Predicting resistance mutations using protein design algorithms

Kathleen M. Frey[a,1], Ivelin Georgiev[b,1,3], Bruce R. Donald[b,c,2], and Amy C. Anderson[a,2]

[a]Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT 06269; [b]Department of Computer Science, Duke University; and [c]Department of Biochemistry, Duke University, Durham, NC 27708-0129

Drug resistance resulting from mutations to the target is an unfortunate common phenomenon that limits the lifetime of many of the most successful drugs. In contrast to the investigation of mutations after clinical exposure, it would be powerful to be able to incorporate strategies early in the development process to predict and overcome the effects of possible resistance mutations. Here we present a unique prospective application of an ensemble-based protein design algorithm, K^*, to predict potential resistance mutations in dihydrofolate reductase from _Staphylococcus aureus_ using positive design to maintain catalytic function and negative design to interfere with binding of a lead inhibitor. Enzyme inhibition assays show that three of the four highly-ranked predicted mutants are active yet display lower affinity (18-, 9-, and 13-fold) for the inhibitor. A crystal structure of the top-ranked mutant enzyme validates the predicted conformations of the mutated residues and the structural basis of the loss of potency. The use of protein design algorithms to predict resistance mutations could be incorporated in a lead design strategy against any target that is susceptible to mutational resistance.

negative design algorithm | DHFR | MRSA | K* design algorithm

Resistance has been observed for even the most reserved antibiotics, sometimes after only brief clinical exposure. One of the most common resistance mechanisms is the accumulation of mutations in an enzyme target, creating an active site that can no longer accommodate the inhibitor yet maintains function. When these resistance mutations are discovered in the clinic, the mutants must be identified and studied, forcing the drug design process to start anew. To address this problem in preclinical drug discovery, resistance mutants are generated and studied in vitro with labor-intensive experiments. In contrast, it would be useful to predict resistance mutations in silico during the very early stages of drug discovery, thus encouraging strategies to overcome these limitations during the design process. In response to this need, we have developed and experimentally tested a protocol to computationally predict resistance mutations in a protein target, using algorithms for positive and negative structure-based protein design.

Protein design algorithms have recently been developed and used to reengineer proteins and enzymes to bind unique ligands. In the case of enzymes, successful design requires catalytic activity as well as binding. A promising approach involves ensemble-based scoring and search algorithms for protein design, which have been applied to modify the substrate specificity of an antibiotic-producing enzyme in the nonribosomal peptide synthetase pathway (1). New algorithms for protein design, an example of which is called K^*, combine a statistical mechanics-derived ensemble-based approach to computing the binding constant with the speed and completeness of a branch-and-bound pruning algorithm (2–4). In addition, efficient deterministic algorithms include provable ε-approximation algorithms for estimating partition functions in order to model binding to arbitrary precision (4). Other examples of successful applications of computational protein design include design of unique enzymes (5, 6), enhancement of antibody affinity (7), and design of a unique protein fold (8).

There have been few previous attempts to predict potential resistance mutations; most of these have been applied retrospectively to predict known mutations that occur in response to clinically used antibiotics. Predictions have identified peptide sequences that would bind more tightly to a mutant HIV protease enzyme (9), profiled tyrosine kinase inhibitors (10), and estimated "vitality values" to account for the catalytic activity of resistant mutants of HIV protease (11). A recent review (12) summarizes structure- and sequence-based attempts to predict known mutations in HIV protease and reverse transcriptase.

Using structure-based design, we have been developing a unique class of propargyl-linked antifolate inhibitors that are active against several pathogenic species of dihydrofolate reductase (DHFR), specifically, methicillin-resistant _Staphylococcus aureus_ (MRSA) DHFR (Sa DHFR) (13, 14). As MRSA has an extensive array of resistance mechanisms, it is critical to consider the likely development of resistance for any new inhibitors. Therefore, given that we have determined high resolution structures of wild-type Sa DHFR bound to these propargyl-linked antifolates (13, 14), we considered this to be an excellent case to apply structure-based protein design algorithms for resistance mutation prediction.

Here, we report a prospective study that uses the protein design algorithm, K^*, to predict resistance mutations in DHFR from MRSA toward a potent propargyl-linked antifolate. Structure-based negative design predicted mutations to destabilize the binding of the inhibitor; positive design predicted mutations that stabilize the native protein function. Intersecting the top-scoring positive and negative designs predicted candidate resistance mutations. Four of the top ten ranked mutant enzymes were created and evaluated. Three of the mutants indeed maintained activity and displayed lower (18-, 9- and 13-fold) affinity for the inhibitor. A crystal structure of the top-ranked mutant bound to the inhibitor shows a conformation of the ligand that clearly has significantly fewer interactions with the protein. This unique and expedient approach to resistance mutation prediction should be useful for the development of inhibitors toward other targets for which drug therapy is limited by mutational resistance.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Results and Discussion

Compound **1** (Fig. 1) is a lead inhibitor of MRSA DHFR ($K_i = 10$ nM) and possesses antibacterial activity against isolates of MRSA in culture. In order to elucidate the basis of its potency, we determined a crystal structure of the enzyme bound to its cofactor, NADPH and compound **1**. Crystals of the wild-type Sa DHFR showed diffraction amplitudes to 2.1 Å (Table 1); the structure was determined with difference Fourier methods using coordinates of Sa (F98Y)DHFR (15). The model reveals the overall characteristic "DHFR fold" that consists of an eight-stranded beta sheet and four alpha helices connected by flexible loop regions. The pyrimidine ring of compound **1** forms two conserved hydrogen bonds with Asp 27 and the 4-amino group forms two additional hydrogen bonds with the backbone carbonyl atoms of residues Leu 5 and Phe 92 (Fig. 1). The propargyl linker and propargylic methyl group form van der Waals contacts with Phe 92 and Leu 20. The methyl group at the C6 position forms van der Waals interactions with Leu 20 and Leu 28 and the meta-biphenyl ring extends into a hydrophobic pocket formed primarily by residues Ile 50, Leu 28, and Leu 54.

