

## RESEARCH ARTICLE

# Thermal Stabilization of Dihydrofolate Reductase Using Monte Carlo Unfolding Simulations and Its Functional Consequences

Jian Tian<sup>1,2</sup>, Jaie C. Woodard<sup>3</sup>, Andrew Whitney<sup>1</sup>, Eugene I. Shakhnovich<sup>1\*</sup>

**1** Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, United States of America, **2** Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China, **3** Graduate Program in Biophysics, Harvard University, Cambridge, Massachusetts, United States of America

 These authors contributed equally to this work.

\* [shakhnovich@chemistry.harvard.edu](mailto:shakhnovich@chemistry.harvard.edu)



## OPEN ACCESS

**Citation:** Tian J, Woodard JC, Whitney A, Shakhnovich EI (2015) Thermal Stabilization of Dihydrofolate Reductase Using Monte Carlo Unfolding Simulations and Its Functional Consequences. *PLoS Comput Biol* 11(4): e1004207. <https://doi.org/10.1371/journal.pcbi.1004207>

**Editor:** James M. Briggs, University of Houston, United States of America

**Received:** December 11, 2014

**Accepted:** February 20, 2015

**Published:** April 23, 2015

**Copyright:** © 2015 Tian et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by NIH grant R01 GM06870 (EIS) [www.nih.gov](http://www.nih.gov); NIH Biophysics training grant (JCW) [www.nih.gov](http://www.nih.gov); National High Technology Research and Development Program of China (863 Program, 2013AA102804) [http://www.most.gov.cn/eng/programmes1/200610/t20061009\\_36225.htm](http://www.most.gov.cn/eng/programmes1/200610/t20061009_36225.htm); and the National Natural

## Abstract

Design of proteins with desired thermal properties is important for scientific and biotechnological applications. Here we developed a theoretical approach to predict the effect of mutations on protein stability from non-equilibrium unfolding simulations. We establish a relative measure based on apparent simulated melting temperatures that is independent of simulation length and, under certain assumptions, proportional to equilibrium stability, and we justify this theoretical development with extensive simulations and experimental data. Using our new method based on all-atom Monte-Carlo unfolding simulations, we carried out a saturating mutagenesis of Dihydrofolate Reductase (DHFR), a key target of antibiotics and chemotherapeutic drugs. The method predicted more than 500 stabilizing mutations, several of which were selected for detailed computational and experimental analysis. We find a highly significant correlation of  $r = 0.65\text{--}0.68$  between predicted and experimentally determined melting temperatures and unfolding denaturant concentrations for WT DHFR and 42 mutants. The correlation between energy of the native state and experimental denaturation temperature was much weaker, indicating the important role of entropy in protein stability. The most stabilizing point mutation was D27F, which is located in the active site of the protein, rendering it inactive. However for the rest of mutations outside of the active site we observed a weak yet statistically significant *positive* correlation between thermal stability and catalytic activity indicating the lack of a stability-activity tradeoff for DHFR. By combining stabilizing mutations predicted by our method, we created a highly stable catalytically active *E. coli* DHFR mutant with measured denaturation temperature 7.2°C higher than WT. Prediction results for DHFR and several other proteins indicate that computational approaches based on unfolding simulations are useful as a general technique to discover stabilizing mutations.

Science Foundation of China (NSFC, Grant no. 31371748)(JT) <http://www.nsf.gov.cn>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author Summary

All-atom molecular simulations have provided valuable insight into the workings of molecular machines and the folding and unfolding of proteins. However, commonly employed molecular dynamics simulations suffer from a limitation in accessible time scale, making it difficult to model large-scale unfolding events in a realistic amount of simulation time without employing unrealistically high temperatures. Here, we describe a rapid all-atom Monte Carlo simulation approach to simulate unfolding of the essential bacterial enzyme Dihydrofolate Reductase (DHFR) and all possible single point-mutants. We use these simulations to predict which mutants will be more thermodynamically stable (*i.e.*, reside more often in the native folded state vs. the unfolded state) than the wild-type protein, and we confirm our predictions experimentally, creating several highly stable and catalytically active mutants. Thermally stable active engineered proteins can be used as a starting point in directed evolution experiments to evolve new functions on the background of this additional “reservoir of stability.” The stabilized enzyme may be able to accumulate a greater number of destabilizing yet functionally important mutations before unfolding, protease digestion, and aggregation abolish its activity.

## Introduction

Protein stability is an important determinant of organismal fitness and is central to the process of enzyme design for industrial applications [1–3]. Most proteins must be folded to carry out their functions *in vitro* or *in vivo*. In addition, non-functional aggregation of unfolded or partially-unfolded proteins can have a deleterious effect on the fitness of an organism and can lead to protein aggregation diseases, which include Alzheimer’s and Huntington’s, in humans [4–6]. Aggregation of poorly folded proteins can also hamper protein production for research and technological purposes [7].

While most mutations in a natural protein are destabilizing [8,9], biological proteins are not generally at their highest possible stability; some mutations will stabilize a protein, increasing the equilibrium population of the folded state [10–12]. This stabilization can be achieved by either slowing the rate of unfolding or speeding the rate of folding, depending on the role of the mutated residue in the folding nucleation process [13,14]. The unfolding temperature,  $T_m$ , at which the free energy of the folded and unfolded states coincide ( $\Delta G = 0$ ) serves as a common measure of protein stability.  $T_m$  is obtainable by experiment and, in theory, from simulation, although current molecular dynamics simulations are limited in their ability to capture full folding or unfolding trajectories of most proteins (except very small fast folding domains [15]) in a tractable amount of simulation time [16].

Several computational methods to predict protein stability or changes in stability upon mutation have been developed and tested [17–19]. However, the performance of these popular methods is still relatively weak [20–22]. Other existing techniques to rationally design proteins with improved stability have involved optimization of charge-charge interactions [23], saturation mutagenesis of residues with high crystallographic B-factors [24], methods based on protein simulation and calculation of free energies [25–27] and comparison to homologous proteins including the ultra-stable proteins of thermophiles [28,29]. We reasoned that better predictions of mutant stability might be obtained by evaluating the unfolding temperature  $T_m$  in realistic yet computationally tractable simulations of protein unfolding.

Here, we use a Monte Carlo protein unfolding approach (MCPU) with an all-atom simulation method and knowledge-based potential developed earlier in our lab [16,30,31] to simulate

unfolding and predict melting temperatures for all possible single point mutants of *E. coli* Dihydrofolate Reductase (DHFR). DHFR is an essential enzyme in bacteria and higher organisms, and it is an important target of antibiotics [32] and anti-cancer drugs [33,34]. Its moderate size (18 kDa) makes it amenable to both simulation and experiment. As described in the Materials and Methods section, the Monte Carlo move set consists of rotations about torsional angles. At high temperature, the higher entropy of unfolded states overcomes the increase in energy due to loss of favorable contacts and torsional preferences, leading to unfolding. We experimentally determine melting temperatures and catalytic activities for several predicted stabilizing mutants, and for mutants combining multiple stabilizing mutations. Our approach allows us to identify several stabilized mutants of DHFR, and our prediction method marks an improvement over existing stability predictors such as Eris [19], FoldX [17], and PopMusic [18]. Simulations of non-DHFR proteins likewise indicated that our method is useful as a general approach to simulate protein unfolding and select stabilizing mutations.

## Results

### Predicting the effects of mutations on protein stability from non-equilibrium unfolding simulations

Ideally, protein stability for any sequence should be predicted in all-atom equilibrium simulations that cover multiple folding-unfolding events to determine equilibrium populations of various states of the protein. However, despite recent progress in *ab initio* simulations of protein folding [15] this goal is not attainable for proteins of realistic size and biological relevance. Currently, non-equilibrium unfolding simulations are within reach for sufficiently large proteins and the question arises whether such simulations can be used to assess mutational effects on protein stability, which is an equilibrium property. The following analysis provides an affirmative answer to this question, under certain assumptions. Although the idea of obtaining equilibrium free energy differences from non-equilibrium measurements is not new [35], and protein stabilities have been calculated from molecular dynamics simulations using the Jarzynski equality, *e.g.*, [36–38], such simulations require application of an external steering force; in the present paper we report the use of multi-temperature Monte-Carlo unfolding simulations in obtaining protein stabilities.

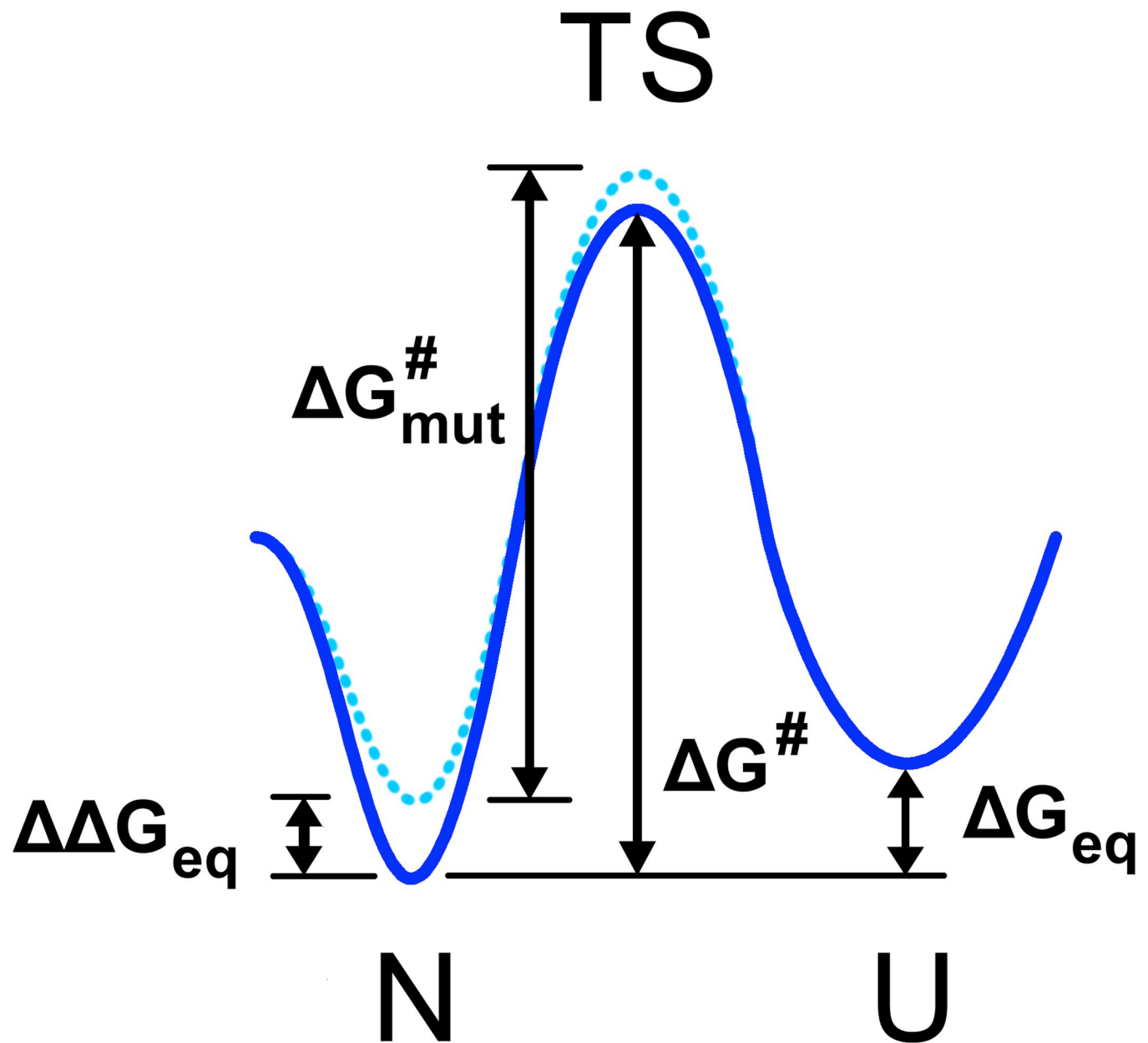
Assuming two-state unfolding kinetics [39–42] we can estimate the characteristic time required to cross the unfolding free energy barrier (in fact it is the time spent in the native state waiting for sufficient thermal fluctuation to cross the barrier) as:

$$\tau_u^{fp} = \tau_0 e^{\frac{\Delta G}{kT}} \quad (1)$$

where  $\tau_u^{fp}$  is first-passage time from the folded to the unfolded state,  $\Delta G^\ddagger$  is the free energy barrier between the folded state and the transition state for unfolding (see Fig. 1) and  $\tau_0$  is the elementary time constant. When simulation time  $\tau_{sim}$  approaches  $\tau_u^{fp}$  unfolding events are observed in simulation. The apparent “melting temperature”, *i.e.*, the temperature at which unfolding events occur in simulations, therefore depends on the simulation time  $\tau_{sim}$  according to Eq. (1):

$$kT_m^{app} = \frac{\Delta G}{\ln\left(\frac{\tau_{sim}}{\tau_0}\right)} \quad (2)$$

This analysis suggests that non-equilibrium first passage unfolding simulations are not suitable to predict the temperature at which a protein would unfold at equilibrium. However the



**Fig 1. Two state diagram depicting the protein native state N, transition state TS, and unfolded state U.** The WT energy landscape is shown in solid blue and the mutant energy landscape in dotted blue. The difference in free energy between WT folded and unfolded states,  $\Delta G_{eq}$ , the change in this quantity upon mutation,  $\Delta\Delta G_{eq}$ , and the height of the transition state barrier relative to the native state for WT ( $\Delta G^{\#}$ ) and mutant ( $\Delta G^{\#}_{mut}$ ) are labeled.

