### Title

Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset




### Journal

ACS Omega, 5(12)

### ISSN

2470-1343



### Authors

Huang, Peishan
Chu, Simon KS
Frizzo, Henrique N
et al.



### Publication Date

2020-03-31

### DOI

10.1021/acsomega.9b04105

Peer reviewed

ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

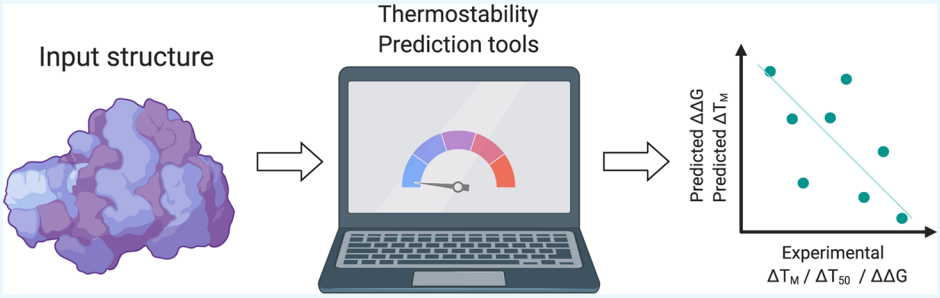
# Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset

Peishan Huang, Simon K. S. Chu, Henrique N. Frizzo, Morgan P. Connolly, Ryan W. Caster, and Justin B. Siegel*

Cite This: *ACS Omega* 2020, 5, 6487−6493

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information




**ABSTRACT:** Engineering proteins to enhance thermal stability is a widely utilized approach for creating industrially relevant biocatalysts. The development of new experimental datasets and computational tools to guide these engineering efforts remains an active area of research. Thus, to complement the previously reported measures of $T_{50}$ and kinetic constants, we are reporting an expansion of our previously published dataset of mutants for $\beta$-glucosidase to include both measures of $T_M$ and $\Delta\Delta G$. For a set of 51 mutants, we found that $T_{50}$ and $T_M$ are moderately correlated, with a Pearson correlation coefficient and Spearman's rank coefficient of 0.58 and 0.47, respectively, indicating that the two methods capture different physical features. The performance of predicted stability using nine computational tools was also evaluated on the dataset of 51 mutants, none of which are found to be strong predictors of the observed changes in $T_{50}$, $T_M$, or $\Delta\Delta G$. Furthermore, the ability of the nine algorithms to predict the production of isolatable soluble protein was examined, which revealed that Rosetta $\Delta\Delta G$, FoldX, DeepDDG, PoPMuSiC, and SDM were capable of predicting if a mutant could be produced and isolated as a soluble protein. These results further highlight the need for new algorithms for predicting modest, yet important, changes in thermal stability as well as a new utility for current algorithms for prescreening designs for the production of mutants that maintain fold and soluble production properties.

## INTRODUCTION

A common goal of enzyme engineering is the enhancement of thermal stability.[1] For industrial applications, improving a protein's robustness to thermal challenges or half-life at elevated temperatures can often be the deciding factor for the commercialization of a biocatalyst.[2−5] Currently, the most common approach for improving thermal stability is through directed evolution methodologies,[6,7] which can be time-consuming, costly, and limited in the ability to search sequence space. Computational design tools to predictably identify single and combinatorial mutations that enhance thermal stability are rapidly developing and growing in popularity.[8−14] However, accurate predictions using computational tools to guide protein stability design remain an active area of research and is not always successful.

The use of large datasets on the mutational effect on protein stability, such as ProTherm[15] now maintained by ProtaBank,[16] is often used to train computational methods for predicting thermal stability. The datasets utilized generally consist of the equilibrium constant of unfolding ($K_u$) or the melting temperature of an enzyme ($T_M$).[17] In our previous study, we determined the thermal stability of 79 $\beta$-glucosidase B (BglB) variants by finding $T_{50}$, a type of kinetic stability that is determined by the temperature at which a mutant's residual activity is reduced by 50% after a heat challenge over a defined time.[4,17,18] When analyzing this set of mutants using two established computational programs (Rosetta $\Delta\Delta G$ and FoldX PSSM) for predicting thermal stability, we found that there was no significant correlation between the predictions and the observed $T_{50}$.[19]

One hypothesis explaining the poor predictive performance of the algorithms with the BglB dataset is that the algorithms

Received: December 2, 2019
Accepted: March 6, 2020
Published: March 20, 2020




https://dx.doi.org/10.1021/acsomega.9b04105
*ACS Omega* 2020, 5, 6487−6493

are evaluated on $T_M$, a direct measure of structural thermal stability. However, the algorithms were being used to predict $T_{50}$, which is an indirect measure of the protein's thermal stability.[17] Alternatively, the poor performance could have come from the narrow $T_{50}$ range (extreme variants are +6.06 and −5.02 °C from the wild type (WT)) as the algorithms are generally benchmarked on larger changes in thermal stability and ±3 °C may be within the error of the currently developed algorithms. In this study, we evaluated both hypotheses.

