



Computational tools help improve protein stability but with a solubility tradeoff

Received for publication, March 3, 2017, and in revised form, July 11, 2017. Published, Papers in Press, July 14, 2017, DOI 10.1074/jbc.M117.784165

Aron Broom, Zachary Jacobi, Kyle Trainor, and Elizabeth M. Meiering¹

From the Department of Chemistry, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Edited by Wolfgang Peti

Accurately predicting changes in protein stability upon amino acid substitution is a much sought after goal. Destabilizing mutations are often implicated in disease, whereas stabilizing mutations are of great value for industrial and therapeutic biotechnology. Increasing protein stability is an especially challenging task, with random substitution yielding stabilizing mutations in only ~2% of cases. To overcome this bottleneck, computational tools that aim to predict the effect of mutations have been developed; however, achieving accuracy and consistency remains challenging. Here, we combined 11 freely available tools into a meta-predictor (meieringlab.uwaterloo.ca/stabilitypredict/). Validation against ~600 experimental mutations indicated that our meta-predictor has improved performance over any of the individual tools. The meta-predictor was then used to recommend 10 mutations in a previously designed protein of moderate thermodynamic stability, ThreeFoil. Experimental characterization showed that four mutations increased protein stability and could be amplified through ThreeFoil's structural symmetry to yield several multiple mutants with >2-kcal/mol stabilization. By avoiding residues within functional ties, we could maintain ThreeFoil's glycan-binding capacity. Despite successfully achieving substantial stabilization, however, almost all mutations decreased protein solubility, the most common cause of protein design failure. Examination of the 600-mutation data set revealed that stabilizing mutations on the protein surface tend to increase hydrophobicity and that the individual tools favor this approach to gain stability. Thus, whereas currently available tools can increase protein stability and combining them into a meta-predictor yields enhanced reliability, improvements to the potentials/force fields underlying these tools are needed to avoid gaining protein stability at the cost of solubility.

Most natural proteins have modest thermodynamic stability, limiting their development as effective biocatalysts, biosensors, and therapeutics (1). Generally, increasing stability tends to increase protein production yields (2, 3), lengthen shelf-life (4–6), and improve survival under challenging solution conditions (pH, proteases, cosolvents, etc.). Increased stability also

allows the use of elevated reaction temperatures, improving reaction rates and reducing unwanted bacterial growth in industrial reactions (7). Increased stability may also decrease local and/or global unfolding rates, thereby reducing the formation of aggregates and, consequently, immunogenicity (4–6, 8). Despite its great significance, the reliable engineering of protein stability via sequence changes is still a challenging and often laborious task (9). Various experimental methods, including directed evolution (10) and deep sequencing (11, 12), are commonly used to improve protein stability in practice, but these are time-consuming and costly (13).

To allow simple and cost-effective protein engineering, numerous computational tools have been developed to predict the effect of mutations on stability. These tools are readily employed by both experts and non-experts due to simple inputs (a PDB² structure) and output (a predicted change in stability). Some tools, like EGAD (14), rely on physical force fields that seek to recapitulate the forces felt by a solvated protein. Others, such as PoPMuSiC (15), use statistical potentials, based on the probability of particular amino acids or atoms being in contact in known structures. Combining some elements of each of the above with empirical terms (such as an explicit term for hydrogen bonding) is often referred to as an empirical potential, employed by well-established tools, such as FoldX (16) and Rosetta (17). Still others, like IMutant3 (18), utilize a set of specific features, such as change in hydrophobicity or size, in conjunction with machine learning techniques. To date, however, only a few of these tools (FoldX (16), Rosetta-ddG (17), and PoPMuSiC (15)) have been employed to improve protein stability experimentally through the use of point mutations (19–31), whereas many have only been tested on existing mutation data sets. Entirely sequence-based approaches, such as consensus design and ancestral reconstruction, have also been applied to the protein engineering problem. Whereas these methods often generate sequences with high stability, functional specificity may be sacrificed (32). Making specific consensus mutations, rather than changing the entire sequence, has also proven to be effective (9). However, the size and quality of the multiple sequence alignments needed for this method can vary considerably from one target to another, and too much or too little diversity can be detrimental (33). Thus, choosing a particular tool when undertaking protein engineering to enhance protein stability can be fraught.

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant (RGPIN-102173 to E. M. M.). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Tables S1–S4 and Figs. S1–S4.

¹ To whom correspondence should be addressed. Tel.: 519-888-4567; E-mail: meiering@uwaterloo.ca.

² The abbreviations used are: PDB, Protein Data Bank; MCC, Matthews correlation coefficient; GuSCN, guanidine thiocyanate.

Tradeoff between protein stability and solubility

To decrease reliance on any one tool and increase the chance of success when using computational methods, which remains notably low (9, 34, 35), we combined diverse tools into a single meta-predictor. Meta-predictors have performed well in related predictions, such as covalent protein modification (36), protein–ligand binding (37), protein–protein interfaces (38), disordered regions of proteins (39, 40), and protein aggregation (41), where the combination of tools with differing and perhaps complementary strengths and weaknesses may allow retention of strengths while ameliorating weaknesses. We show that this stability meta-predictor performs better than any single tool when tested against a large set of experimental mutations from the Protherm database (42). In addition, we tested the meta-predictor by using it to predict mutations to stabilize a previously designed protein, ThreeFoil, which has moderate thermodynamic stability (43), as is often the case in initial design iterations and generally limits further development of protein function (44, 45). Experimental characterization of 10 of the top predicted stabilizing mutations resulted in four stabilizing mutations that combine to yield a substantial improvement in stability. However, the mutation predicted to be the most stabilizing was experimentally the most destabilizing. Moreover, the highly stabilized multiple mutants lost considerable solubility due to increases in surface hydrophobicity. Critically, analysis of the Protherm database and individual prediction tools shows that increased surface hydrophobicity is a common and significant caveat to enhancing stability. Thus, we find that, although increasing accuracy and accounting for solubility warrant further attention, meta-prediction offers an effective, inexpensive, and extensible way to model and improve protein stability, which may be valuable for a great diversity of applications.

Results

Meta-prediction of stability change

There are numerous freely available and easy-to-use tools for predicting the change in protein stability upon point mutation. These include FoldX (16), Rosetta-ddG (17), PoPMuSiC (15), EGAD (14), IMutant3 (18), CUPSAT (46), SDM (47), Hunter (48), MuPro (49), MultiMutate (50), and DFire (51). Examining the individual performance of each of these 11 stability prediction tools against a data set of 605 mutations (supplemental Table S1) from the Protherm database (42), which had not been used in their training or parameterization, revealed significant differences between tools, depending on the type of mutation (e.g. solvent-exposed *versus* buried) (Fig. 1, *a–e*).