In order to identify residues that when mutated may incur a binding penalty for compound **1** but that would maintain binding to the substrate, dihydrofolate, K^* searches using the minDEE/$A * /K^*$ algorithms (1, 4) were performed. For these searches, the algorithm used as input the crystal structure of Sa DHFR:compound **1**:NADPH as well as a model of Sa DHFR bound to NADPH and dihydrofolate, which was adapted from coordinates of a single mutant Sa DHFR bound to the same ligands (15). Ten active site residues (Leu 5, Val 6, Leu 20, Asp 27, Leu 28, Val 31, Thr 46, Ile 50, Leu 54, and Phe 92) were modeled as flexible using rotamers (16) and allowed to keep their wild-type identity or to mutate within a restricted group of amino acids (see *Materials and Methods*) that conserve the amino acid property or correlate with a different DHFR species. In order to comprehensively test the algorithm, the group included mutations that represent both single- and double-nucleotide polymorphisms. While single-nucleotide polymorphisms are the most prevalent mechanism of mutational resistance, more complex mutations do evolve such as those in HIV-1 reverse transcriptase (17–19), hepatitis C protease (20), hepatitis B reverse transcriptase (21), and *Helicobacter pylori* rpoB (22).

Based on the input model, K^* searches designed to identify up to 2-point mutations were performed separately for dihydrofolate and compound **1**. K^* scores are computed as a ratio of Boltzmann-weighted partition functions over rotamer-based conformational ensembles for the bound protein-ligand complex, the free protein, and free ligand (see *Materials and Methods*). Since higher K^* scores imply better affinity, we focused our attention on pairs of mutations that scored highly for dihydrofolate and poorly for compound **1**, particularly on those that returned a score of zero for compound **1** (Table 2) and a resultant score ratio of infinity. A mutant K^* score of 0 for compound **1** implies one

of two scenarios: (1) K^* pruned all rotameric conformations for these mutants; or (2) the computed partition function for the bound protein-ligand complex for a given mutant was significantly less (in the current redesigns: more than eight orders of magnitude) than the product of the partition functions for the free protein and the free ligand. Typically, scenario (1) occurs when significant steric interference exists either within the protein or between the protein and the ligand for all possible rotameric conformations, even after minimization. An initial steric overlap (before minimization) between two rotamers of more than 1.5 Å was considered significant and the corresponding pair of rotamers was pruned from further consideration (*Materials and Methods*). Scenario (2) is typically caused by significantly unfavorable interactions between the protein and the ligand, as opposed to the protein in its unbound form (Table 2). The different mutants with a score of 0 for compound **1** are thus not equivalent with respect to the predicted destabilized interactions with compound **1**. There were 105 mutants with a score ratio of infinity; the top ten exhibited a dihydrofolate score considerably better than the others and were further investigated. These ten mutants fell into two groups: the first group contained mutations at residues 31 and 92, the second group contained mutations at positions 50 and 92 (Table 2). The computed partition functions for the bound protein-ligand complex (with compound **1**) and the free protein (Table 2 *D*, *E*), along with the lowest energy for the respective ensemble of conformations (Table S1) suggest that compound **1** has significantly destabilized interactions with these mutants.

Interestingly, the wild-type sequence was ranked #306 out of 1,173 with dihydrofolate and #369 out of 1,173 with compound **1**, suggesting that a subset of the higher-ranked sequences may have improved binding to dihydrofolate or compound **1**.

Experimental Validation of the Predicted Mutants. To test the validity of the results of the algorithm, we created four of the mutants that show diversity at each of the positions (ranked by the algorithm #1, 3, 7, and 9) using site-directed mutagenesis and purification procedures similar to the wild-type enzyme (13, 14). Enzyme activity assays show that there is only a modest 3-fold reduction in K_M values for dihydrofolate (Table 3) for the Val31/Phe92 mutants. While the k_{cat} and k_{cat}/K_M values are lower than those of the wild-type enzyme (Table 3), the losses are within the range of other clinically observed DHFR mutants. For example, the F57L mutation in *Plasmodium vivax* DHFR (pyrimethamine, cycloguanil and WR99210 resistance) (23), the L22R mutation in human DHFR (methotrexate resistance) (24) and the A16V mutation in *Plasmodium falciparum* DHFR (cycloguanil resistance) (25) suffer 220-, 250-, and 680-fold losses, respectively, in k_{cat}/K_M. The ninth-ranked mutant, Sa (I50W, F92S) DHFR, was not active, suggesting that the Ile50Trp mutation prevents binding to the substrate. The results for the mutants ranked #1, 3, and 7 by the algorithm experimentally validate the positive design