To assess if there was a significant difference in $T_M$ and $T_{50}$, we developed a dataset of 51 BglB mutants (Figure 1) in which

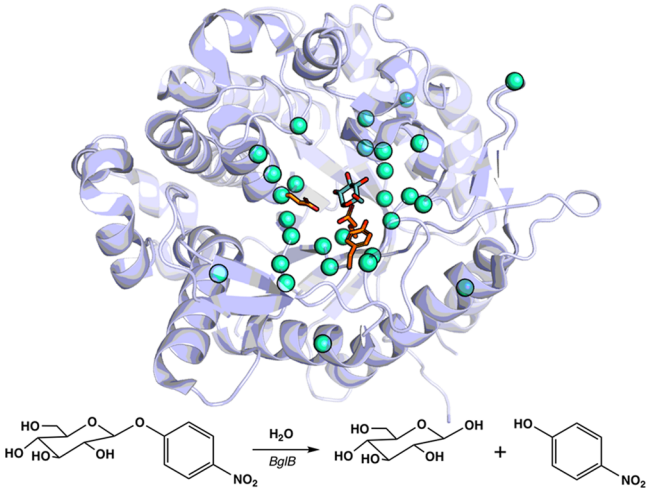



**Figure 1.** Structure of BglB (PDB ID: 2JIE) from the bacterium *Paenibacillus polymyxa*. PyMOL rendering[20] of BglB showing the 28 sequence-positions (teal spheres) of the 51 mutants chosen out of the original 92 previously expressed proteins for the $T_M$ analysis.[19] The reaction scheme of the hydrolysis of 4-nitrophenyl β-D-glucopyranoside by BglB used in the $T_{50}$ study.[19]

both thermal stability measurements, $T_{50}$ and $T_M$, were measured. Interestingly, for the set of 51 measurements, there was only a modest correlation between $T_{50}$ and $T_M$, with a Pearson coefficient correlation (PCC) and Spearman's rank correlation (SRC) of 0.58 and 0.47, respectively. This highlights the difference in the physical properties being measured using these two techniques, $T_M$ being the thermal stability of the protein's structural elements and $T_{50}$ reporting on the thermal stability to irreversible denaturation. However, similar to the previous study,[19] the relationship between the predicted stability with the experimental $T_M$ only results in a weak correlation not only with the previous algorithms evaluated (Rosetta ΔΔ$G$ and FoldX PSSM) but also with five other commonly used methods: ELASPIC, DeepDDG, PoPMuSiC, SDM, and AUTO-MUTE. This result suggests that while the two measurements are reporting on different physical properties, this is not the key factor that led to the low predictive accuracy of established algorithms on this dataset.

To evaluate the second hypothesis, that the changes in thermal stability of the BglB dataset are too small for current algorithms, we investigated the ability of the algorithms to predict if a mutation reduced thermal stability to the point that the protein could no longer be produced and isolated in a soluble form. Analysis of the computational algorithms to predict destabilization to the point where no soluble protein could be isolated showed a significant enrichment based on the calculated energetics of the mutants for several algorithms, the most significant of which is for Rosetta ΔΔ$G$. This supports the hypothesis that the lack of performance on the BglB dataset is due to the narrow range in changes observed for thermal stability. These slight molecular changes, especially interactions that are less than 1 kcal/mol, are challenging to accurately model. This highlights the need for new algorithms for predicting modest, yet important, changes in thermal stability as well as a new utility for current algorithms for prescreening designs for the production of mutants likely to maintain protein structure and be produced as a soluble protein.

## METHODS

**Mutant Selection, Protein Expression, and Purification.** Out of 79 mutants of BglB that were previously characterized with $T_{50}$ data,[19] 51 variants with plasmid readily available were transformed into chemically competent *Escherichia coli* BLR (DE3) cells. The variants were produced and purified, as previously described.[14] Expression was carried out by growing a 5 mL overnight culture in a 50 mL falcon tube with a breathable seal in Terrific broth (TB) medium with kanamycin while shaking at 250 rpm at 37 °C. After the initial overnight culture, cells were spun down and resuspended in fresh TB with kanamycin with 1 mM isopropyl β-D-1-thiogalactopyranoside in a 50 mL falcon tube with a breathable seal and incubated while shaking at 250 rpm at 18 °C for 24 h. Then, the cells were spun down, lysed, and purified using immobilized metal ion affinity chromatography, as previously described.[19] The purity of the protein samples was analyzed using 12−14% SDS-PAGE (Figure SI 1-1), and the yield was assessed based on the A280 for proteins that appeared >75% pure in the SDS-PAGE analysis. Protein samples were considered expressed if they were detectable in the SDS-PAGE analysis and greater than 0.10 mg/mL using A280, as previously described.[19]

**Melting Temperature Assay.** The melting temperature ($T_M$) of BglB was determined using the Protein Thermal Shift (PTS) kit (Applied Biosystems, from Thermo Fisher Scientific). Standard protocols provided by the manufacturer were used. Protein concentrations ranged from 0.1−0.5 mg/mL, and fluorescence reading was monitored with a QuantaStudio 3 system from 20 to 90 °C. The temperature melting curve was first smoothed with a 20 step sliding window average (Script SI 2). $T_M$ was determined from the average of three to four replicates at which the derivative was largest, and all melting curves can be found in Figure SI 3.

**Δ$G$ Calculations from $T_M$.** Calculations were conducted, as previously described.[21] First, we assumed that the protein follows the two-state folding mechanism, a binary conversion of native state to full denaturation. Second, to derive $\Delta G^\circ_{unfolding}$, the fluorescence intensity was first translated into fractions of folded ($P_f$) and unfolded ($P_u$) proteins of the linear portion of the graph at different temperatures starting from the minimum fluorescence ($F_{min}$) to the maximum fluorescence ($F_{max}$) shown in eq 1.

