Inhibitor Specificity via Protein Dynamics: Insights from the Design of Antibacterial Agents Targeted Against Thymidylate Synthase

Stefania Ferrari,[1,2,*] Paola M. Costi,[1]
and Rebecca C. Wade[2,3]
[1]Dipartimento di Scienze Farmaceutiche
Università di Modena e Reggio Emilia
Via Campi 183
41100 Modena
Italy
[2]European Molecular Biology Laboratory
Meyerhofstr 1
D-69112 Heidelberg
[3]EML gGBMH
Villa Bosch, Schloss-Wolfsbrunnenweg 33
D-69118 Heidelberg
Germany

Summary

Structure-based drug design of species-specific inhibitors generally exploits structural differences in proteins from different organisms. Here, we demonstrate how achieving specificity can be aided by targeting differences in the dynamics of proteins. Thymidylate synthase (TS) is a good target for anticancer agents and a potential target for antibacterial agents. Most inhibitors are folate-analogs that bind at the folate binding site and are not species specific. In contrast, α156 is not a folate-analog and is specific for bacterial TS; it has been shown crystallographically to bind in a nonconserved binding site. Docking calculations and crystal structure-based estimation of the essential dynamics of TSs from five different species show that differences in the dynamics of TSs make the active site more accessible to α156 in the prokaryotic than in the eukaryotic TSs and thereby enhance the specificity of α156.

Introduction

The structure-based drug design paradigm is applicable to the discovery of new molecules that inhibit the biological functions of selected macromolecular targets, whose 3D structures are known. The technique has been successful, resulting in clinically useful drugs [1]. However, there are several bottlenecks to its wider application. One of these is the treatment of the flexibility of the receptor macromolecule. Exploration of the conformational space of a macromolecule in drug design is often limited to those conformations whose 3D structures have been determined experimentally, and conformational changes occurring upon ligand binding are frequently partially or fully neglected.

The experimentally determined conformation of a protein often differs in its ligand-bound and -unbound forms. Protein conformational changes can open new ligand binding sites that are difficult to predict a priori

[2–4]. Hydrophobic regions, complementary to drug molecules, which tend to be more hydrophobic than natural protein ligands, can become exposed to the ligand upon binding [5]. The change in conformation of a protein upon ligand binding can be considered as induced fit or as the selection, from a range of preexisting protein conformations around the native state, of a conformation that is complementary to the incoming ligand [6]. An excellent example of these phenomena can be found in antibody binding kinetics, recently investigated by James and coworkers [7]. The necessity of treating proteins not as single experimentally determined structures but as dynamic entities in equilibrium among different conformational states therefore arises [8–10]. Then, if each conformational state can be considered to have a different affinity for the ligand, the experimentally measured affinity is given by a suitably weighted average of the binding energy to different conformations. Binding to a given protein conformation depends not only on the conformation's binding energy but also on its accessibility to the ligand. Protein motions can result in conformational gating, which can affect the specificity and kinetics of binding [11, 12]. In acetylcholinesterase, for example, conformational gating modulates access of the substrate to the active site and is important for enzyme specificity and efficiency [11]. Conformational gating may provide geometric selectivity, e.g., by preventing binding of larger ligands and permitting binding of smaller ligands. Conformational gating also affects the rate of ligand binding and does so in a manner that depends on the relative timescales of the protein gating motions and the process of ligand binding [13].

Treating protein dynamics is a challenge in the comparative analysis of isozymes. Often, isozymes have high sequence homology and a highly conserved structural fold, especially in the catalytic domain. But they can have different molecular properties, including different dynamics that affect binding and catalysis. To date, most structure-based drug design studies aimed at designing species-specific ligands have made use of differential interaction models that consider solely the structure/sequence differences among isozymes [14, 15]. However, the shortcoming of this approach was highlighted recently by the observation of Agard et al. [16] that differences in protein dynamics between the wild-type and a single point mutant of α-lytic protease could explain the different spectrum of ligands hydrolyzed by these enzymes.

In the present paper, we examine the molecular basis of species-specific thymidylate synthase (TS) inhibition by phenolnaphthalein derivatives. TS is an essential enzyme with a fundamental role in DNA synthesis. TS catalyzes the methylation of 2′-deoxyuridine-5′-monophosphate (dUMP) to form 2′-deoxythymidine-5′-monophosphate (dTMP), with a molecule of N^5,N^{10}-methylene-tetrahydrofolate (mTHF) as a cofactor. Crystal structures solved in the presence of reaction intermediate analogs show that structural changes that occur throughout the

*Correspondence: ferrari.stefania@unimore.it

protein are integral to the enzymatic reaction mechanism [17].

dUMP methylation is the final step in the sole de novo biosynthetic pathway of dTMP. The blockage of this process leads to cell death. TS is present in almost all living organisms, is very conserved among different species, and is highly specific in its catalytic function [18]. TS inhibitors can interfere with either the dUMP-enzyme or the mTHF-enzyme interaction. Because of its essential role in cell replication, TS has been considered a good target for anticancer therapy [4]. It is only recently that researchers have started to study TS as an antibacterial drug target [19]. To use TS inhibitors in antibacterial therapy, the inhibitors should have species-specific activity profiles, i.e., they should inhibit bacterial TS without affecting human TS (hTS).

Recently, Stout et al. [19] discovered a new phenolnaphthalein compound (α156) that inhibits *Lactobacillus casei* TS (LcTS) with a K_i value 43 times lower than that against hTS ($K_i = 0.7$ and 30 μM, respectively) (Table 1). The determination of the crystal structure of the binary LcTS-α156 complex [19] revealed that this compound binds in a nonconserved region of the enzyme active site more than 5 Å from the binding site of folate-analog compounds. It interacts with the "small domain" (SD) (Figure 1), which consists of a sequence of 70 residues in LcTS, but is reduced to 32 residues in hTS [20]. The first hypothesis formulated to explain α156 activity data relied on the structural difference in the SD region between LcTS and hTS. That is, it was proposed that α156 is species specific because it binds in a species-specific site of the enzyme where it exploits structural differences in the SD between hTS and LcTS. Further investigation of the inhibition activity data profile of α156 [21] surprisingly revealed that this compound shows the same affinity toward TSs from *Escherichia coli* (EcTS) ($K_i = 0.6$ μM) as from Lc even though the SD region of EcTS is reduced to only 20 residues. This observation prompted these questions: how can α156 discriminate among different TSs and which relations exist between its binding mode and its inhibition activity profile? To address these issues, we deemed it necessary to consider not only static structural differences between TSs of different species but also how structural differences can influence protein dynamics and flexibility and how these enzyme features can affect ligand binding, and in particular the α156 binding site conformation.

In this paper, we report on the evaluation, by experiment and computation, of the inhibition activity profile of α156 and five other phenolnaphthalein derivatives, as well as the nonspecific folate-analog inhibitor **CB3717** [22], toward five different TS isozymes (LcTS, EcTS, hTS, *Pneumocystis carinii* TS [PcTS], and *Leishmania major* TS [LmTS]). The inhibitor binding sites, the protein conformational changes upon ligand binding, and the structural differences among TSs from different species were studied by crystal structure comparison and molecular docking. Because differences in sequence and structure, as deduced from crystal structure comparison alone, were not sufficient to explain the specific activity of α156, we analyzed the structural flexibility of TSs from different species using CONCOORD-essential dynamics [23, 24] analysis, focusing on how differences in the

dynamics of TSs from different species could affect α156 binding and consequently provide an explanation for its species specificity.

