertion(s) and 20% of them have deletion(s) (Supporting Information). Deletions are present in or near the IL region (at residues 138–158, based on hTS numbering for all following descriptions), near the SD region (residues 98–99), in the monomer–monomer interface region (at residues 164–170, 182), in the α helix at the bottom of the active site (residues 223–224, 235), or in the surface region at the back of the active site (residues 279–286). Insertions are present at the N-terminal region, in the Arg 50 loop (residues 45–59), near and in the monomer–monomer interface region (at residues 61–71, 202–209), near the SD (from residues 90–98), near the catalytic site (residues 189–191), in the α helix at the bottom of the active site (residues 226–230), on the surface (at residues 264–267, 274–275, 279–295), and at the C-terminal region (residues 302–308) (Figure 2b and c)

1.1. Sequences with SD and IL of hTS

In this section we give an example of a sequence/structure-based study in the area of protozoan enzymes in comparison with the human enzyme. There are 30 pathogens similar to hTS (percent identity: 50–62; Table 3). Among them, 14 protozoans have a DHFR–TS bifunctional enzyme. All these TS sequences show SD regions of 36 residues (SD percent identity: 47–75) and the eight-residue insertion in the IL with respect to the EcTS and LcTS sequences (IL percent identity: 44–75; see Supporting Information).

The group of pathogens with DHFR–TS includes organisms from five genera: *Cryptosporidium*, *Leishmania*, *Plasmodium*, *Toxoplasma*, and *Trypanosoma*. Only the TS domains of DHFR–TS enzymes were considered in this analysis. The TSs of this group show identity percentages in the range of 51–59% relative to hTS (Supporting Information). In our view, the structural differences between the pathogen and human enzymes, as defined by variability in residues and by differences in protein flexibility, can be exploited to identify specific inhibitors. However, the protozoan pathogens mentioned above are examples in which the evaluation of the differences in flexibility between the pathogen proteins and hTS is not a useful analytical tool. In fact, these three enzymes belong to the same group as hTS. Therefore, the SD and IL have the same length, and the flexibility is expected to be the same. This result was confirmed in previous studies with LmTS and PcTS, which belong to the same hTS group.[19]

The analysis of the residues in the active site (within 5 Å from the bound substrate/ligand, defined on the basis of the X-ray ternary complex structure: PDB code 1HVY) shows that some are not conserved and may be exploitable to specifically inhibit the pathogen enzymes. In *Cryptosporidium* spp. only three (F80A, G217C, and V223S) out of 32 active site residues are not conserved (Figure 3b). In *Leishmania* spp., G217 is mutated to C (Figure 3b) while in *Leishmania tropica* L192 is also replaced by a V residue. In *Plasmodium* spp., only G217 is mutated to C (Figure 3b); two other mutations (P193L and M311L) with respect to hTS active site residues are present only in *Plasmodium vivax*. In *Toxoplasma gondii* G217 is mutated to C, and K308 is mutated to Q. In *Trypanosoma* spp., G217 is replaced by C, and K308 is mutated to S; moreover, in *Trypanosoma cruzi*, G222 is switched to A. Thus, it is evident that in DHFR–TS enzymes, the G217C substitution is common among the pathogens. Analyses of the available crystallographic structures of TS from *Cryptosporidium hominis*, *Leishmania major*, and *Plasmodium falciparum* in relation to hTS[44,48–50] allow a deeper understanding of the role that these variable residues may play in the development of species-specific antiparasitic drugs (Figure 2d). In particular, F80 is involved in the quadrupole–dipole interaction with the glutamic tail of the cofactor; the mutation F80A in *Cryptosporidium hominis* TS may suggest that bulkier or less polar moieties may replace the glutamic tail of MTHF in species-specific drugs. There are additional regions with insertions or deletions exploitable for species-specific drug targeting (Figures 2e and 3b). Among these sequences there are only short (1–2 amino acids long) insertions in two regions: near or in the monomer–monomer interface region,

or on the surface. An example of this principle applied to *Cryptosporidium hominis*, *Leishmania major*, and *Plasmodium falciparum* is given.

