

Evolutionary Analysis As a Powerful Complement to Energy Calculations for Protein Stabilization

Koen Beerens,^{†,∇} Stanislav Mazurenko,[†] Antonin Kunka,[†] Sergio M. Marques,^{†,‡} Niels Hansen,[§] Milos Musil,^{†,||} Radka Chaloupkova,^{†,‡} Jitka Waterman,[⊥] Jan Brezovsky,^{†,‡} David Bednar,^{†,‡} Zbynek Prokop,^{*,†,‡} and Jiri Damborsky^{*,†,‡}

[†]Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic

[‡]International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic

[§]Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, D-70569 Stuttgart, Germany

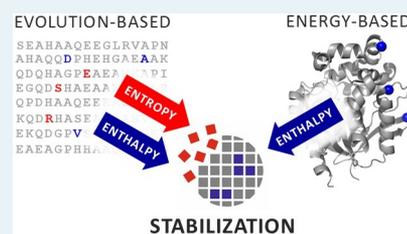
^{||}Department of Information Systems, Faculty of Information Technology, Brno University of Technology, 612 66 Brno, Czech Republic

[⊥]Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, United Kingdom

Supporting Information

ABSTRACT: Stability is one of the most important characteristics of proteins employed as biocatalysts, biotherapeutics, and biomaterials, and the role of computational approaches in modifying protein stability is rapidly expanding. We have recently identified stabilizing mutations in haloalkane dehalogenase DhaA using phylogenetic analysis but were not able to reproduce the effects of these mutations using force-field calculations. Here we tested four different hypotheses to explain the molecular basis of stabilization using structural, biochemical, biophysical, and computational analyses. We demonstrate that stabilization of DhaA by the mutations identified using the phylogenetic analysis is driven by both entropy and enthalpy contributions, in contrast to primarily enthalpy-driven stabilization by mutations designed by the force-field calculations. Comprehensive bioinformatics analysis revealed that more than half (53%) of 1 099 evolution-based stabilizing mutations would be evaluated as destabilizing by force-field calculations. Thermodynamic integration considers both folded and unfolded states and can describe the entropic component of stabilization, yet it is not suitable for predictive purposes due to its high computational demands. Altogether, our results strongly suggest that energetic calculations should be complemented by a phylogenetic analysis in protein-stabilization endeavors.

KEYWORDS: protein stabilization, thermostability, evolutionary analysis, force-field calculations, computational tools, entropy, enthalpy, thermodynamic integration



INTRODUCTION

Proteins are used in an ever-expanding list of biotechnological applications, ranging from household products and technical industries, to food and animal feed, to fine chemicals and biopharmaceuticals.¹ Not only do proteins provide high specificity and activity but they also are environmentally friendlier than typical chemical synthesis protocols. However, their application can be hampered by limited stability because production-line conditions, including high temperatures, extreme pH, or the presence of organic solvents or proteases are often far from the natural conditions for which proteins were evolved.² Protein engineering can be used to improve natural proteins via directed evolution³ or computational prediction of hotspots,⁴ as well as prediction of single-point⁵ or multiple-point mutations.⁶ Recently, the computational approaches have been increasingly applied for protein stabilization as they allow fast and focused redesign of existing proteins, requiring only limited resources and experimental

effort. However, some of the beneficial mutations observed in these studies were not computationally predicted to be beneficial. Thus, despite the increasing success rates of computational methods,^{5–10} a deeper understanding of the mechanistic basis for protein stability is needed to enable new and improved algorithms for computational tools.

We recently reported a new computational method called FireProt⁶ that was used to direct the development of an improved variant of the haloalkane dehalogenase DhaA (Supporting Information Methods). This approach searches for mutations likely to stabilize a protein of interest via energy calculations of potential point mutants, called “energy-based mutations”, and via phylogenetic analysis to identify residues that have drifted from more stable consensus sequences, called

Received: April 29, 2018

Revised: July 14, 2018

Published: August 31, 2018

“evolution-based mutations”. FireProt then combines both of these calculations using smart filtering to allow the design of highly stabilized multiple-point mutants.⁶ In our previous work, we experimentally characterized a triple-point mutant, DhaA101, containing three evolution-based mutations selected following the back-to-consensus approach. Even though the impact of these mutations had been predicted by various computational tools as destabilizing, neutral, or slightly stabilizing at best, the triple-point mutant displayed apparent melting temperature eight degrees higher than the wild type.

Here we explore potential explanations for the discrepancy between our computational and experimental data through consideration of multiple hypotheses, including (H1) failed stability prediction due to incorrect modeling of the evolution-based mutations by the energy-based approach, (H2) unanticipated stabilization due to changes in the oligomeric state, (H3) changes in formation of protein–ligand complexes, and (H4) entropy-driven stabilization (Figure 1). To

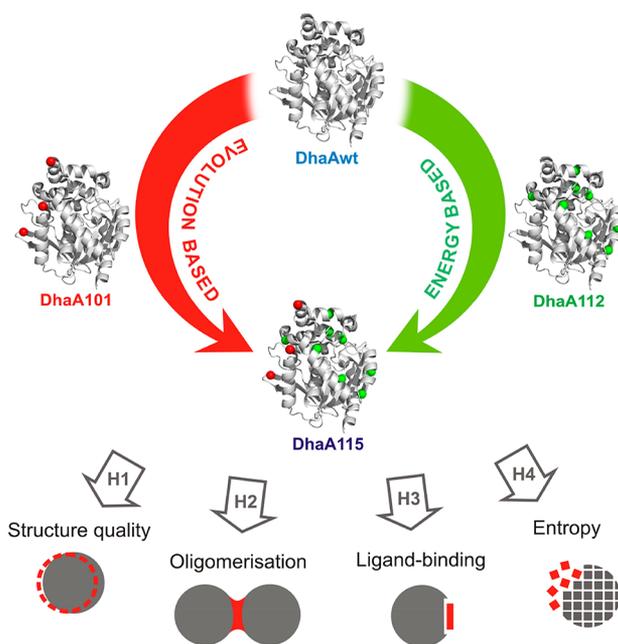


Figure 1. Overview of studied haloalkane dehalogenase DhaA variants and schematic representation of different working hypotheses (H1–H4). DhaA101 contains three evolution-based mutations predicted by phylogenetic analysis, DhaA112 contains eight energy-based mutations predicted by force-field calculations, and DhaA115 combines all 11 mutations.

investigate these hypotheses, we have determined the crystal structure of DhaA101 and examined the biophysical properties of this protein variant and its deconstructed single-point mutants. Additionally, more accurate molecular-modeling approaches were applied to perform free-energy predictions to explore the role of nonproteinaceous components in the studied system. Overall, our data highlight the strong added value of computational tools, such as FireProt, which combine energy-based force-field calculations with evolutionary analysis.⁶ This combination indirectly considers both enthalpic and entropic factors for protein stabilization, allowing larger stability improvements in contrast to cases where only force-field calculations are applied, naturally biasing predictions toward the enthalpy changes. Moreover, evolution-based methods have a good potential to uncover stabilizing

mutations with additional beneficial features, such as increased activity,¹¹ as we have also observed.