$$P_f = 1 - \frac{F - F_{min}}{F_{max} - F_{min}} \quad (1)$$

By taking a two-state folding−unfolding model, the equilibrium constant of unfolding ($K_u$) at different temperatures is then given by


https://dx.doi.org/10.1021/acsomega.9b04105
ACS Omega 2020, 5, 6487−6493

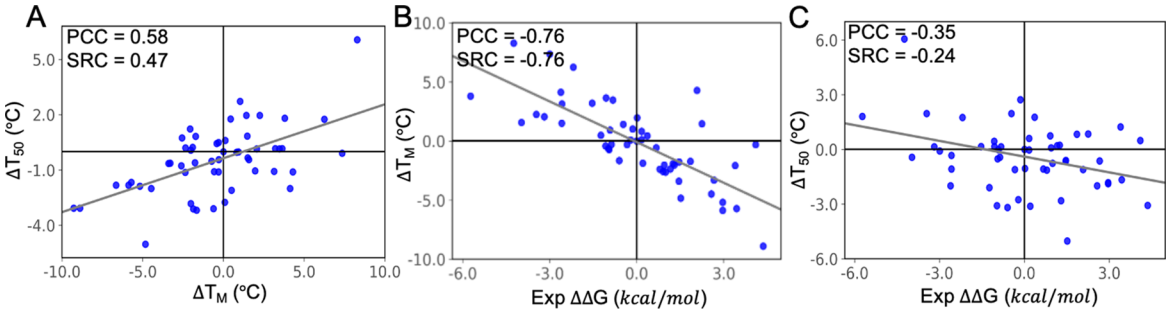



**Figure 2.** Comparison of two different experimental thermal stability datasets and experimentally derived $\Delta\Delta G$. (A) Relationship for each mutant between $T_{50}$ and $T_M$. The PCC of 0.58 illustrates that the two methods are modestly positively correlated with mutations that are in the extreme ends of the temperature range (±5 °C). (B) Evaluation of $\Delta T_M$ with experimentally derived $\Delta\Delta G$ shows the two qualities are highly correlated (PCC = −0.76), unlike (C) where the relationship between $\Delta T_{50}$ and experimentally derived $\Delta\Delta G$ has a PCC of −0.35.

$$K_u = \frac{P_f}{P_u} \tag{2}$$

We plotted ln $K_u$ against $1/T$ using the van't Hoff method shown in eq 3 (Script SI 2), the $-{}^{\Delta H}/_{R}$ is defined by the slope, ${}^{\Delta S}/_{R}$ is the $y$-intercept, $T$ is temperature, and $R$ is the ideal gas constant.

$$\ln K_u = -\frac{\Delta H}{RT} + \frac{\Delta S}{R} \tag{3}$$

The Gibbs free energy of protein-unfolding can then be determined using eq 4, where $\Delta G^\circ_{unfolding}$ is the unfolding energy at a $T_{RT}$ of 298 K. All calculations can be found in Script SI 2.

$$\Delta G^\circ_{unfolding} = \Delta H - T_{RT}\Delta S \tag{4}$$

**Molecular Modeling.** Seven popular, readily accessible, and recently developed molecular modeling methods, many of them force-field and machine-learning-based, were evaluated for their ability to recapitulate the experimental data: Rosetta $\Delta\Delta G$,[22] FoldX,[23] ELASPIC,[24] DeepDDG,[25] PoPMuSiC,[26] SDM,[27] and AUTO-MUTE (DDG).[28] The crystal structure of BglB (PDB ID: 2JIE) was used across seven different algorithms. First, using a previously described method,[19] the 2JIE structure was used as input to the Rosetta $\Delta\Delta G$ application and run, as previously described (Script SI 5). Briefly, 50 poses of the WT and the mutant were generated for which 15 energy terms were reported from the score function used.[22] The three lowest system energy scores out of 50 from WT and the mutant were averaged to give the final Rosetta $\Delta\Delta G$ score. Second, for the FoldX position-specific scoring metric (PSSM) protocol, the 2JIE structure was first minimized for any potential inaccurate rotamer assignment using the RepairPDB application.[23] The repaired PDB structure was mutated with single-point mutants and then modeled using FoldX PSSM. The model was scored based on 17 terms within the FoldX force-field.[23] Third, the ELASPIC protocol first constructed a homology model of the WT using the crystal structure, sequence, molecular, and energetics information. Using the standard procedure described, the FoldX algorithm was used to construct the mutant model. The final mutational change is predicted using Stochastic Gradient Boosted Decision Trees based on the energetic, chemical, and structural features from FoldX.[24,29] Fourth, using a curated dataset derived from the Protherm database,[15] DeepDDG used their previously described shared residue pair neural network structure to make a prediction of stability.[25] The DeepDDG output indicated that >0 kcal/mol could be considered stable, whereas <0 kcal/mol could be considered unstable. Fifth, PoPMuSiC estimated the stability of the WT structure and mutants using 13 statistically potential terms, and an additional two terms that account for the volume differences of the residues between WT and the mutant.[26] Sixth, the SDM method evaluated mutational changes using a statistical potential energy function based on environment-specific substitution tables. These tables consisted of data such as structural information, solvent accessibility of the sidechain, and hydrogen bonding.[27] Lastly, similar to SDM, the seventh method, AUTO-MUTE, which predicts for $\Delta T_M$ and $\Delta\Delta G$, utilized a statistical potential to calculate the environmental changes of the residue compared to the WT.[28] The protocol was performed using tree regression at 23 °C and pH 7.5.

Apart from predicting $\Delta\Delta G$, two additional methods were used to evaluate the algorithms' ability to predict $\Delta T_M$ changes. As mentioned above, the AUTO-MUTE prediction using the Stability Changes ($\Delta T_M$) protocol was performed with tree regression. Also, HotMuSiC was used to evaluate the mutational effect with the temperature-dependent potential and other statistical potential terms such as solvent accessibility, structural, and sequence-based information.[30]

Pearson correlation coefficient (PCC) and Spearman's rank correlation (SRC) analyses were performed between their respective $\Delta\Delta G$ ($\Delta\Delta G = \Delta G_{mutants} - \Delta G_{WT}$) or the change in total system energy (ΔTSE) of the nine computational methods. Additionally, the available individual features within the Rosetta $\Delta\Delta G$ and FoldX PSSM force field were further evaluated against the $T_M$ dataset for correlation.

Finally, ΔTSE was evaluated against mutants that could be isolated as a soluble protein and those that lost structural integrity and either precipitated or were degraded and therefore could no longer be isolated as a soluble protein (nonisolated). A Student's $t$-test was used to obtain $p$-values for the nine computational methods. The two categories between isolated and nonisolated protein were treated as an independent sample using an unequal variance.