Table 1. Statistics of data collection and refinement

Complex	Sa (WT)/ cmpd 1/NADPH	Sa (V31Y,F92I)/ cmpd 1/NADPH
PDB ID	3F0Q	3LG4
Space group	P6$_1$22	P6$_1$
Unit cell (a,b,c in Å)	$a = b = 79.16$, $c = 108.80$	$a = b = 88.752$, $c = 103.167$
Resolution, (last shell, Å)	25.9 - 2.10 (2.29 - 2.08)	42.8 - 3.15 (3.23 - 3.15)
Completeness, % (last shell,%)	87.9 (95.3)	91.2 (99.7)
Redundancy (last shell)	10.40 (12.2)	5.0 (5.4)
Rsym (last shell)	0.050 (0.143)	0.061 (0.379)
$\langle I/\sigma \rangle$ (last shell)	8.6 (2.7)	14.5 (3)
Refinement statistics		
R-factor/R_{free}	0.196; 0.231	0.26; 0.292
Rms deviation bond lengths (Å), angles (°)	0.010; 1.386	0.007; 1.219
Average B factors (Å^2): overall; NADPH; compound 1	20.6; 16.7; 21.8	88.7; 94.7; 68.5
Residues in most favored regions, allowed regions (%)	90.4, 9.6	87.5; 12.5
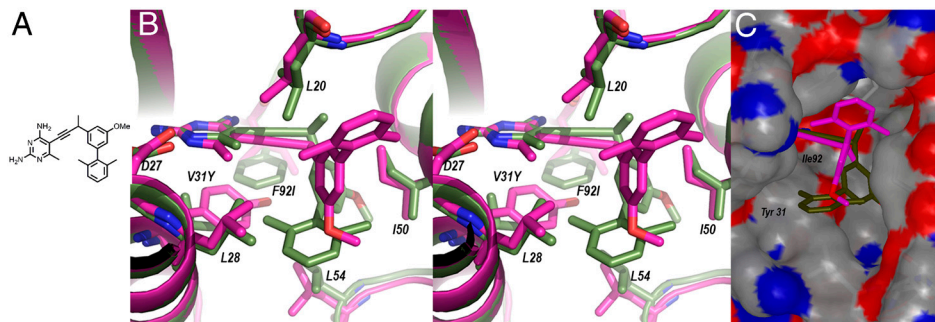
Fig. 1. Stereo view of the superposition of the ternary crystal structures of Sa (WT) DHFR (*green*) and Sa(V31Y/F92I) DHFR (*magenta*) bound to NADPH and compound **1** (*A*) as *B*) stick models and *C*) surface rendering.

component of the K^* algorithm and its success in predicting mutants that retain catalytic activity.

In order to assess the results of the negative design component of the algorithm, K_i values were measured for the wild-type and Sa (V31F, F92I) DHFR enzymes with compound **1**. Dixon plots show that the inhibitor binds competitively (Figs. S1 and S2); analysis of the plots yields K_i values of 7.5 nM and 128 nM for the wild type and mutant, respectively. K_i values were also calculated from IC_{50} values and K_M values (26) for all active mutants (Table 4). The top-ranked resistance mutant, Sa (V31Y, F92I) DHFR, shows the greatest (18-fold) loss in affinity for compound **1** (Table 4). Mutants Sa (V31Y, F92S) and Sa (V31F, F92L) DHFR also show significant (9-fold and 13-fold, respectively) losses in potency, suggesting that the algorithm is also successful in its negative design component. The success of the algorithm prompted the investigation of a structure of the mutant to determine why the resistance mutations at positions 31 and 92 retain activity but lose affinity for compound **1**.

Determination of a Crystal Structure of Sa (V31Y, F92I) DHFR, NADPH and Compound 1. Crystals of Sa (V31Y, F92I) DHFR showed diffraction amplitudes to 3.15 Å (Table 1); the structure of the mutant was determined using difference Fourier methods based on the wild-type structure bound to NADPH and compound **3** (Table S2) as a model (PDB ID: 3FQC) (13). There is a high degree of similarity between Sa (wild-type) and Sa (V31Y, F92I) DHFR, reflected in a root mean square deviation for 157 C_α atoms of 0.355 Å. The similarity of the enzymes is also reflected in their melting temperatures, as determined by circular dichroism (T_m wild-type = 42.5 °C, T_m Sa(V31Y,F92I) = 36.3 °C, graphs shown in Fig. S3). The Sa (V31Y, F92I) DHFR mutant

structure exhibits the standard, extended conformation of NADPH, in contrast to the alternate conformation observed in several structures of the Sa (F98Y) DHFR mutant (13). In contrast to the wild-type structure in which the ligand fully occupies the site, compound **1** binds the mutant active site with 50% occupancy, suggesting that the V31Y and F92I mutations affect ligand binding.

Despite the moderate resolution of the data for the mutant enzyme, the electron density maps revealed significant structural details including side chain and ligand orientations in the active site that disclose the basis of the lower affinity of compound **1** (Fig. 1). Strong hydrophobic interactions made with the native Phe92 residue and propargyl linker of compound **1** are reduced with the mutation to Ile92. The Val31Tyr mutation introduces steric bulk in the active site that interferes with the 2-methyl substitution on the distal phenyl ring, causing the substituted biphenyl of the ligand to contort around the propargyl position and reorient by approximately 60°. Reorientation positions the two phenyl rings outside the main hydrophobic pocket, causing the loss or reduction of strong hydrophobic interactions with residues Leu 28, Val 31, Leu 54, and Phe 92. In the new position, the distal phenyl ring maintains interactions only with Leu 20. While it appears that the mutant enzyme may have bound the opposite enantiomer compared to that bound in the wild-type structure, the resolution of the electron density does not permit exact evaluation.