RESULTS

Deconstructing Contributions to DhaA101 Stability.

We first constructed the single-point mutants corresponding to the three mutations in DhaA101 to determine whether the triple-point mutant might have unique properties that would explain the observed stabilization. Each of the single mutations had previously been predicted to be mildly stabilizing at best by FoldX;⁶ additional calculations using various force fields showed that these approaches consistently failed to reproduce the stabilizing effects of the combined mutations (Table S1). To determine whether individual mutations might display a different outcome from their combination, we followed protein unfolding upon thermal denaturation of the single-point mutants by differential scanning calorimetry (DSC), previously shown to be an appropriate method for this purpose.^{12,13} In contrast to force-field calculations, DSC data showed stabilization for two of three evolution-based mutants (Table 1). These data thus confirmed that the prediction failure is already occurring at the level of the individual mutations, encouraging us to undertake more detailed investigations of these sites and their contributions in the structure–function relationships of DhaA101.

Correctly Predicted Tertiary Structure of DhaA101.

To test the first hypothesis that incorrect modeling of the mutant structure led to the misassignment of the roles of the three evolution-based mutations, we determined the crystal structure of DhaA101 to look for unpredicted changes in the backbone and side-chain orientations (Figure 2a). The structure of DhaA101 was solved to the resolution of 0.99 Å by molecular replacement (Table S2). The search model used for the structure determination was DhaA14 (PDB ID 3G9X).¹⁴ The final structural model contains residues 4–296 and shows that the enzyme exists as a monomer in the crystal. The overall structure of DhaA101 resembles that of DhaAwt, namely, the typical α/β hydrolase core domain and a helical cap domain. Circular dichroism (CD) spectra also indicated no secondary structure changes of the mutants.⁶

On the basis of the superimposition of the crystal structures and the Rosetta predicted structure of DhaA101, it was clear that the mutations do not induce significant changes in the protein’s backbone and that the mutant’s backbone was thus correctly modeled (Figure 2c). The RMSD values of the backbone atoms for the overlap of both crystal structures and the overlay of both DhaA101 structures are 0.104 and 0.152 Å, respectively. Comparison of the modeled and crystal structures of DhaA101 further reveals that two of three mutations (E20S and A155P) were correctly predicted (Figure 2d and e). Only the side chain of the newly introduced arginine (F80R) was partially misplaced in the model (Figure 2f). The C_ζ – C_ζ distance d between the new Arg and the naturally present Arg204 in the model was 6.6 Å, while in the determined structure these are stacked closely together, with a C_ζ – C_ζ distance of only 3.8 Å. In addition, the angle between the planes formed by the guanidinium groups of these arginines (θ) is 35.1° in the crystal structure compared to 52.3° in the modeled structure. Hence, the geometry of the arginine pair observed in the crystal structure is close to parallel ($d \leq 4.0$ Å but not $\theta \leq 30^\circ$),¹⁵ whereas that in the model is neither parallel nor perpendicular (neither $60^\circ \leq \theta \leq 90^\circ$ nor 5.0 Å $\leq d \leq 6.0$ Å).¹⁵ The repulsions between the two positively

Table 1. Stability and Functional Characteristics of the DhaA Variants

variant	mutations	T_m^b (°C)	ΔT_m^b (°C)	specific activity ^c (nmol·s ⁻¹ ·mg ⁻¹)
DhaAwt	^a	50.4 ± 0.5	^a	18.0 ± 0.1
evolution-based triple-point variant ^d				
DhaA101	E20S + F80R + A155P	58.4 ± 0.2	+8.0	49.3 ± 0.6
evolution-based single-point variants ^d				
DhaA123	E20S	57.6 ± 0.2	+7.2	26.7 ± 0.5
DhaA124	F80R	49.9 ± 0.3	-0.5	27.3 ± 1.3
DhaA125	A155P	51.5 ± 0.2	+1.1	25.1 ± 1.3
energy-based multiple-point variant ^e				
DhaA112	C128F + T148L + A172I + C176F + D198W + V219W + C262L + D266F	65.1 ± 0.1	+14.7	5.5 ± 0.1
combined multiple-point variant ^{d,e}				
DhaA115	E20S + F80R + C128F + T148L + A155P + A172I + C176F + D198W + V219W	73.5 ± 0.1	+23.1	5.6 ± 0.1
DhaA115 monomer	+ C262L + D266F	73.4 ± 0.1	+23.0	ND
DhaA115 dimer		73.4 ± 0.1	+23.0	ND

^aNot applicable. ^bDetermined by differential scanning calorimetry. ^cActivity determined with 1-iodohexane at 37 °C and pH 8.6. ^dDesigned based on phylogenetic analysis. ^eDesigned based on force-field calculations; ^f T_m , apparent thermal transition midpoint or melting temperature; ND, not determined.

charged side chains of the arginines are most likely the reason for the partially incorrect positioning of the newly introduced arginine side chain in the model (Figure 2f). We concluded that the modeled and crystal structures showed an overall very good agreement, suggesting that the source of the discrepancy must lay elsewhere. Determination of the crystal structures of the individual mutants would likely not reveal any additional valuable information for our study. Therefore, we decided to focus on the other hypotheses.

Ligands and Ions That Do Not Impact Protein Stability. While examining the crystal structure of DhaA101, we further noted strong electron density for potential ligands in the active site and at two positions on the protein surface. Three ligands were subsequently modeled in the structure of DhaA101 (Figure 2a). The ligand bound in the enzyme active site was interpreted as 2-(*N*-morpholino)ethanesulfonic acid (MES) and occupies two alternative conformations A and B with occupancies of 0.6 and 0.4, respectively (Figure 2b). The MES position in the active site is stabilized by interactions with the halide-stabilizing amino acids (Trp107 and Asn41), nucleophile (Asp106), and Phe149. Phe149 adopts two different conformations depending on the conformations of MES in the enzyme active site. Both ligands located on the protein surface were identified as di(hydroxyethyl)ether (PEG), designated as PEG1 and PEG2 (Figure S1). The cocrystallization of these ligands with DhaA101 could play a role in stabilizing the protein if these molecules were bound by DhaA101 more generally. However, all three ligands were components of the crystallization solution and not present in the buffers used for the activity and thermostability assays, suggesting they are non-native and highly unlikely to influence the observed stability.