## RESULTS

**Evaluating the Relationship between $T_M$ and $T_{50}$.** To the best of our knowledge, there has not been a large dataset (>50 data points) directly comparing the $T_M$ and $T_{50}$ relationship for a single set of protein mutants uniformly produced and characterized. It is important to distinguish both $T_M$ and $T_{50}$ methods since the measurements are quantifying


https://dx.doi.org/10.1021/acsomega.9b04105
ACS Omega 2020, 5, 6487−6493

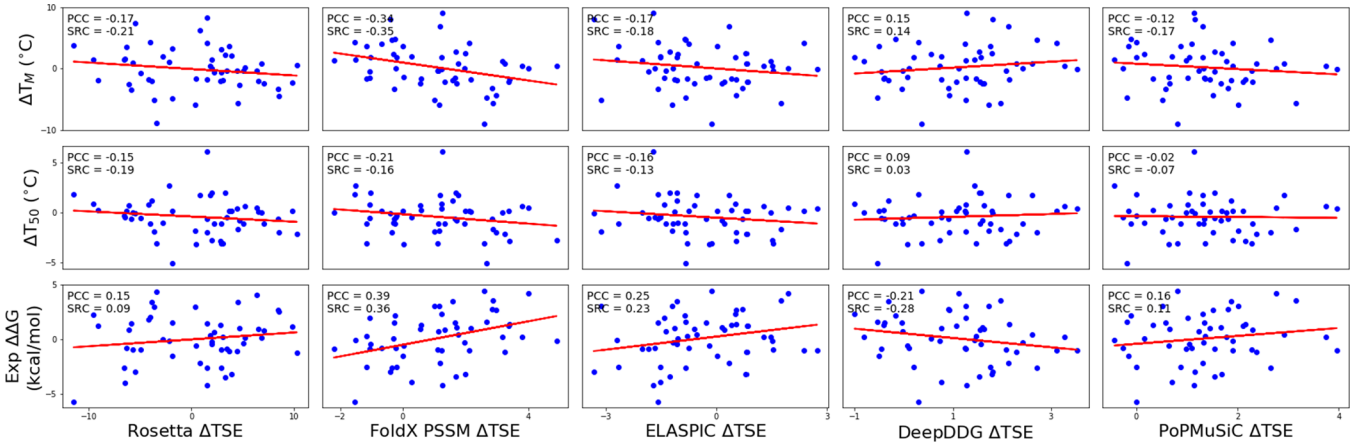



**Figure 3.** Evaluation of the algorithms ΔTSE versus the experimentally derived $\Delta\Delta G$ and the $T_M$ and $T_{50}$ datasets. The Pearson correlation coefficient and Spearman's rank correlation for each performance against three types of experimental data were determined. Five representative comparisons are illustrated above, with four additional algorithms, SDM, AUTO-MUTE (DDG), AUTO-MUTE ($\Delta T_M$), and HoTMuSiC provided in Figure SI 1-3. No algorithm resulted in a significant correlation between the calculated energies and the observed $T_M$, $T_{50}$, or $\Delta\Delta G$ for this dataset.

and reporting different structural and functional properties. $T_M$ is the temperature at which half the enzyme is found in the unfolded state over the folded state.[17,31] This is often evaluated through denaturation assays, where the thermodynamic measurements ($\Delta G_{unfolding}$) can be obtained.[31] This method is generally a lower throughput method as purified protein is required to obtain an accurate measurement for the structural properties for the mutant being evaluated. $T_{50}$ measures the temperature of half-inactivation that leads to irreversible unfolding,[11,32] and it is determined by the reduction of half of the enzymatic activity due heat challenges.[17] This is a very common assay for protein engineering due to its compatibility with high-throughput assays and the ability to use cell lysates to evaluate function.

To complement our previously measured dataset of $T_{50}$, 51 of the 92 expressed proteins with available plasmids[19] were selected and evaluated for $T_M$ using the Protein Thermal Shift assay to compare $T_{50}$ and $T_M$. The WT BglB $T_M$ was determined to be 45.97 ± 1.03 °C, while the previously determined $T_{50}$ was 39.9 ± 0.1 °C.[19] When evaluating the entire dataset, the $T_M$ ranged between 37.1 and 54.3 °C, slightly larger than what was observed for $T_{50}$, which was between 34.9 and 46.0 °C (Figure SI 1-2). The variant that had the highest $T_M$ in this dataset was E167A, with a $\Delta T_M$ of 54.3 °C (+8.33 °C), which was also observed to have a similar increase in $T_{50}$ compared with the WT (+6.06 °C).[19] The variant that had the lowest $T_M$ in this dataset was found to be E225A, with a $\Delta T_M$ of −8.9 °C, which had a corresponding $T_{50}$ of −3.1 °C.

The relationship between $T_{50}$ and $T_M$ is plotted in Figure 2A. The PCC and SRC of 0.58 and 0.47, respectively, indicate that the two methods are moderately positively correlated. Correlation between methods increased in cases where mutations resulted in extremely stable and unstable products, for example, E167A and E225A, respectively. This is an expected result for small changes (<3 °C) in thermal stability; the differences in measurement methods would be expected to play a more significant role than for larger changes (>5 °C). The evaluation of $\Delta T_M$ and $\Delta T_{50}$ with experimentally derived $\Delta\Delta G$ is also plotted in Figure 2B,C, respectively. The PCC and SRC show that the $T_M$ method and experimentally derived $\Delta\Delta G$ are strongly correlated (PCC and SRC of −0.76), compared to those between $\Delta\Delta G$ and $T_{50}$ (PCC of −0.35 and SRC of −0.24).