Among the other TSs showing SD and IL regions as long as those of hTS, *Legionella pneumophila* and *Nocardia farcinica* TSs show lower percent identities to hTS (<45%; see Supporting Information) and could be selectively targeted by drugs designed ad hoc. In particular, *Legionella pneumophila* TS shows six mutations among the active site residues: F80H, N112W, P193M, G217A, K308M, and M311I; *Nocardia farcinica* TS also has six residues in the active site that are not conserved with respect to hTS: F80H, N112W, P193A, G217A, M309A, and M311V.

All the other organisms bearing TSs with SD and IL regions as long as those of hTS show percent identities in range of 50–62%. To elucidate how these TSs can be inhibited in a species-specific manner, a detailed residue-based analysis of all the differences in and near the active site and of all the insertions and deletions of these TS sequences should be carried out. However, a full in-depth analysis of all the available TS sequences is beyond the scope of the present work.

1.2. Sequences with SD and IL of EcTS

Forty-four pathogens have TSs similar to EcTS (Table 3). All these sequences show SD regions of 24 residues (SD percent identity: 50–100) and ILs of eight residues, as that of EcTS (HIL percent identity: 38–100; see Supporting Information). They show an identity percentage versus EcTS that varies from 57 to 99%. In particular, *Aeromonas hydrophila*, *Enterobacter* sp., *Salmonella* spp. (*enterica*, *paratyphi*, *typhi*, *typhimurium*), *Serratia proteamaculans*, *Shigella* spp. (*boydii*, *dysenteriae*, *flexneri*, *sonnei*), and *Yersinia* spp. (*bercovieri*, *enterocolitica*, *frederiksenii*, *intermedia*, *mollaretii*, *pestis*, *peseudotuberculosis*) show an identity percentage higher than 84%.

All these sequences show mutations, with respect to hTS, in residues involved in substrate and cofactor binding. These mutations are: F80H/Y, N112W, P193A/H, G217C/A, V223L/I, K308A/R, M309A/G, M311V/I, and A312S.

1.3. Sequences with SD and IL of LcTS

Ten pathogens of the genera *Bacillus*, *Enterococcus*, *Listeria*, and *Staphylococcus* have TSs similar to LcTS (Table 3). All of these sequences show an SD region of 74 residues (SD percent identity: 61–69) as well as an IL of eight residues, as that of LcTS (HIL percent identity: 13–63; see Supporting Information). There are no gaps and no relevant insertions. They show an identity percentage versus LcTS that varies from 50 to 69%. F80P/S, N112W, G217A, V223I, M309A, M311I/V, and A312S are the mutations that involve residues in the substrate and cofactor binding sites.

2. FDTS Sequences

From the analysis of the available crystal structures,[28,34,41] the FDTS structure consists of 10 α helices and five β strands. Two main regions are present: 1) an outer region, formed by the random-coil region and helices α5, α6, and α7 (residues 87–154 in TmFDTS, residues 104–179 in MtFDTS, residues 87–162 in Pbcv1FDTS), and 2) a main core, formed by the remaining portions (Figures 2 f, g, and 5). The outer region, as suggested from the sequence alignment, is not a highly conserved region of the protein.[41] A comparison of these three structures reveals that there are three other regions (residues 17–19, 28–43, and 67–68 in TmFDTS) in which there are different insertions. MtFDTS has 15-, 3-, and 1-residue insertions in the three cited regions, respectively. Pbcv1FDTS has an eight-residue insertion in the first cited region.

Several organisms of the genera *Bacillus* (*anthracis*, *cereus*, *weihenstephanensis*), *Clostridium* (*difficile*), *Corynebacterium* (*diphtheriae*, *jeikeium*), and *Mycobacterium* (*avium*, *bovis*, *flavescens*, *leprae*, *tuberculosis*, *ulcerans*) possess both TS and FDTS (Table 1). Moreover, 34 other organisms possess only FDTS (Table 2).

FDTS primary sequences of pharmaceutically interesting organisms were compared with TmFDTS, MtFDTS, and Pbcv1FDTS. The results show that with only few exceptions, the percent identity varies from 11 to 33%. This range is particularly low, suggesting that FDTS is a less highly conserved enzyme than TS. Only FDTS from *Corynebacterium* spp. (*diphtheriae*, *jeikeium*) show 53–60% identity toward MtFDTS. In contrast, FDTS from *Chlamydia trachomatis* and *Chlamydophila pneumoniae* show no more than 10% identity toward the three FDTS sequences considered.