Notably, there are many cases where specific tools are particularly well or poorly suited to certain types of mutations. For instance, tools using physical or empirical potentials (EGAD, FoldX, and Rosetta-ddG) appear to more accurately predict mutations that increase hydrophobicity than those that reduce it or leave it essentially unaltered (Fig. 1*a*, *red versus green* and *blue bars*). By contrast, many of the tools using statistical potentials (CUPSAT, SDM, PoPMuSiC, and MuPro) show the opposite trend (Fig. 1*a*). Interestingly, whereas all tools are better at predicting changes to buried residues compared with partially buried or exposed residues (Fig. 1*c*, *red versus green* and *blue bars*), some tools (EGAD, FoldX, Rosetta-ddG, DFire, and

PoPMuSiC) do fairly well overall, whereas others like MultiMutate and IMutant3 are not reliable for surface-exposed residues (Fig. 1*c*, *blue bars*).

To utilize all of these tools to maximum effect, a meta-predictor was constructed such that the weight given to the prediction from any particular tool was based on that tool's performance against similar types of mutations from the training set of 605 mutations (see "Experimental procedures"). Cross-validation (see "Experimental procedures") showed that the meta-predictor has improved performance as more individual tools are incorporated (Fig. 1*f*). The performance of the meta-predictor was superior to any individual tool based on numerous statistical measures, including correlation coefficients, precision, accuracy, and error (Fig. 2 and Table 1). Thus, combining any number of stability prediction tools weighted according to their individual performance is a useful and extensible strategy for improving performance and may be valuable for many applications, such as predicting stability changes for disease-causing mutations and identifying mutations to stabilize a protein of biotechnological importance. A web-server for the meta-predictor can be found at meieringlab.uwaterloo.ca/stabilitypredict/.³

Improving ThreeFoil stability

In an effort to improve the thermodynamic stability of ThreeFoil (which is ~ 6 kcal/mol) (43), the meta-predictor was used to predict the change in thermodynamic stability ($\Delta\Delta G$) for all point mutations (excluding cysteine) to each of ThreeFoil's 141 residues, excluding 21 residues probably involved in carbohydrate or sodium binding function (supplemental Fig. S1). Because ThreeFoil possesses 3-fold sequence and structural symmetry, the predicted $\Delta\Delta G$ values were averaged across symmetric positions. Thus, small structural variations between symmetric positions can provide an ensemble-like representation, improving prediction accuracy (52). The outcome was a total of 720 predictions, 83 (12%) of which were expected to be stabilizing. At some positions there were multiple predicted stabilizing mutations; these may be unidentified functional sites (53–55) or areas where the initial design was poor (56). To test diverse positions, only the most stabilizing mutation at each residue was considered. An exception was the selection of two mutations at Glu-66, a completely solvent-exposed position for which 10 of the 18 possible substitutions were predicted to stabilize and where poor agreement between Rosetta and consensus design was observed during the initial development of ThreeFoil (56). In total, 10 point mutations predicted to stabilize were tested experimentally (Fig. 3*a*). These mutations were made in ThreeFoil's second symmetric subdomain, and the effect of each mutation was determined using kinetic unfolding and folding measurements (Fig. 3*b*, Table 2, and supplemental Fig. S2).

The experimental effects of the mutations on stability were mixed (Fig. 3*c*). Four mutations (K53V, A62V, D85P, and D49N) increased the folding rate, stabilizing ThreeFoil by 0.8, 0.5, 0.2, and 0.2 kcal/mol, respectively (Table 2 and Fig. 3 (*green mutations*)). Three mutations (A68G, R90L, and D93P) had no

³ Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

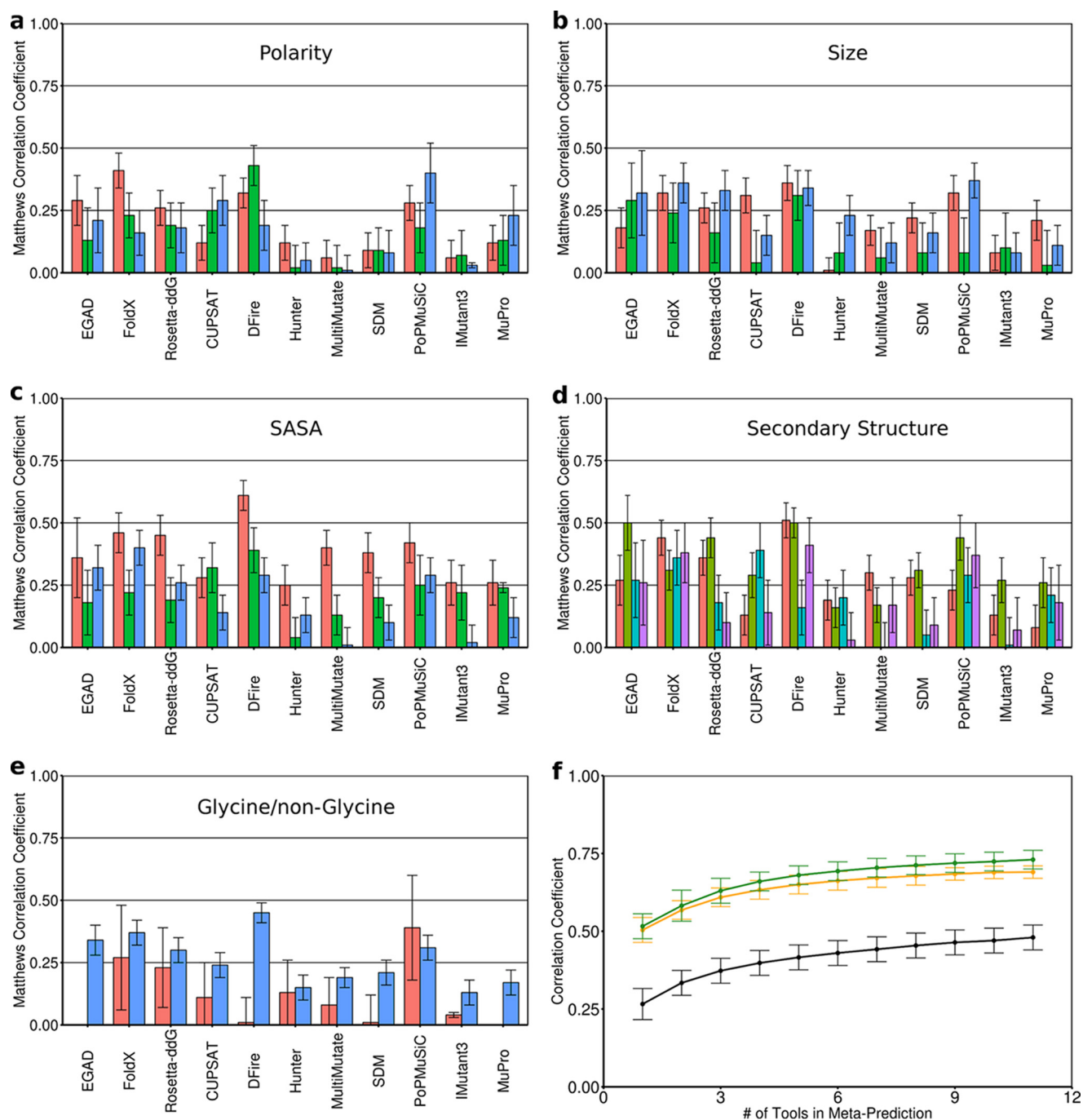


Figure 1. Building the meta-predictor. The performance of each tool as measured by MCC (96) is shown for different classifications of point mutations. *a*, polarity: decreased (red), similar (green), increased (blue). *b*, size: smaller (red), similar (green), larger (blue). *c*, solvent-accessible surface area of the WT residue: buried (red), partially exposed (green), exposed (blue). *d*, secondary structure of the backbone at the mutated position: helical (red), strand (green), turn (cyan), unstructured (purple). *e*, whether the mutation is to/from a glycine (red) or non-glycine (cyan). *f*, the more tools that were combined in the meta-predictor, the better the average performance, shown for MCC (black), ρ (orange), and R (green). Error bars (*a–e*), S.D. from 1000 analyses using 50% of the mutation data set. Individual predictors are grouped based on methodology: physical force field (EGAD), empirical potentials (FoldX and Rosetta), statistical potentials without machine learning (CUPSAT, DFire, Hunter, MultiMutate, and SDM), statistical potentials with machine learning (PoPMuSiC), and feature-based with machine learning (IMutant3 and MuPro).

significant effect on stability, although in the case of A68G and D93P, the predicted effect was also small (Table 2). The remaining three mutations (E66Y, E66L, and Q78I), which had been predicted to be the most stabilizing, resulted in faster unfolding and stability losses of 0.6, 0.9, and 2.7 kcal/mol, respectively. Consistent with mutations being, by design, outside ThreeFoil's carbohydrate binding sites, none of the stabilizing mutations

affected carbohydrate binding affinity (supplemental Fig. S3). Predictions of the individual tools differed considerably, and no single tool was reliable (supplemental Fig. S4). Overall, the 40% success rate of the meta-predictor compares favorably with an expected rate of ~2% for random mutagenesis (based on recent deep sequencing experiments; supplemental Table S2 (12, 57–59)).

Tradeoff between protein stability and solubility

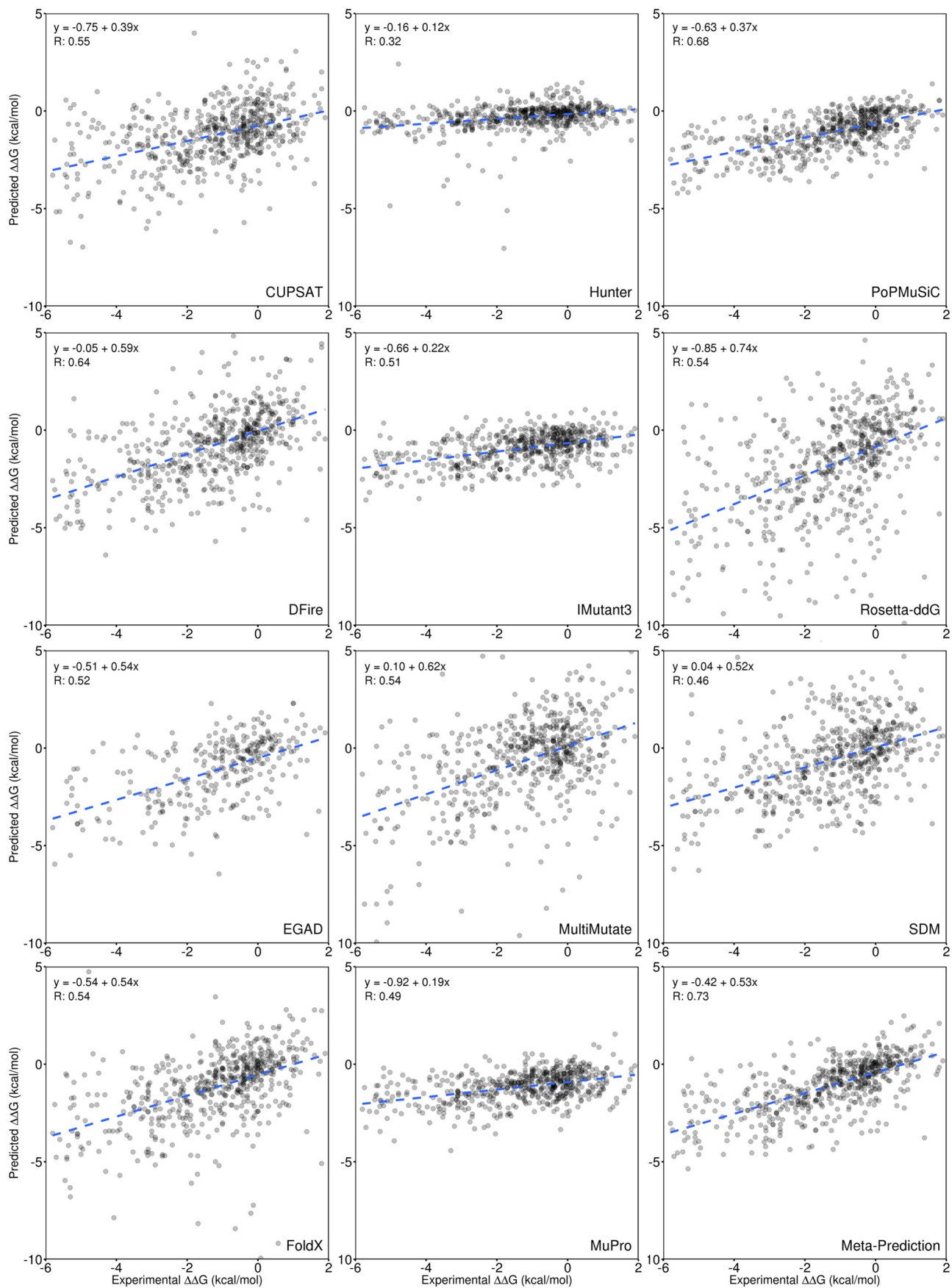


Table 1
Comparison of individual tools with the meta-predictor

Each tool was tested on a data set of 605 mutations covering protein structural classes of α , β , and mixed. Performance was measured using correlation coefficients (MCC and Pearson R) as well as precision (the percentage of mutations predicted to stabilize that do so experimentally), accuracy (the percentage of mutations correctly classified as stabilizing or destabilizing), and S.E. For MCC, precision, and accuracy, mutations predicted or experimentally determined to have an effect <0.2 kcal/mol were not included, thus eliminating a significant source of noise due to uncertainty in small $\Delta\Delta G$ values. Reported values are the mean from 1000 tests using a randomly selected 50% of the data set. For the meta-predictor, this allowed cross-validation by training the weights using 50% of the data set while testing on the remaining 50%. The best result for each metric (shown in boldface type) was obtained using the meta-predictor.