Results and Discussion

Inhibition Activity Profiles

Enzymatic inhibition assays against TS from two bacterial species (LcTS, EcTS) and three nonbacterial species (hTS, PcTS, and LmTS) were carried out for six phenolnaphthalein derivatives as well as for the folate-analog **CB3717** (see Experimental Procedures). The inhibition activity profiles reported in Table 1 show that all the phenolnaphthalein derivatives have species-specific activity toward bacterial species, whereas they show much weaker or undetectable inhibition against eukaryotic species (human, Pc, and Lm). The specificity index between LcTS and hTS varies from 20 to over 306 for the most specific compound (2). For PcTS, the specificity index ranges from 12 to over 251 (1), and for LmTS it ranges from 10 to over 1546 (α156).

In contrast, the folate-analog **CB3717** shows a nonspecies-specific inhibition activity profile, inhibiting the TSs of all species tested at approximately the same level ($K_i = 0.03$–0.24 μM).

The species-specific activity profile of these phenolnaphtalein compounds can be understood through the analysis of their binding site in TS enzymes.
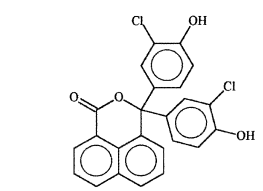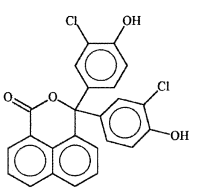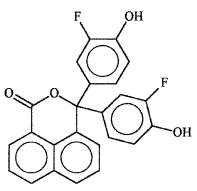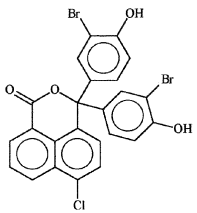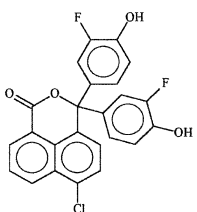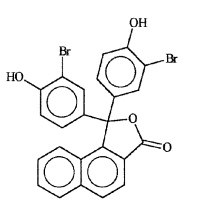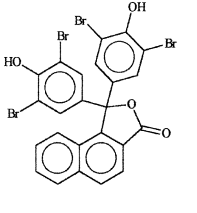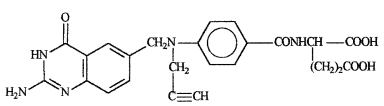
Binding Sites of Inhibitors

The crystal structures of the LcTS-α156 [19] and LcTS-dUMP-**CB3717** [22] complexes show that α156 binds in a poorly conserved site that is distinct from the substrate (folate) binding site in which **CB3717** binds (Figure 1). The interaction with less conserved subsites of the enzyme, such as the SD region, allows α156 to exploit the differences between TSs from different species.

Crystal structure analysis of hTS, LcTS, EcTS, PcTS, and LmTS [19, 22, 26–32] (see Table S1 in the Supplemental Data available online at http://www.chembiol.com/cgi/content/full/10/12/1183/DC1) reveals the greater similarity of the ternary folds of hTS, PcTS, and LmTS to each other than to the bacterial TSs (LcTS and EcTS). These two groups are distinguished most prominently in the SD and the dimer interface loop (HIL) (Figure 1).

On the other hand, it might be expected that the α156 binding sites in LcTS and EcTS share common features that result in their similar K_i values for the phenolnaphthalein derivatives. However, the large SD that is present in LcTS compared to the smaller SD in eukaryotic TSs is apparently not the only critical factor for binding, since EcTS lacks the majority of the SD. The variation in the SD suggests that, beyond structural and sequence features of TS in different species, there should be another key feature that differentiates eukaryotic TSs from bacterial TSs and is exploited by phenolnaphthalein compounds in their interaction with the enzyme.

An important aspect to consider is the changes in the enzyme conformation induced by ligand binding. Comparison of the LcTS-α156 binary complex and the

Table 1. Inhibition Activity Profile of Six Selective Phenolnaphthalein Derivatives and of the Nonselective Folate-Analog CB3717 against TS Enzymes from Five Different Species

		K_i (μM)				
		LcTS	EcTS	hTS	PcTS	LmTS
α156		0.7[a]	0.6[b]	30[a]	13	>1082[c]
1		1.6[a]	1.1	>402[c]	>402[c]	>1775[c]
2		0.8	4.1	>245[c]	28	>1082[c]
3		1.8	1.1	>245[c]	22	>1082[c]
4		6.6	0.9	>132[c]	>132[c]	63[d]
5		1.7	8.5	>245[c]	32.7[d]	41[d]
CB3717		0.06	0.06	0.03	0.24	0.08

[a] [25].
[b] [21].
[c] Calculated assuming 2% inhibition at the maximum solubility concentration (27 μM for compound 4; 50 μM for compounds α156, 2, 3, and 5; and 82 μM for compound 1).
[d] Calculated from the inhibition percentage measured at the maximum solubility concentration (16% and 35%, respectively, for compounds 4 and 5, versus LmTS; 13% for compound 5 versus PcTS).
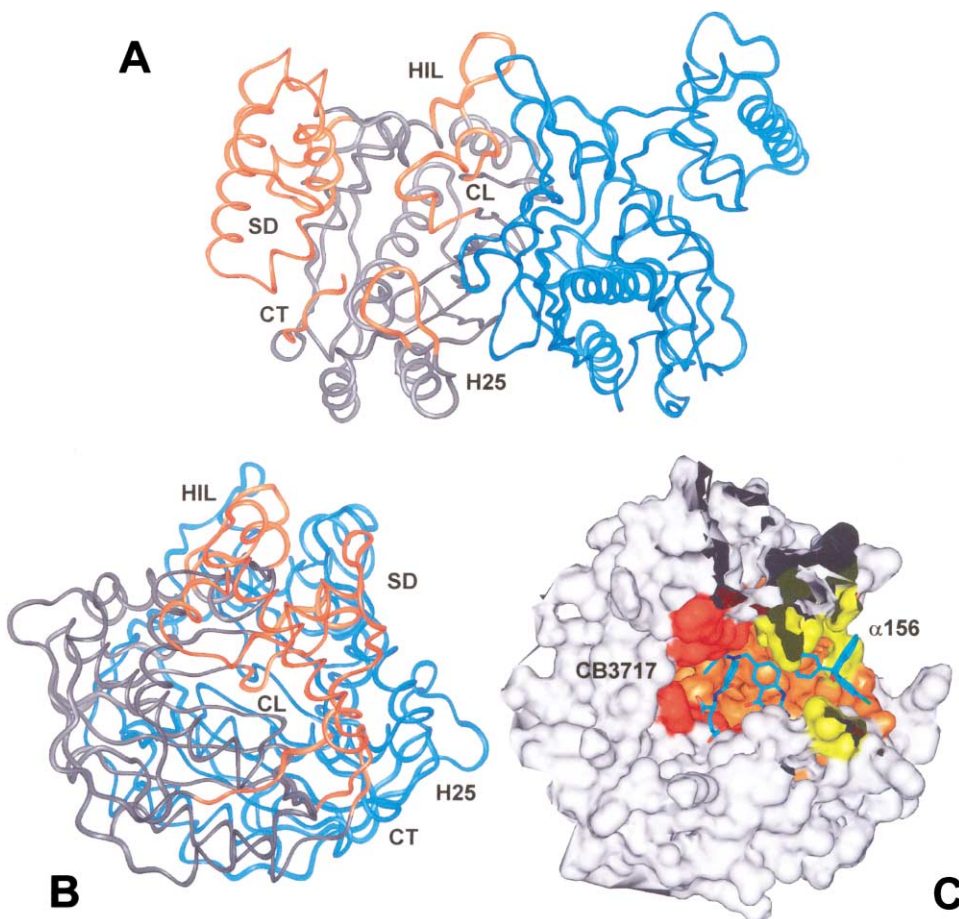
Figure 1. LcTS Structure and Inhibitors' Binding Sites