<https://doi.org/10.1371/journal.pcbi.1004207.g001>

effect of mutations on stability can be predicted from unfolding simulations. In order to see this we note that the mutational effect on protein stability  $\Delta\Delta G$  is related to the change in the unfolding free energy barrier  $\Delta\Delta G^{\#}$ , the difference between the WT barrier height and the mutant barrier height, shown in Fig 1.

$$\Delta\Delta G_i = (1 - \varphi_i)\Delta\Delta G_i^{eq} \quad (3)$$

where  $i$  denotes the mutated amino acid and  $\varphi_i$  is the  $\varphi$ -value for residue  $i$  which determines the fraction of interactions that this residue forms in the folding/unfolding transition state

[40,43,44]. We therefore obtain

$$k\Delta T_m^{app}(i) = \frac{(1 - \varphi_i)\Delta\Delta G_i^{eq}}{\ln\left(\frac{\tau_{sim}}{\tau_0}\right)} \quad (4)$$

where  $\Delta T_m^{app}(i) = T_m^{app}(i) - T_m^{app}(WT)$  is the shift in apparent unfolding temperature upon a specific mutation in the  $i$ -th residue. Introducing the relative (to WT) unfolding temperature  $\Delta T_m^{rel}(i) = \Delta T_m^{app}(i)/T_m^{app}(WT)$  we get

$$\Delta T_m^{rel}(i) = \frac{(1 - \varphi_i)\Delta\Delta G_i^{eq}}{\Delta G} \quad (5)$$

*i.e.* the mutational shift in observed unfolding temperature, normalized to the observed unfolding temperature of the wild-type at the same simulation condition does not depend on the simulation length, provided that the simulation is sufficiently equilibrated *in the native basin* so that the rules of transition state theory apply. The analysis of extensive kinetic and equilibrium data for multiple proteins shows that for the majority of mutations (except for a small fraction of residues that participate in the folding nucleus)  $\varphi_i \approx 0.24$  with remarkable accuracy and consistency [45]. We get therefore

$$\Delta T_m^{rel}(i) = 0.76 \frac{\Delta\Delta G_i^{eq}}{\Delta G} \quad (6)$$

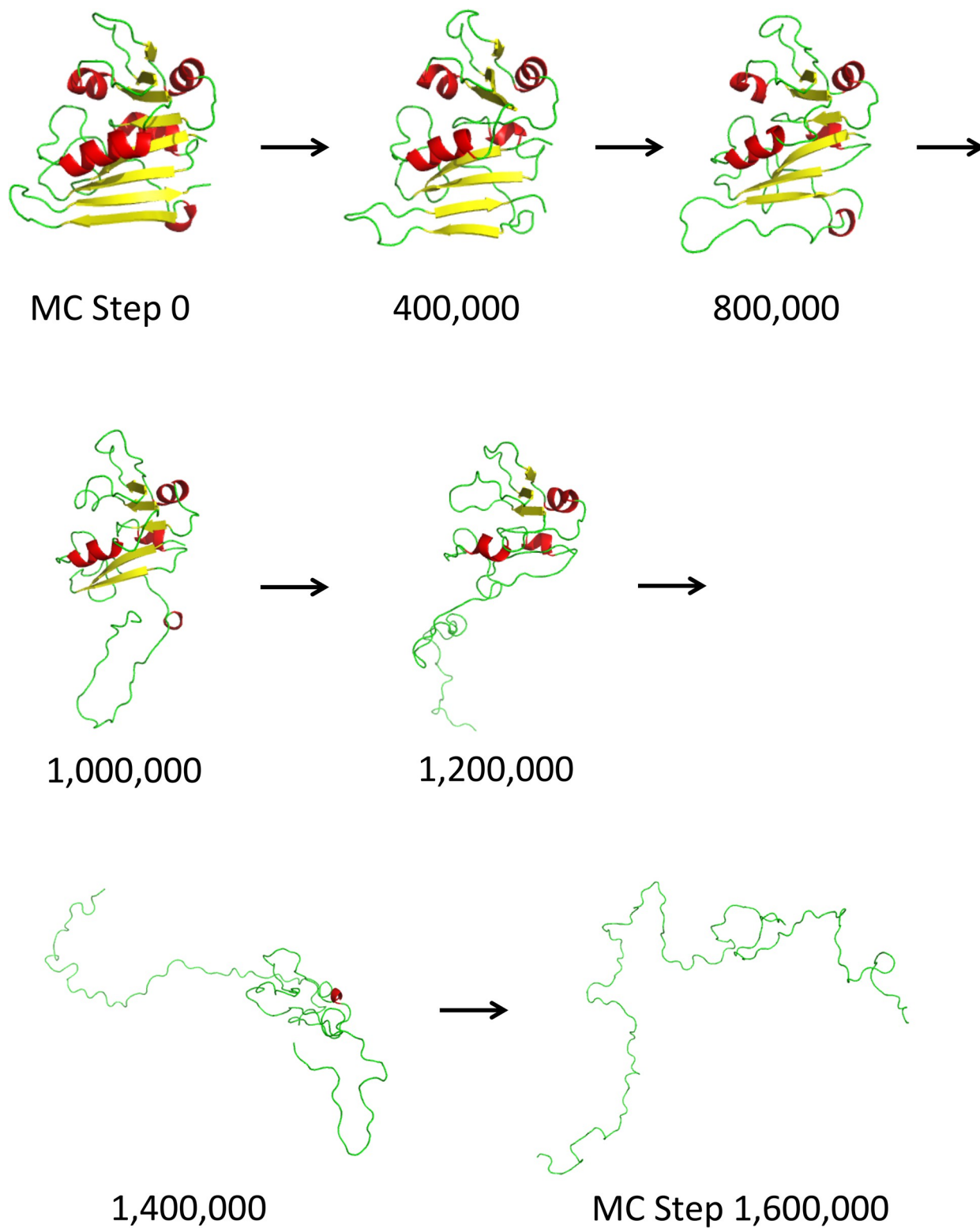
*i.e.*  $\Delta T_m^{rel}(i)$  is independent of simulation time and proportional to the equilibrium free energy effect of mutations, provided that simulations have equilibrated in the native basin of attraction.

### Monte Carlo protein unfolding simulation

We ran MCPU on DHFR (PDB ID: 4DFR) at a range of temperatures, to generate simulated unfolding curves. Unfolding steps of a sample trajectory are shown in Fig. 2, and a flowchart of the simulation method is shown in S1 Fig. The protein was subject to a brief MD energy minimization, beginning from the WT crystallographic native state, followed by unfolding simulations at each of 32 different temperatures using all-atom Monte-Carlo (see [Materials and Methods](#) section). As shown in figures S2 Fig—;S4 Fig, the RMSD and total energy increased and the number of contacts decreased as each simulation proceeded, and with increasing temperature. (Here, temperature is given in arbitrary simulation units.) Plots of RMSD and contact number vs. temperature showed sigmoidal behavior, with a clearly identifiable transition temperature, and the melting temperature ( $T_m$ ) could be determined by fitting to a sigmoidal function (Fig. 3). Plots of energy vs. temperature (S5 Fig) were sigmoid-like, but with an additional rise in energy at low to intermediate temperatures, perhaps indicating pre-melting to a dry-molten globule state with loosened side chains but native-like topology [46,47]. This deviation from sigmoidal behavior becomes clearer as the simulation length is increased (S6 Fig).

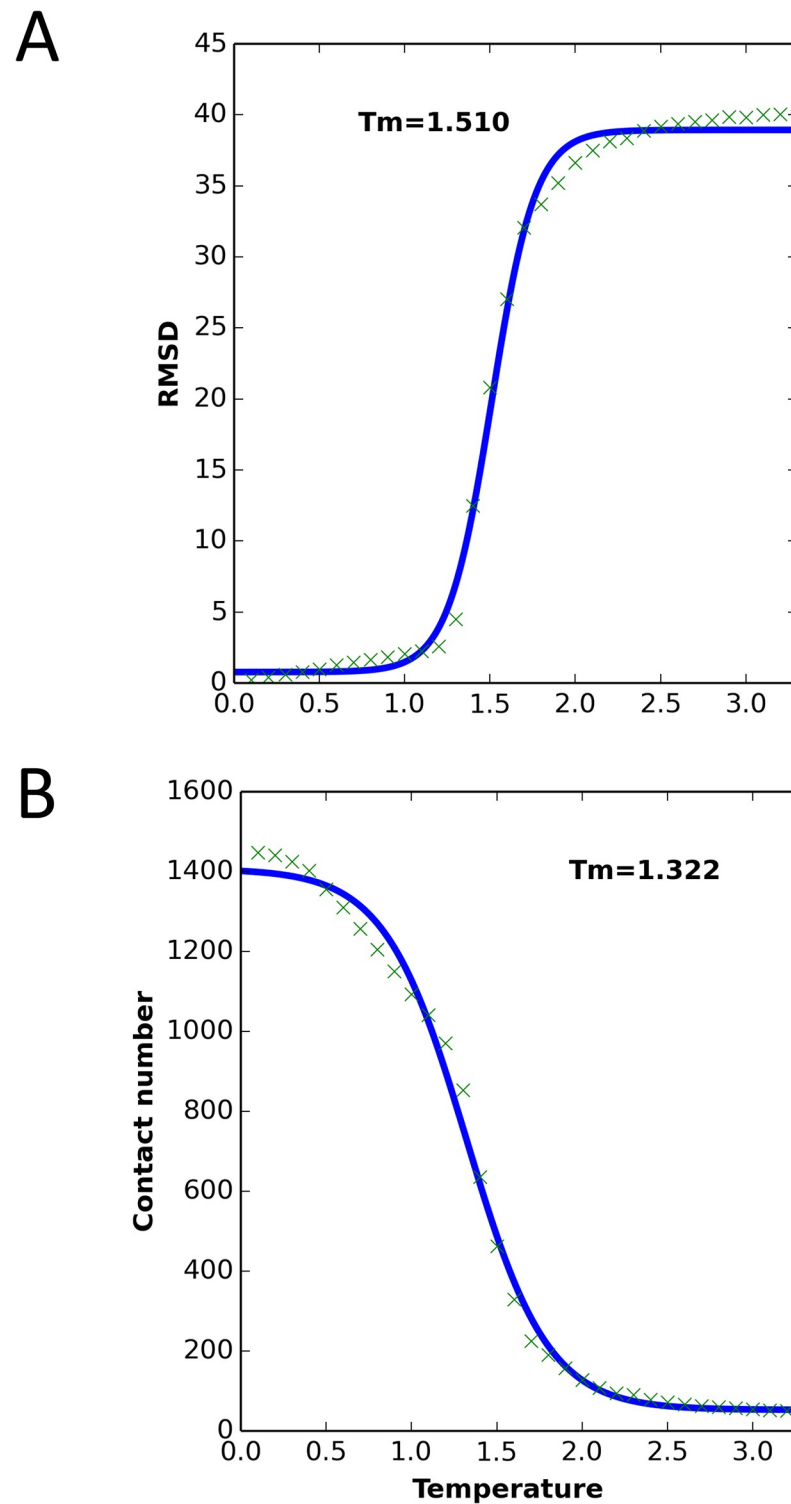
### Computational identification of stabilizing single point mutations

All possible single point mutations of DHFR ( $159 * 19 = 3,021$ ) were simulated with the Monte Carlo protein unfolding simulation protocol. The simulated  $T_m$  values were calculated as described above. Of the 3,021 mutations, 523 mutations (17.3%) were predicted to have a stabilizing effect according to all three metrics (energy, contacts, and RMSD), while 42.1% of mutations had a destabilizing effect according to all three metrics. These predictions are in good agreement with statistical analysis of published experimental data and FoldX predictions [8,12].