Tool	MCC	R	Precision	Accuracy	S.E.
			%	%	kcal/mol
EGAD	0.34	0.52	50	74	1.61
FoldX	0.38	0.54	52	78	1.78
Rosetta-ddG	0.32	0.54	46	75	2.34
CUPSAT	0.24	0.55	44	75	1.77
DFire	0.43	0.64	49	76	1.84
Hunter	0.16	0.32	34	68	1.89
MultiMutate	0.19	0.54	32	62	2.34
SDM	0.26	0.46	37	68	1.96
PoPMuSiC	0.33	0.68	59	79	1.32
IMutant3	0.14	0.51	41	75	1.52
MuPro	0.18	0.49	57	78	1.52
Meta-predictor	0.48	0.73	63	82	1.29

Taking advantage of symmetry

By using the most stabilizing mutations and not only combining them, but repeating them in all three symmetric sub-domains (Fig. 4), ThreeFoil's symmetry was leveraged to achieve substantial stabilization. Two multiple mutants were made, the first (MMut1) incorporating K53V, A62V, and D85P at all symmetric sites, and the second (MMut2) adding D49N at each symmetric site. Both multiple mutants stabilize the protein by >2 kcal/mol and increase the denaturation midpoint as a result of greatly increased refolding rates (Fig. 3 (*b* and *c*), *blue mutants*). For instance, whereas the wild-type has a refolding half-life in the absence of denaturant of several hours, the multiple mutants have refolding half-lives of 1 min or less under the same conditions. Although the stabilization and improvement in refolding rates represent a marked improvement, both multiple mutants are considerably less soluble (0.1–0.2 mg/ml) than the original protein (>20 mg/ml) (Table 3). Given the nature of the individual mutations (K6V/K53V/K100V and D38P/D85P/D132P replace charged and exposed side chains with hydrophobic ones, A15V/A62V/A109V increases the size of a hydrophobic side chain, and D2N/D49N/D96N replaces a charged and exposed side chain with a polar one), these results highlight a potential problem wherein selecting for the singular characteristic of improved stability has given the unintended outcome of lower solubility, probably due to increased surface hydrophobicity.

Stabilizing mutations tend to increase surface hydrophobicity

Whereas protein function often relies on having at least some stability (maintenance of the native form of the protein), it is also crucial to have enough solubility (protein available in solu-

tion) to perform that function adequately. Examining the source of reduced solubility of the multiple mutants, and minor reduction in solubility of the single mutants, reveals critical problems with the potentials/force fields underlying current protein stabilization tools.

A diverse set of factors that may contribute to reduced protein solubility have been recognized, from specific structural features like exposed hydrophobic patches (60) and regions with high propensity to form amyloid-like structures (61) to overall properties, such as net charge (62) and amino acid content (63, 64). Many solubility prediction tools exist that take these factors into account. Applying six tools available online showed they successfully predicted the greatly reduced solubility of the multiple mutants but could not discriminate well the differences in solubility between the single mutants (Table 3). Notably, however, simply accounting for the change in hydrophobicity of the mutation (measured by change in solvation free energy of the side chains) (65) gave an equally good prediction of the multiple mutants' poor behavior, as measured by both linear and rank correlation coefficients (Table 3). Indeed, 9 of the 10 mutations are to more hydrophobic residues, and this apparent preference for hydrophobic mutations, even on the protein surface, suggested a potential general problem with computational prediction.

Whereas a few experimental studies have suggested that increasing hydrophobicity at the protein surface improves stability (66–70), and the favoring of increased hydrophobicity has been reported for a few computational tools (14, 71, 72), our analysis indicates that this is, in fact, a widespread phenomenon. Examining the Protherm data set shows that stabilizing mutations on the protein surface are often observed when making substitutions to more hydrophobic residues, whereas destabilizing mutations tend to be substitutions to more hydrophilic residues (Fig. 5, *Protherm*). Critically, we find that 10 of the 11 individual tools have a bias toward recommending hydrophobic mutations on the protein surface as stabilizing. Overall, this results in predicted stabilizing mutations increasing solvation free energy by 0.84 kcal/mol on average (Fig. 5), similar to mutating alanine to valine (65). Although the tools may be recapitulating a real tendency for hydrophobic surface mutations to stabilize, the fact that protein design attempts most frequently fail because of poor solubility (73) makes this a critical concern for protein engineering and design.

Prediction tools underestimate the importance of buried polar groups

In addition to favoring hydrophobic mutations on the protein surface, the individual tools and meta-predictor may underestimate the importance of buried polar groups. This is suggested by the failure of the Q78I mutation, predicted to stabilize by seven of the individual tools (CUPSAT, DFire, FoldX, MultiMutate, PoPMuSiC, Rosetta-ddG, and SDM, with the remaining predicting essentially no change) (supplemental Fig.

Figure 2. Comparison of individual tools with the meta-predictor. The predicted $\Delta\Delta G$ for a point mutation is compared with the experimentally determined value for 605 mutations from the Protherm database (42). This comparison is performed for each of the individual prediction tools and the meta-predictor (*bottom right*). The equations for the linear fit are given as well as the Pearson (R) correlation coefficient based on the mean of 1000 samples of 50% of the data, as shown in Table 1.