LcTS dimer with one monomer in gray and the other in light blue ([A] and [B], coordinates from PDB files 1TSL); in the first monomer, important regions (loop around H25 [H25], small domain [SD], dimer interface loop [HIL], catalytic loop [CL], and C terminus [CT]) are colored in orange. In (B), the enzyme is rotated about the vertical axis by ca. 90° compared with its orientation in (A). The Connolly surface of LcTS (C) is shown with **CB3717** (folate-analog, cyan; PDB identifer, 1LCA) and α**156** (cyan; PDB identifier, 1TSL) binding sites. The active site surface (orange, 10 Å around the dUMP binding site), the folate and folate-analog binding site (red, 4 Å around folate binding site), and the α**156** binding site (yellow, 4 Å around α**156** binding site) are shown. The SD region has been removed to make the complete active site visible. The enzyme orientation is the same as in (B).

apo-enzyme structures [19, 22, 26] reveals that α**156** does not induce significant conformational changes in the enzyme upon binding (the structures can be superimposed with a Cα rmsd of 0.35 Å). In contrast, when the ternary complex with dUMP and a folate-analog inhibitor is formed, the enzyme undergoes major conformational changes [22, 28, 29]. These involve the C terminus (CT), the loop around H25 (H25), and the SD (Figure 1) [22]. This suggests that these flexible regions can move most easily to accommodate the binding of different ligands. These regions constitute the binding site of α**156** (Figure 1), and therefore, this site could be affected by conformational changes of the enzyme.

Comparison of the LcTS-α**156** binary complex with the LcTS-**CB3717**-dUMP ternary complex [19, 22] shows that the conformations of R23, H106, and V316 affect the binding of α**156**. In the ternary noncovalent complex, all these residues are rotated so that the crystallographic binding site of α**156** is not accessible. Be-

cause the α**156** binding site is not accessible, we consider the noncovalent ternary complex of LcTS [22] to be a closed form of TS, even though the closed conformation of TS is usually defined as the covalent ternary complex of TS in the literature [2]. The α**156** crystallographic binding site is also not accessible in the structures of the hTS ternary complex deposited in the Protein Data Bank [33] until now. This is due to R50 and T51 moving toward V311 and F117 and other residues (A111, N112, S120, and L121) restricting the binding site of α**156**. This is true also for the structure 1I00 [30], a noncovalent ternary structure of hTS. Comparison of the ternary complex structures, with dUMP and a folate-analog inhibitor (such as **CB3717** or **ZD1694**), of the TS enzymes from different species (Lc, Ec, Pc, Lm, and human [22, 28, 29, 31, 32]) shows that the conformation adopted by the enzyme is mostly conserved. However, the degree of closure could be different for the different species and could be the key to differentiating eukary-
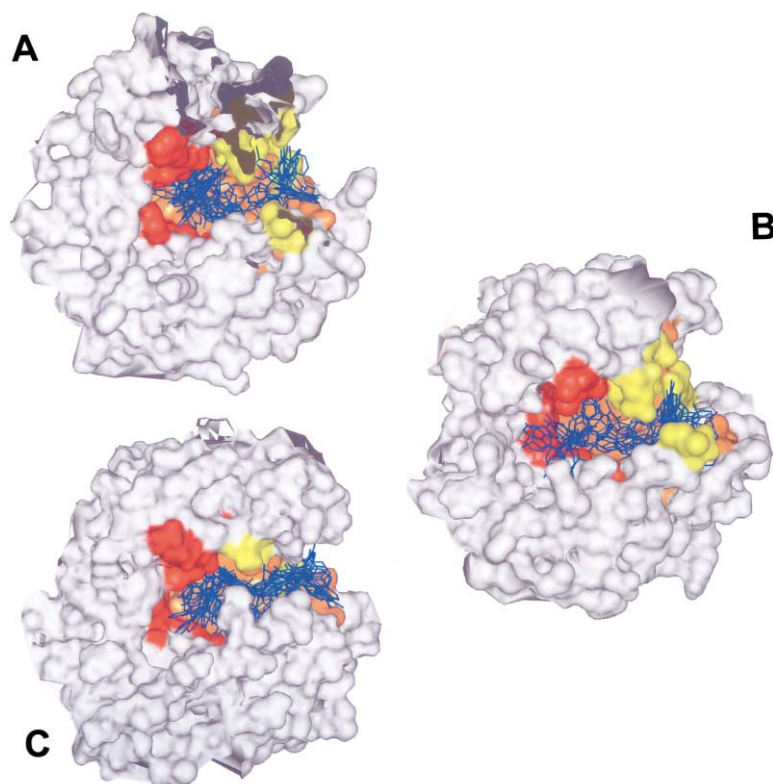
Figure 2. Autodock-Bound Conformations

The bound conformations obtained with Autodock (blue) are shown for LcTS (A), EcTS (B), and hTS (C). The protein surface is shown with the active site (orange), the folate (red), and the α156 (yellow) binding sites, as in (C), from the same viewpoint. In (A), the SD region has been removed to make the complete active site visible.

otic TSs from bacterial TSs and explaining the basis of the species-specific activity of α156 and phenolnaphthalein derivatives.

Molecular Docking

The docking program, AutoDock 3.0.5 [34], was used to identify favorable binding sites for α156 in TSs from bacterial and eukaryotic species. To explore the whole binding site of TS, docking was performed for open protein conformations only (see Experimental Procedures).

For the four isozymes, LmTS, hTS, LcTS, and EcTS, AutoDock predicts different possible binding modes either in the folate binding site or in the α156 crystallographic binding site (Figure 2). The folate binding site is computed to be energetically more favorable than the latter binding site (by about 1–1.7 kcal/mol in Lc, Ec, and human TS and about 3 kcal/mol in LmTS) (Table 2).

The identification of docked conformations at the fo-

late binding site is not in contrast with the crystallographic conformation seen in LcTS-α156 complex. In fact, site-directed mutagenesis and enzyme kinetic experiments [19] have previously pointed to a multiple binding mode for α156 to TS.