**Fig 2. A sample WT DHFR unfolding trajectory at simulation temperature 1.5 (arbitrary simulation units).** In MC simulations, separation of the C-terminal beta hairpin from the rest of the protein (steps 1,000,000 through 1,200,000) is an early event in the unfolding process.

<https://doi.org/10.1371/journal.pcbi.1004207.g002>



**Fig 3. WT DHFR unfolding curves from MC simulations, averaged over 2,000,000 simulation steps, with 50 replications. The  $T_m$  value was calculated based on the sigmoidal fit (solid blue line). (A) RMSD vs. simulation temperature. (B) Number of contacts vs. simulation temperature.**

<https://doi.org/10.1371/journal.pcbi.1004207.g003>



The simulated  $T_m$  values evaluated using RMSD, total energy, and number of contacts are strongly correlated, as shown in Fig. 4A. The distribution of predicted melting temperatures (averaged over the 3 metrics) for all 3021 point mutants is shown in Fig. 4B. Next, we selected a subset of predicted stabilizing mutations for subsequent in depth computational and experimental analysis. To that end we selected the loci where multiple mutations were consistently predicted as stabilizing. Out of this set we selected one mutation at each loci which were predicted as most stabilizing. As a result we arrived at 23 single predicted stabilizing point mutants shown in S1 Table, which we deemed most promising for subsequent in depth computational and experimental analysis. Furthermore, five stabilizing mutations at different sites within DHFR, shown in Fig. 5, were combined to form the multiple mutants listed in Table 1, with the rationale that the combination of individual stabilizing mutants often yields more stable proteins, and these mutants were likewise subjected to computational and experimental analysis.

### Computational test of the theoretical analysis

First we test two predictions that emerge from the theoretical analysis of unfolding simulations. The first prediction is that the apparent unfolding temperature decreases as the length of the unfolding simulation increases (Equation 4). Secondly and most importantly the mutational change in relative (normalized to WT) apparent unfolding temperature is a) robust with respect to simulation time provided that simulations have equilibrated in the native basin and b) directly proportional to the effect of mutations on equilibrium protein stability (Equation 6). We test these predictions using MCPU simulations and experiment.

We carried out two sets of MCPU simulations of different lengths: 2,000,000 and 20,000,000 steps for the 23 predicted stabilizing mutants, 15 mutants studied previously by experiment [48] (the complete set of single mutants is listed in S1 Table), and the 5 stabilizing multiple mutants combining individual mutations listed in Table 1, and compared their predicted absolute and relative simulated unfolding temperatures (Fig. 6). Indeed both predictions of our theoretical analysis are confirmed, *i.e.*, the apparent unfolding temperature decreases with simulation time (Fig. 6A) while the relative unfolding temperature  $\Delta T_m^{rel}$  is remarkably independent of simulation time (Fig. 6B). We note that due to the nature of the energy function used in our simulations, there is no obvious mapping of simulation temperature to real absolute temperature (*i.e.*, in Celsius or Kelvin). Conversion of simulation temperature to physical temperature would require use of experimental data (*e.g.*, WT unfolding temperature and deviation of temperatures over all mutants) and therefore would not provide a completely simulation- or theory-based prediction. Furthermore, as noted above, the apparent absolute value of the transition temperature in the Monte-Carlo unfolding approach depends on simulation time. Therefore, we used relative melting temperature,

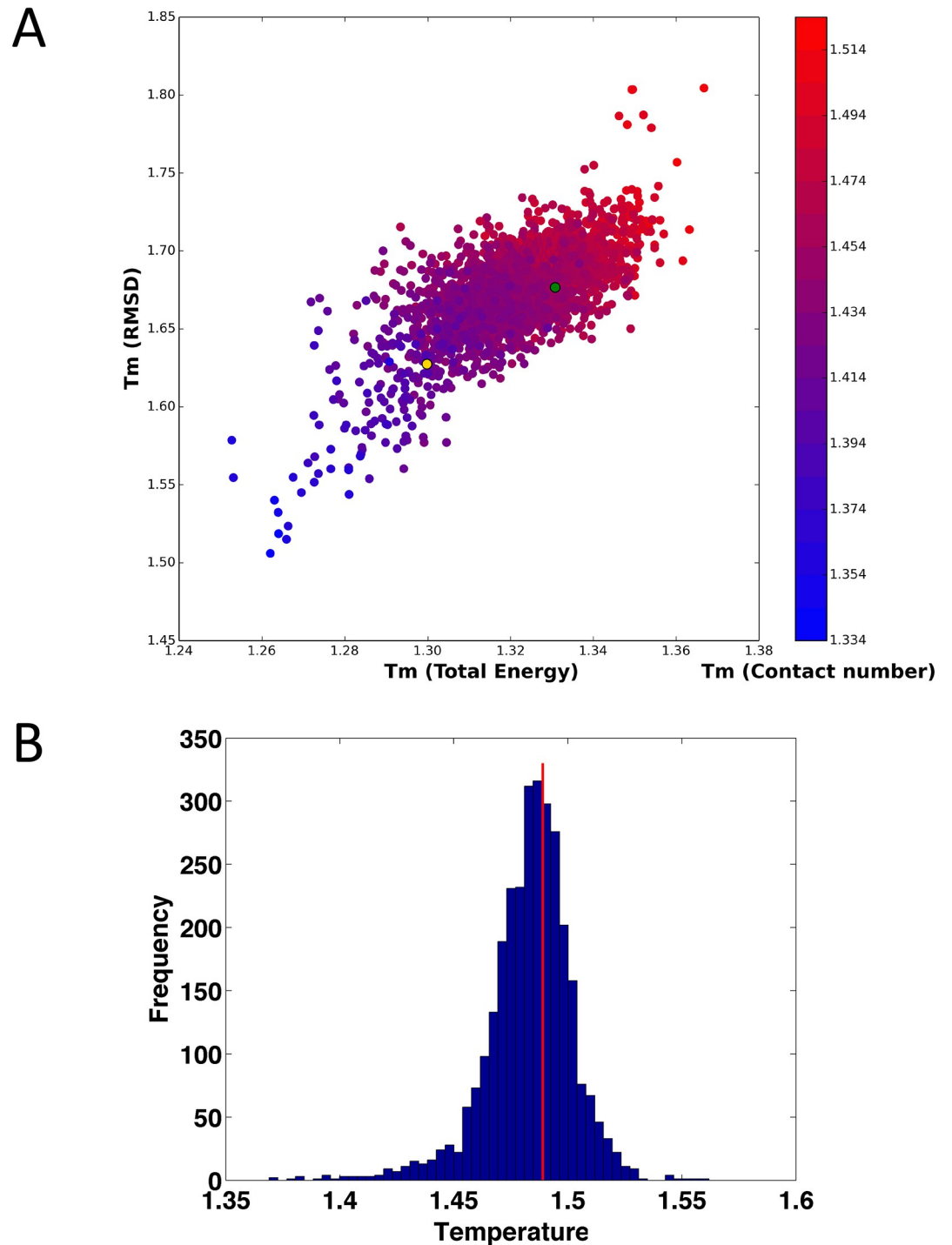
$\Delta T_m^{rel}(i) = \Delta T_m^{app}(i) / T_m^{app}(WT)$ , when comparing simulation results with experimental results.

As mentioned, the simulated  $T_m$  values evaluated using RMSD, total energy, and number of contacts are strongly correlated in our simulations as shown in Fig. 4A and S1 Table. In what follows we define the computational unfolding temperature  $T_m$  as averaged over  $T_m$  values determined using these three criteria.

### Experimental characterization of predicted mutants

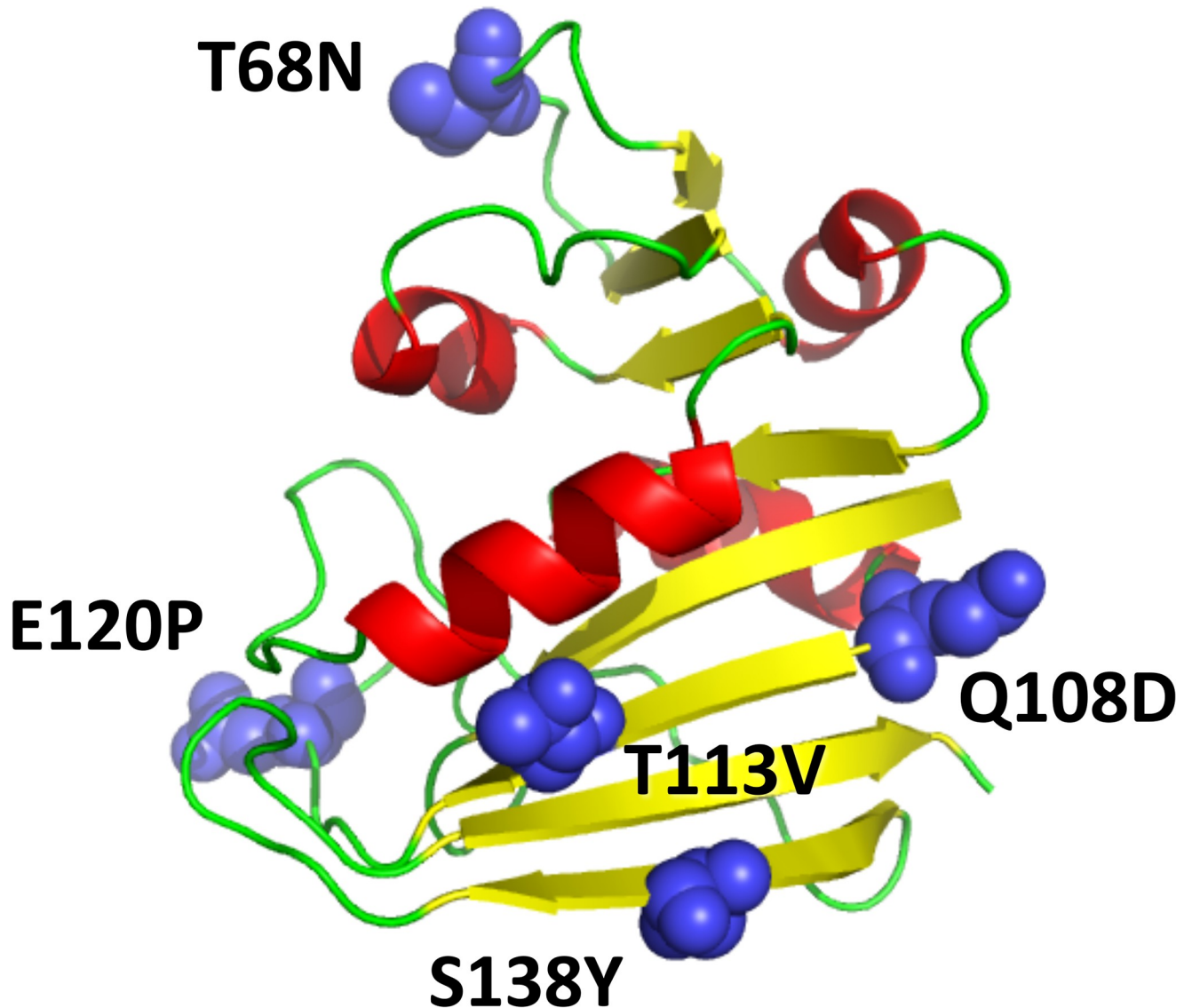
We cloned, expressed, and purified the 23 single point mutants of DHFR listed in S1 Table, as well as the multiple mutants listed in Table 1 (see Materials and Methods). The biophysical properties of the mutants were measured and compared with WT DHFR, as shown in S2 Table. As many studies have shown that oligomerization can alter protein stability [23,48,49], we first tested whether mutations induce oligomerization and/or aggregation using the gel filtration





**Fig 4. Simulated  $T_m$  values, based on RMSD, Total Energy and Contact number.** (A) Scatter plot of  $T_m$  (RMSD) vs.  $T_m$  (Total energy), with  $T_m$  (contact number) represented by color (see color bar to right of plot). The green ball denotes WT and the gold ball denotes the destabilized mutant I155A. The correlation coefficients of simulated  $T_m$  between RMSD and total energy, RMSD and Contact number, and Contact number and total energy were 0.68, 0.79 and 0.84, respectively. (B) Histogram of  $T_m$  values, determined by averaging the values obtained from RMSD, energy, and contact number. The vertical red line denotes WT  $T_m$ .