Broadening our consideration of small molecules that might influence protein stability further, we noted that solvated ions had not been taken into account during our initial force-field calculations and could potentially be relevant factors that influence protein stability.¹⁶ Indeed, the distribution of the charges on the protein surface is a crucial determinant of protein–ion interactions. E20S and F80R mutations removed a negative surface charge and introduced a positive surface charge, respectively. The occupancy of cations calculated for 200-ns-long molecular dynamics (MD) simulations revealed

significant differences in the regions with the highest density of sodium ions compared to the wild type (Figure 3a) and the evolution-based triple-point mutant DhaA101 (Figure 3c), over the entire simulations. Deconvoluted single-point mutants were also analyzed and MD trajectories revealed significant differences due to E20S and F80R mutations (Figure 3b and c), whereas no changes were observed for the third mutation (A155P). The largest change in sodium ion density was observed for the E20S mutation, which had high sodium ion density around the wild type glutamate, but which was strongly reduced in the presence of the serine mutation, even showing a long-distance influence. Similarly, the introduced positive charge of F80R also significantly reduced the high density of sodium ions in its region (Figure 3c). On the other hand, the occupancy of the chloride ions is much more dispersed than that of sodium ions. Despite a slight chloride ion density increase around the R80 residue in the mutant, this effect is negligible compared to the effect of sodium ion density. These observations were corroborated by the radial distribution functions of the ions around the proteins (Figure S2) and suggest that the different interactions with the sodium ions in solution might be related to the observed stabilization. DSC thermograms were, therefore, also determined at increasing concentrations of sodium chloride. The results obtained did not reveal any clear dependence of the stability of those proteins on the ion concentrations (Table S3), thus rejecting this hypothesis.

Partial Oligomerization That Does Not Influence Stability. Our initial energetic calculations on the evolution-based mutations were carried out using monomeric structures. However, if the introduced mutations at the surface of the protein are involved in the formation of multimeric structures, then our energetic predictions using monomers could be misleading. Therefore, we examined the oligomeric state of DhaA101 as well as two mutants that contained eight energy-based mutations: (i) DhaA112 designed by the energy-based protocol of the FireProt tool⁶ and (ii) DhaA115 combining three evolution-based and eight energy-based mutations (Table 1). Analysis by native polyacrylamide gel electrophoresis demonstrated that the mutants were mainly present as monomers, although a limited amount of higher-order oligomer formation was also observed (Figure S3a). The

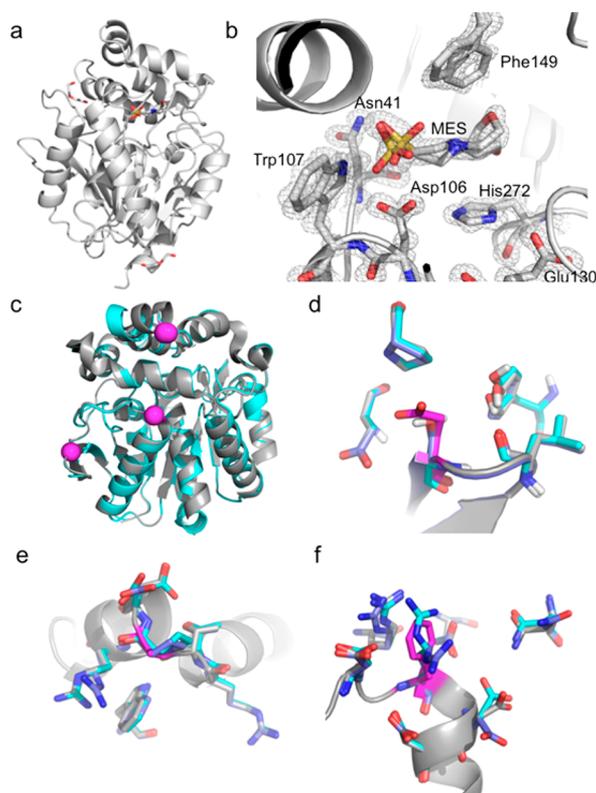


Figure 2. Crystal structure of DhaA101. (a) Overall structure of DhaA101 (PDB ID 5FLK) shown in cartoon representation and the three ligands identified in the crystal structure shown as sticks. (b) Expanded view of the DhaA101 active site. $2Fo-Fc$ electron density map contoured at 1σ is shown for the active site residues with bound 2-(*N*-morpholino)ethanesulfonic acid (MES) ligand. The residues of catalytic pentad and MES are represented as sticks. Two alternative conformations of MES and Phe149 illustrate how the side chain affects binding modes of MES inside the active site. (c–f) Comparison of the crystal structure of DhaA101 with the crystal structure of DhaAwt and the theoretical model of DhaA101. (c) Superimposition of the crystal structures of DhaAwt (PDB ID 4E46) and DhaA101 (PDB ID 5FLK). The root-mean-square deviation (RMSD) of the alignment of the crystal structures is 0.104 Å. The C_{α} of mutations are represented as magenta spheres. (d–f) Expanded view of the overlays of the introduced mutations E20S (d), A155P (e), and F80R (f) and their close surroundings. (c–f) The crystal structure of DhaAwt is given in gray with the respective mutated residues in magenta sticks. The modeled and crystal structures of DhaA101 are shown in dark blue and cyan, respectively. The mutated and interacting residues are shown in sticks.

dimeric form of DhaA101 was not present in sufficient quantities to allow isolation, but the dimeric and monomeric forms of DhaA115 were successfully separated by gel permeation chromatography, as was confirmed by native polyacrylamide gel electrophoresis (Figure S3b and c). The lack of a rapid equilibrium between the monomeric and oligomeric states allowed the DSC analysis of the separated monomer and dimer fractions to determine whether the minor dimer population ($\sim 8\%$) was skewing the observed melting temperature (Figure S3b). Stability analysis of the separated monomer and dimer fractions via DSC revealed the same melting temperatures for the different oligomerization states (Table 1). For the DhaA115 dimer, small additional heat capacity was observed before the main peak, most likely resulting from dimer dissociation (Figure S3d). Therefore, we

ruled out oligomerization as the source of the computational mismatch.