For each calculated binding site, multiple orientations were predicted for α156. These multiple orientations result from the pseudo-3-fold symmetry of the α156 molecule. In Figures 2A, 2B, and 2C, the range of bound conformations obtained by AutoDock in Lc, Ec, and human TS are shown. Two of the binding modes computed for α156 in LcTS are compared with the crystallographic binding sites of **CB3717** and α156 in Figures 3A and 3B [19, 22].

In the folate binding site (Figure 3A), α156 forms three hydrogen bonds, through its two hydroxyl groups and its carbonyl group, with D221, E60, and W85. D221 and E60 are conserved among all the species, whereas W85 is conserved between Lc and Ec TSs but is replaced by

Table 2. Results of Docking Compound α156 to LcTS, EcTS, LmTS, and hTS using AutoDock 3.0.5

	Folate			α156		
	%LCP[a]	ΔG_{bind}[b]	$<\Delta G_{bind}>$[c]	%LCP[a]	ΔG_{bind}[b]	$<\Delta G_{bind}>$[c]
HTS	30	−8.47	−7.77	52	−7.41	−5.65
LmTS	36	−10.62	−7.12	46	−7.60	−7.49
LcTS	38	−8.41	−8.08	30	−7.06	−7.01
EcTS	34	−9.48	−7.57	30	−7.75	−7.27

[a] Percentage of ligand docking positions predicted at the specified site (folate or α156).
[b] Estimated free energy of binding (kcal/mol) of the best-scored ligand conformation at the specified site.
[c] Average estimated free energy of binding for docked ligand configurations in the whole binding site (see Experimental Procedures).
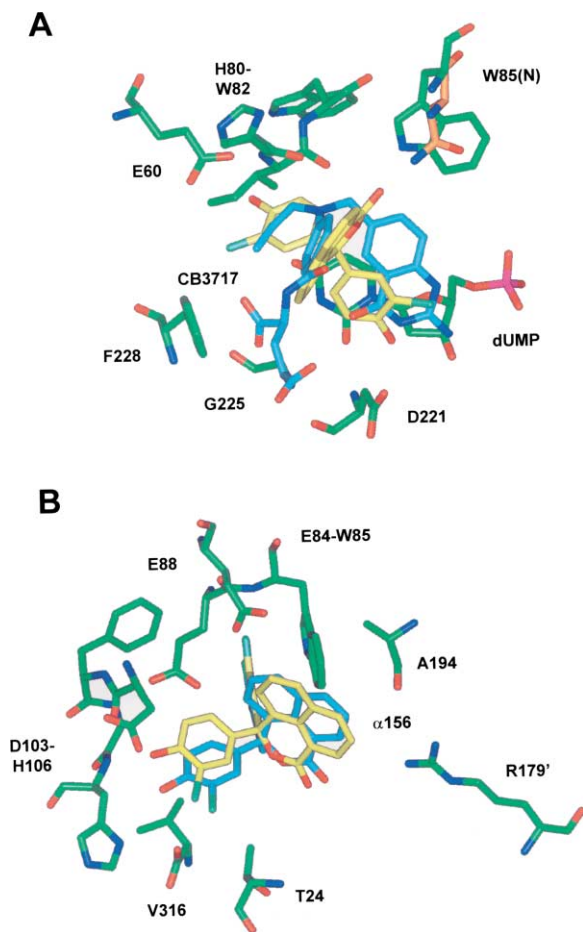
Figure 3. α156 Autodock-Bound Conformations Overlaid on CB1717 and α156 Crystallographic-Bound Conformations

Two LcTS-bound conformations (yellow) obtained with Autodock are overlaid on bound conformations from crystal structures (cyan) of CB3717 ([A], 1LCA) and α156 ([B], 1TSL). In (A) matched to W85 there is the corresponding asparagine residue from hTS (orange, coordinates from PDB files 1HVY).

an asparagine in Pc, Lm, and human TSs. However, an asparagine in this position is probably able to form a hydrogen bond with the carbonyl group of the ligand so that this intermolecular interaction would be conserved (Figure 3A). The ligand also forms van der Waals interactions with dUMP, H80, I81, W82, G225, and F228. All of these residues except for H80 are conserved among the five TSs. H80 is substituted by lysine in hTS, LmTS, and PcTS and threonine in EcTS. Since H80 interacts through its main chain with the ligand, the variation of this residue does not affect the possibility for the residue to interact with the ligand. In conclusion, the folate binding site is highly conserved in all the species, and it is anticipated that α156 can most probably bind in the folate binding site in all the species.

In the binding mode predicted by AutoDock for α156 in its crystallographic binding site in LcTS (Figure 3B), the ligand makes two hydrogen bonds through its hydroxyl groups with D103 and E84. The ligand also makes van der Waals interactions with T24, W85, E88, F104, G105, H106, A194, V316, and R179′. In the crystal struc-

ture, the interaction with D103 is mediated through a well-ordered water molecule, and the carbonyl group of α156 interacts with a second well-ordered molecule [19]. These water molecules were not considered during the calculation, and this fact provides an explanation for the slight differences between the crystallographic binding mode and the predicted one. Unlike the folate binding site, this binding site is highly species specific. In particular, LcTS residues D103, F104, G105, and H106 are in the SD and are absent in other species. E84 is substituted by an alanine in Lm and hTS and a glycine in PcTS. W85 (which also contributes to the folate binding site, see above) is replaced by an asparagine in Lm, Pc, and hTS. E88 is substituted by an arginine in Lm, Pc, and hTS, and it is conserved in EcTS but has a different orientation because of the big differences in the nearby SD region. Other residues are conserved but are part of highly flexible regions, such as the catalytic loop (CL), CT, and H25.

Despite the differences in the α156 crystallographic binding site among Lc, Ec, Lm, and hTS, AutoDock predicts binding positions for the ligand in this site in all the enzymes, although with slightly different binding modes.

The docking procedure correctly reproduced experimentally observed multiple binding modes for α156. This supports its use in further calculations to locate potential ligand binding sites. On the other hand, the AutoDock-estimated free energies of binding of the ligand to the two identified sites indicate that the crystallographic binding site of α156 is energetically less favorable than the folate binding site in all isozymes. This shows that the dominant conformation—as observed in the crystal structure with LcTS—cannot be easily detected from the ensemble of possible conformations resulting from the docking to the protein conformation in this crystal structure. This docking study nevertheless suggests that the inhibition activity of α156 toward TS is the result of the binding of the ligand in two possible sites in multiple orientations.

Protein Flexibility

As the binding pattern of α156 toward the four enzymes (LcTS, EcTS, LmTS, and hTS) predicted by AutoDock does not explain its species-specific inhibition profile, the structural flexibility of TSs from five different species was investigated. Computational techniques were applied to explore the conformational space and identify principal components of motion of these protein structures.