Among the 36 residues known to be involved in substrate/cofactor binding, only four of them (R78, H79, R81, and R174; TmFDTS sequence numbering) are invariant, whereas seven other residues (H53, F64, S88, R90, Y91, E182, and A187) are invariant in all the sequences analyzed (Table 5).

Summary and Outlook

A knowledge-based grouping of the TS protein family is presented herein. TS and FDTS, included in the family, represent key enzymes in cell replication. The former is a well-known anticancer drug target, the latter is present specifically in pathogenic organisms and heralds a new approach in anti-infective chemotherapy targeting. They are structurally very different and deserve separate considerations. The IL and SD regions of TS proteins have been identified from previous studies as essential for protein flexibility and specificity in the interaction with drugs. The TS enzymes are grouped into three ensembles based on similarities of the SD and IL regions of hTS, EcTS, and LcTS. Structure-based discovery of specific inhibitors requires the identification of putative drug target regions on TSs from pathogenic organisms. This work provides an updated structural study of the TS protein family and gives essential information for a detailed and accurate analysis of the 3D structural differences between pathogen TS and hTS. This study should prove useful to guide sub-domain drug targeting.

The authors currently carry out regular updates of the TS/FDTS sequence collection, which is available, together with all

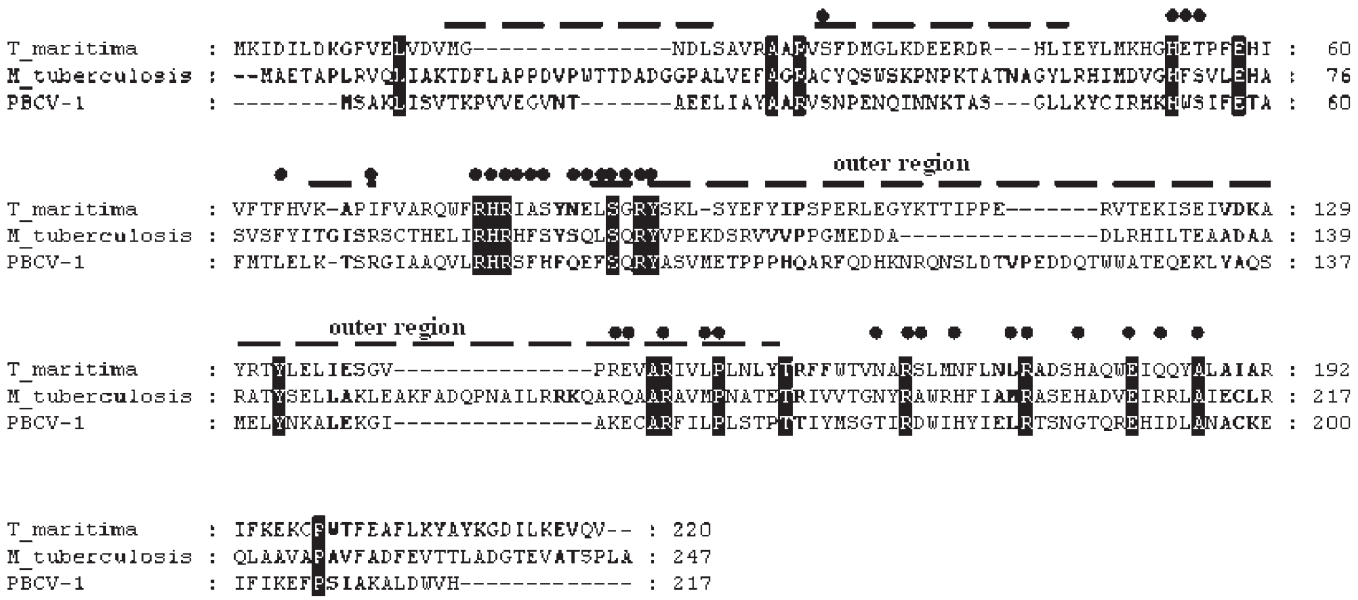

Figure 5. *Thermotoga maritima*, *Mycobacterium tuberculosis*, and *PBCV-1* FDTS sequence alignment: identical residues among these three enzymes are highlighted in dark boxes. The outer region and other species-specific insertions are indicated with dashed lines; known active site residues are indicated (●).