Calorimetry Identification of Changes in Entropy of Evolution-Based Mutants. Finally, we have carried out the analysis of calorimetric data to investigate the potential role of entropy in the stabilization. The deconvolution of the DSC data allowed the determination of the apparent melting temperature (T_m) and the calorimetric enthalpy change (ΔH_{cal}) of each variant (Figure 4). The introduction of the energy-based mutations (DhaA112) to the wild type substantially increased both the ΔH_{cal} and T_m , as indicated by an increased area under the DSC thermogram peak and its shift to higher temperatures, respectively (Figure 4a), and thus serves as a positive control for enthalpic stabilization. The introduction of the evolution-based mutations (DhaA101), however, had only a minor effect on the area under the peak while still shifting it to higher temperatures. Combination of the two sets of mutations (DhaA115) resulted in the combination of the observed effects, i.e., further increase in T_m with the change in ΔH_{cal} similar to that of the energy-based mutations alone (DhaA112). A similar picture appears when analyzing the thermograms of single- and double-point precursors of DhaA101: a mixed direction of changes in T_m and enthalpy differences was observed upon the introduction of the corresponding mutations (Figures 4b and S4). These observations illustrate that changes in T_m induced by evolutionary mutations are poorly correlated with the corresponding changes in experimentally obtained enthalpy, which is in strong contrast to energy-based mutations. These observations imply a significant role of entropy for the evolution-based mutations.

Calculation of the entropy change from DSC thermograms is difficult. Privalov and Dragan suggested using the Kirchhoff's relation $\partial\Delta H/\partial T = \Delta C_p$ and $\partial\Delta S/\partial T = \Delta C_p/T$, where ΔC_p stands for the difference between the heat capacities of the folded and unfolded protein states.¹⁷ Following their procedure, we calculated the apparent entropy changes (ΔS_{cal}) by integration of the melting curves (Figure S4). The evolutionary-based mutations decrease ΔS_{cal} (DhaAwt versus DhaA101; DhaA112 versus DhaA115), whereas an increase in ΔS_{cal} is observed upon introduction of the energy-based mutations as expected due to the common enthalpy–entropy compensation¹⁸ (DhaAwt versus DhaA112; DhaA101 versus DhaA115). Thus, the result of the direct integration of the DSC curves supports our conclusions regarding the role of entropy in the stabilization of DhaA by the evolution-based mutations. Nevertheless, the calculated values of the entropy changes are only approximates because we noted a significant sensitivity of the method¹⁷ to the baseline and peak border determination, as well as different degrees of reversibility. A small decrease in reversibility of unfolding was observed after introduction of the evolutionary-based mutations (86% for DhaAwt and 65% for DhaA101), whereas a more significant decrease in reversibility was observed upon introduction of the energy-based mutations ($<20\%$ for DhaA112 and DhaA115).

Computational Thermodynamic Integration Accounting for Entropic Stabilization. In an attempt to describe the thermal stabilization observed in the single-point mutants, we have performed alchemical free-energy calculations¹⁹ based on molecular-dynamics simulations. The morphing of one amino acid into the other was carried out gradually by coupling the Hamiltonians of the two states by a coupling parameter and performing equilibrium simulations at

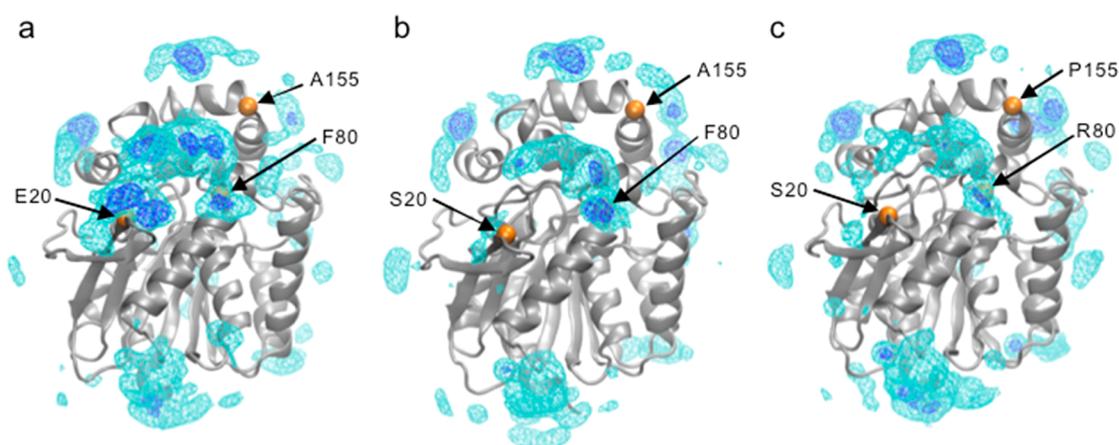


Figure 3. Occupancy density of sodium ions over molecular dynamics simulations with DhaAwt and two DhaA variants. Ion densities displayed above the structures of DhaAwt (a), the single-point mutant DhaA123 (E20S) mutant (b), and the triple mutant DhaA101 (E20S + F80R + A155P) (c). The blue occupancy surfaces correspond to the 0.025 isovalue, and the cyan ones correspond to the 0.005 isovalue.