First, the CONCOORD [24] method was used to generate an ensemble of structures representing the available conformational space of the protein as determined from the crystal structure. The structures are generated by a random search for structures that satisfy distance constraints derived from analysis of the geometry and strengths of interatomic interactions in the crystal structure. The CONCOORD ensemble of structures (Cα atoms only) was then subjected to an essential dynamics [23, 35] analysis to identify the principal components (or eigenvectors) of motion described by the CONCOORD ensemble. These eigenvectors are ranked by decreasing eigenvalues. The eigenvalues quantify the contribution
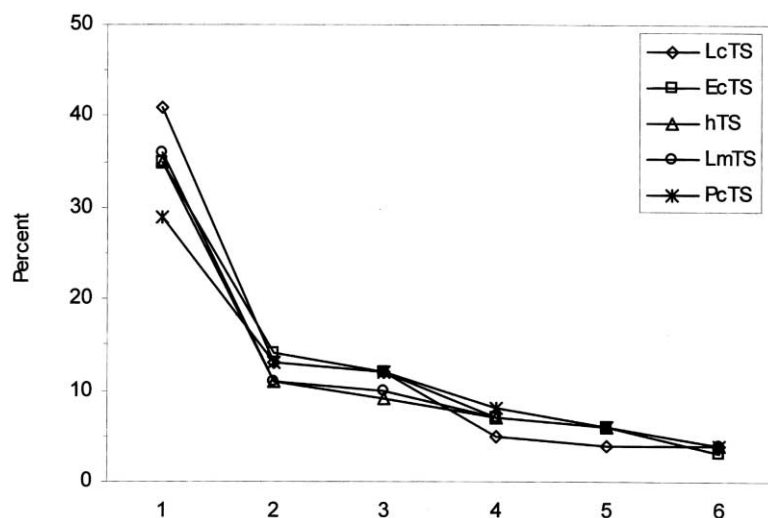
Figure 4. Proportion of the Overall Motion Described by the First Six Eigenvectors Identified by Essential Dynamics

Analysis of the CONCOORD structural ensemble for TSs from five organisms.

of the eigenvectors to describing the whole CONCOORD ensemble. Typically, a small number of eigenvectors account for most of the variation in conformation in the CONCOORD ensemble. The CONCOORD essential dynamics analysis highlights differences in the flexibility of TSs in different species that aid explanation of the species-specific inhibition activity of α156. In particular, clear differences emerge between bacterial (Lc and Ec) and eukaryotic (human, Lm, and Pc) TSs.

CONCOORD Essential Dynamics Analysis

The first eigenvector detected in the essential dynamics analysis plays the major role in describing the overall conformational space explored by the enzyme (accounting for 29%–41% of the overall motion, Figure 4). The motion along this eigenvector does not affect the binding site of α156. Despite the structural differences among TS species in the SD region, this eigenvector is almost the same in all the enzymes. The only region significantly different is the HIL region. This loop is shorter in LcTS and EcTS than in the other enzymes considered, and it is turned toward the intermonomer interface where it moves together with the whole interface region. In hTS, LmTS, and PcTS, this loop is longer and turned toward the SD region. As a consequence, in the latter enzymes, the HIL and SD regions show the same motion along the first eigenvector.

The second and third eigenvectors together describe 20%–26% of the whole conformational space explored (Figure 4). Both these vectors show a different type of motion in LcTS and EcTS compared with hTS, LmTS, and PcTS. In LcTS and EcTS, both vectors show that the SD, the H25, the CT, and the CL regions move together in the same direction. The motion along these vectors does not affect the accessibility of the α156 binding site. On the other hand, in hTS, PcTS, and LmTS, motion along these vectors results in closing (and subsequent re-opening) of the α156 binding site because the top (SD and CL) and the bottom (H25 and CT) of the site move toward each other. Figure 5 shows the second eigenvector for LcTS (Figure 5A) and hTS (Figure 5B). For each enzyme, two conformations along the motion, one in

blue and the other in red, were superimposed in order to highlight how the different regions of the α156 binding site move.

The fourth through sixth eigenvectors are less important in describing the conformational space of the proteins ($\leq$8%). In all enzymes, they correspond to α156 binding site closing/deforming motions. Thus, LcTS and EcTS also have α156 binding site closing (and re-opening) motions, but these are modest compared to those in the eukaryotic TSs.

Summarizing, in hTS, LmTS, and PcTS, closure/deformation of the α156 binding site is one of the principal components of the flexibility of these structures as highlighted in the second and the third eigenvectors; in LcTS and EcTS, this motion is a minor component of the flexibility of the protein structures that is apparent only in eigenvectors with lower eigenvalues. As a consequence, the α156 binding site is more accessible in the LcTS and EcTS structures than in the hTS, LmTS, and PcTS structures.

AutoDock calculations were performed for representative conformations of Lc, Ec, and human TSs extracted from the CONCOORD ensembles (see Experimental Procedures) in order to obtain an estimate of how much the enzyme flexibility affects the docking of α156. The results show that α156 binds to 100% of the representative conformations of LcTS and EcTS but only a subset of the representative conformations in hTS, corresponding to 64% of the whole CONCOORD ensemble of 500 structures. In the remaining 36% of conformations of hTS, the ligand cannot reach the active site and binds at the surface of the enzyme.

The closed conformations from the CONCOORD ensemble in hTS resemble the crystal structure as far as the type of shifts at the SD, CT, and H25 regions are concerned but do not reproduce the changes in conformation of these regions themselves that are seen in the experimental structure. The representative structures from the CONCOORD ensemble for LcTS all permit access of α156. Thus, the main conformations sampled in the CONCOORD ensemble do not close to the extent of the crystal structure of LcTS.
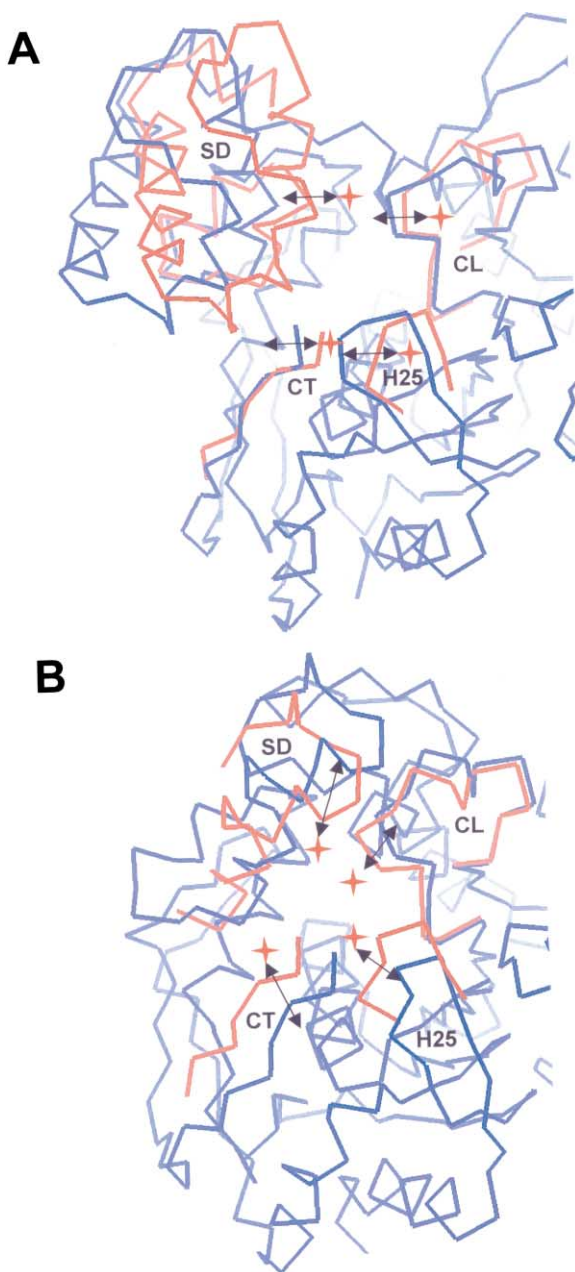
Figure 5. Comparison of the Motion along the Second Eigenvector of the CONCOORD Ensembles

Two conformations representing motion along the second eigenvector are shown for LcTS (A) and hTS (B). For clarity, only the small domain (SD), catalytic loop (CL), C terminus (CT), and loop around H25 (H25) are shown in the second conformation (red). The double-headed arrows indicate the directions of motion, with the stars showing the directions moved in simultaneously by different regions of the proteins.