Table 5. FDTS residues involved in FAD and dUMP binding, and their conservation within pathogen FDTS sequences.[a]

Residue[b]		Residue[b]	
S30	C (48%), S (40%), Y (6%), L (5%)	G89	Q (32%), A (26%), T (16%), G (11%), S (10%), L (5%)
H53	H (93%), L (5%)	R90	R (92%), L (5%), C
E54	T (37%), A (19%), E, F, S (10%), G, M (5%), R, V, W	Y91	Y (93%), F, S, T
T55	S (61%), T (27%), A, C (5%), F	R143	R (48%), L (16%), I (10%), N (8%), A, K (6%), E (5%)
F64	F (92), L, M, V, Y	E144	E (61%), D (24%), Q (14%)
I70	R (56%), I (39%), F (5%)	R147	R (69%), K (24%), Q(5%), H
R74	R (43%), H, Q (24%), A (5%), T	L150	L (73%), F, M (10%), V (6%)
Q75	Q (53%), E (42%), D (5%)	P151	P (79%), T (21%)
R78	R (100%)	N163	N (82%), D (16%), T
H79	H (100%)	R165	R (66%), H (29%), N (5%)
R80	R (100%)	S166	S (39%), N (34%), A (14%), G (10%), D, T
I81	T (35%), I (27%), H (24%), G (6%), M, S, V	N169	H (47%), N (39%), D (6%), W (5%), R
A82	A (61%), F (23%), T (8%), L (5%), G, S	L173	L (66%), M (13%), H (10%), R (6%), V
S83	S (79%), R (8%), N (6%), T (5%), H	R174	R (100%)
N85	S (56%), N (37%), Q (6%)	H178	H (37%), S (14%), G, K (11%), R (10%) N, Q (5%), A, T, Y
E86	E (45%), V (31%), Q (24%)	E182	E (95%), N, S
L87	L (32%), Y (24%), Q (16%), K (5%), V (6%), R (5%), E, F, I, S	Q184	R (74%), Q (13%), A (6%), E, H, I
S88	S (95), Q (5%)	A187	A (97%), S

[a] As of May 31, 2007. [b] TmFDTS sequence numbering.

the alignments, at http://cdm.unimo.it/home/dipfarm/costi.mariapaola/TS-FDTS.htm.

Keywords: drug discovery · drug targets · sequence analysis · specificity · thymidylate synthases

Acknowledgements

This work was supported by MPC FIRB RBAU01S38Z and MPC WHO WHOA50599. The authors thank Dr. D. A. Matthews for the crystal structure of LmTS, and Dr. A. Venturelli for a critical reading of the manuscript. The authors thank the European FP6 project LIGHTS (http://www.lights-eu.org) for supporting this work in the hTS structural study.

[1] B. J. Druker, *Trends Mol. Med.* **2002**, *8(Suppl. 4)*, S14–8.
[2] B. Fingleton, *Curr. Pharm. Des.* **2007**, *13*, 333–346.
[3] H. R. Saibil, *Nat. Struct. Biol.* **2000**, *7*, 711–714.
[4] A. Engel, D. J. Mueller, *Nat. Struct. Biol.* **2000**, *7*, 715–718.
[5] T. E. Fisher, P. E. Marszalek, J. M. Fernandez, *Nat. Struct. Biol.* **2000**, *7*, 719–724.
[6] S. Weiss, *Nat. Struct. Biol.* **2000**, *7*, 724–729.
[7] P. R. Selvin, *Nat. Struct. Biol.* **2000**, *7*, 730–734.
[8] W. L. Hubbell, D. S. Cafiso, C. Altenbach, *Nat. Struct. Biol.* **2000**, *7*, 735–739.
[9] R. Ishima, D. A. Torchia, *Nat. Struct. Biol.* **2000**, *7*, 740–743.