**Evaluating Computational Stability Tools Using the BglB $T_M$ Dataset.** The computational evaluation of protein stability of the current experimental $T_M$ dataset was analyzed in the same manner as our previous study on $T_{50}$.[19] An energetically evaluated model for each mutant was generated using established computational methods and subsequently plotted as a function of $T_M$ to evaluate the calculated energies related to the observed $T_M$. The PCC and SRC for the most commonly assessed term, the ΔTSE, was found to be highest for FoldX PSSM (PCC of −0.34 and SRC of −0.35) with $\Delta T_M$ (Figure 3). Similarly, the FoldX PSSM correlations with experimentally derived $\Delta T_{50}$ data were found to be −0.21 and −0.16 for PCC and SRC, respectively. The overall relationship between the ΔTSE and the $\Delta T_M$ thermal stability dataset slightly improved for FoldX, DeepDDG, PoPMuSiC, and AUTO-MUTE (Figure SI 1-3), while Rosetta $\Delta\Delta G$ and ELASPIC remained relatively unchanged with no significant correlation. Interestingly, SDM was the only method where the correlation with $\Delta T_{50}$ is stronger than that of $\Delta T_M$ (Figure SI 1-3).

An analysis of individual energetic term from Rosetta $\Delta\Delta G$ and FoldX PSSM did not uncover any specific feature in either method's energetic evaluation that was strongly correlated with the $T_M$ dataset, as was previously observed for the $T_{50}$ dataset[19] (Figure SI 4). The strongest PCC for $T_M$ against any of the available energetic terms was 0.39 for the Δbackbone clash term from FoldX PSSM and −0.31 for the Omega energy term from Rosetta $\Delta\Delta G$. To be consistent with the previous performance assessment, we also evaluated the algorithms on experimentally derived $\Delta\Delta G$ in this dataset (Figure 3). The PCC and SRC of 0.39 and 0.36, respectively, between experimental $\Delta\Delta G$ and ΔTSE for FoldX PSSM outperformed six other algorithms that were compared. The correlation between experimental $\Delta\Delta G$ with ΔTSE was not unexpected as $T_M$ showed a correlation with $\Delta\Delta G$ with a PCC and SRC of −0.76 (Figure 2B). Analysis of AUTO-MUTE and HotMuSiC to predict for $\Delta T_M$ revealed no significant correlation with the experimental $\Delta T_M$ (Figure SI 1-3).


https://dx.doi.org/10.1021/acsomega.9b04105
ACS Omega 2020, 5, 6487−6493

Based on this analysis, it is apparent that the general performance of all given methods at best only weakly correlates with the experimentally determined effects of the mutations. This data fails to support the hypothesis that the lack of a previously observed correlation of these established computational tools with observed changes in thermal stability in the BglB dataset is due to the difference in the physical property being measured.

**Prediction of Mutant Soluble Expression.** The current dataset consists primarily of modest changes in thermal stability of <5 °C, calculated to be ±4 kcal/mol of the WT, and therefore may be challenging for current computational methods to predict. However, this change has only been analyzed in a fraction of the 129 mutants tested in the overall BglB dataset. Of the 129 mutants, only 92 were found to be produced and isolated in a soluble form. All purification procedures are conducted at ∼20 °C. Since the WT has a $T_{50}$ of 39.9 °C, any reduction in $T_{50}$ of >18 °C would result in a loss of structural stability from which insoluble aggregates or proteolytic degradation would readily occur during production and purification. In this case, the proteins would no longer be able to be isolated in a soluble form similar to the WT protein. Therefore, it seemed pertinent to evaluate if any of the nine algorithms could differentiate variants in this dataset that could be isolated as a soluble protein versus those that were not able to be separated as a soluble protein.

For this evaluation, all of the previously reported 129 mutants were assessed using the nine algorithms following the same methods used for $T_{50}$ and $T_M$. A mutant was generally considered soluble if it was observed on an SDS-PAGE analysis and had an A280 >0.1 mg/mL. The WT protein produced using the methods described generally resulted in an average A280 of 1.5 mg/mL, which would provide a >10-fold change in yield for mutants having an A280 less than 0.1 mg/mL. While factors other than thermal stability can affect production and isolation of soluble protein, in this case, it is assumed that the primary factor that decreases soluble protein yield is from denaturation of the mutant protein either during expression or purification. The results of this analysis are presented in Figure 4.

Of the nine algorithms evaluated, Rosetta ΔΔ*G*, FoldX, DeepDDG, PoPMuSiC, and SDM can capture the enrichment of mutants isolated as a soluble protein. The differences were evaluated for statistical significance using the Student's *t*-test, and the highest among the top five methods was shown for Rosetta ΔΔ*G* with a *p*-value of $1.0 \times 10^{-5}$. In contrast, enrichment was lower for ELASPIC, AUTO-MUTE (DDG), AUTO-MUTE ($\Delta T_M$), and HotMuSiC with *p*-values of 0.06, 0.38, 0.65, and 0.07, respectively.

A few outliers were observed in all methods, except for ELASPIC (Figure SI 1-4). For example, the mutant G15N for both Rosetta ΔΔ*G* and FoldX PSSM was identified as severely energetically unfavorable, which is consistent with the observation that this variant was not able to be isolated as a soluble protein.

## DISCUSSION

Both $T_M$ and $T_{50}$ are methods commonly used to quantify different physical aspects of protein thermal stability; however, to date, there has been relatively little experimental data collected to empirically evaluate the relationship of these two measurements. Using a dataset of 51 protein mutants, we observed that there is a moderate positive correlation (PCC of 0.58 and SRC of 0.47) between these two properties. The theory comparing two methods has been extensively described in the work of Hei and Clark.[33] Briefly, $T_{50}$ can only be used to assess the temperature at which half of the protein is irreversibly unfolded. Meanwhile, $T_M$ provides information on the folded state of the protein regardless of whether or not the unfolding events are irreversible. Therefore, it is not surprising that there is only a moderate correlation between the relationship of $T_{50}$ and $T_M$.