Significance

A crucial problem in the design of antimicrobial agents is to achieve sufficient specificity for proteins of the target organism so that side effects in humans are avoided. In this paper, we report on how species-spe-

cific enzyme inhibition can be achieved by exploiting not only static structural differences in protein structure but also differences in protein dynamics and flexibility across species. We studied a set of inhibitors of TS, a highly conserved enzyme that is a potential target for antibacterial agents. We measured inhibition constants for six phenolnaphthalein analogs. In contrast to nonspecific folate-analog inhibitors, all these compounds showed species specificity, inhibiting bacterial species better than eukaryotic species. We used computational techniques to explore the basis for the species specificity of the phenolnaphthalein analogs. In molecular docking calculations, it was possible for the phenolnaphthalein derivatives to bind at both the classical folate binding site and a second more specific binding site in all the TS proteins studied. However, our analysis of protein flexibility shows that binding site closing motions in eukaryotic TSs appear to make the $\alpha156$ binding site less accessible than in the bacterial TSs. Our results thus suggest that the TS species specificity of phenolnaphthalein derivatives results not only from differences in structure but also from differences in protein dynamics. The insights obtained from the present study show how protein dynamics can influence ligand specificity and point to the importance of exploiting species-dependent variations in protein dynamics in the structure-based design of specific compounds.

Experimental Procedures

Experimental Inhibition Activity Profile Evaluation

All reactants used for enzyme purification and inhibition assays were at the maximum commercially available purity grade. 6R-N^5,N^{10}-methylen-5,6,7,8-tetrahydrofolate was kindly provided by EPROVA (Schaffhausen, Switzerland). **CB3717** was a gift of R.M. Stroud (UCSF).

Chemistry

Compounds $\alpha156$ (referred to as 5c in [25] and as 4 in [19]) and **1–5** were synthesized at the University of Modena and Reggio Emilia (M.P. Costi) and at the University of Milano (D. Barlocco). Their purity grade was the maximum obtainable. Synthesis of these compounds will be described elsewhere (our unpublished data).

Enzyme Purification

All the strains and the plasmids were provided by D.V. Santi at the University of California, San Francisco.

Expression and purification of EcTS, LcTS, PcTS, and hTS was done following known procedures [21, 35–39]. The LmTS (pE1) plasmid [40] was expressed in Ec stain BL21(DE3)pLysS following the procedure described by Hanan [41] and purified as reported in the literature [40]. The purified enzymes were stored at −80°C in 10 mM phosphate buffer (pH 7.0) and 0.1 mM EDTA (ethylene-diamine-tetraacetic acid).

Enzyme activity was determined spectrophotometrically (Beckman DU640 spectrometer thermostated with HAAKE F3C) by steady-state kinetic analysis, following the increasing absorbance at 340 nm due to the oxidation reaction of N^5,N^{10}-methylenetetrahydrofolate to dihydrofolate [42]. Assays were performed at 20°C in the standard assay buffer formed by TES [N-tris(hydroxymethyl) methyl-2-aminoethane-sulfonic acid] (50 mM) at pH 7.4, MgCl$_2$ (25 mM), formaldehyde (6.5 mM), EDTA (1 mM), and 2-mercaptoethanol (75 mM). Reaction solution (1 ml) was formed by standard assay buffer (pH 7.4), dUMP (120 μM), 6R-N^5,N^{10}-methylen-5,6,7,8-tetrahydrofolate (140 μM), and enzyme (0.07 μM).

Inhibition Assays

Stock solutions of the inhibitors were freshly prepared in DMSO (dimethyl sulfoxide). The inhibition pattern for all of the compounds was determined by steady-state kinetic analysis of the dependence

of the TS enzyme activity on folate concentration at varying inhibitor concentrations. The apparent inhibition constant (K_i) was obtained by applying a suitable equation [43] for competitive inhibitors, introducing the appropriate experimental K_m values (Michaelis-Menten constant: 10 μM for LcTS, 5 μM for EcTS, 8 μM for hTS, 16 μM for PcTS, 94 μM for LmTS). The specificity index of one compound toward TS from one species (xTS) is obtained from the ratio between its $K_{i(xTS)}$ and $K_{i(LcTS)}$. Each experiment was repeated at least three times, and no individual measurement differed more than 20% from the mean.

Computational Studies
Structural Comparison of TS Enzymes from Different Species and Binding Site Analysis
Crystal Structure Analysis. Crystal structures of LcTS (4TSM, 2TDM, 1THY, 1TSL, 1TSM, 1LCA) [19, 22, 26], EcTS (3TMS, 2TSC) [26, 28], and hTS (1HVY, 1HZW, 1I00) [29, 30] were taken from the PDB [33]. The crystal structures of the LcTS-phenolphtalein complex (TSPTH) [27] and PcTS (PCTS) [32] were provided by R.M. Stroud (UCSF) and that of LmTS (LMTS) [31] was provided by D.A. Matthews (Agouron Pharmaceuticals). When the asymmetric unit in the crystal was a monomer, the dimer was created with InsightII [44]. In some crystal structures (4TSM, 1THY, 1LCA, 2TSC, LMTS), some residues on the surface of the enzyme were lacking some atoms. Sequence errors, such as a glutamic acid indicated as glutamine, were found in some protein structures. For the structures considered only for visual analysis, the crystal structures were used "as is"; in all other cases, the protein structures were corrected. The affected residues were corrected by adding missing atoms and, then, minimizing the residue conformation in the context of the enzyme structure with AMBER 6.0 [45]. Only corrected residues, ligands, water molecules, and all hydrogen atoms of the system were allowed to move during the minimization. Parameterization of the protein-ligand complexes was performed with the xLEaP module of AMBER 6.0 [45] using a modified version of the parm94.dat force field. Atomic partial charges for dUMP were assigned on the basis of literature data [46]; atomic partial charges for the ligands were assigned using the electrostatic potential charges (ESP) computed with the semiempirical MOPAC (MNDO) program interfaced to InsightII [44].