K_M values suggest that active sites mutated at the Val31 and Phe92 positions retain affinity for the substrate, dihydrofolate. In order to understand why these mutations allow substrate binding, we compared the Sa (V31Y, F92I) DHFR crystal structure bound to compound **1** and NADPH with the lowest energy predicted structure for Sa (V31Y, F92I) DHFR bound to dihydrofolate (Fig. 2*A*). In contrast to compound **1**, which relies on the interaction between the propargyl linker and Phe 92, dihydrofolate lacks a propargyl linker and is minimally affected by the Ile 92 mutation. In addition, the para-aminobenzoic acid moiety of dihydrofolate can be accommodated in the limited space near the Tyr 31 mutation. This accommodation differs from compound **1**, where the biphenyl changes orientation to avoid steric repulsion with Tyr 31. Although a crystal structure for the Sa (V31F, F92L) DHFR mutant was not obtained, it appears that the dihydrofolate molecule could also be accommodated in the space near the Phe 31 mutation.

Table 2. The top ten [a]mutants as ranked by the *score ratio* criterion are shown with the respective K^* scores with [b]dihydrofolate (DHF)

| | | | Compound 1 | |
| | | | Partition function[c] | |
Rank	Mutations[a]	DHF K^* Score[b]	FP[d]	PL[e]
1	**V31Y/F92I**	4.30×10^{40}	6.6×10^{292}	2.1×10^{132}
2	V31Y/F92V	3.81×10^{40}	2.9×10^{301}	4.3×10^{143}
3	**V31Y/F92S**	3.13×10^{40}	1.2×10^{331}	2.4×10^{173}
4	V31Y/F92A	2.94×10^{40}	1.3×10^{330}	4.4×10^{171}
5	V31Y/F92M	6.77×10^{38}	3.4×10^{332}	1.0×10^{169}
6	V31Y/F92L	6.38×10^{38}	2.3×10^{320}	2.7×10^{150}
7	**V31F/F92L**	6.01×10^{33}	2.3×10^{327}	2.3×10^{310}
8	I50W/F92M	7.70×10^{32}	8.8×10^{340}	5.0×10^{327}
9	**I50W/F92S**	2.74×10^{32}	1.9×10^{339}	3.2×10^{325}
10	I50W/F92A	2.10×10^{32}	1.9×10^{338}	3.9×10^{323}

For each of the mutants, the computed [c]partition function for the [d]free protein (FP) and [e]protein-ligand (PL) complex (with compound **1**) are also shown; all of the top 10 mutants had a K^* score of 0 for compound **1**. Mutants shown in bold were selected for experimental validation.

Table 3. Kinetic parameters for the wild-type and mutant enzymes

Enzyme	K_M, units in μM	k_{cat}, units in 1/s	k_{cat}/K_M (fold decrease)
Sa (WT)	14.5 ± 3.5	31	2.14 (1.00)
Sa (V31Y, F92I)	43 ± 2.6	2.8	0.06 (36)
Sa (V31Y, F92S)	58 ± 3.0	1.4	0.02 (107)
Sa (V31F, F92L)	45 ± 4.3	0.31	0.007 (306)

Table 4. Inhibition assays for enzymes and compound 1

Enzyme	K_i, nM	Fold loss*
Sa wild-type (WT)	10	1.0
Sa (V31Y,F92I)	180	18
Sa (V31Y, F92S)	87	8.7
Sa (V31F, F92L)	130	13

*Fold loss = K_i value for enzyme/K_i value for WT.

To assess whether the K^* algorithm predicted similar rotamer conformations as were observed in the crystal structure, the crystal structure of Sa (V31Y, F92I) DHFR was compared with the lowest energy predicted structure (Fig. 2B). The Tyr 31 mutation observed in the crystal structure has the same conformation as each of the top ten members of the ensemble of predicted structures. Although the rotamer for Ile 92 in the crystal structure differs from the predicted structure, the predicted residue is located in the same location and space as both the mutant Sa (V31Y, F92I) and Sa (WT) DHFR structures. Similarity of rotamer conformations suggests that the algorithm is accurate in predicting not only the identity of residues targeted for mutation, but also the proper orientation of mutated residues.

The conformation of compound **1** from the top K^*-predicted structure lacks the reorientation of the biphenyl observed in the Sa (V31Y, F92I) DHFR crystal structure. This reoriented conformation is significantly out of the range of ligand conformations input to the software to be modeled by K^*. For example, in the K^* input ligand rotamer (see *Materials and Methods* section) closest to the ligand conformation in the Sa (V31Y, F92I) DHFR crystal structure, the biphenyl of compound **1** is rotated by approximately 60°.