<https://doi.org/10.1371/journal.pcbi.1004207.g004>



**Fig 5.** The 3D structure of DHFR (PDB ID 4DFR), with residues that were altered in the stabilized quintuple mutant shown in blue.

<https://doi.org/10.1371/journal.pcbi.1004207.g005>

method [48,50] and light scattering. The results indicated that all of the 23 mutants were monomeric at studied concentrations except for E154V, which appeared aggregation-prone. We excluded E154V from the subsequent analysis.

As shown in S2 Table, all single mutants are catalytically active except for D27F. D27 is known to be a key catalytic residue of *E. coli* DHFR [51].

For each mutant we obtained two measures of stability: the apparent melting temperature determined by Differential Scanning Calorimetry (DSC) and the urea midpoint unfolding concentration ( $C_m$ ) determined by monitoring chemical denaturation by Circular Dichroism (CD) with subsequent fitting to a two-state model (see Materials and Methods). Both measures of stability were highly correlated, despite the fact that thermal unfolding was irreversible (S7 Fig). Of the selected 22 single point mutations, 10 mutations were stabilizing, according to their  $T_m$  or  $C_m$  values (S2 Table). Given that statistically most random mutations are destabilizing with only a small fraction (less than 18%) stabilizing [8,12], this statistically significant

**Table 1. The simulated and experimental results of the selected single point mutants and WT.**

Mutation(s)	$T_m$ (DSC)	$C_m$ (CD)	$k_{cat}$	$k_{cat}/K_m$	Simulated $T_m$
WT	54.1	3.09	24.60	14.07	1.358 ± 0.004
T113V	58.0	3.28	13.67	10.86	1.389 ± 0.004
Q108D	55.7	3.18	24.60	10.35	1.361 ± 0.004
S138Y	55.6	3.33	24.51	9.33	1.366 ± 0.004
D116F	55.5	3.43	24.80	9.53	1.369 ± 0.004
T68N	55.5	3.26	29.36	13.32	1.367 ± 0.003
E120P	55.3	3.25	30.02	13.91	1.371 ± 0.004
T68N,Q108D,T113V,E120P,S138Y	61.3	3.52	32.63	12.20	1.400 ± 0.004
T113V,E120P,S138Y	58.5	3.49	31.13	13.10	1.384 ± 0.004
T68N,Q108D,E120P,S138Y	56.4	3.47	22.80	10.94	1.377 ± 0.003
T68N,Q108D	55.8	3.14	17.99	15.24	1.366 ± 0.003
E120P,S138Y	55.6	3.29	16.01	10.81	1.371 ± 0.004

Note: The data were averaged over 50 replications. 2,000,000 MC steps were simulated in total, and the last 1,000,000 steps were used to calculate  $T_m$ .

The units:  $T_m$ : °C,  $C_m$ : M,  $k_{cat}$ : s<sup>-1</sup>,  $k_{cat}/K_m$ : s<sup>-1</sup> μM<sup>-1</sup>

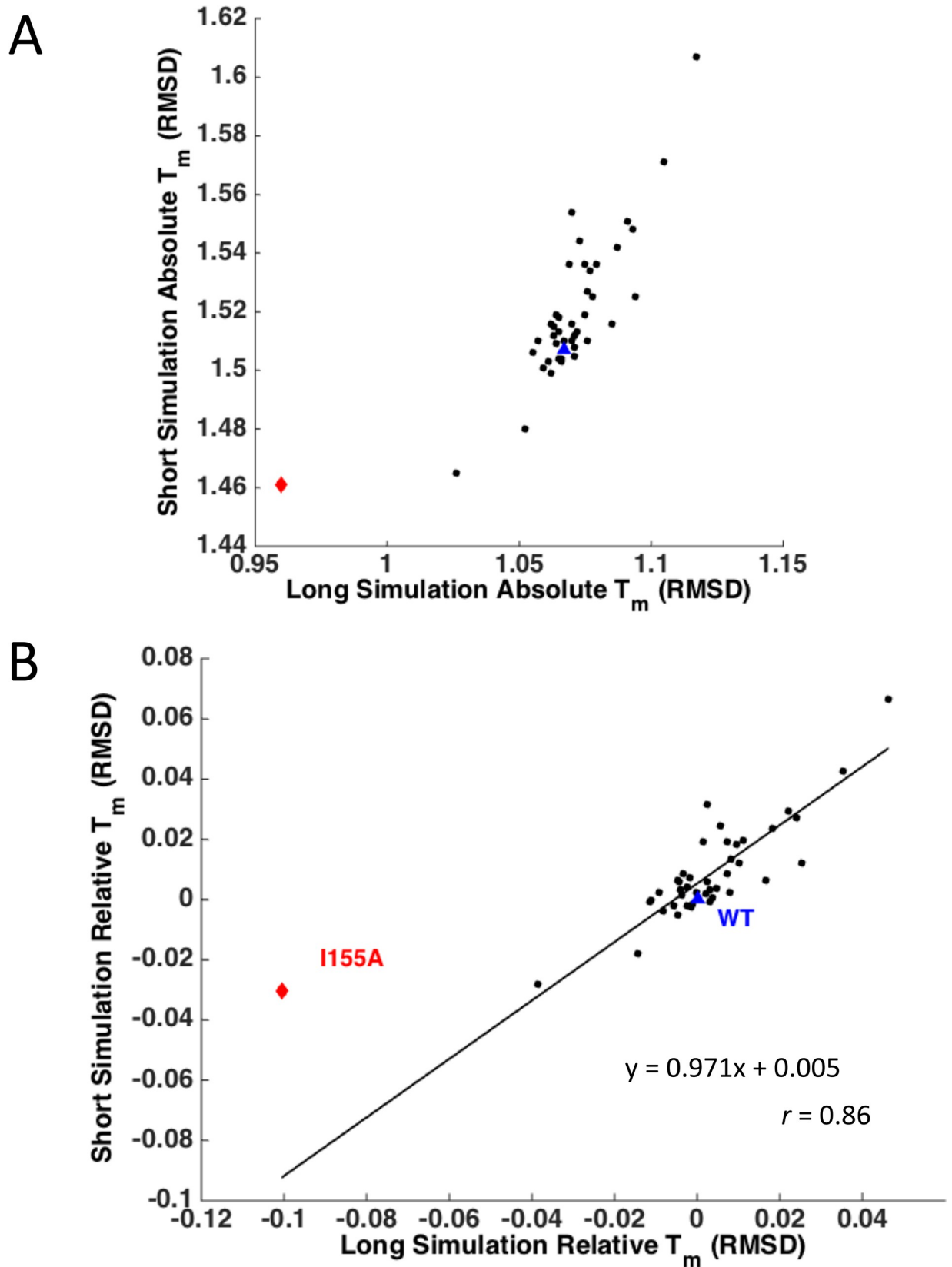
<https://doi.org/10.1371/journal.pcbi.1004207.t001>

result ( $p = 0.002$  under the null hypothesis that mutations are random) indicates that MCPU is an effective method for selecting stability-enhancing mutations.

As expected, combinations of single stabilizing mutations led to more stable multiple mutant variants, [24,25,52] as predicted by simulation. In particular, the stability of the quintuple mutant (T68N,Q108D,T113V,E120P,S138Y) was found to be substantially higher than that of the wild type protein (Table 1), with  $T_m$  7.2°C higher than WT, and  $C_m$ , the urea concentration at the mid-unfolding point, was 0.43M higher than WT. All multiple mutants were catalytically active, and the quintuple mutant and triple mutant (T113V,E120P,S138Y) were found to be more catalytically active than WT. We note that while combination of stabilizing mutations generally increases stability, the effect is less than additive (S8 Fig); for instance, the quintuple mutant is about 4°C less stable than predicted under the assumption of additive  $\Delta T_m$  (a 7.2°C stability increase vs. predicted 9.6°C).

We computationally predicted relative unfolding temperatures of 15 DHFR mutants published earlier [48] and added these mutants to the set for analysis resulting in 42 mutants in total. The correlation coefficient between experimental relative  $T_m$  and simulated relative  $T_m$  for the 42 mutants was about 0.65, as shown in Fig. 7A. To address the issue that both simulated  $T_m$  and DSC measurements are not strictly at equilibrium, we plotted the relation between simulated  $T_m$  and equilibrium measurement of stability in chemical denaturation by urea. The denaturation mid-transition urea concentration  $C_m$  and computationally determined unfolding temperature exhibit even a slightly higher correlation of  $r = 0.68$  (Fig. 7B), demonstrating that our non-equilibrium simulation method shows good agreement with the equilibrium measurement of urea denaturation, as predicted by Equation 6.

We also used the dataset to evaluate the effect of the number of replications and the number of MC steps on the performance of the method. As shown in Fig. 8A, the prediction accuracy is sensitive to the number of replications. To achieve reliable  $T_m$  predictions, at least 20 replications should be used. However, the number of MC steps did not greatly affect prediction accuracy, provided simulations were run for at least ~ 200,000 steps (see Fig. 8B). In the context of the theory developed in the earlier section: “Predicting the effects of mutations on protein stability from non-equilibrium unfolding simulations,” this initial equilibration period may allow



**Fig 6. Correlation between  $T_m$  values for simulations of different lengths.** WT is shown as a blue triangle and mutant I155A as a red diamond. (A)  $T_m$  calculated from simulation RMSD, for short (2,000,000-step) and long (20,000,000-step) simulations. Simulation  $T_m$  is clearly smaller for long simulations, in which the protein has more time to unfold. (B) Relative  $T_m$  normalized to WT, for short and long simulations. Remarkably, the points fall nearly on the line  $y = x$ , with a correlation of 0.86, with one distinct outlier I155A.

<https://doi.org/10.1371/journal.pcbi.1004207.g006>

time for equilibration within the native basin, after which simulation length does not appreciably affect the consistency of results with equilibrium stability measurements.

### Stability and activity do not trade-off for DHFR

It has been proposed that stability imposes a constraint on protein function leading to stability-activity tradeoffs [53,54]. Our data, however, paints a different picture for DHFR—of a weak *positive* correlation between  $T_m$  and  $k_{cat}$  or  $k_{cat}/K_M$  ( $r = 0.46$ ,  $p = 0.02$  and  $r = 0.41$ ,  $p = 0.03$  respectively) with one notable outlier D27F, where the stabilizing mutation is made right in the active site (Fig. 9). The D27F mutant has high thermal stability but, as noted above, is not catalytically active, indicating that there is in fact a stability-activity trade-off for this active-site residue.