Molecular Docking. The program AutoDock 3.0.5 [34] was used for docking. The protein structure of LcTS was taken from the crystal structure of the LcTS-α156 complex 1TSL [19], after removal of the ligand α156 and addition of a molecule of dUMP in place of the phosphate ion. The structure and coordinates of the dUMP molecule were taken from the ternary complex structure of LcTS, 1LCA [22]. For EcTS, the crystal structure of the unbound enzyme 3TMS [26] was used. A molecule of dUMP, taken from the ternary complex structure of EcTS, 2TSC [28], was added in place of the phosphate ion. For LmTS, only the ternary complex crystal structure LMTS [31] was available. A first trial was done with this structure, but it was not possible to completely explore the binding site of α156 since the region of the enzyme corresponding to the crystallographic binding site of α156 in LcTS is not accessible in the ternary complex of LmTS. An open structure of LmTS was obtained from LMTS by removing the molecule of CB3717 and adding a molecule of α156. The complex was then minimized and the enzyme conformation was forced to allow the binding of α156 in correspondence with its crystallographic binding site. This structure was used for docking, after removal of the molecule of α156. For hTS, there are seven crystal structures available in the PDB [33] and one of these, 1HWZ, is in the unbound state [30]. Unfortunately, this structure is lacking the last seven residues at the C terminus, so we added the last seven residues by taking them from the 1HVY [29] structure together with a molecule of dUMP from the 1I00 [30] structure. The complex was locally minimized to remove bad geometries at the graft point.

To run AutoDock, dimeric protein structures were prepared as described above. Polar hydrogen atoms were added with the WHATIF program [47]. Solvation parameters and atomic partial charges were assigned to the protein using the programs q.kollua and addsol, included in the AutoDock program suite. For dUMP (which was considered part of the target for docking), the solvation parameters and atomic partial charges were added manually on the basis of similarity to protein atoms and literature data, respectively [46].

The α156 structure was taken from the crystal structure of the LcTS-α156 complex [19]. Atomic partial charges were assigned on the basis of the Amber force field within the InsightII program [44]; for the atoms Cl_1/Cl_2 and C_{15}/C_{21}, (crystal structure numbering, X_n, from pdb files) charges of –0.15e and 0.15e, respectively, were assigned.

A cubic grid with sides 37.5 Å (for Lc, Ec, and Lm TSs) or 22.5 Å (for human TS) long, centered on the TS active site, was defined. These grids are large enough to cover all the enzymatic active site. A Lamarckian genetic algorithm was used to generate 50 bound conformations for α156 in each protein studied. All nonterminal rotatable bonds were allowed to rotate during the calculation. All other settings were kept as default.

On the basis of the clustering histogram output from the AutoDock program [34], the lowest energy conformation of each cluster was selected. The selected conformations were grouped on the basis of their binding sites. Two main binding sites were considered, the folate and the α156 binding site. For each binding site, multiple binding modes were considered. The average estimated free energy of binding ($<\Delta G_{bind}>$, Table 2) for each binding site was calculated as the simple mean of the estimate free energy of binding of each conformation found in that binding site. All the conformations that were in other binding sites were not considered.
Protein Flexibility Analyses
Structures 1TSL, 1LCA, 3TMS, 2TSC, 1HVY, LMTS, and PCTS [19, 22, 26, 28, 29, 31, 32] were used in flexibility studies. All ligands and water molecules were removed. For 1LCA, 2TSC, and LMTS, the corrected structures were considered.

Essential Dynamics. The program CONCOORD [24] was first used to generate an ensemble of 500 protein conformations. Then the ESSDYN module of the program WHATIF [47] was used to analyze the set of conformations and calculate its principal components.

Essential dynamics analysis is based on calculation of the eigenvectors that are solutions of the atomic fluctuations covariance matrix. Only Cα atoms were considered.

Essential dynamics analysis was performed on LcTS, EcTS, LmTS, PcTS, and hTS crystal structures. All parameters were kept at their default values. Trials were done on LcTS and EcTS to verify that the conformational state of the starting structure (apo- or ternary complex structure) did not influence the results of calculation (data not shown).

The CONCOORD ensembles of Lc, Ec, and human TS were analyzed with NMRCLUST [48]. Forty-six, thirty-eight, and forty clusters of different conformations were obtained for Lc, Ec, and human TSs, respectively. From the most populated clusters representative of 325, 370, and 375 out of 500 CONCOORD conformations, 13, 14, and 17 conformations were selected for Lc, Ec, and human TSs, respectively. These conformations were prepared and used in AutoDock calculations as described above. The number of conformations in each cluster was used to estimate the percentage of inhibitor binding for the whole ensemble.

References

1. Kubinyi, H. (1998). Structure-based design of enzyme inhibitors and receptor ligands. Curr. Opin. Drug Discov. Devel. *1*, 4–15.

2. Fritz, T.A., Tondi, D., Finer-Moore, J.S., Costi, M.P., and Stroud, R.M. (2001). Predicting and harnessing protein flexibility in the design of species-specific inhibitors of thymidylate synthase. Chem. Biol. *131*, 1–15.

3. Anderson, A.C., O'Neil, R.H., Surti, T.S., and Stroud, R.M. (2001). Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. Chem. Biol. *88*, 1–13.

4. Stout, T.J., and Stroud, R.M. (1996). The complex of anti-cancer therapeutic, BW1843U89, with thymidylate synthase at 2.0 Å resolution: implications for a new mode of inhibition. Structure *1*, 67–77.

5. Davis, M., and Teague, T.J. (1999). Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. Angew. Chem. Int. Ed. Engl. *38*, 736–749.

6. Ma, B., Shatsky, M., Wolfson, H.J., and Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci. *11*, 184–197.

7. James, L.C., Roversi, P., and Tawfik, D.S. (2003). Antibody multi-specificity mediated by conformational diversity. Science *299*, 1362–1367.

8. Schnecke, V., and Kuhn, L.A. (1999). Database screening for HIV protease ligands: the influence of binding-site conformation and representation on ligand selectivity, p. 242–251. In T. Lengauer, R. Schneider, P. Bork, D. Brutlag, J. Glasgow, H.W. Mewes, and R. Zimmer (eds) Proceeding of the Seventh International Conference on Intelligent Systems for Molecular Biology, Menlo Park, CA.

9. Lorber, D.M., and Shoichet, B.K. (1998). Flexible ligand docking using conformational ensembles. Protein Sci. *7*, 938–950.

10. Carlson, H.A. (2002). Protein flexibility is an important component of structure-based drug discovery. Curr. Pharm. Des. *8*, 1571–1578.

11. Zhou, H.X., Wlodek, S.T., and McCommon, J.A. (1998). Conformation gating as a mechanism for enzyme specificity. Proc. Natl. Acad. Sci. USA *95*, 9280–9283.

12. Yuan, P., Marshall, V.P., Petzold, G.L., Poorman, R.A., and Stockman, B.J. (1999). Dynamics of stromelysin/inhibitor interactions studied by ^{15}N NMR relaxation measurements: comparison of ligand binding to S_1-S_3 and S'_1-S'_3 subsites. J. Biomol. NMR *15*, 55–64.

13. Northrup, S.H., and McCammon, J.A. (1982). Rate theory for gated diffusion-influenced ligand binding to proteins. J. Phys. Chem. *86*, 2314–2321.

14. Costi, M.P., Tondi, D., Rinaldi, M., Barlocco, D., Pecorari, P., Soragni, F., Venturelli, A., and Stroud, R.M. (2002). Structure-based studies on species-specific inhibition of thymidylate synthase. Biochim. Biophys. Acta *1587*, 206–214.