The goal of the K^* algorithm was to block binding to compound **1** by introducing mutations to DHFR, based on the selected set of input ligand conformations. Using this input model, the algorithm successfully identified mutations that have significantly diminished binding to compound **1**. The ligand flexibility model used by K^* significantly improves on models where ligands are treated as rigid or in which rigid ligand rotamers are used. However, due to the expense of the computation, the number of rotamers and the span of ligand conformations were still significantly limited. While, ideally, the entire feasible conformation space around a ligand should be evaluated by the computational procedure, this would come at the cost of a combinatorial increase in compute times. Optimal balance should thus be sought between the computational requirements and the comprehensiveness of the flexibility models.

Nevertheless, it is interesting to determine how a ligand conformation mimicking the one observed in the crystal structure of the Sa (V31Y, F92I) mutant complex would be scored by our algorithm. To this end, we set up and performed a K^* score computation for the Sa (V31Y, F92I) mutant, with the ligand dihedrals constrained to a conformation close to the one observed in the mutant crystal structure. Interestingly, the computed lowest energy (approx. −179.4 kcal/mol) in the K^* conformation ensemble for the bound protein-ligand complex was virtually identical to the lowest energy of the original K^* ensemble (approx. −180.0 kcal/mol, see Table S1). Thus, even with the ligand conformation mimicking the one observed in the crystal structure complex, K^* predicts the Sa (V31Y, F92I) mutant to have significantly destabilized interactions with compound **1** as compared to the free protein (see Table S1).

Fig. 2. Stereo views of the superposition of the Sa(V31Y/F92I) DHFR crystal structure (*magenta*) with A) model of dihydrofolate in the active site (*cyan*) and B) the lowest energy model from the predicted ensemble (*yellow*).

Elucidating Structure-Activity Relationships of Additional Antifolate Compounds. The structural evidence suggests that the Val31Tyr and Phe92Ile mutations decrease affinity by reducing hydrophobic interactions with Phe 92 and introducing steric bulk at Val 31. In order to validate these hypotheses, we tested the activity of four other propargyl-based antifolates for the mutant enzymes and probed the structure-activity relationships. All four additional antifolates (compounds **2**, **3**, **4**, and **5** shown in Table S2 with inhibition values) are potent inhibitors for the wild-type enzyme, but lose potency for all of the mutant enzymes. Compound **2**, with an unsubstituted distal phenyl ring relative to the dimethyl substitution on compound **1**, loses only 7-fold affinity for the Sa (V31Y, F92I) DHFR enzyme, validating that the unsubstituted phenyl is less affected by mutations at the Val 31 position. In contrast, compound **4**, with an ethyl group at the C6 position of the diaminopyrimidine ring, is affected by the Val 31 mutation (exhibiting 10.5-, 16-, and 28-fold losses with each of the enzymes mutated at Val 31) because the ethyl group has extensive van der Waals contacts with Val 31.

Mutations at the Phe 92 position affect the interactions with the propargyl substitutions on the compounds. As such, these mutations reduce the affinity of compound **3**, with a *gem*-dimethyl at the propargylic position, by 7–18-fold. Compound **5**, with a para-biphenyl ring instead of the meta-biphenyl of compound **1**, has lower affinity for the wild-type enzyme ($K_i = 24$ nM); analysis of cocrystal structures shows that the para-biphenyl ring juts out of the active site (13). Interestingly, the orientation of the para-biphenyl ring in these structures is similar to the new orientation adopted by the meta-biphenyl ring of compound **1** in the Sa (V31Y, F92I) DHFR structure.

Conclusions

Predicting potential resistance mutations in an enzyme target in silico would enable the design of compounds early in the drug design process that overcome these limitations. Towards that goal, we used the protein design algorithm, K^*, to predict mutations in DHFR from *S. aureus*. Positive design selected mutations that maintain binding to the substrate, dihydrofolate; negative design selected mutations that lower the affinity for a lead inhibitor and the intersection of these sets resulted in a ranked series of double mutants. Four of the ten top-ranked mutants were chosen for experimental validation. Excitingly, as predicted, three of the mutants maintain catalytic activity and show lower affinity for the inhibitor. A crystal structure of the top-ranked mutant further validates the predicted conformations of the mutated residues and reveals a conformation of the ligand with many fewer interactions than are apparent in the wild-type structure.

Materials and Methods

K^*. The following mutations (one-letter amino acid codes) were allowed for the ten active site residues: Leu5, Val6, Leu20, Leu28, Val31, and Ile50—AVLIMFWY; Asp27—DE; Phe92—AVLIMFWYS. Residues Thr46 and Leu54 were modeled using rotamers but were not allowed to mutate. K^* performed 2-point mutation searches, in which any two of the ten active site residues were allowed to simultaneously mutate, while the other eight residues were allowed to change rotamers. These searches yielded a total of 1,173 candidate mutants (that included the wild-type, 1-, and 2-point mutations), corresponding to 4.37×10^{10} (DHFR:NADPH:dihydrofolate) and 2.49×10^{12} (DHFR:NADPH:compound **1**) conformations, for which K^* scores were computed (see below). The use of a factor of ~100 more conformations to predict mutants with reduced binding to compound **1** reflects the need for greater conformational flexibility for negative design.

NADPH and a steric shell of residues within 5 Å from the active site or 8 Å from dihydrofolate were included as part of the input structure. The substrate was modeled as flexible using rotamers and was allowed to rotate/translate. Modal values from the Penultimate rotamer library (16) were used for the amino acid side chains. Ligand rotamers were defined for dihydrofolate and compound **1** by sampling sets of rotatable bonds (see SI Text). For all rotamers, each dihedral was allowed to minimize within ±9° from its initial value.