### Evolutionary analysis

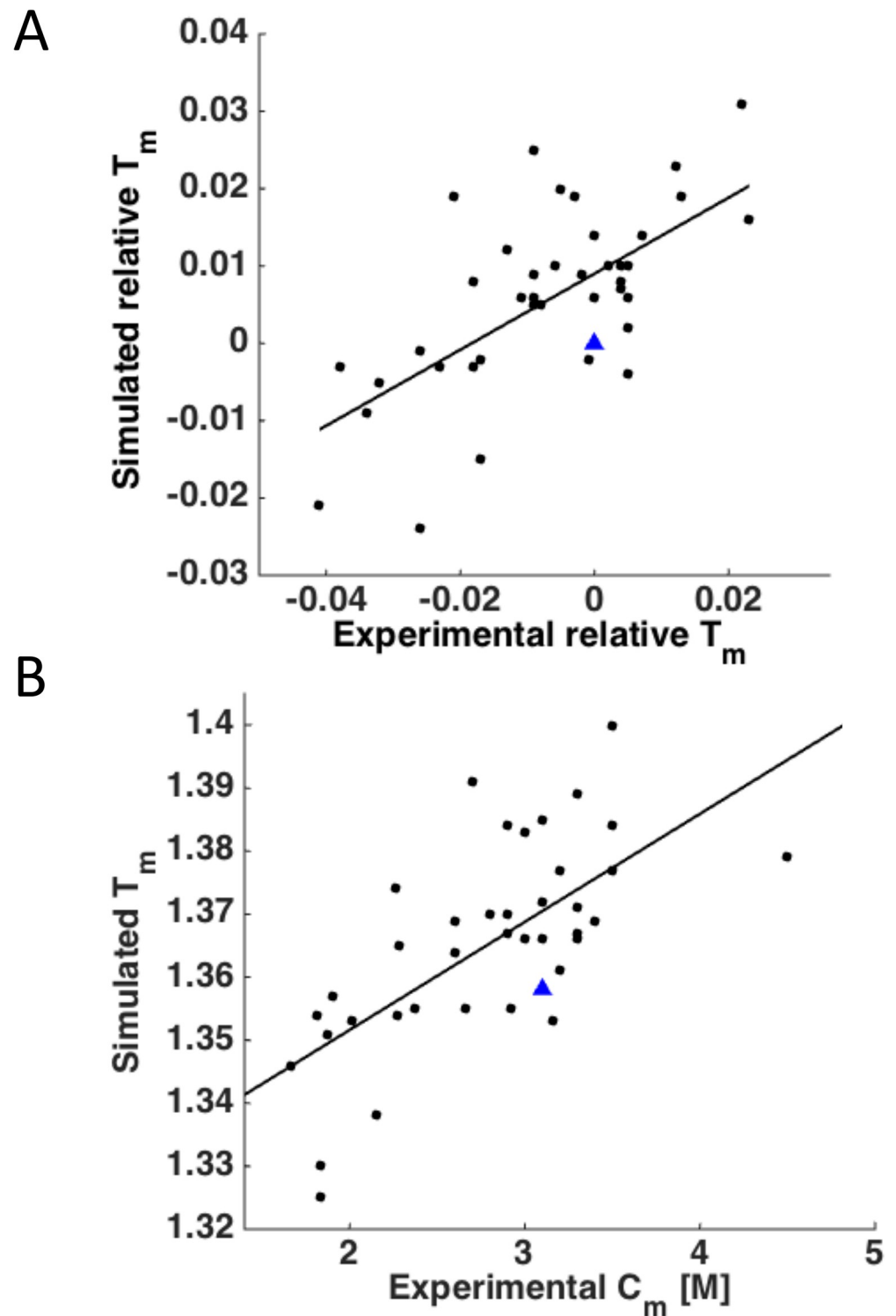
Using an alignment of 290 bacterial DHFRs, we determined the DHFR consensus sequence (S9 Fig). Mutation of a non-consensus residue to the consensus residue generally resulted in protein stabilization [29]. In 4/16 of the experimentally stabilizing mutations, a residue was changed to the consensus residue, while only 2/29 destabilizing mutations resulted from a change to consensus. Likewise, in 18/29 destabilizing mutations, a residue was changed away from the consensus residue, while this was true for only 5/16 of stabilizing mutations.

### Simulated melting temperatures by residue

We compared the minimum and maximum simulated  $T_m$  values obtainable by mutating a single residue to any of the 19 other amino acids (Fig. 10A). There is a weak positive correlation between minimum and maximum melting temperatures ( $r = 0.30$ ,  $p = 10^{-4}$ ). Apparently, protein loci where mutations can cause significant stabilization are statistically less susceptible to destabilizing mutations and vice versa, which may be expected: once a residue is already at its most stabilizing amino acid variant, the protein cannot be stabilized further by mutation. Distinct outliers correspond to the loci with the strongest stabilizing or destabilizing effects of mutations. Interestingly, these outliers, which may represent structural weak spots in DHFR, tend to fall on the interface connecting the C-terminal beta hairpin and the rest of the protein (Fig. 10B). This is in fact the interface that is the first to dissociate in the Monte Carlo simulations (see Fig. 2).

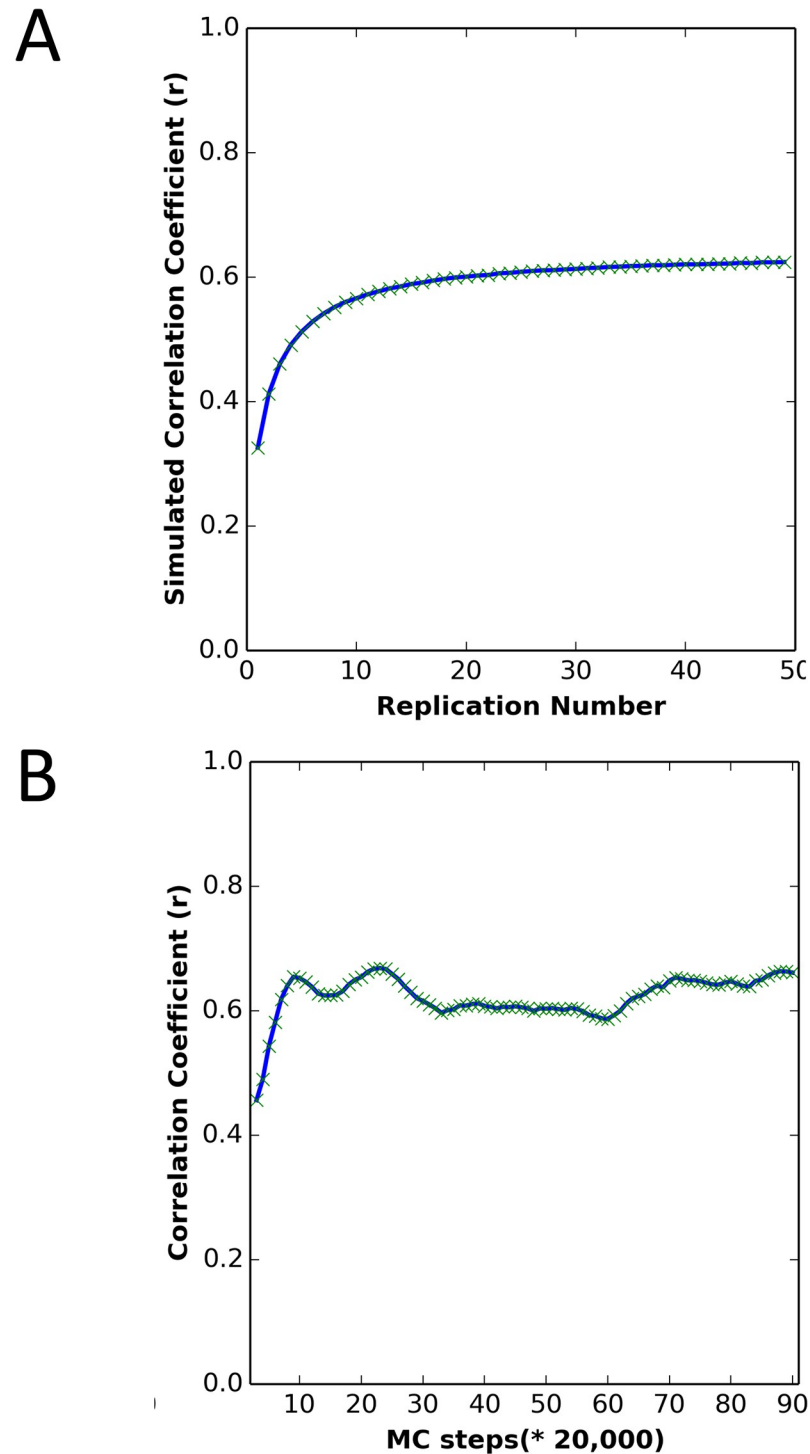
### Comparison with other methods

We compared our computational DHFR predictions with four popular approaches to predict the effect of a mutation on protein stability: FoldX [17], Eris [26], PopMusic [55], and SDM [56]. (S3 Table). The MCPU performs better than these methods on DHFR mutants. PopMusic shows also strong performance with highly statistically significant  $r = 0.55$  between theory and experiment, however the limitation of this method is that it can consider only single point mutations. To further evaluate MCPU performance we tested it on four additional proteins from four different SCOP structural classes: the Cro repressor protein from bacteriophage lambda (PDB-ID 5CRO), the *B. Subtilis* major cold shock protein (1CSP), *E. coli* Thioredoxin



**Fig 7. Correlation between the relative simulated and experimental  $T_m$  values.** (A) Plot of simulated  $T_m$  vs. experimental  $T_m$ . The relative  $T_m$  values were calculated by normalizing to WT:  $(T_m(\text{mutant}) - T_m(\text{wild type})) / T_m(\text{wild type})$ . Experimental values from this study and from Bershtein *et al.* [48] are included. WT is shown as a blue triangle.  $r = 0.65$ ,  $p = 3 \times 10^{-6}$ . (B) Plot of simulated  $T_m$  vs. experimental  $C_m$ .  $r = 0.68$ ,  $p = 6 \times 10^{-7}$ .

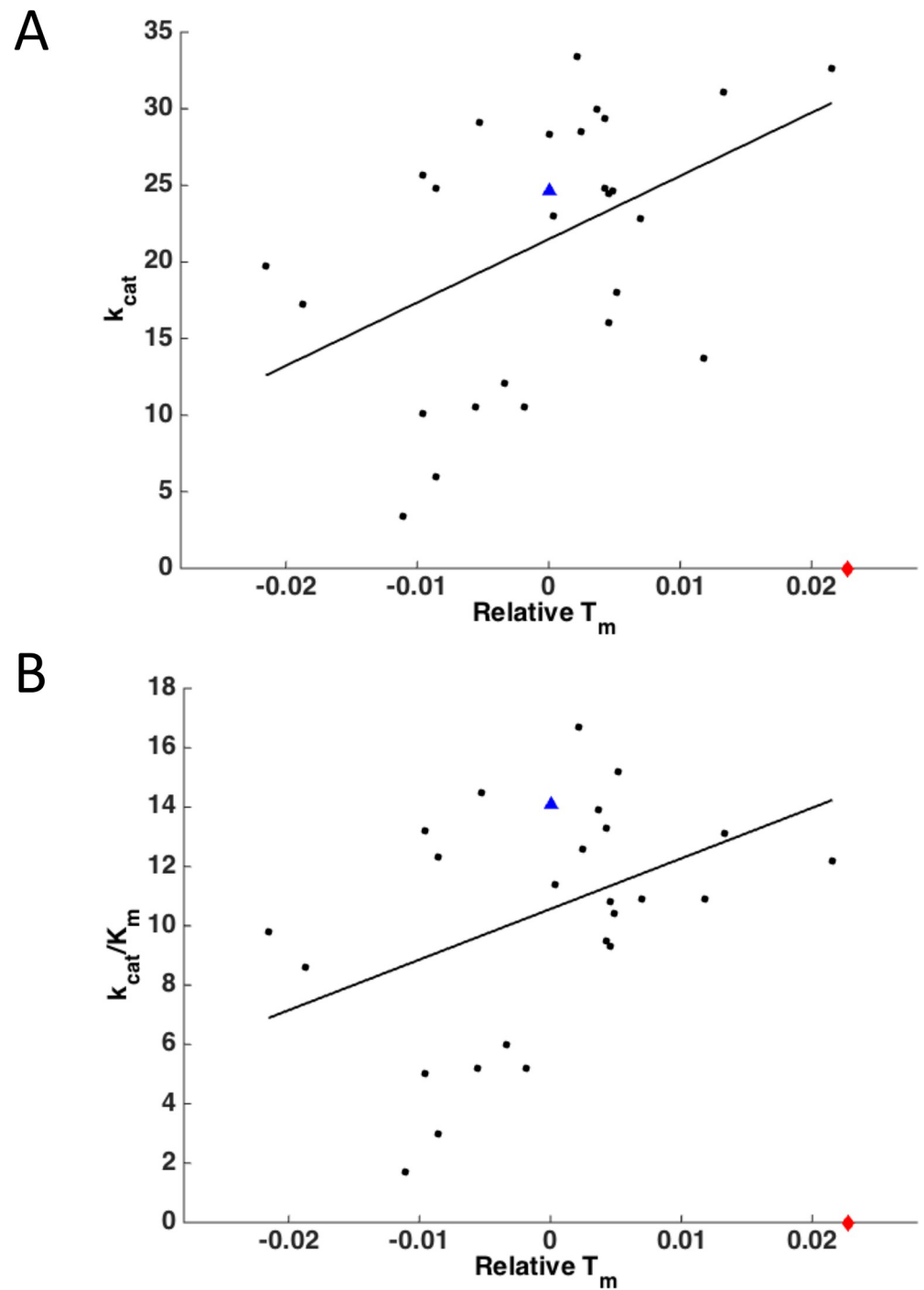
<https://doi.org/10.1371/journal.pcbi.1004207.g007>



**Fig 8. The effect of replication number and number of MC steps on simulation predictive power.** (A) Correlation between simulated  $T_m$  and experimental  $T_m$ , averaging over different numbers of replications, for the DHFR wild type and mutants. Each protein was simulated for 2,000,000 MC steps, following MD minimization and equilibration at low temperature. (B) Correlation between the simulated  $T_m$  and experimental  $T_m$  with different numbers of MC steps and 50 replications, for the DHFR wild type and mutants. Each protein was first simulated for the number of steps given on the x-axis, and the next 100,000 steps were averaged in determining the simulated  $T_m$ .