15. Gschwend, D.A., Sirawaraporn, W., Santi, D.V., and Kuntz, I.D. (1997). Specificity in structure-based drug design: identification of a novel, selective inhibitor of Pneumocystis carinii dihydrofolate reductase. Proteins *29*, 59–67.

16. Ota, N., and Agard, D.A. (2001). Enzyme specificity under dynamic control II: principal component analysis of α-litic protease using global and local solvent boundary conditions. Protein Sci. *10*, 1403–1414.

17. Stroud, R.M., and Finer-Moore, J.S. (2003). Conformational dynamics along the enzymatic reaction pathway: thymidylate synthase, "the movie." Biochemistry *42*, 239–247.

18. Carreras, C.W., and Santi, D.V. (1995). The catalytic mechanism and structure of thymidylate synthase. Annu. Rev. Biochem. *64*, 721–762.

19. Stout, T.J., Tondi, D., Rinaldi, M., Barlocco, D., Pecorari, P., Santi, D.V., Kuntz, I.D., Stroud, R.M., Shoichet, B.K., and Costi, M.P. (1999). Structure-based design of inhibitors specific for bacterial thymidylate synthase. Biochemistry *38*, 1607–1617.

20. Hardy, L.W., Finer-Moore, J.S., Montfort, W.R., Jones, M.O., Santi, D.V., and Stroud, R.M. (1987). Atomic structure of thymidylate synthase: target for rational drug design. Science *235*, 448–455.

21. Cardilicchia, I. (1998/1999). La Timidilato Sintasi come nuovo target biologico nelle malattie iperproliferative. Preparazione di una bio-library e applicazioni. Degree Thesis, University of Modena and Reggio Emilia, Modena, Italy.

22. Finer-Moore, J.S., Fauman, E.B., Foster, P.G., Perry, K.M., Santi, D.V., and Stroud, R.M. (1993). Refined structure of substrate-bound and phosphate-bound thymidylate synthase from Lactobacillus casei. J. Mol. Biol. *232*, 1101–1116.

23. Amadei, A., Lissen, A.B.M., and Berendsen, J.C. (1993). Essential dynamics of proteins. Proteins *17*, 412–425.

24. De Groot, B.L., Van Aalten, D.M.F., Scheek, R.M., Amadei, A., Vriend, G., and Berendsen, H.J.C. (1997). Prediction of protein conformational freedom from distance constraints. Proteins *29*, 240–251.

25. Costi, M.P., Rinaldi, M., Tondi, D., Pecorari, P., Barlocco, D., Ghelli, S., Stroud, R.M., Santi, D.V., Stout, T.J., Musiu, C., et al. (1999). Phthalein derivatives as a new tool for selectivity in thymidylate synthase inhibition. J. Med. Chem. *42*, 2112–2124.

26. Perry, K.M., Fauman, E.B., Finer-Moore, J.S., Montfort, W.R., Maley, G.F., Maley, F., and Stroud, R.M. (1990). Plastic adaptation toward mutations in proteins: structural comparison of thymidylate synthases. Proteins *8*, 315–333.

27. Shoichet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D., and Perry, K.M. (1993). Structure-based discovery of inhibitors of thymidylate synthase. Science *259*, 1445–1450.

28. Montfort, W.R., Perry, K.M., Fauman, E.B., Finer-Moore, J.S., Maley, G.F., Hardy, L., Maley, F., and Stroud, R.M. (1990). Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dUMP and an anti-folate. Biochemistry *29*, 6964–6977.

29. Phan, J., Koli, S., Minor, W., Dunlap, R.B., Berger, S.H., and Lebioda, L. (2001). Human thymidylate synthase is in the closed conformation when complexed with dUMP and Raltitrexed, an antifolate drug. Biochemistry *40*, 1897–1902.

30. Almog, R., Waddling, C.A., Maley, F., Maley, G.F., and Van Roey, P. (2001). Crystal structure of a deletion mutant of human thymidylate synthase Δ(7–29) and its ternary complex with Tomudex and dUMP. Protein Sci. *10*, 988–996.

31. Knighton, D.R., Kan, C.C., Howland, E., Janson, C.A., Hostomska, Z., Welsh, K.M., and Matthews, D.A. (1994). Structure of and kinetic channelling in bifunctional dihydrofolate reductase-thymidylate synthase. Nat. Struct. Biol. *1*, 186–194.

32. Anderson, A.C., O'Neil, R.H., DeLano, W.L., and Stroud, R.M. (1999). The structural mechanism for half-the-sites reactivity in an enzyme, thymidylate synthase, involves a relay of changes between subunits. Biochemistry *38*, 13829–13836.

33. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242.

34. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. *19*, 1639–1662.

35. Maley, G.F., and Maley, F. (1988). Properties of a defined mutant of Escherichia coli thymidylate synthase. J. Biol. Chem. *263*, 7620–7627.

36. Kealey, J.T., and Santi, D.V. (1992). Purification methods for recombinant Lactobacillus casei thymidylate synthase and mutants: a general, automated procedure. Protein Expr. Purif. *4*, 380–385.

37. Santi, D.V., Edman, U., Minkin, S., and Greene, P.J. (1991). Purification and characterization of recombinant Pneumocystis carinii thymidylate synthase. Protein Expr. Purif. *2*, 350–354.

38. Edman, U., Edman, J.C., Lundgren, B., and Santi, D.V. (1989). Isolation and expression of the Pneumocystis carinii thymidylate synthase gene. Proc. Natl. Acad. Sci. USA *84*, 6503–6507.

39. Banerjee, C.K., Bennett, L.L., Jr., Brockman, R.W., Santi, B.P., and Temple, C., Jr. (1982). A convenient procedure for purification of thymidylate synthase from L1210 cells. Anal. Biochem. *121*, 275–280.

40. Grumont, R., Sirawaraporn, W., and Santi, D.V. (1988). Heterologous expression of the bifunctional thymidylate synthase-dihydrofolate reductase from Leishmania major. Biochemistry *27*, 3776–3784.

41. Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. *166*, 557–580.

42. Pogolotti, A.L., Jr., Danenberg, P.V., and Santi, D.V. (1986). Ki-

netics and mechanism of interaction of 10-propargyl-5,8-didea-zafolate with thymidylate synthase. J. Med. Chem. *29*, 478–482.

43. Segel, I.H. (1975). Enzyme kinetics. In Behaviour and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems (New York: John Wiley and Sons, Inc.), pp. 105.

44. InsighII (2000). Accelrys Inc., www.accelrys.com

45. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham, T.E., III, Ross, W.S., Simmerling, C.L., Darden, T.A., Merz, K.M., Stanton, R.V., Cheng, A.L., et al. (1999). AMBER 6. University of California, San Francisco.

46. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Jr., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. *117*, 5179–5197.

47. Vriend, G. (1990). WHATIF: A molecular modeling and drug design program. J. Mol. Graph. *8*, 52–56.

48. Sutcliffe, M (1996). NMRCLUST. Departments of Biochemistry and Chemistry, University of Leicester, UK. http://neon.chem.le.ac.uk/nmrclust/.