For each of the candidate mutants, positive design (dihydrofolate as substrate) and negative design (compound **1** as substrate) computations were performed and scored. The ratio of the K^* scores for a given mutant with dihydrofolate and compound **1** was used to rank the candidate mutants; mutants with high ratios were predicted to be good dihydrofolate and poor compound **1** binders. The energy function consisted of the Amber vdW, electrostatic, and dihedral energies (27), and the EEF1 implicit pairwise solvation model (28) (see SI Text).

A K^* score for a given mutant was computed as the ratio of the partition functions for the bound protein-ligand complex and the free protein and free ligand. Partition functions were computed over Boltzmann-weighted ensembles of conformations: a partition function q for a given ensemble S of conformations was computed as: $q = \sum_{s \in S} \exp(-E_s/RT)$, where E_s is the energy of conformation s, T is the temperature in Kelvin, and R is the gas constant. For each partition function computation, initial rotamer pruning based on the MinDEE algorithm (4) was first applied as a preprocessing step to reduce the number of candidate rotamers and rotamer-based conformations. MinDEE is provably accurate with continuously flexible rotamers defined over a bounded (but continuous) region of side-chain conformation space. Following, A^* enumerated the remaining unpruned conformations in order of increasing lower bounds on their energies. For computational efficiency, a provably accurate ϵ-approximation algorithm was applied to guarantee the accuracy of the computed partition function based only on a small fraction of the remaining unpruned conformations (4). Using this ϵ-approximation algorithm, the A^* enumeration could be provably halted once the computed partition function was guaranteed to be within ϵ from the full partition function (when all rotameric conformations are taken into account). An ϵ value of 0.03 guaranteed that the computed partition functions were at least 97% of the corresponding full partition functions (1, 4). The MinDEE/A^* and K^* algorithms return a gap-free list of predictions in the order of the predicted score (either empirical molecular mechanics energy [MinDEE], or free energy based on molecular ensembles [K^*]).

Mutant Enzyme Preparation. Sa DHFR mutants ranked #1, 3, 7, and 9 by the K^* algorithm were prepared with site-directed mutagenesis using the DNA encoding Sa (WT) DHFR as a template. All clones were verified by DNA sequencing. Mutant enzymes were expressed in *E. coli* BL21(DE3) cells and purified using procedures previously adapted for the wild-type protein (13).

Enzyme Assays. Enzyme activity and inhibition assays were performed in triplicate for each DHFR mutant by monitoring the rate of NADPH oxidation by the DHFR enzyme at an absorbance of 340 nm (13). Kinetic parameters were measured by performing triplicate enzyme activity assays at varying substrate concentrations of dihydrofolate and analyzed with Lineweaver-Burk plots. K_M values, in addition to the obtained IC$_{50}$ values, were used to calculate K_i values for each enzyme and inhibitor (26). K_i values were also determined experimentally for compound **1** and the wild-type and Sa (V31Y, F92I) enzymes using four concentrations of substrate (50, 100, 150, and 200 μM) and four concentrations of inhibitor near the IC$_{50}$ value (50, 75, 100, and 150 nM for wild-type and 150, 300, 450, and 600 nM for Sa(V31Y,F92I)DHFR). The inhibition data were analyzed with Dixon plots.

Enzyme Stability by Circular Dichroism. Temperature-induced unfolding experiments were performed for both Sa (WT) and Sa (V31Y, F92I) DHFR by increasing the temperature from 5 °C to 70 °C and monitoring circular dichroism at 222 nm. CD measurements were taken at 1 °C temperature increments with an equilibration time of 2 min. Protein concentrations used in these experiments were 12 μM for Sa (WT) DHFR and 17 μM for Sa (V31Y, F92I) DHFR. To account for the difference in enzyme concentration, the data points were converted from millidegrees to molar ellipticity and then plotted to determine T_m values.

Crystallization. Both Sa DHFR and Sa (V31Y, F92I) DHFR were crystallized using hanging-drop vapor diffusion. The purified enzymes (12 mg/mL) were incubated with compound **1** (1 mM) and NADPH (2 mM) for 2 h on ice. Crystals of the protein:ligand:NADPH complex were optimized in a crystallization solution containing 15% PEG MW10,000, 150 mM sodium acetate, and 100 mM MES pH 6.5.

Data Collection and Refinement. Diffraction data with amplitudes extending to 2.1 Å or 3.15 Å were collected at National Synchrotron Light Source (beamline X29 or X25) for complexes of Sa DHFR and Sa (V31Y, F92I) DHFR, respectively. Data were indexed and scaled using HKL 2000 (29). Programs Coot (30) and Refmac (31) were used to build and refine the structure until

an acceptable R_{cryst} and R_{free} were achieved. The geometry of the structure was validated using Procheck (32) and Ramachandran plots. Data collection and refinement statistics are reported in Table 1.

Software. All software is freely available open-source upon publication.