[10] G. W. Daughdrill, P. Narayanaswami, S. H. Gilmore, A. Belczyk, C. J. Brown, *J. Mol. Evol.* **2007**, *65*, 277–288.
[11] E. Chu, M. A. Callender, M. P. Farrell, J. C. Schmitz, *Cancer Chemother. Pharmacol.* **2003**, *52(Suppl. 1)*, 80–9.
[12] M. P. Costi, D. Tondi, M. Rinaldi, D. Barlocco, P. Pecorari, F. Soragni, A. Venturelli, R. M. Stroud, *Biochim. Biophys. Acta Mol. Basis Dis.* **2002**, *1587*, 206–214.
[13] D. Leduc, F. Escartin, H. F. Nijhout, M. C. Reed, U. Liebl, S. Skouloubris, H. Myllykallio, *J. Bacteriol.* **2007**, *189*, 8537–8545.
[14] J. L. Tonkinson, P. Marder, S. L. Andis, R. M. Schultz, L. S. Gosset, C. Shih, L. G. Mendelsohn, *Cancer Chemother. Pharmacol.* **1997**, *39*, 521–531.
[15] R. D. Ladner, *Curr. Protein Pept. Sci.* **2001**, *2*, 361–370.
[16] S. J. Welsh, S. Hobbs, G. W. Aherne, *Eur. J. Cancer* **2003**, *39*, 378–387.
[17] M. P. Costi, M. Rinaldi, D. Tondi, P. Pecorari, D. Barlocco, S. Ghelli, R. M. Stroud, D. V. Santi, T. J. Stout, C. Musiu, E. M. Marangiu, A. Pani, D. Congiu, G. A. Loi, P. La Colla, *J. Med. Chem.* **1999**, *42*, 2112–2124.
[18] T. J. Stout, D. Tondi, M. Rinaldi, D. Barlocco, P. Pecorari, D. V. Santi, I. D. Kuntz, R. M. Stroud, B. K. Shoichet, M. P. Costi, *Biochemistry* **1999**, *38*, 1607–1617.
[19] S. Ferrari, M. P. Costi, R. C. Wade, *Chem. Biol.* **2003**, *10*, 1183–1193.
[20] D. Tondi, U. Slomczynska, M. P. Costi, D. M. Watterson, S. Ghelli, B. K. Shoichet, *Chem. Biol.* **1999**, *6*, 319–331.
[21] D. Tondi, A. Venturelli, S. Ferrari, S. Ghelli, M. P. Costi, *J. Med. Chem.* **2005**, *48*, 913–916.
[22] M. P. Costi, A. Gelain, D. Barlocco, S. Ghelli, F. Soragni, F. Reniero, T. Rossi, A. Ruberto, C. Guillou, A. Cavazzuti, C. Casolari, S. Ferrari, *J. Med. Chem.* **2006**, *49*, 5958–5968.
[23] J. L. Dynes, R. A. Firtel, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 7966–7970.
[24] A. Chernyshev, T. Fleischmann, A. Kohen, *Appl. Microbiol. Biotechnol.* **2007**, *74*, 282–289.
[25] J. Zhong, S. Skouloubris, Q. Dai, H. Myllykallio, A. G. Barbour, *J. Bacteriol.* **2006**, *188*, 909–918.
[26] P. Sampathkumar, S. Turley, C. H. Sibley, W. G. J. Hol, *J. Mol. Biol.* **2006**, *360*, 1–6.
[27] J. Griffin, C. Roshick, E. Iliffe-Lee, G. McClarty, *J. Biol. Chem.* **2004**, *280*, 5456–5467.
[28] P. Sampathkumar, S. Turley, J. E. Ulmer, H. G. Rhie, C. H. Sibley, W. G. J. Hol, *J. Mol. Biol.* **2005**, *352*, 1091–1104.
[29] D. Leduc, S. Graziani, G. Lipowski, C. Marchand, P. Le Marechal, U. Liebl, H. Myllykallio, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7252–7257.
[30] M. Y. Galperin, E. V. Koonin, *Nat. Biotechnol.* **2000**, *18*, 609–613.
[31] H. Myllykallio, G. Lipowski, D. Leduc, J. Filee, P. Forterre, U. Liebl, *Science* **2002**, *297*, 105–107.
[32] A. Chernyshev, T. Fleischmann, E. M. Kohen, S. A. Lesley, A. Kohen, *Chem. Commun.* **2007**, 2861–2863.
[33] A. Mason, N. Agrawal, M. T. Washington, S. A. Lesley, A. Kohen, *Chem. Commun.* **2006**, 1781–1783.