Tradeoff between protein stability and solubility

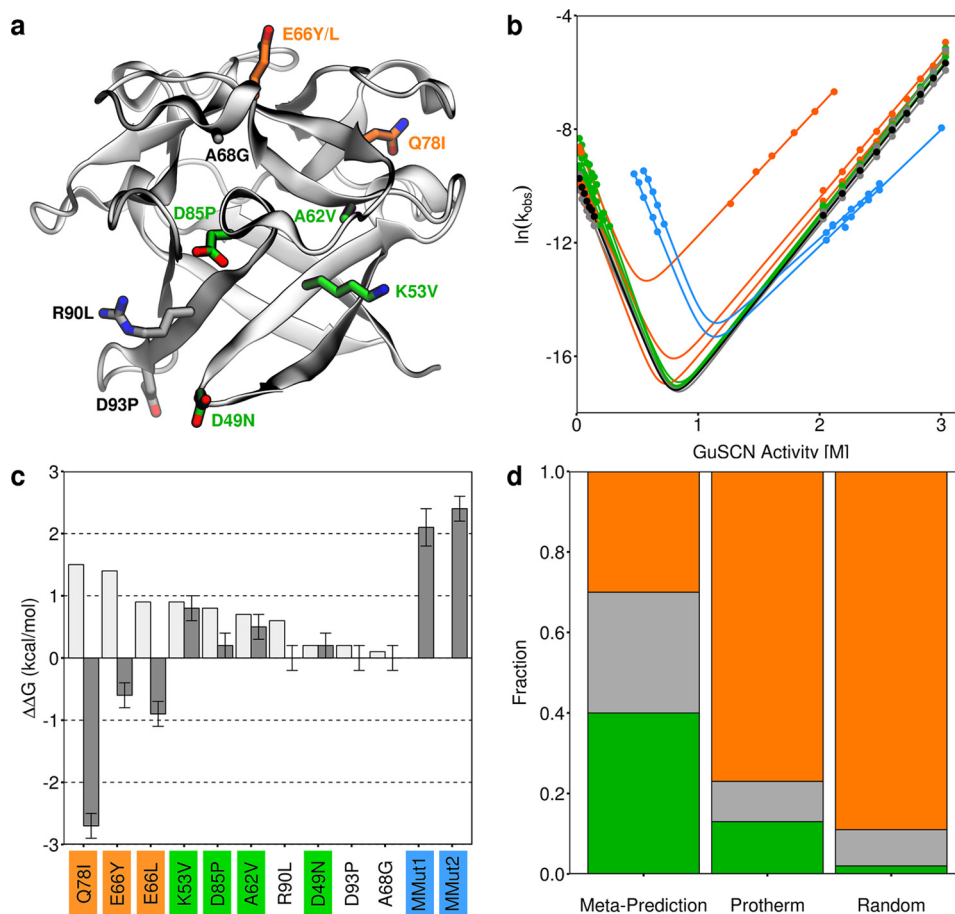


Figure 3. Experimental characterization of mutations to ThreeFoil. *a*, ThreeFoil's structure with sites of mutation shown as sticks. Stabilizing mutations are shown in green, destabilizing mutations in orange, and neutral in gray. *b*, folding and unfolding kinetics of ThreeFoil and mutants. The same colors are used as in *a*, with wild type shown in black and the highly stabilized multiple mutants in blue. Solid lines, fits of the data to a two-state unfolding model using Equation 3. *c*, predicted (light gray) and experimentally determined (dark gray) change in thermodynamic stability for ThreeFoil mutants, $\Delta\Delta G_{U-F}$. Positive values indicate increased stability, calculated using folding and unfolding rate constants in the absence of denaturant (see "Experimental procedures" and Table 2). *d*, fractions of mutations that are stabilizing (green, $\Delta\Delta G_{U-F} \geq 0.2$ kcal/mol), neutral (gray, 0.2 kcal/mol $> \Delta\Delta G_{U-F} > -0.2$ kcal/mol), or destabilizing (orange, $\Delta\Delta G_{U-F} \leq -0.2$ kcal/mol) determined for ThreeFoil mutants designed using the meta-predictor, the Protherm database, and random mutagenesis (from deep sequencing studies (12, 57–59); see supplemental Table S2). Error bars, uncertainty from linear fit using Origin 5 software.

Table 2

Kinetic parameters of ThreeFoil mutants

Equations for determining kinetic values are given under "Experimental procedures." Single mutants are listed in order of experimental stability changes, with mutations confirmed to stabilize experimentally ($\Delta\Delta G_{U-F} \geq 0.2$ kcal/mol) listed in boldface type. MMut1, K6V/K53V/K100V, A15V/A62V/A109V, D38P/D85P/D132P; and MMut2, MMut1 plus D2/49/96N. Error estimates from the data fit used Origin version 5.

Mutant	$\ln(k_f)$	$\ln(k_u)$	m_f	m_u	C_{mid}	$\Delta\Delta G_{U-F}$
			$kcal\ mol^{-1}\ M^{-1}$	$kcal\ mol^{-1}\ M^{-1}$	M	$kcal/mol$
WT	-9.57 ± 0.06	-22.0 ± 0.2	-6.22 ± 0.33	3.20 ± 0.05	0.79 ± 0.02	0
K53V	-8.07 ± 0.05	-21.9 ± 0.3	-7.10 ± 0.24	3.24 ± 0.08	0.80 ± 0.03	0.8 ± 0.2
A62V	-8.66 ± 0.05	-22.0 ± 0.3	-6.78 ± 0.28	3.25 ± 0.07	0.79 ± 0.03	0.5 ± 0.2
D85P	-9.14 ± 0.06	-21.9 ± 0.3	-5.97 ± 0.33	3.19 ± 0.07	0.83 ± 0.03	0.2 ± 0.2
D49N	-9.07 ± 0.03	-21.9 ± 0.1	-6.37 ± 0.17	3.23 ± 0.03	0.80 ± 0.02	0.2 ± 0.2
A68G	-9.59 ± 0.05	-22.1 ± 0.2	-6.12 ± 0.29	3.28 ± 0.04	0.79 ± 0.02	0.0 ± 0.2
R90L	-9.99 ± 0.05	-22.4 ± 0.2	-5.78 ± 0.29	3.34 ± 0.04	0.81 ± 0.02	0.0 ± 0.2
D93P	-9.45 ± 0.06	-21.9 ± 0.2	-6.26 ± 0.36	3.13 ± 0.05	0.79 ± 0.03	0.0 ± 0.2
E66Y	-9.73 ± 0.07	-21.2 ± 0.3	-6.76 ± 0.38	3.09 ± 0.06	0.69 ± 0.03	-0.6 ± 0.2
E66L	-9.58 ± 0.06	-20.6 ± 0.2	-5.60 ± 0.31	3.02 ± 0.05	0.76 ± 0.02	-0.9 ± 0.2
Q78I ^a	-8.40 ± 0.13	-16.3 ± 0.1	-6.22 ± 0.33	2.70 ± 0.03	0.52 ± 0.03	-2.7 ± 0.2
MMut1	-4.50 ± 0.3	-20.4 ± 0.3	-6.34 ± 0.31	2.46 ± 0.07	1.08 ± 0.05	2.1 ± 0.3
MMut2	-3.15 ± 0.17	-19.6 ± 0.2	-6.75 ± 0.16	2.32 ± 0.06	1.08 ± 0.03	2.4 ± 0.2

^a Only a single point was determined on the refolding branch for Q78I due to aggregation at intermediate denaturant concentrations. The m_f from WT was used to estimate $\ln(k_f)$, and values affected by this approximation are given in italic type.

S4) and estimated by the meta-predictor to give ~ 1.5 kcal/mol of stability, yet resulting in much faster unfolding than WT, and a loss of ~ 3 kcal/mol of stability (Fig. 3, *b* and *c*). A single buried hydrogen bond has been estimated to contribute ~ 1 kcal/mol

to protein stability (74). Because Gln-78 makes three buried hydrogen bonds (one side chain–side chain and two side chain–backbone) (Fig. 6), the loss of ~ 3 kcal/mol of stability may be accounted for by the loss of these specific interactions.