1. Chen C, Georgiev I, Anderson A, Donald B (2009) Computational structure-based redesign of enzyme activity. *Proc Natl Acad Sci USA* 106:3764–3769.
2. Georgiev I, Donald B (2007) Dead-end elimination with backbone flexibility. *Bioinformatics* 23:i185–194.
3. Georgiev I, Keedy D, Richardson J, Richardson D, Donald B (2008) Algorithm for backrub motions in protein design. *Bioinformatics* 24:i196–204.
4. Georgiev I, Lilien R, Donald B (2008) The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *J Comput Chem* 29:1527–1542.
5. Jiang L, et al. (2008) De Novo computational design of retro-aldol enzymes. *Science* 319:1387–1391.
6. Rothlisberger D, et al. (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453:190–195.
7. Lippow S, Wittrup K, Tidor B (2007) Computational design of antibody affinity improvement beyond in vivo maturation. *Nat Biotechnol* 25:1171–1176.
8. Kuhlman B, et al. (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–1368.
9. Altman M, Nalivaika E, Prabu-Jeyabalan M, Schiffer C, Tidor B (2008) Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins* 70:678–694.
10. Verkihiver G (2007) Computational proteomics of biomolecular interactions in the sequence and structure space of the tyrosine kinome: Deciphering the molecular basis of the kinase inhibitors selectivity. *Proteins* 66:912–929.
11. Ishikita H, Warshel A (2008) Predicting drug-resistant mutations of HIV protease. *Angewandte Chemie International Edition* 47:697–700.
12. Cao Z, et al. (2005) Computer prediction of drug resistance mutations in proteins. *Drug Discov Today* 10:521–529.
13. Frey K, et al. (2009) Crystal structures of wild-type and mutant methicillin-resistant *Staphylococcus aureus* dihydrofolate reductase reveal an alternative conformation of NADPH that may be linked to trimethoprim resistance. *J Mol Biol* 387:1298–1308.
14. Frey K, Lombardo M, Wright D, Anderson A (2010) Towards the understanding of resistance mechanisms in clinically isolated, trimethoprim-resistant, methicillin-resistant *Staphylococcus aureus* dihydrofolate reductase. *J Struc Biol* 170:93–97.
15. Dale G, et al. (1997) A single amino acid substitution in *Staphylococcus aureus* dihydrofolate reductase determines trimethoprim resistance. *J Mol Biol* 266:23–30.
16. Lovell S, Word J, Richardson J, Richardson D (2000) The penultimate rotamer library. *Proteins* 40:389–408.
17. Hu Z, et al. (2006) Fitness comparison of thymidine analog resistance pathways in human immunodeficiency virus type 1. *J Virol* 80:7020–7027.
18. Larder B, Coates K, Kemp S (1991) Zidovudine-resistant human immunodeficiency virus selected by passage in cell culture. *J Virol* 65:5232–5236.
19. Larder B, Kemp S (1989) Multiple mutations in HIV-1 reverse transcriptase confer high-level resistance to zidovudine (AZT). *Science* 246:1155–1158.
20. Rong L, Dahari H, Ribeiro R, Perelson A (2010) Rapid emergence of protease inhibitor resistance in hepatitis C virus. *Science Translational Medicine* 2.
21. Yatsuji H, et al. (2006) Emergence of a novel lamivudine-resistant hepatitis B virus variant with a substitution outside the YMDD motif. *Antimicrob Agents Chemother* 50:3867–3874.
22. Wueppenhorst N, Stueger H, Kist M, Glocker E (2009) Identification and characterization of triple- and quadruple-resistant *Helicobacter pylori* clinical isolates in Germany. *J Antimicrob Chemother* 63:648–653.
23. Leartsakulpanich U, et al. (2002) Molecular characterization of dihydrofolate reductase in relation to antifolate resistance in *Plasmodium vivax*. *Mol Biochem Parasitol* 119:63–73.
24. Ercikan-Abali E, et al. (1996) Active site-directed double mutants of dihydrofolate reductase. *Cancer Res* 56:4142–4145.
25. Sirawaraporn W, Sathitkul T, Sirawaraporn R, Yuthavong Y, Santi D (1997) Antifolate-resistant mutants of *Plasmodium falciparum* dihydrofolate reductase. *Proc Natl Acad Sci USA* 94:1124–1129.
26. Cheng Y, Prusoff W (1973) Relationship between the inhibition constant (Ki) and the concentration of inhibitor which causes 50 percent inhibition (I50) of an enzymatic reaction. *Biochem Pharmacol* 22:3099–3108.
27. Cornell W, et al. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197.
28. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. *Proteins* 35:133–152.
29. Otwinowski Z, Minor W (1997) Macromolecular Crystallography. *Method Enzymol*, eds CW Carter and RM Sweet 276 (Academic Press, New York), 307–326.
30. Emsley P, CoWTan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
31. Murshudov G, Vagin A, Dodson E (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
32. Laskowski R, MacArthur M, Moss D, Thornton J (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.

Supporting Information

Frey et al. 10.1073/pnas.1002162107

SI Text

SI Materials and Methods. ***Ligand rotamer definitions.*** As a modeling improvement (vs. redesigns in which substrates are modeled with fixed geometries), ligand rotamers were defined for dihydrofolate and compound **1** by sampling sets of rotatable bonds. Based on the input dihydrofolate structure and eighteen folate structures from the Protein Data Bank (PDB), eight dihydrofolate rotamers defined over ten rotatable bonds were selected. In the case of compound **1**, 512 rotamers defined over four rotatable bonds were selected. The choice of substrate rotamers aimed at improving the accuracy of the model, while keeping the redesigns computationally feasible.