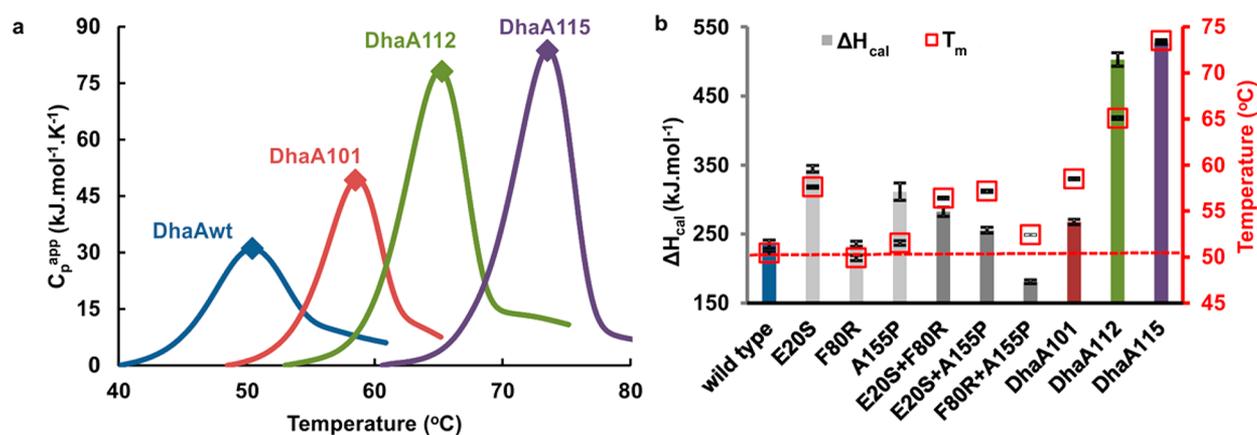


Figure 4. Thermostability analysis of DhaA variants. (a) Calorimetric curves of DhaA variants show increasing apparent melting temperatures (T_m , marked with \blacklozenge) upon introduction of evolutionary (red), energy-based mutations (green), and their combination (purple) compared to the wild type (blue). The area under the calorimetric curve, corresponding to calorimetric enthalpy (ΔH_{cal}), is large for variants containing energy-based mutations and comparable to the wild type in the case of DhaA101, which carries evolutionary-based mutations. (b) ΔH_{cal} calculated via integration of the peak area under the DSC thermograms using linear baseline subtraction versus T_m for single- and double-point evolutionary variants. Although the single-point mutants demonstrate mixed effects in terms of enthalpy changes, the double-point variants follow the pattern of DhaA101: an increase in melting temperatures with minor increase or even decrease of enthalpy change as compared to the wild type. The wild type T_m level is indicated by the dotted red line.

each intermediate state. The free-energy differences between the intermediate states were summed up to yield the free-energy difference between the physical end states²⁰ (Table S6 and Figure S8). The difference in stability between native DhaA and the mutant with the charge-conserving mutation A155P shows essentially no temperature dependence. To explain the failure of all computational methods in predicting the stabilizing effect of the A155P mutation, we speculate that the key reason may lie in the existence of an intermediate state, which contributes to reshape the energy landscape of unfolding in ways that were not modeled in this work. The other two mutations were predicted to be stabilizing by thermodynamic integration (TI), with the E20S mutant being significantly more stable, which corresponds well with the experimental T_m value. This mutation displays a significant temperature dependence of the $\Delta\Delta G$ values. The strongest stabilization effect is predicted at 300 K, and then $\Delta\Delta G$ increases at 325 K (Figure S8). Although the values for this mutant have large uncertainties, the trends suggest that at lower temperatures the stabilization was dominated by entropy, but at higher

temperatures the enthalpy contribution plays a more important role, which is again in good agreement with the experimental results described earlier.

Evolutionary Analysis during Protein Stabilization.

Back-to-consensus analysis is a well-established evolutionary method frequently used for protein stabilization.^{21–26} To explore the relationships between back-to-consensus analysis and force-field-based calculations, we have studied 103 unique protein structures available in the ProTherm²⁷ and the HotMuSiC²⁸ data sets and carried out parallel back-to-consensus analysis and energetic evaluation using Rosetta. To construct the back-to-consensus data set, BLASTp search with the e-value 10^{-12} was used for collection of homologous sequences with at least 35% identity. Sequences were clustered by CD-HIT at 90% identity threshold, and the sequences showing the identity >90% were filtered out. Maximum of 200 sequences were randomly selected for the construction of the alignment. Back-to-consensus mutations were identified in the alignment by both simple consensus and frequency ratio approaches. For the validation set, only the records with Gibbs

free energies changes of folding <-0.5 kcal·mol⁻¹ and >0.5 kcal·mol⁻¹ were included, omitting neutral mutations and in this way increasing the reliability of the analysis. In total, we have extracted 1328 mutations, from which 256 (19.3%) mutations were stabilizing and 1072 (80.7%) mutations were destabilizing. A total number of 1099 potentially stabilizing mutations was identified for these 103 proteins by back-to-consensus analysis, from which 515 (46.9%) were predicted as stabilizing by the energy-based approach, while in more than half of the cases (53.1%), the back-to-consensus and force-field methods disagreed in the identification of the stabilizing mutations (Table S7). The experimental and back-to-consensus data sets have overlapped in nine of the cases. While all nine mutations were correctly assigned as stabilizing by the evolutionary analysis, only six of the mutations were predicted as stabilizing by the force-field calculations. This bioinformatics analysis highlights the complementary information derived from the energy-based and evolution-based analyses.

DISCUSSION

Our combined structural, biophysical, and computational analysis allowed resolution of the factors that led to the failure of our energy-based calculations to predict the outcome of evolution-based mutations. The first hypothesis—that the mutant structure was incorrectly modeled—was ruled out by good agreement between the model and crystal structure of both the protein backbone and the positions of the E20S and A155P side chains. Only the side chain from the newly introduced arginine (F80R) was partially incorrectly positioned in the model, most likely due to prediction of a repulsive interaction between the new arginine R80 and the neighboring R204. We can conclude that the modeled structure resembled the crystal structure well and that the minor differences observed cannot explain the incorrect force-field predictions.

Our second hypothesis—that oligomerization was influencing stability—can also be ruled out. We do observe a small fraction of oligomeric protein, presumably because the change in surface charge of the more stable variants could remove potential repulsive interaction between monomers that prevent the wild type from oligomerization. However, the thermostability analysis of the isolated dimer and monomer fractions of DhaA115 revealed no improved stability due to dimerization. The small additional heat-capacity peak observed for the DhaA115 dimer most likely results from dimer dissociation. Because only a limited fraction ($<8\%$) of the protein was present as dimers in the original sample, the heat-capacity effect of the dimer fraction is insignificant.

Similarly, we are able to discard our third hypothesis—that the interaction between the mutant protein and ligands or other ions altered the stability in an unpredictable way. Because the energy-based calculations were performed in the absence of any small molecules, it was possible that significant interactions could have been missed in the modeling. Indeed, we identified electron density of ligands on the protein surface and in the active-site cavity of the crystal structure of DhaA101. However, all three molecules originate from the crystallization buffer and therefore are not present in the samples analyzed for stability by DSC. Therefore, the stabilizing effect of these ligands can be omitted. While the modeling revealed potential differences in the interactions of ions with the mutant constructed, DSC experiments with

varying salt concentrations did not reveal any clear relationship between stability and altered ion concentrations. We can thus conclude that the evolution-inspired stabilization does not result from the binding of ligands or ions to the protein structures.