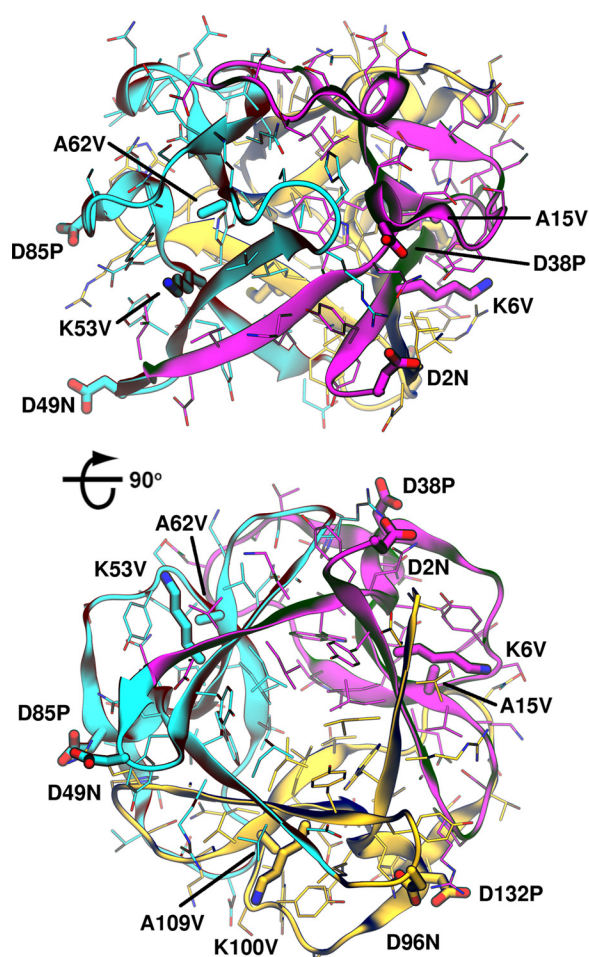


Figure 4. Taking advantage of symmetry to improve stability. ThreeFoil structure (PDB entry 3PG0) is shown from a *side view*, perpendicular to the 3-fold axis of symmetry (*top*) and rotated 90° to look along the axis of symmetry toward the capping β -hairpin triplets (*bottom*). The first, second, and third subdomains are shown in *magenta*, *cyan*, and *yellow*, respectively. Sites of mutations used for MMut1 (K6V/K53V/K100V, A15V/A62V/A109V, and D38P/D85P/D132P) and MMut2 (same as MMut1 plus D2N/D49N/D96N) are shown as *thick sticks*.

Whereas buried hydrophobic groups (e.g. $-\text{CH}_2-$) may contribute a similar ~ 1 kcal/mol to stability (74), isoleucine has only 1 additional $-\text{CH}_2-$ group compared with glutamine, suggesting that the buried hydrogen bonds were either undervalued or unrecognized by most of the prediction tools. In the case of physical or empirical force fields, this could arise from solvation terms or parameterization that penalizes the burial of polar or charged residues (75), whereas statistical potentials may inherently undervalue certain interactions (76–78). Here, the mutation of a buried polar sidechain to a hydrophobic one illustrates the dramatic failure that results when computational methods either do not recognize the presence of buried hydrogen bonds or inaccurately predict their contribution to protein stability.

Discussion

Stabilizing mutations give multiple desirable attributes to proteins: enhanced capacity for directed evolution and design of novel functions (44, 45), increased resistance to high temperature and harsh solution conditions (7, 9), and reduced immunogenicity caused by local or global unfolding and subsequent

aggregation (4–6, 8). The meta-prediction approach used here, which incorporated 11 diverse computational tools for predicting $\Delta\Delta G$ upon mutation, was better than any single tool at predicting the effects of more than 600 experimentally characterized point mutations. The meta-predictor also successfully identified stabilizing mutations to a previously designed protein, ThreeFoil. Most mutations to ThreeFoil resulted in stability changes close to the predicted values (Fig. 3), with a success rate of 40%, comparing favorably with other studies (23, 24, 26, 27, 30). Moreover, the stabilizing single mutations could be combined to yield highly stabilized multiple mutants. Because the meta-predictor weighs the contribution of each tool according to its performance on different mutation types (increased size, decreased polarity, etc.), it is readily extensible, allowing for future improvements using complementary or improved individual tools.

The experimental results are also notable because functional residues, which are often sources of instability and easy targets for improvement, were excluded (53–55). Indeed, binding function was maintained upon mutation. The amplification of stability (>2 kcal/mol) gained by leveraging the sequence and structural symmetry of ThreeFoil (56), demonstrates the tractability of engineering using symmetric (79) or repeat protein scaffolds (80, 81), both of which are abundant in nature (82).

Although the stabilization offered by the multiple mutants is substantial, there was a considerable loss in solubility, probably resulting from increases in surface hydrophobicity. All but 1 of the 10 tested mutations either introduced a hydrophobic amino acid or changed a charged residue to a polar one, and the majority of these changes occurred on the protein surface. Although it may seem counterintuitive that hydrophobic mutations at solvent-exposed positions would stabilize, it may be that even on the protein surface, hydrophobic side chains can bury relatively more surface area when in the native rather than the denatured state (67). Here, an analysis of more than 200 experimentally characterized point mutations on protein surfaces reveals that when computational tools attempt to improve protein stability, they often rely on the stabilizing effect of hydrophobic mutations at the cost of solubility. Even the tools that are the most widely used and/or perform the best (high correlation coefficients and/or low error) fall into this trap, such as Rosetta, FoldX, DFire, and PoPMuSiC. Although protein solubility may be generally improved by introducing a high density of similarly charged residues on the protein surface, this leads to decreases in stability (62). However, specifically designing optimized networks of surface electrostatics using a small number of mutations has been shown to improve both stability and foldability, illustrating that stability and solubility tradeoffs are not mandatory (83).

The tendency of the prediction tools to favor hydrophobic mutations is not limited to the protein surface. Mutation of a buried glutamine (Gln-78) to isoleucine was predicted to be the most stabilizing, with the majority of individual tools predicting considerable stabilization and the remainder predicting a nearly neutral effect. Mutations to valine, leucine, and methionine at the same position were also predicted to stabilize. Surprisingly, the Q78I mutation was the most destabilizing mutation tested, dramatically accelerating unfolding. The results for

Tradeoff between protein stability and solubility

Table 3

Experimental and predicted solubility of ThreeFoil variants

Values for each solubility prediction tool were normalized to be 1.0 for WT. In cases where the tool's output is a lower value when solubility is decreased, the score is simply normalized as $\text{Mut}_{\text{score}}/\text{WT}_{\text{score}}$. In cases where the tool's output is a higher value when solubility is decreased, the normalized score is $(1/\text{Mut}_{\text{score}})/(1/\text{WT}_{\text{score}})$ (alternatively, $\text{WT}_{\text{score}}/\text{Mut}_{\text{score}}$). Thus, all predictions of reduced solubility relative to WT have scores < 1.0. The following output values were used from each tool: Aggrescan3D (99), total score; CamSol (100), intrinsic solubility; PASTA (101), number of amyloid-forming regions; TANGO (102), Agg parameter; ZipperDB (103), number of amyloid-prone regions; Zyggregator (104), intrinsic aggregation propensity. Hydrophobicity is defined by the change in side chain solvation free energy, as determined by Wimley *et al.* (65).