[34] S. Graziani, J. Bernauer, S. Skouloubris, M. Graille, C-Z Zhou, C. Marchand, P. Decottignies, H. van Tilbeurgh, H. Myllykallio, U. Liebl, *J. Biol. Chem.* **2006**, *281*, 24048–24057.
[35] N. Agrawal, S. A. Lesley, P. Kuhn, A. Kohen, *Biochemistry* **2004**, *43*, 10295–10301.
[36] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
[37] C. Combet, C. Blanchet, C. Geourjon, G. Deléage, *Trends Biochem. Sci.* **2000**, *25*, 147–150.
[38] *GeneDoc: Analysis and Visualization of Genetic Variation*: K. B. Nicholas, H. B. Nicholas, Jr., D. W. Deerfield II, *EMBNET News* **1997**, *4*, 14; http://www.psc.edu/biomed/genedoc (accessed January 28, 2008).
[39] S. A. Lesley, P. Kuhn, A. Godzik, A. M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H. E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L. S. Brinnen, M. D. Miller, T. M. McPhilips, A. M. Miller, D. Scheibe, J. M. Canaves, C. Guda, L. Jaroszewski, T. L. Selby, M. A. Elsliger, J. Woley, S. S. Taylor, K. O. Hodgson, I. A. Wilson, P. G. Schultz, R. C. Stevens, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11664–11669.
[40] P. Kuhn, S. A. Lesley, I. I. Mathews, J. M. Canaves, L. S. Brinen, X. Dai, A. M. Deacon, M. A. Elsliger, S. Eshaghi, R. Floyd, A. Godzik, C. Grittini, S. K. Grzechnik, C. Guda, K. O. Hodgson, L. Jaroszewski, C. Karlak, H. E. Klock, E. Koesema, J. M. Kovarik, A. T. Kreusch, D. McMullan, T. M. McPhillips, M. A. Miller, M. Miller, A. Morse, K, Moy, J. Ouyang, A. Robb, K. Rodrigues, T. L. Selby, G. Spraggon, R. C. Sevens, S. S. Taylor, H. van der Beden, J. Velasquez, J. Vincent, X. Wang, B. West, G. Wolf, J. Wooley, I. A. Wilson, *Proteins Struct. Funct. Genet.* **2002**, *49*, 142–145.
[41] I. I. Mathews, A. M. Deacon, J. M. Canaves, D. McMullan, S. A. Lesley, S. Agarwalla, P. Kuhn, *Structure* **2003**, *11*, 677–690.
[42] M. P. Costi, S. Ferrari, A. Venturelli, S. Calò, D. Tondi, D. Barlocco, *Curr. Med. Chem.* **2005**, *12*, 2241–2258.
[43] L. W. Hardy, J. S. Finer-Moore, W. R. Montfort, M. O. Jones, D. V. Santi, R. M. Stroud, *Science* **1987**, *235*, 448–455.
[44] J. Phan, S. Koli, W. Minor, R. B. Dunlap, S. H. Berger, L. Lobioda, *Biochemistry* **2001**, *40*, 1897–1902.
[45] K. M. Perry, E. B. Fauman, J. S. Finer-Moore, W. R. Montfort, G. F. Maley, F. Maley, R. M. Stroud, *Proteins Struct. Funct. Genet.* **1990**, *8*, 315–333.
[46] R. M. Stroud, J. S. Finer-Moore, *Biochemistry* **2003**, *42*, 239–247.
[47] J. S. Finer-Moore, D. V. Santi, R. M. Stroud, *Biochemistry* **2003**, *42*, 248–256.
[48] A. C. Anderson, *Acta Crystallogr. Sect. F* **2005**, *61*, 258–262.
[49] J. Yuvaniyama, P. Chitnumsub, S. Kamchonwongpaisan, J. Vanichtanankul, W. Sirawaraporn, P. Taylor, M. S. Walkinshaw, Y. Yuthavong, *Nat. Struct. Biol.* **2003**, *10*, 357–365.
[50] D. R. Knighton, C. C. Kan, E. Howland, C. A. Janson, Z. Hostomska, K. M. Welsh, D. A. Matthees, *Nat. Struct. Biol.* **1994**, *1*, 186–194.

Received: August 19, 2007
Revised: January 16, 2008
Published online on February 12, 2008