In our last hypothesis, we have considered that the evolution-based mutations caused changes in the protein entropy of unfolding. One of the downsides of the energy-based prediction methods is that the calculations of the variation of the Gibbs free energy ($\Delta\Delta G$) are primarily based on the enthalpic components, whereas the entropic effects are difficult to calculate²⁹ and are generally approximated or neglected,³⁰ often resulting in incorrect predictions of entropy-driven mutations.⁶ The entropic effects require extensive conformational sampling, while tested computational methods work well for rigid systems that involve changes in an enthalpic term. Hence, we compared the calorimetric enthalpy levels of the different DhaA variants. It is clear from the data that the evolutionary-based mutations have only limited effect on the enthalpy difference, whereas the energy-based mutations clearly show enthalpy-driven stabilization. This suggests an entropic nature of the stabilization due to the mutations inferred by phylogenetic analysis. Although approximate, entropy calculations revealed that the entropy change curves vary in both the respective position and the slope for the evolutionary triple mutant. Moreover, DhaA variants with single- and double-point evolutionary mutations also showed a more pronounced entropy-related effect as compared to the energy-based mutants.

The entropy-change curves of the energy-based and combined mutants were shifted to higher values compared to the wild type, which is in agreement with the enthalpy–entropy compensation.³¹ Nonetheless, accurate estimation of the entropy of large systems currently remains very challenging, and the results can be biased by available sampling.²⁹ TI was the computational method providing the most consistent result with the experimental observations. However, recent large-scale screening studies show that, especially for charge-changing mutations, the predictive capabilities of alchemical free-energy simulations are less reliable.^{32,33} For complex systems such as the one studied here, the results are most likely influenced by both sampling issues and force-field bias. Implementations of alchemical transformations making use of graphics processing units³⁴ will greatly enhance the accessible time scales of MD simulations. While this reduces the sampling problem in free-energy simulations, the reduction of the force-field bias remains a topic of active research in the coming years, making these calculations demanding if several force fields are to be probed. Moreover, thermodynamic stability (as probed by TI) and thermal stability (as probed by DSC) might not always be directly correlated.^{35,36} Therefore, the fact that TI was yet unable to predict the stabilization observed for the A155P mutation provides striking evidence of the power of the evolutionary-inspired methods as valuable complementary tools to the energy-based methods for predicting protein stabilization.

It is also worth noting that one more factor is desirable in computational analyses—a detailed (un)folding pathway. On the one hand, intermediates on unfolding pathways that may change upon adding mutations may affect observed thermodynamic stability.^{37–39} On the other hand, analysis of transition states may shed light on the changes to energy barriers

separating native and denatured conformations and reflecting kinetic stability, which is of particular interest when protein denaturation is irreversible.^{39–41} Such activation free-energy changes ($\Delta\Delta G^\ddagger$) can similarly be decomposed into enthalpy and entropy contributions, as well as structure-unfolding and solvation-barrier terms.^{42,43} In our case, unfolding of DhaAwt proceeds via intermediate state followed by an irreversible step.⁴⁴ However, because the first step is reversible and contributes the most to the observed calorimetry signal in both DhaAwt and DhaA101 cases, our conclusions about the effects of the evolutionary-based mutations seem robust in this respect. The dramatic decrease in reversibility caused by the energy-based mutations indicates that the irreversible step becomes rate-limiting for unfolding of DhaA112 and DhaA115. Sanchez-Ruiz characterized this scenario as “global unfolding”, in which the mutation effects on thermodynamic stability immediately affect kinetic stability.⁴¹

CONCLUSIONS

In this Article, we report the thorough analysis of the evolutionary mutations of the haloalkane dehalogenase DhaA identified by FireProt.⁶ Evolution-based mutations can be challenging to explain because natural selection can be influenced by a diverse set of parameters, including stability and biological function, as well as accessibility of native states and folding pathways.¹¹ Thus, we embarked on this mechanistic study to understand the basis for the incorrect predictions of several evolution-based mutations as destabilizing by the force-field calculations. Experimental proof of the entropic contribution to the stabilizing effect of these mutations shows the incompleteness of the current force-fields methods, especially with respect to the entropic effects. Indeed, several examples of entropic stabilization are known for specific substitutions, e.g., G to A,⁴⁵ X to P,⁴⁵ and R to K,⁴⁶ but their effects remain hard to predict computationally.²⁹ It also shows that protein stabilization can be achieved via both enthalpy and entropy changes, while the latter can only be predicted via evolutionary approaches or extremely computationally demanding calculations. This observation argues for mutual complementarity of the evolution-based and energy-based approaches and their parallel use for stability predictions as in FireProt.⁶ Including the entropic term in the $\Delta\Delta G$ predictions by the molecular mechanics combined with the generalized Born and surface area continuum solvation method did not lead to significant improvement of the stabilization predictions. Understanding mutations stabilizing proteins primarily by entropic contributions could lead to improvements of the protein design methods, such as FoldX,⁷ Rosetta,⁸ ERIS,⁹ or CUPSAT;¹⁰ toward that end, we put forward the mutants described in this work as suitable case systems for future development of improved predictive techniques.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscatal.8b01677.

Experimental procedures and additional data; predicted effects of the mutations on the stability of DhaA (PDB ID 4E46); diffraction data collection and refinement statistics; effect of ions on stability of mutants determined using DSC; primers used for site-directed mutagenesis; B-factors calculated for the backbone

atoms of the mutated residues over the MD simulations; relative free-energy differences for the evolutionary single mutants; protein stability predictions of the mutations identified by the back-to-consensus analysis and evaluated by Rosetta; expanded view of two PEG molecules in the crystal structure of DhaA101; radial distribution function, $RDF(r)$, of the Na^+ and Cl^- ions and water around the protein for DhaAwt and the mutants that introduced a change of charge; analysis of oligomer formation of DhaA variants and thermostability analysis of DhaA115 oligomers; entropy differences of DhaA variants; thermostability analysis of DhaA evolutionary-based single- and double-point variants; flexibility of DhaAwt and the single mutants during the MD simulations; water-oxygen density maps relative to the bulk solvent obtained with GIST analysis for DhaAwt and the single-point mutants; temperature dependence of relative free-energy differences obtained from TI for the single mutants (PDF)

Dataset statistics, experimental data set, and back-to-consensus data set (XLSX)

Accession Codes

The refined model of DhaA101 and its corresponding atomic coordinates and experimental structure factors were deposited in the database under Protein Data Bank (PDB) code 5FLK.

AUTHOR INFORMATION

Corresponding Authors

*zbynek@chemi.muni.cz.

*jiri@chemi.muni.cz.