<https://doi.org/10.1371/journal.pcbi.1004207.g008>

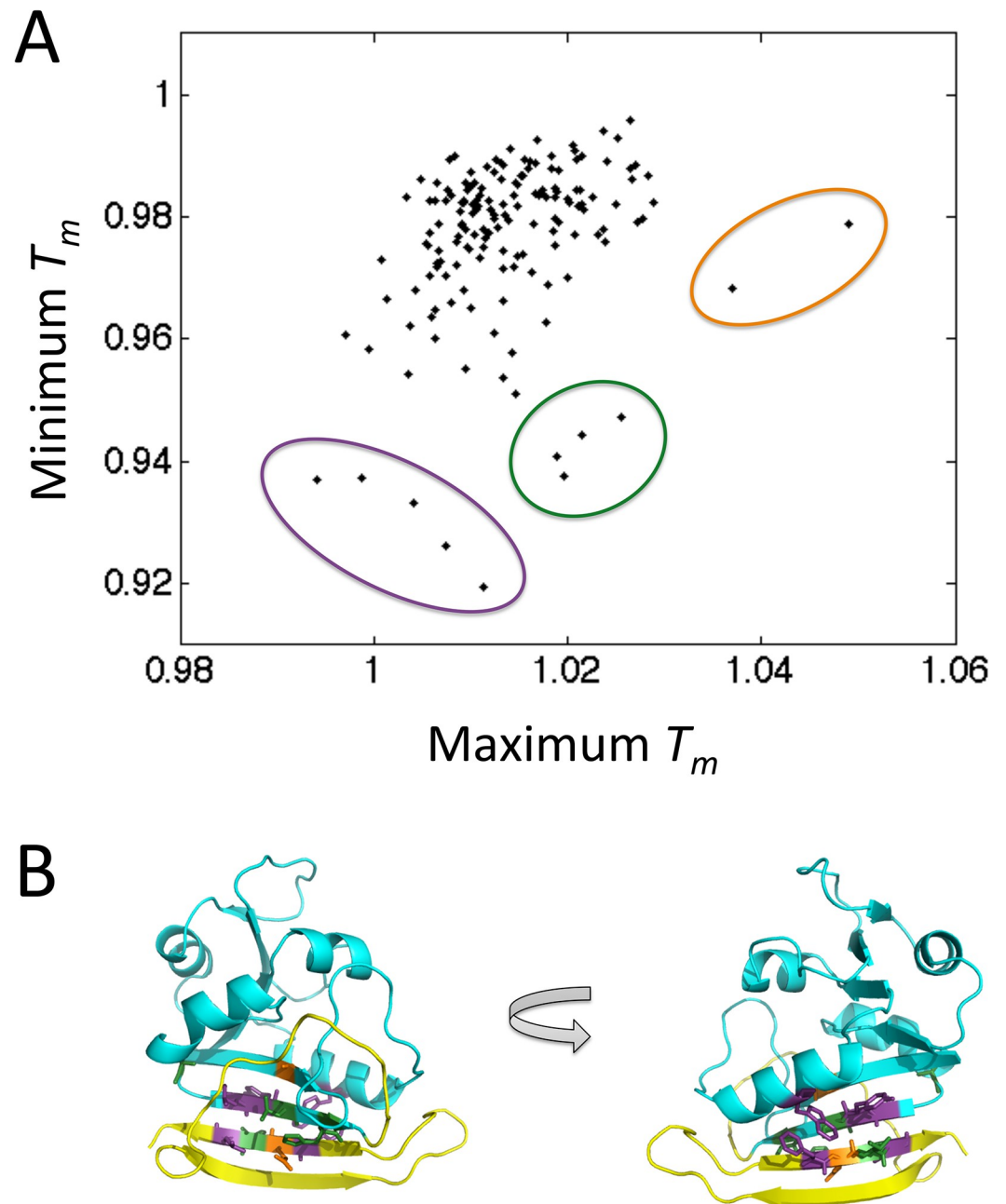




**Fig 9. Correlation between DHFR activity and stability.** WT is shown as a blue triangle; D27F is shown as a red diamond at zero activity. (A) Plot of  $k_{cat}$  vs. experimental relative  $T_m$ .  $r = 0.46$ ,  $p = 0.02$  (excluding outlier D27F). (B) Plot of  $k_{cat}/K_m$  vs. experimental relative  $T_m$ .  $r = 0.41$ ,  $p = 0.03$  (excluding outlier D27F).

<https://doi.org/10.1371/journal.pcbi.1004207.g009>

(2TRX), and Gln-25 ribonuclease T1 from *Aspergillus oryzae* (1RN1). Our predictions were compared with Eris and SDM. We did not compare MCPU results with FoldX and PopMusic as these mutations were selected in the training dataset for the two methods. As shown in



**Fig 10. Maximum stabilization and destabilization induced by mutations at each residue position.** (A) Plot comparing the minimum and maximum simulated  $T_m$  values, for each residue across all 19 simulated mutants.  $T_m$  is normalized to WT, by dividing each  $T_m$  value by the simulated WT  $T_m = 1.489$  simulation units. Outliers are circled in purple (left), green (middle) and orange (right). (B) DHFR with outlier residues colored according to the color scheme from (A). Purple: residues F153, W30, Y111, L156, L110. Green: residues A107, I155, L112, H114. Orange: residues A6, E154. Excluding outlier residues, the C-terminal beta hairpin is colored yellow, and the rest of the protein is colored cyan.

<https://doi.org/10.1371/journal.pcbi.1004207.g010>

Table 2, the correlation coefficient between MCPU predictions and the experimental  $T_m$  values, averaged over all proteins, is about 0.71, which is higher than that provided by Eris (-0.05), for which predictions were quite poor for both DHFR and other proteins, and SDM (0.63). If we consider only the binary prediction of whether a mutation is stabilizing or

Table 2. Simulation results on non-DHFR proteins.

SCOP	Length	PDB	Mutant	Real T <sub>m</sub>	MCPU	Eris	SDM	Native energy
All alpha proteins	66	5CRO	Y26D	54.0	0.878	-6.958	-0.690	-2044.4
	66	5CRO	Y26H	49.5	0.869	-3.425	-0.660	-1908.2
	66	5CRO	Y26L	46.0	0.868	-3.458	0.300	-1891.8
	66	5CRO	WT-5CRO	39.5	0.869	0.000	0.000	-1886.0
	66	5CRO	Y26W	37.5	0.871	-1.200	0.350	-1872.7
All beta proteins	67	1CSP	A46E	48.6	1.015	1.203	0.020	-1624.9
	67	1CSP	E3L	62.7	1.033	2.271	-0.460	-1536.7
	67	1CSP	E3R	69.6	1.023	1.638	-0.650	-1878.0
	67	1CSP	E66L	66.4	1.042	2.510	1.320	-1568.1
	67	1CSP	WT-1CSP	53.6	1.022	0.000	0.000	-1596.3
Alpha and beta proteins (a/b)	109	2TRX	D26I	98.0	1.135	2.847	4.290	-2478.5
	109	2TRX	WT-2TRX	87.0	1.107	0.000	0.000	-2558.4
	109	2TRX	T66L	85.0	1.124	3.540	2.180	-2552.0
	109	2TRX	T77V	82.0	1.124	-0.980	1.960	-2541.9
	109	2TRX	C35A	73.0	1.104	-14.670	-2.040	-2574.7
Alpha and beta proteins (a+b)	125	1RN1	V16A	44.5	0.887	2.033	-1.530	-1934.1
	125	1RN1	V16S	36.9	0.876	3.027	-4.410	-1920.1
	125	1RN1	V78S	34.6	0.878	3.870	-4.450	-1943.8
	125	1RN1	V89S	29.6	0.877	3.414	-4.100	-1904.8
	125	1RN1	WT-1RN1	51.5	0.899	0.000	0.000	-1929.7
				Error number	5	7	8	7
				Error rate	0.313	0.438	0.500	0.438
				<i>r</i>	0.708	-0.053	0.635	-0.348

Error number and error rate describe the number and fraction of mutations not predicted in the correct direction (stabilizing vs. destabilizing)

<https://doi.org/10.1371/journal.pcbi.1004207.t002>

destabilizing, MCPU can correctly predict 11 out of 16 mutations, while Eris and SDM correctly classify 9 and 8 mutations respectively.

### Entropy of the native state is an important contributor to stability

The theoretical analysis of the unfolding simulations relates the effect of mutations on the equilibrium between folded and unfolded states to the effect of mutations on free energy of the folded and transition states. It is widely believed that in the low-entropy folded state energetic factors dominate. If so that would imply that we can get an equally good correlation between prediction and experiment by estimating the mutational effect on energy of the native state as is the case for most empirical methods. To that end we evaluated the correlation between the energy of the minimized (after long MC equilibration) native state and the experimental  $T_m$  and found only a weak correlation with experimental melting temperatures (Table 2, last column), indicating that protein entropy, which is accounted for in the MCPU, in addition to enthalpy, is important in determining protein stability.

### Discussion

Estimates of protein stability using Molecular Dynamics are prohibitive for all but the smallest protein domains. However using MCPU we were able to efficiently explore stabilities of all possible point mutants for an essential enzyme of a typical size (159 amino acids) in a manageable amount of computational time (approx. one hour for every 1,000,000 MC steps).

Although the use of rapid Monte Carlo simulations reduces simulation time and allows for a greater number of replicates, our method to predict stability effects of mutations based on non-equilibrium unfolding simulations represents a general approach that could be modified for use with conventional MD simulations, especially given the current rate of improvement in simulation speed and accuracy.

Since our method involves protein unfolding simulations and not equilibrium simulations of both folding and unfolding processes, we expect it to be especially useful for predicting mutations that mostly affect the rate of protein unfolding as highlighted in our theoretical analysis. Low  $\phi$ -value residues, which acquire contacts with other residues late in the folding process and lose contacts early in the unfolding process [14] constitute the majority of residues in proteins, with  $\phi$ -value roughly constant around 0.24 as noted in [45]. Combining this observation with assumptions of transition state theory, we found that for the majority of residues (those not part of the folding nucleus [14,57] exhibiting anomalously high  $\phi$ -values) the observed simulation  $T_m$  relative to WT is proportional to the equilibrium stability change  $\Delta\Delta G$ , as verified by simulation and experiment. We establish that relative  $T_m$  is independent of simulation length, demonstrating that non-equilibrium simulations can in fact be used to quantify relative protein stability.

Many of the experimentally verified stabilizing mutations in DHFR predicted by MCPU are found in the C-terminal beta hairpin region, which is the first to unfold in simulations, prior to the main unfolding event encompassing the entire structure (see Fig. 2). It has been shown that the source of ultra-stability in hyperthermophiles generally arises from slowing the unfolding rate, rather than increasing the folding rate [28], so our method may be particularly suitable for discovering biologically relevant stabilizing mutations. In addition, our results might be particularly applicable to *in vivo* studies, where protease digestion and/or aggregation proceed from the partially-unfolded state. We note, however, that some stabilizing residues predicted by MCPU lie in the region of the protein that is late to unfold in simulations, including I61V, which raises the experimental melting temperature by 1.7°C. These mutants, along with the destabilized outlier I155A for which relative  $T_m$  depends on simulation length (Fig. 6), are appealing candidates for further study, as they may reflect a breakdown in the simplifying assumptions of 2-state kinetic theory for proteins.

It has been hypothesized that there exists a tradeoff between enzyme activity and stability, since certain regions of an enzyme must be sufficiently flexible to promote catalysis [53,54]. This conclusion was reached in [53,58], based on the exploration of stability effects of mutations in the active site of beta-lactamase [53] and rubisco [58]. Fersht and coauthors also found several stabilizing mutations in the active site of Barnase rendering the protein inactive [59]. While we observe a similar effect with the D27F mutation in DHFR, Fig. 9 shows that exploring only mutations in the active site provides a biased view on the tradeoff between activity and stability. Rather a vast majority of mutations throughout the protein show a qualitatively opposite trend. The likely explanation of the distinction between an apparent tradeoff when mutations are made in the active site and the opposite trend for mutations outside of the active site is that “carving” an active site requires special selection of catalytic amino acids, which could indeed have a destabilizing effect, overall. However our observation of a small positive correlation argues against an obligate relation between global protein dynamics and activity for DHFR, at least for the aspects of dynamics that are correlated with stability. Warshel and colleagues reached a similar conclusion in their theoretical analysis of the role of dynamics for DHFR and other proteins in [60]. This point has likewise been made by Bloom *et al.* [11], who noted that a number of proteins have been stabilized experimentally without loss of activity, and Taverna and Goldstein argued that marginal stability is an inherent property of proteins

due to the high dimensionality of sequence space and not due to a requirement of reduced stability in order to generate sufficient flexibility [61].