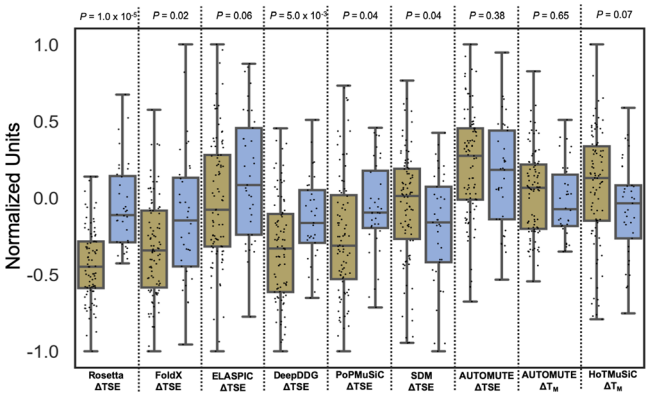



**Figure 4.** Computational prediction for the effect on mutant soluble protein production using nine different algorithms. From left to right: Rosetta ΔΔ*G*, FoldX PSSM, ELASPIC, DeepDDG, PoPMuSiC, SDM, AUTO-MUTE (DDG), AUTO-MUTE ($\Delta T_M$), and HoTMuSiC of soluble (green) and nonisolated protein (blue). In this case, mutants that resulted in a significant (>10-fold) decrease in yield of purified soluble protein are considered nonisolatable. Significance in population differences was determined using a Student's *t*-test. The units (ΔTSE and $\Delta T_M$) of all algorithms are individually normalized between 1 to −1. For visualization purposes, outliers were omitted after normalization. Each graph without normalization and with outliers can be found in Figure SI 1-4 and all raw values in Figure SI 4.

Mutants with extreme stability changes, such as E164A (>6 °C), usually exhibit a similar magnitude of change in $T_M$ and $T_{50}$ results. However, the majority of the mutants show a change of ∼3 °C or less in this $T_M$ and $T_{50}$ dataset being analyzed, a range in which the relationship between $T_M$ and $T_{50}$ appears to be weaker. Therefore, analysis with larger datasets with more extreme stability changes may reveal an even stronger correlation between these two properties.

The relationship between $\Delta T_M$ and the experimentally derived ΔΔ*G* of this dataset (PCC and SRC of −0.76) is not expected to reach a perfect correlation since it is dependent on the temperature at which ΔΔ*G* was evaluated, as described by Pucci et al.[34] For example, the ΔΔ*G* evaluated at $T_M$ of the WT will yield a correlation closer to −1 and $\Delta\Delta G_{(25°C)}$ will lead to a lower correlation (−0.68).[34]

Consistent with our previous analysis, we found a lack of performance using established computational tools when predicting $T_M$ and $T_{50}$ from the WT for this dataset. According to Jia et al., stability prediction using the experimentally derived free energy change of unfolding ΔΔ*G* (kcal/mol) outperforms the prediction using $\Delta T_M$ (°C).[35] However, in this case, we saw no significant change in the predictive performance for all seven computational tools compared to the experimentally derived free energy change. In addition, we found $T_M$ and ΔΔ*G* to be strongly correlated with this dataset, which may suggest that the improved performance is only relevant for more diverse datasets composed of different proteins as opposed to mutants of a single protein.


https://dx.doi.org/10.1021/acsomega.9b04105
ACS Omega 2020, 5, 6487−6493

While none of the computational methods demonstrated a strong predictive power for the mutants in this study, Rosetta ΔΔ*G*, FoldX PSSM, ELASPIC, and PoPMuSiC all have previously been shown to have high correlations with experimental data (PCC between 0.69 to 0.83).[22,26,29,36] This dataset with an experimental ΔΔ*G* range of ±∼4 kcal/mol is within the majority of the mutants observed in the algorithms that were typically evaluated on (+8 to −5 kcal/mol).[17,25,26,29] One potential reason for the lack of performance could be that the structure used in this dataset has a ligand bound structure, and often, the structures used in the development of the methods were apoprotein structures. However, using the PDBFlex database,[37] a clustering of five available PDBs of BglB from the bacterium *P. polymyxa* showed an average RMSD of 0.234 and a maximum RMSD of 0.274, thus making BglB a rigid structure. As there are no significant structural changes between the apo-form and holo-form of the protein, it is unlikely that the exact structure used for this study resulted in the low level of performance by the algorithms. Another possibility is that the protein evaluated here (BglB) is an outlier when compared to the proteins used to develop the algorithms in terms of its structure−function relationship. However, a related study to our analysis has been conducted for human superoxide dismutase 1 in which a low correlation is observed between experimental and predicted stability.[38] This further validates that current algorithms have limited utility for proteins outside of those they were benchmarked on. This limitation hindered by an over-representation of protein families such as lysozyme, tryptophan synthase, and ribonuclease in curated datasets is often utilized in benchmarking.[39] Thus, this highlights the importance of generating high-quality and diverse datasets of more proteins for evaluating and training new computational tools.

This study underlines the need for new computational tools that can more accurately predict modest changes, rather than major changes, in thermal stability. This becomes particularly important because single-point mutants often increase thermal stability by a few degrees at a time, while major changes are more often produced from the synergistic effect of combining multiple mutations.[11,40−42] Furthermore, as larger datasets of protein mutants with explicitly measured biophysical properties are generated, opportunities to explore combinations of molecular modeling and machine learning methods will become practical. These algorithms and datasets will enable the development of robust predictors of thermal stability.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.9b04105.