ORCID

Jiri Damborsky: 0000-0002-7848-8216

Present Address

[∇]Centre for Synthetic Biology, Unit for Biocatalysis and Enzyme Engineering, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, BE-9000 Ghent, Belgium

Author Contributions

All authors designed experiments, interpreted data, and contributed to the writing of the paper. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors express their sincere thanks to Dr. Catherine Goodman for her critical comments on the manuscript. The authors kindly thank Prof. Huimin Zhao, the assigned Associate Editor, and the anonymous referees for their constructive feedback and helpful comments. Dr. Maria Pechlaner and Prof. Chris Oostenbrink are acknowledged for providing a modified version of the GROMOS program package. The work was supported by the Grant Agency of the Czech Republic (17-24321S and P503/12/0572), the Czech Ministry of Education of the Czech Republic (LO1214, LQ1605, LM2015051, LM2015047, and LM2015055), and European Union (316345, 720776, and 722610). K.B. and J.B. were supported by the “Employment of Best Young Scientists for International Cooperation Empowerment” (CZ.1.07/2.3.00/30.0037). A.K. was supported by the Brno Ph.D. Talent Scholarship Holder, which is funded by the Brno City Municipality. M.M. was supported by the grant awarded to the

Faculty of Information Technologies (FIT-S-17-3964). S.M.M. was supported by the SoMoPro II Programme (Project BIOGATE, Nr. 4SGA8519) cofinanced by the REA Grant Agreement no. 291782 and the South Moravian Region. N.H. was supported by the German Research Foundation (DFG) within the Cluster of Excellence in Simulation Technology (EXC 310/2) at the University of Stuttgart. The MetaCentrum and CERIT-SC are acknowledged for providing access to computing facilities (LM2010005, CZ.1.05/3.2.00/08.0144). TI calculations were performed using the computational facilities bwUniCluster and BinAC funded by the Ministry of Science, Research and Arts and the Universities of the State of Baden-Württemberg, Germany, within the framework program bwHPC and the DFG (INST 39/963-1 FUGG).

REFERENCES

- (1) Adrio, J.; Demain, A. Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules* **2014**, *4*, 117–139.
- (2) Bommarius, A. S. Biocatalysis: A Status Report. *Annu. Rev. Chem. Biomol. Eng.* **2015**, *6*, 319–345.
- (3) Palackal, N.; Brennan, Y.; Callen, W. N.; Dupree, P.; Frey, G.; Goubet, F.; Hazlewood, G. P.; Healey, S.; Kang, Y. E.; Kretz, K. A.; Lee, E.; Tan, X.; Tomlinson, G. L.; Verruto, J.; Wong, V. W. K.; Mathur, E. J.; Short, J. M.; Robertson, D. E.; Steer, B. A. An Evolutionary Route to Xylanase Process Fitness. *Protein Sci.* **2004**, *13*, 494–503.
- (4) Koudelakova, T.; Chaloupkova, R.; Brezovsky, J.; Prokop, Z.; Sebestova, E.; Hesselner, M.; Khabiri, M.; Plevaka, M.; Kulik, D.; Kuta Smatanova, I.; Rezacova, P.; Ettrich, R.; Bornscheuer, U. T.; Damborsky, J. Engineering Enzyme Stability and Resistance to an Organic Cosolvent by Modification of Residues in the Access Tunnel. *Angew. Chem., Int. Ed.* **2013**, *52*, 1959–1963.
- (5) Wijma, H. J.; Floor, R. J.; Jekel, P. A.; Baker, D.; Marrink, S. J.; Janssen, D. B. Computationally Designed Libraries for Rapid Enzyme Stabilization. *Protein Eng., Des. Sel.* **2014**, *27*, 49–58.
- (6) Bednar, D.; Beerens, K.; Sebestova, E.; Bendl, J.; Khare, S.; Chaloupkova, R.; Prokop, Z.; Brezovsky, J.; Baker, D.; Damborsky, J. FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLoS Comput. Biol.* **2015**, *11*, e1004556.
- (7) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More than 1000 Mutations. *J. Mol. Biol.* **2002**, *320*, 369–387.
- (8) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins: Struct., Funct., Genet.* **2011**, *79*, 830–838.
- (9) Yin, S.; Ding, F.; Dokholyan, N. V. Eris: An Automated Estimator of Protein Stability. *Nat. Methods* **2007**, *4*, 466–467.
- (10) Parthiban, V.; Gromiha, M. M.; Schomburg, D. CUPSAT: Prediction of Protein Stability upon Point Mutations. *Nucleic Acids Res.* **2006**, *34*, W239–W242.
- (11) Baker, D.; Agard, D. A. Kinetics versus Thermodynamics in Protein Folding. *Biochemistry* **1994**, *33*, 7505–7509.
- (12) Johnson, C. M. Differential Scanning Calorimetry as a Tool for Protein Folding and Stability. *Arch. Biochem. Biophys.* **2013**, *531*, 100–109.
- (13) Wen, J.; Arthur, K.; Chemmalil, L.; Muzammil, S.; Gabrielson, J.; Jiang, Y. Applications of Differential Scanning Calorimetry for Thermal Stability Analysis of Proteins: Qualification of DSC. *J. Pharm. Sci.* **2012**, *101*, 955–964.
- (14) Stsiapanava, A.; Dohnalek, J.; Gavira, J. A.; Kutý, M.; Koudelakova, T.; Damborsky, J.; Kuta Smatanova, I. Atomic Resolution Studies of Haloalkane Dehalogenases DhaA04, DhaA14 and DhaA15 with Engineered Access Tunnels. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 962–969.
- (15) Lee, D.; Lee, J.; Seok, C. What Stabilizes Close Arginine Pairing in Proteins? *Phys. Chem. Chem. Phys.* **2013**, *15*, 5844–5853.
- (16) Shukla, D.; Schneider, C. P.; Trout, B. L. Complex Interactions between Molecular Ions in Solution and Their Effect on Protein Stability. *J. Am. Chem. Soc.* **2011**, *133*, 18713–18718.
- (17) Privalov, P. L.; Dragan, A. I. Microcalorimetry of Biological Macromolecules. *Biophys. Chem.* **2007**, *126*, 16–24.
- (18) Liu, L.; Yang, C.; Guo, Q. X. A Study on the Enthalpy-Entropy Compensation in Protein Unfolding. *Biophys. Chem.* **2000**, *84*, 239–251.
- (19) Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
- (20) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
- (21) Lehmann, M.; Loch, C.; Middendorf, A.; Studer, D.; Lassen, S. F.; Pasamontes, L.; van Loon, A. P. G. M.; Wyss, M. The Consensus Concept for Thermostability Engineering of Proteins: Further Proof of Concept. *Protein Eng., Des. Sel.* **2002**, *15*, 403–411.
- (22) Amin, N.; Liu, A. D.; Ramer, S.; Aehle, W.; Meijer, D.; Metin, M.; Wong, S.; Gualfetti, P.; Schellenberger, V. Construction of Stabilized Proteins by Combinatorial Consensus Mutagenesis. *Protein Eng., Des. Sel.* **2004**, *17*, 787–793.
- (23) Watanabe, K.; Ohkuri, T.; Yokobori, S.; Yamagishi, A. Designing Thermostable Proteins: Ancestral Mutants of 3-Isopropylmalate Dehydrogenase Designed by Using a Phylogenetic Tree. *J. Mol. Biol.* **2006**, *355*, 664–674.
- (24) DiTursi, M. K.; Kwon, S.-J.; Reeder, P. J.; Dordick, J. S. Bioinformatics-Driven, Rational Engineering of Protein Thermostability. *Protein Eng., Des. Sel.* **2006**, *19*, 517–524.
- (25) Polizzi, K. M.; Chaparro-Riggers, J. F.; Vazquez-Figueroa, E.; Bommarius, A. S. Structure-Guided Consensus Approach to Create a More Thermostable Penicillin G Acylase. *Biotechnol. J.* **2006**, *1*, 531–536.
- (26) Vázquez-Figueroa, E.; Chaparro-Riggers, J.; Bommarius, A. S. Development of a Thermostable Glucose Dehydrogenase by a Structure-Guided Consensus Concept. *ChemBioChem* **2007**, *8*, 2295–2301.
- (27) Bava, K. A.; Gromiha, M. M.; Uedaira, H.; Kitajima, K.; Sarai, A. ProTherm, Version 4.0: Thermodynamic Database for Proteins and Mutants. *Nucleic Acids Res.* **2004**, *32*, 120D–121.
- (28) Pucci, F.; Bourgeas, R.; Rooman, M. Predicting Protein Thermal Stability Changes upon Point Mutations Using Statistical Potentials: Introducing HoTMuSiC. *Sci. Rep.* **2016**, *6*, 23257.
- (29) Meirovitch, H.; Cheluvraja, S.; White, R. P. Methods for Calculating the Entropy and Free Energy and Their Application to Problems Involving Protein Flexibility and Ligand Binding. *Curr. Protein Pept. Sci.* **2009**, *10*, 229–243.
- (30) Baxa, M. C.; Haddadian, E. J.; Jumper, J. M.; Freed, K. F.; Sosnick, T. R. Loss of Conformational Entropy in Protein Folding Calculated Using Realistic Ensembles and Its Implications for NMR-Based Calculations. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15396–15401.
- (31) Cooper, A.; Johnson, C. M.; Lakey, J. H.; Nöllmann, M. Heat Does Not Come in Different Colours: Entropy-Enthalpy Compensation, Free Energy Windows, Quantum Confinement, Pressure Perturbation Calorimetry, Solvation and the Multiple Causes of Heat Capacity Effects in Biomolecular Interactions. *Biophys. Chem.* **2001**, *93*, 215–230.
- (32) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angew. Chem., Int. Ed.* **2016**, *55*, 7364–7368.
- (33) Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *J. Mol. Biol.* **2017**, *429*, 948–963.
- (34) Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and Flexible GPU Accelerated