A straightforward explanation for the weak yet statistically significant *positive* correlation between activity and stability observed in our case might be that more stable proteins have greater effective concentration of the folded (*i.e.* active) form. It is also important to note that a weak yet statistically significant positive correlation between activity and stability for DHFR can be revealed only when stabilizing mutations are included in the analysis. Our earlier study [48] analyzed a smaller set of primarily destabilizing mutants and did not reveal any statistically significant trend (positive or negative) in the stability-activity relation for DHFR.

The development of highly-stabilized DHFR mutants through our combined *in silico*–*in vitro* approach opens up promising avenues for new *in vivo* studies. It has been postulated that protein stability places a fundamental constraint on the evolutionary pathways available to a protein [29,62] which has particular significance in the development of antibiotic resistance: higher protein stability can provide the microorganism with an increased capacity to evolve to evade antibiotic drugs [63] or, more generally, with capacity to evolve new functions [62]. We plan to use an approach developed in our lab [48] to endogenously introduce stabilized DHFR mutants into the bacterial chromosome and we will evaluate mutant fitness relative to wild-type using growth rates and competition experiments. These experiments will allow us to assess whether an evolutionary trade-off exists between stability and fitness *in vivo*, particularly in the presence of antibiotics.

We plan to apply MCPU to predict stability effects of mutations in proteins other than DHFR, in particular to develop highly stabilized mutants. Comprehensive experimental analysis of fitness and/or stability effects of mutations [64] could be useful in assessing the predictive capabilities of this method. In addition to predicting mutant stabilities, MCPU can provide atomic-detail molecular trajectories to rationalize the stability effects of mutations; such analysis is left to future study.

## Materials and Methods

### Monte Carlo simulations

We employed an all-atom Monte Carlo simulation program incorporating a knowledge-based potential, described in previous publications [16,31,65]. Briefly, the energy function is a sum of contact energy, hydrogen-bonding, torsional angle, and sidechain torsional terms, with an additional term describing orientation of nearby aromatic residues. The move set consists of rotations about  $\bar{I}$ ,  $\bar{I}$ , and  $\bar{I}$  dihedral angles, with bonds and angles held fixed. Moves are accepted or rejected according to the Metropolis criterion.

Mutations were introduced into the protein using the program Modeller v9.2 [66]. An initial minimization was carried out in NAMD [67] for 5,000 steps, using the default minimization algorithm and `par_all27_prot_lipid.inp` parameter file (without waters). An additional minimization step was carried out by running the Monte Carlo simulation program at low temperature (0.100 in simulation units) for 2,000,000 steps. A 2,000,000-step simulation was then run at each of 32 temperatures, averaging over all 2,000,000 steps to obtain Energy, RMSD, and number of contacts. These results were averaged over 50 simulations, for each temperature. Data was then plotted and fit to a sigmoid to obtain the computationally-predicted melting temperature, for each of Energy, RMSD, and number of contacts. To assess dependence of melting temperature on simulation length, longer simulations of 20,000,000 steps were carried out with 30 replications, averaging over the final 2,000,000 steps. For DHFR, 1,000,000 steps took approximately one hour of simulation time, on a single CPU.

**Bioinformatics Analysis.** DHFR protein sequences from 290 bacterial species were aligned using the program MUSCLE and online server [68]. MATLAB, with the Bioinformatics Toolbox, was used to create sequence logo representations and to determine the consensus sequence.

### Effect of number of replications on simulation accuracy

We evaluated the effect of the number of MC simulation replications on the prediction results. As shown in S9 Fig, the prediction accuracy is sensitive to the number of replications, but converges to a constant value after approximately 20 replications. In addition, we saw that increasing the number of MC steps beyond 2,000,000 steps does not increase prediction accuracy when the protein has been simulated with at least 20 replications, despite the fact that not all simulations have converged by 2,000,000 steps (S2 Fig—;S4 Fig).

### Simulation analysis

Sigmoidal fits were accomplished using the module “Sigmoidal, 4PL” using the software program Prism 6. The sigmoid function has the form:

$$Y = \text{Bottom} + (\text{Top}-\text{Bottom}) / (1 + 10^{((\text{LogIC}_{50}-X) * \text{HillSlope})})$$

### Method availability

The tool is accessible from Shakhnovich lab website <http://faculty.chemistry.harvard.edu/shakhnovich/software>

### Site-directed protein mutagenesis of DHFR

The wild type dhfr gene was cloned in a pET24 expression vector under the inducible T7 promoter, then transformed into BL21(DE3) cells [69]. Single point mutations of DHFR were constructed based on a two-step PCR-mutagenesis strategy [70], in which the template for the PCR is the plasmid of WT DHFR. The multiple-mutant variants of DHFR were constructed based on the same method with the single point mutation, but the template of PCR was the plasmid of the corresponding dhfr mutant. To verify the mutations of dhfr, DNA sequencing was performed at the GENEWIZ Incorporation (MA, U.S.). The verified plasmids were transformed into competent *E. coli* BL21(DE3) cells for expression.

### Protein expression and purification

WT DHFR and all mutants used in this study were cloned into a pET24 expression vector and overexpressed in the BL21(DE3) pLys *E. coli* strain.

A single colony of the transformed *E. coli* carrying the wild type or mutation dhfr was cultured in Luria-Bertani liquid medium containing 50 µg/mL kanamycin (LB-kana) at 30°C overnight, and then inoculated to fresh LB-kana (1:100 dilution) and incubated again at 30°C. When the OD<sub>600</sub> of the culture reached 0.6, isopropyl β-D-1-thiogalactopyranoside (final concentration, 0.4 mM) was added. Cultures were incubated for an additional 12–16 h at 25°C. The cells were then collected by centrifugation and disrupted by sonication. The recombinant proteins were purified with Ni-NTA Superflow (QIAGEN, U.S.) according to the manufacturer's instructions. Then, the collected protein sample was run with Superdex 75pg Column and was desalted with the desalting Column in ÄKTA protein purification system (GE Healthcare, U.S.). The final concentration of the purified protein was determined using the BCA protein assay kit (PIERCE CHEMICAL, USA) or the NanoDrop instrument (GE Healthcare, U.S.).

## Enzyme kinetics

DHFR kinetic parameters were measured by the progress-curve kinetics, essentially as described [69,71]. A Scientific stopped flow apparatus, RX.2000 Rapid kinetics system (Applied Photophysics, UK) was used with absorbance monitoring at 340 nm, under single-turnover conditions. NADPH was preincubated with DHFR for 5 min in syringe 1 at the temperature 25°C in a thermostated syringe compartment, and then the reaction was initiated by rapidly mixing the contents with dihydrofolate (DHF) from syringe 2. The final assay conditions are 25 nM DHFR, 120  $\mu$ M NADPH(D), and 25  $\mu$ M DHF in MTEN buffer (50 mM 2-(N-morpholino)ethanesulfonic acid, 25 mM tris(hydroxymethyl)aminomethane, 25 mM ethanolamine, and 100 mM sodium chloride, pH 7.6). The kinetics parameters ( $k_{\text{cat}}$  and  $K_M$ ) were derived from progress-curves analysis using Global Kinetic explorer [72].

## Stability measurements

Thermal stability was characterized by differential scanning calorimetry (DSC), essentially as described in references [69,73]. Briefly, DHFR proteins in Buffer A (10 mM potassium-phosphate buffer pH 7.8 supplemented with 0.2 mM EDTA and 1 mM beta-mercaptoethanol) were subjected to a temperature increase of 1°C/min between 20 to 90°C (nano-DSC, TA instruments, U.S.), and the evolution of heat was recorded as a differential power between reference (buffer A) and sample (120  $\mu$ M protein in buffer A) cells. The resulting thermogram (after buffer subtraction) was used to derive apparent thermal transition midpoints ( $T_m$  app). Thermal unfolding appeared irreversible for all DHFR proteins tested [48], and the two state scaled model provided in NanoAnalyze software (TA Instruments, U.S.) was used to fit the  $T_m$  app value. The mutants constructed in this study and the ones published earlier [48] were determined with different DSC instruments with slightly different calibration leading to a small offset of about 2°C for the WT DHFR for earlier published data[48].

Urea unfolding was used to measure stability of the DHFR mutants against chemical denaturation. Proteins (25  $\mu$ M in buffer A) were diluted in urea (0.2 mM increments up to a final urea concentration between 0 and 6 M), preequilibrated overnight at 25°C for 3 hours, and the change in the folded fraction was monitored by a circular dichroism signal at far-uv wavelength (221 nm) at 25°C (J-710 spectropolarimeter, Jasco). Fitting to a two-state model was used to derive the chemical transition midpoint ( $C_m$ ).

## Supporting Information

**S1 Fig. The MCPU flow chart.**

(TIF)

**S2 Fig. The RMSDs of DHFR wild type and mutants I115A and I155T vs. Monte Carlo step at temperatures from 0.1 to 3.2. RMSD is averaged over 50 replications.**

(TIFF)

**S3 Fig. The total energy of DHFR wild type and mutants I115A and I155T vs. Monte Carlo step at temperatures from 0.1 to 3.2. Total energies are averaged over 50 replications.**

(TIFF)

**S4 Fig. The number of contacts for DHFR wild type and mutants I115A and I155T vs. Monte Carlo step at temperatures from 0.1 to 3.2. Number of contacts were averaged over 50 replications.**

(TIFF)



**S5 Fig. The unfolding curves of DHFR and mutants I115A and I155T.** Data points are averaged over the 2,000,000 step simulation and 50 separate runs.

(TIFF)

**S6 Fig. Energy vs. Temperature simulated melting curve for WT DHFR and mutants I115A and I155T, averaging over the last 2,000,000 steps of a 20,000,000- step simulation, and 30 separate runs.** The mutant melting temperatures show the same trend as in [S5 Fig](#), although all melting temperatures are shifted to lower values (since the protein is given more time to unfold at each temperature), and the curve deviates more notably from a sigmoid. (A) Overlaid data points from all three mutants. (B) Fits to a sigmoidal function (blue line), for each of the three mutants.

(TIFF)

**S7 Fig. Correlation between urea denaturation midpoint,  $C_m$ , and melting temperature  $T_m$ , for WT and 22 mutants predicted to be stabilizing by MCPU.**  $r = 0.81$ ,  $p = 2 \times 10^{-6}$ . Blue triangle denotes WT, red diamond denotes D27F.

(TIF)

**S8 Fig. Protein stabilization for multiple mutants: deviation from additivity.** Change in experimental melting temperature relative to WT is predicted by summing melting temperature changes of individual mutants. This predicted  $\Delta T_m$  is plotted relative to the observed  $\Delta T_m$  (blue circles).  $r = 0.80$ ,  $p = 0.06$ . Red line denotes predicted  $\Delta T_m =$  observed  $\Delta T_m$ .

(TIFF)

**S9 Fig. Sequence alignment and sequence entropy for 290 bacterial DHFRs.** Sequence Logo for the alignment, generated using the MATLAB Bioinformatics Toolbox.

(TIF)

**S1 Table. The simulated  $T_m$  values of the selected single point mutations and WT DHFR.**

(DOCX)

**S2 Table. The experimental results for all single point mutations of DHFR.**

(DOCX)

**S3 Table. Stability predictions for DHFR mutants, using MCPU/other methods.**

(DOCX)

## Acknowledgments

We thank Bharat Adkar and Shimon Bershtein for help with the purification and Biophysical characterization of the DHFR mutants and Muyoung Heo for help and guidance in the use of the Monte Carlo simulations.

## Author Contributions

Conceived and designed the experiments: EIS JCW JT. Performed the experiments: JT JCW. Analyzed the data: EIS JT JCW. Contributed reagents/materials/analysis tools: AW. Wrote the paper: EIS JT JCW.

## References

1. Serohijos AW, Shakhnovich EI (2014) Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr Opin Struct Biol* 26C: 84–91.