Figure SI 1-1, SDS-PAGE images for 51 BglB mutants and WT; Figure SI 1-2, distribution analysis of temperatures observed for $T_M$ and $T_{50}$; Figure SI 1-3, evaluation of AUTO-MUTE, SDM, and HoTMuSiC on the thermal stability dataset; and Figure SI 1-4, evaluation of nine computational methods on protein expression (PDF)

Script SI 2, Jupyter notebook for all thermal stability data acquisitions with all $T_M$ raw data files (ZIP)

Figure SI 3, images of $T_M$ fluorescence graphs, derivative graphs, and van't Hoff plot for 51 mutants and WT (PDF)

Figure SI 4, Rosetta ΔΔ*G* and FoldX PSSM correlations graphs with $\Delta T_M$; excel files of all the parameters from data acquisition; excel file of the total system energy for DeepDDG, ELASPIC, PoPMuSiC, SDM, and AUTO-MUTE (DDG); and AUTOMUTE ($\Delta T_M$) and HoTMuSiC ($\Delta T_M$) (ZIP)

Script SI 5, example files for Rosetta_ddg_monomer run (ZIP)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Justin B. Siegel** − *Genome Center, Department of Biochemistry & Molecular Medicine, and Department of Chemistry, University of California, Davis 95616, California, United States*; Email: jbsiegel@ucdavis.edu

### Authors

**Peishan Huang** − *Biophysics Graduate Group, University of California, Davis 95616, California, United States*; orcid.org/0000-0002-8417-5793

**Simon K. S. Chu** − *Biophysics Graduate Group, University of California, Davis 95616, California, United States*

**Henrique N. Frizzo** − *Genome Center, University of California, Davis 95616, California, United States*

**Morgan P. Connolly** − *Microbiology Graduate Group, University of California, Davis 95616, California, United States*

**Ryan W. Caster** − *Genome Center, University of California, Davis 95616, California, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.9b04105


### Author Contributions

P.H., S.K.S.C., and H.N.F. curated the data. P.H. and S.K.S.C. conducted the investigation. P.H., S.K.S.C., M.P.C., R.W.C., and J.B.S. provided the methodology. S.K.S.C. and P.H. provided the software. P.H. wrote the original draft. P.H., J.B.S., S.K.S.C., and M.P.C. wrote, reviewed, and edited the paper.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the University of California Davis, the National Institutes of Health (R01 GM 076324-11), the National Science Foundation (award nos. 1827246, 1805510, and 1627539), and the National Institute of Environmental Health Sciences of the National Institutes of Health (award no. P42ES004699). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, National Institute of Environmental Health Sciences, National Science Foundation, or UC Davis.

## ■ REFERENCES

(1) Iyer, P. V.; Ananthanarayan, L. Enzyme Stability and Stabilization-Aqueous and Non-Aqueous Environment. *Process Biochem.* **2008**, *43*, 1019−1032.

(2) Turner, P.; Mamo, G.; Karlsson, E. N. Potential and Utilization of Thermophiles and Thermostable Enzymes in Biorefining. *Microb. Cell Fact.* **2007**, *6*, 9.

(3) Ferdjani, S.; Ionita, M.; Roy, B.; Dion, M.; Djeghaba, Z.; Rabiller, C.; Tellier, C. Correlation between Thermostability and