Binding Free Energy Calculations within the Amber Molecular Dynamics Package. *J. Comput. Chem.* **2018**, *39*, 1354–1358.

(35) Pucci, F.; Bourgeas, R.; Rooman, M. High-Quality Thermodynamic Data on the Stability Changes of Proteins Upon Single-Site Mutations. *J. Phys. Chem. Ref. Data* **2016**, *45*, 023104.

(36) Watson, M. D.; Monroe, J.; Raleigh, D. P. Size-Dependent Relationships between Protein Stability and Thermal Unfolding Temperature Have Important Implications for Analysis of Protein Energetics and High-Throughput Assays of Protein–Ligand Interactions. *J. Phys. Chem. B* **2018**, *122*, 5278–5285.

(37) Fersht, A. R.; Daggett, V. Protein Folding and Unfolding at Atomic Resolution. *Cell* **2002**, *108*, 573–582.

(38) Tsytlonok, M.; Itzhaki, L. S. The How's and Why's of Protein Folding Intermediates. *Arch. Biochem. Biophys.* **2013**, *531*, 14–23.

(39) Lim, S. A.; Hart, K. M.; Harms, M. J.; Marqusee, S. Evolutionary Trend toward Kinetic Stability in the Folding Trajectory of RNases H. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 13045–13050.

(40) Romero-Romero, S.; Costas, M.; Rodríguez-Romero, A.; Fernández-Velasco, D. A. Reversibility and Two State Behaviour in the Thermal Unfolding of Oligomeric TIM Barrel Proteins. *Phys. Chem. Chem. Phys.* **2015**, *17*, 20699–20714.

(41) Sanchez-Ruiz, J. M. Protein Kinetic Stability. *Biophys. Chem.* **2010**, *148*, 1–15.

(42) Costas, M.; Rodríguez-Larrea, D.; De Maria, L.; Borchert, T. V.; Gómez-Puyou, A.; Sanchez-Ruiz, J. M. Between-Species Variation in the Kinetic Stability of TIM Proteins Linked to Solvation-Barrier Free Energies. *J. Mol. Biol.* **2009**, *385*, 924–937.

(43) Rodríguez-Larrea, D.; Minning, S.; Borchert, T. V.; Sanchez-Ruiz, J. M. Role of Solvation Barriers in Protein Kinetic Stability. *J. Mol. Biol.* **2006**, *360*, 715–724.

(44) Mazurenko, S.; Kunka, A.; Beerens, K.; Johnson, C. M. C. M.; Damborsky, J.; Prokop, Z. Exploration of Protein Unfolding by Modelling Calorimetry Data from Reheating. *Sci. Rep.* **2017**, *7*, 16321.

(45) Lehmann, M.; Pasamontes, L.; Lassen, S. F.; Wyss, M. The Consensus Concept for Thermostability Engineering of Proteins. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **2000**, *1543*, 408–415.

(46) Berezovsky, I. N.; Chen, W. W.; Choi, P. J.; Shakhnovich, E. I. Entropic Stabilization of Proteins and Its Proteomic Consequences. *PLoS Comput. Biol.* **2005**, *1*, e47.