Mutant	Solubility mg/ml	Aggrescan3D	CamSol	PASTA	TANGO	ZipperDB	Zyggregator	Hydrophobicity
WT	23.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00
K53V	22.5	0.97	0.78	0.98	0.01	0.94	1.08	0.95
A62V	21.1	1.00	0.98	1.00	1.00	0.88	0.99	0.98
D85P	16.9	0.96	0.99	1.00	0.99	1.00	0.92	0.94
D49N	25.0	0.98	0.99	1.00	1.00	0.94	0.91	0.95
A68G	21.9	1.00	1.01	1.00	1.00	1.00	1.00	1.01
R90L	16.8	0.92	0.87	1.00	0.42	1.00	1.08	0.95
D93P	21.2	0.93	0.98	1.00	1.00	1.00	0.92	0.94
E66Y	19.9	0.95	0.98	1.00	1.00	1.00	0.90	0.93
E66L	17.0	0.96	0.99	1.00	1.00	1.00	0.90	0.92
Q78I	1.4	1.00	0.96	0.91	0.02	1.00	0.99	0.97
MMut1	0.2	0.79	0.28	0.86	0.00	0.63	1.01	0.72
MMut2	0.1	0.76	0.15	0.86	0.01	0.58	0.73	0.65
R		0.71	0.75	0.93	0.75	0.69	0.30	0.75
ρ		0.60	0.63	0.58	0.64	0.20	0.15	0.59

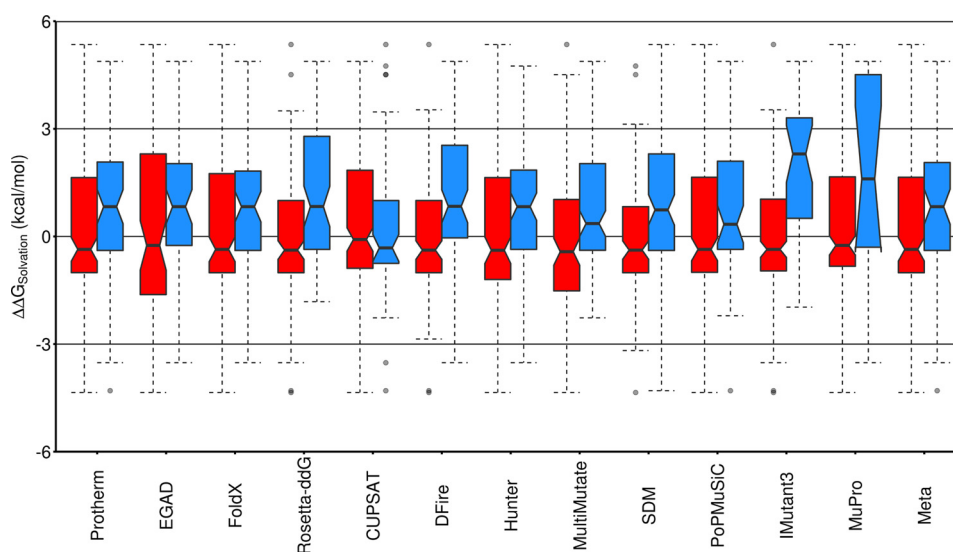


Figure 5. Stabilizing mutations on the protein surface favor increasing hydrophobicity. The changes in side chain solvation free energy, $\Delta\Delta G_{\text{solvation}}$ (kcal/mol) (65) for all 229 solvent-exposed mutations in the Protherm-derived data set (positive $\Delta\Delta G_{\text{solvation}}$ indicates a mutation to a more hydrophobic residue with reduced solubility in water (65)) are shown as box plots. *Red boxes*, mutations predicted to destabilize ($\Delta\Delta G_{U-F} \leq -0.2$ kcal/mol); *blue boxes*, mutations predicted to stabilize ($\Delta\Delta G_{U-F} \geq 0.2$ kcal/mol). For the Protherm-derived data set (*far left*), *red* indicates mutations experimentally determined to destabilize, and *blue* indicates mutations experimentally determined to stabilize relative to the wild-type protein. The *notched region* around the median (*horizontal line*) of each box represents the 95% confidence interval for the median, indicating that for Protherm and the majority of the predictors, mutations on the protein surface that increase stability tend to be to more hydrophobic amino acids, whereas mutations that decrease stability tend to be to more hydrophilic amino acids (CUPSAT is the singular exception). The *colored region* of each box includes the middle 50% of the data, with *dashed whiskers* showing 1.5 times the interquartile range above and below the middle 50% (the interquartile range is the difference between the *top* and *bottom* of the *colored region*, or the difference between the 25th and 75th percentiles). Outliers (mutations *outside* the *whiskers*) are shown as *semitransparent single points*.

Gln-78 point to an imbalance in force field terms (75) or systematic errors and biases in statistical potentials (76–78); properly recognizing the key role of buried polar side-chains that can hydrogen-bond (84–86) is key to future improvements in prediction reliability.

Accurate and reliable prediction of point mutations is critical to protein design and engineering. Because destabilizing mutations, particularly buried ones, often have larger effects than stabilizing ones (87), even a single incorrectly predicted mutation, like Q78I here, can jeopardize the stability of the entire protein and counteract a larger number of stabilizing mutations. Thus, getting the details right when predicting

changes in stability will allow confident engineering of multiple mutants without testing individual mutations and result in reliable success during *de novo* design (9, 34, 35). For design targets with homologous sequences, incorporating information from multiple-sequence alignments may improve reliability (31, 79, 88–90). However, the availability of such sequences, and thus the benefit of their use, will differ from one target to another. Furthermore, choosing the appropriate level of sequence diversity can be challenging (33). Predictions may also be improved by accounting for backbone flexibility, either by building this directly into the tool itself (17) or through multistate design, which could