2. Liberles DA, Teichmann SA, Bahar I, Bastolla U, Bloom J, et al. (2012) The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci* 21: 769–785. <https://doi.org/10.1002/pro.2071> PMID: 22528593
3. Shakhnovich E (2006) Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem Rev* 106: 1559–1588. PMID: 16683745
4. Aguzzi A, O'Connor T (2010) Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov* 9: 237–248. <https://doi.org/10.1038/nrd3050> PMID: 20190788
5. Dobson CM (2003) Protein folding and misfolding. *Nature* 426: 884–890. PMID: 14685248
6. Chiti F, Dobson CM (2009) Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* 5: 15–22. <https://doi.org/10.1038/nchembio.131> PMID: 19088715
7. Chennamsetty N, Voynov V, Kayser V, Helk B, Trout BL (2009) Design of therapeutic proteins with enhanced stability. *Proc Natl Acad Sci U S A* 106: 11937–11942. <https://doi.org/10.1073/pnas.0904191106> PMID: 19571001
8. Zeldovich KB, Chen PQ, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104: 16152–16157. PMID: 17913881
9. Kumar MD, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, et al. (2006) ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res* 34: D204–206. PMID: 16381846
10. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, et al. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America* 104: 1777–1782. PMID: 17267599
11. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* 103: 5869–5874. PMID: 16581913
12. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The Stability Effects of Protein Mutations Appear to be Universally Distributed. *J Mol Biol* 369: 1318–1332. PMID: 17482644
13. Fersht AR (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc Natl Acad Sci U S A* 92: 10869–10873. PMID: 7479900
14. Abkevich VI, Gutin AM, Shakhnovich EI (1994) Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* 33: 10026–10036. PMID: 8060971
15. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, et al. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* 9: 356–369. <https://doi.org/10.1038/nrg2344> PMID: 18398418
16. Yang JS, Wallin S, Shakhnovich EI (2008) Universality and diversity of folding mechanics for three-helix bundle proteins. *Proc Natl Acad Sci U S A* 105: 895–900. <https://doi.org/10.1073/pnas.0707284105> PMID: 18195374
17. Bar-Even A, Noor E, Lewis NE, Milo R (2010) Design and analysis of synthetic carbon fixation pathways. *Proc Natl Acad Sci U S A* 107: 8889–8894. <https://doi.org/10.1073/pnas.0907176107> PMID: 20410460
18. Gilis D, Rooman M (2000) PoPMuSiC, an algorithm for predicting protein mutant stability changes: application to prion proteins. *Protein Eng* 13: 849–856. PMID: 11239084
19. Yin S, Ding F, Dokholyan NV (2007) Eris: an automated estimator of protein stability. *Nat Methods* 4: 466–467. PMID: 17538626
20. Khan S, Vihinen M (2010) Performance of protein stability predictors. *Hum Mutat* 31: 675–684. <https://doi.org/10.1002/humu.21242> PMID: 20232415
21. Potapov V, Cohen M, Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel* 22: 553–560. <https://doi.org/10.1093/protein/gzp030> PMID: 19561092
22. Thiltgen G, Goldstein RA (2012) Assessing predictors of changes in protein stability upon mutation using self-consistency. *PLoS One* 7: e46084. <https://doi.org/10.1371/journal.pone.0046084> PMID: 23144695
23. Neher RA, Shraiman BI (2011) Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* 188: 975–996. <https://doi.org/10.1534/genetics.111.128876> PMID: 21625002
24. Reetz MT, Carballeira JD, Vogel A (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angew Chem Int Ed Engl* 45: 7745–7751. PMID: 17075931

25. Tian J, Wang P, Huang L, Chu X, Wu N, et al. (2013) Improving the thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 using a computationally aided method. *Appl Microbiol Biotechnol* 97: 2997–3006. <https://doi.org/10.1007/s00253-012-4411-7> PMID: 23001009
26. Alfarano P, Varadamsetty G, Ewald C, Parmeggiani F, Pellarin R, et al. (2012) Optimization of designed armadillo repeat proteins by molecular dynamics simulations and NMR spectroscopy. *Protein Sci* 21: 1298–1314. <https://doi.org/10.1002/pro.2117> PMID: 22767482
27. Korkegian A, Black ME, Baker D, Stoddard BL (2005) Computational thermostabilization of an enzyme. *Science* 308: 857–860. PMID: 15879217
28. Luke KA, Higgins CL, Wittung-Stafshede P (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J* 274: 4023–4033. PMID: 17683332
29. Mustonen V, Lassig M (2009) From fitness landscapes to seascape: non-equilibrium dynamics of selection and adaptation. *Trends Genet* 25: 111–119. <https://doi.org/10.1016/j.tig.2009.01.002> PMID: 19232770
30. Xu JB, Huang L, Shakhnovich EI (2011) The ensemble folding kinetics of the FBP28 WW domain revealed by an all-atom Monte Carlo simulation in a knowledge-based potential. *Proteins-Structure Function and Bioinformatics* 79: 1704–1714. <https://doi.org/10.1002/prot.22993> PMID: 21365688
31. Yang JS, Chen WW, Skolnick J, Shakhnovich EI (2007) All-atom ab initio folding of a diverse set of proteins. *Structure* 15: 53–63. PMID: 17223532
32. Boehr DD, McElheny D, Dyson HJ, Wright PE (2006) The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* 313: 1638–1642. PMID: 16973882
33. Neradil J, Pavlasova G, Veselska R (2012) New mechanisms for an old drug: DHFR- and non-DHFR-mediated effects of methotrexate in cancer cells. *Klin Onkol* 25 Suppl 2: 2S87–92. PMID: 23581023
34. El-Subbagh HIH, G.S.; El-Messery S.M., Al-Rashood S.T.; Al-Omary F.A.; Abulfadi Y.S.; Shabayek M. I. (2014) Nonclassical antifolates, part 5. Benzodiazepine analogs as a new class of DHFR inhibitors: synthesis, antitumor testing and molecular modeling study. *Eur J Med Chem* 3: 234–245.
35. Jarzynski C (1997) Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys Rev E Stat Nonlin Soft Matter Phys* 56: 5018–5035.
36. Imparato A, Luccioli S, Torcini A (2007) Reconstructing the free-energy landscape of a mechanically unfolded model protein. *Phys Rev Lett* 99: 168101. PMID: 17995298
37. Echeverria I, Amzel LM (2010) Helix propensities calculations for amino acids in alanine based peptides using Jarzynski's equality. *Proteins* 78: 1302–1310. <https://doi.org/10.1002/prot.22649> PMID: 20014024
38. Schulten KT, E.; Park S.; Khalili-Araghi F. (2003) Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *Journal of Chemical Physics* 119: 3559–3566.
39. Eyre-Walker A, Woolfit M, Phelps T (2006) The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173: 891–900. PMID: 16547091
40. Fersht A (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. New York: W.H. Freeman. xxi, 631 p. p.
41. Shakhnovich EI, Finkelstein AV (1989) Theory of cooperative transitions in protein molecules. I. Why denaturation of globular protein is a first-order phase transition. *Biopolymers* 28: 1667–1680. PMID: 2597723
42. Gutin AM, Abkevich VI, Shakhnovich EI (1998) A protein engineering analysis of the transition state for protein folding: simulation in the lattice model. *Fold Des* 3: 183–194. PMID: 9562547
43. Matouschek A, Kellis JT Jr., Serrano L, Fersht AR (1989) Mapping the transition state and pathway of protein folding by protein engineering. *Nature* 340: 122–126. PMID: 2739734
44. Fersht AR, Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci USA* 101: 7976–7981. PMID: 15150406
45. Naganathan AN, Munoz V (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc Natl Acad Sci U S A* 107: 8611–8616. <https://doi.org/10.1073/pnas.1000988107> PMID: 20418505
46. Baldwin RL, Frieden C, Rose GD (2010) Dry molten globule intermediates and the mechanism of protein unfolding. *Proteins* 78: 2725–2737. <https://doi.org/10.1002/prot.22803> PMID: 20635344
47. Finkelstein AV, Shakhnovich EI (1989) Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 28: 1681–1694. PMID: 2597724
48. Bershtein S, Mu W, Shakhnovich EI (2012) Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A* 109: 4857–4862. <https://doi.org/10.1073/pnas.1118157109> PMID: 22411825

49. Bershtein S, Mu WM, Serohijos AWR, Zhou JW, Shakhnovich EI (2013) Protein Quality Control Acts on Folding Intermediates to Shape the Effects of Mutations on Organismal Fitness. *Mol Cell* 49: 133–144. <https://doi.org/10.1016/j.molcel.2012.11.004> PMID: 23219534
50. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, et al. (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* 150: 389–401. <https://doi.org/10.1016/j.cell.2012.05.044> PMID: 22817898
51. Ohmae E, Miyashita Y, Tate S, Gekko K, Kitazawa S, et al. (2013) Solvent environments significantly affect the enzymatic function of *Escherichia coli* dihydrofolate reductase: comparison of wild-type protein and active-site mutant D27E. *Biochim Biophys Acta* 1834: 2782–2794. <https://doi.org/10.1016/j.bbapap.2013.09.024> PMID: 24140567
52. Reetz MT, Prasad S, Carballeira JD, Gumulya Y, Bocola M (2010) Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J Am Chem Soc* 132: 9144–9152. <https://doi.org/10.1021/ja1030479> PMID: 20536132
53. Beadle BM, Shoichet BK (2002) Structural bases of stability-function tradeoffs in enzymes. *J Mol Biol* 321: 285–296. PMID: 12144785
54. DePristo MA, Weinreich DM, Hartl DL (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 6: 678–687. PMID: 16074985
55. Milo R, Last RL (2012) Achieving Diversity in the Face of Constraints: Lessons from Metabolism. *Science* 336: 1663–1667. <https://doi.org/10.1126/science.1217665> PMID: 22745419
56. Worth CL, Preissner R, Blundell TL (2011) SDM—;a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res* 39: W215–222. <https://doi.org/10.1093/nar/gkr363> PMID: 21593128
57. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254: 260–288. PMID: 7490748
58. Studer RA, Christin PA, Williams MA, Orengo CA (2014) Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A*.
59. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224: 771–782. PMID: 1569556
60. Adamczyk AJ, Cao J, Kamerlin SC, Warshel A (2011) Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proc Natl Acad Sci U S A* 108: 14115–14120. <https://doi.org/10.1073/pnas.1111252108> PMID: 21831831
61. Taverna DM, Goldstein RA (2002) Why are proteins marginally stable? *Proteins* 46: 105–109. PMID: 11746707
62. Romero PA, Arnold FH (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10: 866–876. <https://doi.org/10.1038/nrm2805> PMID: 19935669
63. Brown NG, Pennington JM, Huang W, Ayvaz T, Palzkill T (2010) Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM beta-lactamases. *J Mol Biol* 404: 832–846. <https://doi.org/10.1016/j.jmb.2010.10.008> PMID: 20955714
64. Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, et al. (2013) Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci U S A*.
65. Xu J, Huang L, Shakhnovich EI (2011) The ensemble folding kinetics of the FBP28 WW domain revealed by an all-atom Monte Carlo simulation in a knowledge-based potential. *Proteins* 79: 1704–1714. <https://doi.org/10.1002/prot.22993> PMID: 21365688
66. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, et al. (2006) Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics* Chapter 5: Unit 5.6.
67. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802. PMID: 16222654
68. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32: 1792–1797. PMID: 15034147
69. Bershtein S, Mu WM, Shakhnovich EI (2012) Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci U S A* 109: 4857–4862. <https://doi.org/10.1073/pnas.1118157109> PMID: 22411825
70. Kirsch RD, Joly E (1998) An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res* 26: 1848–1850. PMID: 9512562
71. Kim HS, Damo SM, Lee SY, Wemmer D, Klinman JP (2005) Structure and hydride transfer mechanism of a moderate thermophilic dihydrofolate reductase from *Bacillus stearothermophilus* and comparison to its mesophilic and hyperthermophilic homologues. *Biochemistry* 44: 11428–11439. PMID: 16114879

72. Johnson KA, Simpson ZB, Blom T (2009) FitSpace explorer: an algorithm to evaluate multidimensional parameter space in fitting kinetic data. *Anal Biochem* 387: 30–41. <https://doi.org/10.1016/j.ab.2008.12.025> PMID: [19168024](https://pubmed.ncbi.nlm.nih.gov/19168024/)
73. Lopez MM, Makhatadze GI (2002) Differential scanning calorimetry. *Methods Mol Biol* 173: 113–119. PMID: [11859754](https://pubmed.ncbi.nlm.nih.gov/11859754/)