Stability of Glycosidases in Ionic Liquid. *Biotechnol. Lett.* **2011**, *33*, 1215−1219.
(4) Xie, Y.; An, J.; Yang, G.; Wu, G.; Zhang, Y.; Cui, L.; Feng, Y. Enhanced Enzyme Kinetic Stability by Increasing Rigidity within the Active Site. *J. Biol. Chem.* **2014**, *289*, 7994−8006.
(5) Wu, I.; Arnold, F. H. Engineered Thermostable Fungal Cel6A and Cel7A Cellobiohydrolases Hydrolyze Cellulose Efficiently at Elevated Temperatures. *Biotechnol. Bioeng.* **2013**, *110*, 1874−1883.
(6) Kazlauskas, R. J.; Bornscheuer, U. T. Finding Better Protein Engineering Strategies. *Nat. Chem. Biol.* **2009**, *5*, 526.
(7) Denard, C. A.; Ren, H.; Zhao, H. Improving and Repurposing Biocatalysts via Directed Evolution. *Curr. Opin. Chem. Biol.* **2015**, *25*, 55−64.
(8) Borgo, B.; Havranek, J. J. Automated Selection of Stabilizing Mutations in Designed and Natural Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1494−1499.
(9) Jacak, R.; Leaver-Fay, A.; Kuhlman, B. Computational Protein Design with Explicit Consideration of Surface Hydrophobic Patches. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 825−838.
(10) 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.* **2002**, *15*, 403−411.
(11) Goldenzweig, A.; Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **2018**, *87*, 105−129.
(12) Cruz, L.; Urbanc, B.; Borreguero, J. M.; Lazo, N. D.; Teplow, D. B.; Stanley, H. E. Solvent and Mutation Effects on the Nucleation of Amyloid $\beta$-Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 18258−18263.
(13) Dehghanpoor, R.; Ricks, E.; Hursh, K.; Gunderson, S.; Farhoodi, R.; Haspel, N.; Hutchinson, B.; Jagodzinski, F. Predicting the Effect of Single and Multiple Mutations on Protein Structural Stability. *Molecules* **2018**, *23*, 251.
(14) Tokuriki, N.; Stricher, F.; Schymkowitz, J.; Serrano, L.; Tawfik, D. S. The Stability Effects of Protein Mutations Appear to Be Universally Distributed. *J. Mol. Biol.* **2007**, *369*, 1318−1332.
(15) Gromiha, M. M.; An, J.; Kono, H.; Oobatake, M.; Uedaira, H.; Prabakaran, P.; Sarai, A. ProTherm, Version 2.0: Thermodynamic Database for Proteins and Mutants. *Nucleic Acids Res.* **2000**, *28*, 283−285.
(16) Wang, C. Y.; Chang, P. M.; Ary, M. L.; Allen, B. D.; Chica, R. A.; Mayo, S. L.; Olafson, B. D. ProtaBank: A Repository for Protein Design and Engineering Data. *Protein Sci.* **2018**, *27*, 1113−1124.
(17) Polizzi, K. M.; Bommarius, A. S.; Broering, J. M.; Chaparro-Riggers, J. F. Stability of Biocatalysts. *Curr. Opin. Chem. Biol.* **2007**, 220.
(18) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5869−5874.
(19) Carlin, D. A.; Hapig-Ward, S.; Chan, B. W.; Damrau, N.; Riley, M.; Caster, R. W.; Bethards, B.; Siegel, J. B. Thermal Stability & Kinetic Constants for 129 Variants of a Family 1 Glycoside Hydrolase Reveal That Enzyme Activity & Stability Can Be Separately Designed. *PLoS One* **2017**, *12*, No. e0176255.
(20) Schrödinger, L. *The PyMOL Molecular Graphics System* http://www.pymol.org.
(21) Wright, T. A.; Stewart, J. M.; Page, R. C.; Konkolewicz, D. Extraction of Thermodynamic Parameters of Protein Unfolding Using Parallelized Differential Scanning Fluorimetry. *J. Phys. Chem. Lett.* **2017**, *8*, 553−558.
(22) 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., Bioinf.* **2011**, *79*, 830−838.
(23) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The FoldX Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382−W388.
(24) Witvliet, D. K.; Strokach, A.; Giraldo-Forero, A. F.; Teyra, J.; Colak, R.; Kim, P. M. ELASPIC Web-Server: Proteome-Wide Structure-Based Prediction of Mutation Effects on Protein Stability and Binding Affinity. *Bioinformatics* **2016**, *32*, 1589−1591.
(25) Cao, H.; Wang, J.; He, L.; Qi, Y.; Zhang, J. Z. DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. *J. Chem. Inf. Model.* **2019**, *59*, 1508−1514.
(26) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. PoPMuSiC 2.1: A Web Server for the Estimation of Protein Stability Changes upon Mutation and Sequence Optimality. *BMC Bioinf.* **2011**, *12*, 151.
(27) Worth, C. L.; Preissner, R.; Blundell, T. L. SDM - A Server for Predicting Effects of Mutations on Protein Stability and Malfunction. *Nucleic Acids Res.* **2011**, *39*, W215−W222.
(28) Masso, M.; Vaisman, I. I. AUTO-MUTE 2.0: A Portable Framework with Enhanced Capabilities for Predicting Protein Functional Consequences upon Mutation. *Adv. Bioinf.* **2014**, *2014*, 278385.
(29) Berliner, N.; Teyra, J.; Çolak, R.; Lopez, S. G.; Kim, P. M. Combining Structural Modeling with Ensemble Machine Learning to Accurately Predict Protein Fold Stability and Binding Affinity Effects upon Mutation. *PLoS One* **2014**, *9*, No. e107353.
(30) Pucci, F.; Bourgeas, R.; Rooman, M. Predicting Protein Thermal Stability Changes upon Point Mutations Using Statistical Potentials: Introducing HoTMuSiC. *Sci. Rep.* **2016**, *6*, 23257.
(31) Musil, M.; Konegger, H.; Hon, J.; Bednar, D.; Damborsky, J. Computational Design of Stable and Soluble Biocatalysts. *ACS Catal.* **2019**, *9*, 1033−1054.
(32) Colón, W.; Church, J.; Sen, J.; Thibeault, J.; Trasatti, H.; Xia, K. Biological Roles of Protein Kinetic Stability. *Biochemistry* **2017**, *56*, 6179−6186.
(33) Hei, D. J.; Clark, D. S. Estimation of Melting Curves from Enzymatic Activity−Temperature Profiles. *Biotechnol. Bioeng.* **1993**, *42*, 1245−1251.
(34) 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*, No. 023104.
(35) Jia, L.; Yarlagadda, R.; Reed, C. C. Structure Based Thermostability Prediction Models for Protein Single Point Mutations with Machine Learning Tools. *PLoS One* **2015**, *10*, e0138022.
(36) 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.
(37) Hrabe, T.; Li, Z.; Sedova, M.; Rotkiewicz, P.; Jaroszewski, L.; Godzik, A. PDBFlex: Exploring Flexibility in Protein Structures. *Nucleic Acids Res.* **2016**, *44*, D423−D428.
(38) Kepp, K. P. Computing Stability Effects of Mutations in Human Superoxide Dismutase 1. *J. Phys. Chem. B* **2014**, *118*, 1799−1812.
(39) McGuinness, K. N.; Pan, W.; Sheridan, R. P.; Murphy, G.; Crespo, A. Role of Simple Descriptors and Applicability Domain in Predicting Change in Protein Thermostability. *PLoS One* **2018**, *13*, e0203819.
(40) Korkegian, A.; Black, M. E.; Baker, D.; Stoddard, B. L. Computational Thermostabilization of an Enzyme. *Science* **2005**, *308*, 857−860.
(41) Wakabayashi, H.; Griffiths, A. E.; Fay, P. J. Combining Mutations of Charged Residues at the A2 Domain Interface Enhances Factor VIII Stability over Single Point Mutations. *J. Thromb. Haemostasis* **2009**, *7*, 438−444.
(42) Li, G.; Fang, X.; Su, F.; Chen, Y.; Xu, L.; Yan, Y. Enhancing the Thermostability of Rhizomucor Miehei Lipase with a Limited Screening Library by Rational-Design Point Mutations and Disulfide Bonds. *Appl. Environ. Microbiol.* **2018**, *84*, No. e02129.


https://dx.doi.org/10.1021/acsomega.9b04105
*ACS Omega* 2020, 5, 6487−6493