Positive and negative design. For each of the candidate mutants, two separate computations were performed: a positive design and a negative design. Positive design was performed with dihydrofolate as substrate and aimed at identifying mutants with similar or improved binding to dihydrofolate, as determined by the respective K^* scores. Negative design was performed for compound **1** as substrate and aimed at identifying dihydrofolate reductase (DHFR) mutants with significantly decreased or ablated binding to compound **1**, as determined by the respective

K^* scores. The ratio of the K^* scores for a given mutant with dihydrofolate and compound **1** was then used to rank the candidate mutants; mutants with high ratios were ranked higher since such mutants were predicted to be good dihydrofolate and poor compound **1** binders.

Energy function. A distance-dependent dielectric of 6, a solvation energy scaling factor of 0.05, and a vdW radii scaling factor of 0.95 were used. While the computed energies are in units of kcal/mol (see, e.g., Table 2), the computed values are only used for the relative comparison of candidate conformations and mutations and are not a quantitative measure of biophysical energies. Rotameric conformations with initial steric overlap (before minimization) of more than 1.5 Å for a pair of atoms were pruned from consideration, and their energies were not computed; hydrogens were not considered in these steric checks.

Determining the unbound structures for K^*. The unbound structures were obtained by removing the ligand from the starting bound structure.
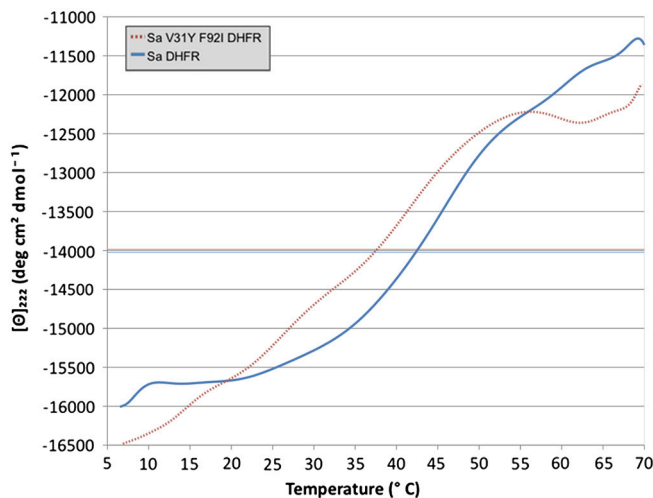
Fig. S1. Circular dichroism data shown as molar ellipticity at 222 nm versus temperature.

Fig. S2. Dixon plot for determining the K_i value of the wild-type enzyme and compound **1**.

Fig. S3. Dixon plot for determining the K_i value of Sa (V31Y, F92I) DHFR and compound **1**.

Table S1. The top ten [a]mutants as ranked by the *score ratio* criterion are shown with the respective K^* scores with [b]dihydrofolate (DHF).

			Compound **1**			
			Lowest energy[c]		Partition function[d]	
Rank	Mutations[a]	DHF K^* Score[b]	FP[e]	PL[f]	FP[g]	PL[h]
1	**V31Y/F92I**	4.30×10^{40}	−399.0	−180.0	6.6×10^{292}	2.1×10^{132}
2	V31Y/F92V	3.81×10^{40}	−410.8	−195.5	2.9×10^{301}	4.3×10^{143}
3	**V31Y/F92S**	3.13×10^{40}	−450.8	−236.0	1.2×10^{331}	2.4×10^{173}
4	V31Y/F92A	2.94×10^{40}	−449.9	−233.8	1.3×10^{330}	4.4×10^{171}
5	V31Y/F92M	6.77×10^{38}	−453.2	−229.9	3.4×10^{332}	1.0×10^{169}
6	V31Y/F92L	6.38×10^{38}	−436.6	−204.9	2.3×10^{320}	2.7×10^{150}
7	**V31F/F92L**	6.01×10^{33}	−446.0	−423.2	2.3×10^{327}	2.3×10^{310}
8	I50W/F92M	7.70×10^{32}	−464.7	−447.5	8.8×10^{340}	5.0×10^{327}
9	**I50W/F92S**	2.74×10^{32}	−462.1	−444.1	1.9×10^{339}	3.2×10^{325}
10	I50W/F92A	2.10×10^{32}	−461.0	−441.2	1.9×10^{338}	3.9×10^{323}

For each of the mutants, the computed [c]lowest energy (in kcal/mol) and [d]partition function for the [e,g]free protein (FP) and [f,h]protein-ligand (PL) complex (with compound **1**) are also shown; all of the top 10 mutants had a K^* score of 0 for compound **1**. Mutants shown in bold were selected for experimental validation

Table S2. Inhibition assays (K_i values in nM) for enzymes and propargyl-linked antifolates

Compound	Sa V31Y, F92I	Sa V31Y, F92S	Sa V31F, F92L	Sa WT	Fold loss (V31Y, F92I/WT)	Fold loss (V31Y, F92S/WT)	Fold loss (V31F, F92L/WT)
1	180	87	130	10	18	8.7	13
2	60	49	100	8.5	7.1	5.8	11.8
3	65	160	81	9.0	7.2	17.8	9.0
4	100	150	270	9.5	10.5	15.8	28.4
5	86	140	270	24	3.6	5.8	11.3