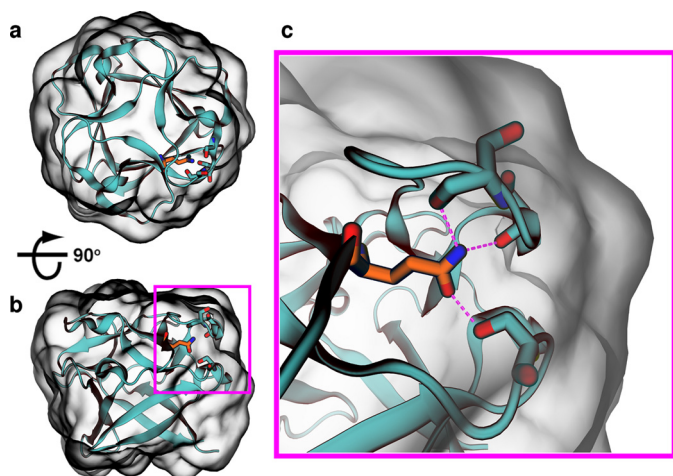


Figure 6. Thermodynamic contributions of buried hydrogen bonds are poorly predicted. In ThreeFoil, the wild-type glutamine at position 78 is completely buried from solvent yet forms three hydrogen bonds. *a* and *b*, views of the entire protein are shown with a 90° rotation to illustrate the buried nature and position of Gln-78. *c*, enlarged view illustrating hydrogen bonds. Two hydrogen bonds are with backbone oxygens, and one is with the oxygen of a serine side chain. Bonds to the glutamine's nitrogen involve the backbone oxygens from serines 22 and 24, whereas the single bond to the glutamine's oxygen involves the side chain oxygen of serine 59. Hydrogen bonds are shown as magenta dashed lines; Gln-78 is shown as orange sticks; and Ser-22, Ser-24, and Ser-59 are shown as cyan sticks.

improve the accuracy of existing predictors, although at the cost of additional computational resources (91). Further improvements in the reliability of meta-prediction approaches may be gained by incorporating these additional tools and methods.

In summary, the meta-prediction approach presented here combines individual predictors to improve performance, as demonstrated in testing against a large data set of experimentally characterized mutations and by finding new stabilizing mutations to a previously designed protein while retaining its function. This approach, wherein the weight given to any one tool's prediction is proportional to its performance on similar types of mutations, is extensible to future tools and offers a simple method to gain additional reliability when improving protein stability. Critically, however, the protein engineering tools used here have a marked tendency to improve protein stability by mutating residues on a protein's surface to more hydrophobic side chains, which commonly results in a loss of solubility. This loss of solubility is a key issue that must be resolved for protein engineering to be employed widely and reliably, with a transformative impact on biotechnology.

Experimental procedures

Stability prediction by individual tools

The following individual tools were used for $\Delta\Delta G$ predictions: CUPSAT (web server: cupsat.tu-bs.de/)³ (46), DFIRE2 (stand-alone executable, version 1.1; wild-type and mutant structures were generated using SCWRL4 (92)) (51), EGAD (stand-alone library) (14), FoldX (stand-alone executable, version 3.0) (16), Hunter (stand-alone executable) (48), IMutant3 (stand-alone executable, version 3.0.7) (18), MultiMutate (stand-alone executable) (50), MuPro (stand-alone executable, version 1.1) (49), PoPMuSiC (web server, version 2.0; currently

only version 2.1 is available) (15), Rosetta-ddG (stand-alone executable (ddg_monomer) using protocol 16, which incorporates backbone flexibility, Rosetta version 3.5) (17), and SDM (web server: mordred.bioc.cam.ac.uk/~sdm/sdm.php)³ (47). For predictions of point mutations to ThreeFoil, the PDB structure 3PG0 was used. For predictions of mutations in the Protherm database (42), the structure associated with the particular entry in the database was used and reduced to the residues indicated in the database (*i.e.* for multidomain proteins). For prediction tools requiring temperature or pH as part of the input, a temperature of 27 °C was used and a pH of 8.1 for consistency with experimental conditions used in testing ThreeFoil. Note that values of the change in free energy for the mutants is based on the convention of measuring the difference in free energy changes from folded to unfolded,

$$\Delta\Delta G_{U-F} = \Delta G_{U-F(\text{mut})} - \Delta G_{U-F(\text{WT})} \quad (\text{Eq. 1})$$

where $\Delta G_{U-F} = G_U - G_F$.

Thus, positive $\Delta\Delta G_{U-F}$ values represent increased stability (output from individual tools is converted to this convention).

Curation of the Protherm database

We searched the Protherm database (42) for unique point mutations with experimentally measured $\Delta\Delta G$ values where the temperature and pH of the experiment was between 20 and 30 °C and between 5 and 9, respectively. Additionally, most of the prediction tools cannot account for co-factors or prosthetic groups, so proteins with such groups were also removed. This left 1663 point mutations in total. Testing the prediction tools against mutations used for their training or parameterization may give misleading results. To avoid this, we removed all 1058 mutations used in the training or parameterization of any of the tools, leaving 605 mutations to 60 proteins, including all major structural classes (results of testing against the 1058 removed mutations are given as supplemental Table S3).

A known problem with the Protherm database is entry of $\Delta\Delta G$ values with the wrong sign (42). To correct such entries, we manually reviewed the publications reporting the 605 mutations. In addition to fixing numerous cases where the sign was entered incorrectly, we found several cases where the value reported in the publication was in kJ/mol but had been entered in kcal/mol. A table summarizing the 605 curated mutations is included as supplemental Table S1. As above, we use $\Delta\Delta G_{U-F}$ with positive values indicating increased stability.

Building the meta-predictor

For the 605-mutation data set (supplemental Table S1), predictions were made using each of the 11 tools. Mutations were then categorized based on five criteria: 1) change in polarity as measured by ΔG of solvation (less, the same, more); 2) change in size (smaller, the same, larger); 3) solvent-accessible surface area of the wild-type residue (buried, partially exposed, fully exposed); 4) secondary structure at the site of mutation (α , β , turn, unstructured); and 5) whether the mutation was to or from glycine (because performance on mutations involving glycine tended to differ most dramatically from other amino acids).

Tradeoff between protein stability and solubility

Measures of solvation free energy were taken from the work of Wimley *et al.* (65), with a difference of >0.1 kcal/mol ($\Delta\Delta G_{\text{solvation}}$) required to consider the mutation more or less polar. Measures of size were taken from the work of Darby and Creighton (93), and a difference of at least 19 \AA^3 (approximately the size of a CH_3 group) was considered larger or smaller. Exposure of the WT residues was calculated using VMD (www.ks.uiuc.edu/Research/vmd/)³ (94), by dividing the solvent-accessible surface area of the amino acid in the PDB structure by that of the fully exposed amino acid. Ranges for the ratios were as follows: buried, ≤ 0.05 ; partially buried, $0.05\text{--}0.20$; exposed, ≥ 0.20 . The secondary structure was calculated using DSSP (95).

Predictions from each tool were weighted by the Matthews correlation coefficient (96) (MCC) for each type of mutation (Fig. 1). The MCC values were determined through cross-validation, in which the data set of 605 mutations was split into halves, with one half used to determine MCC values as weights and the other half used to test overall performance (Fig. 1*f*); this was repeated 1000 times. Final reported performance values are an average of those 1000 tests (Fig. 1*f* and Table 1), and the scatter plot in Fig. 2 shows the average value for each mutation when it was not used to train the weights. Overall, the meta-predictor score was calculated as follows,

$$\Delta\Delta G_{\text{meta}} = \frac{\sum_i^{\text{tools}} \sum_j^{\text{types}} w_{i,j} \times \Delta\Delta G_{\text{tool}_i}}{\sum_i^{\text{tools}} \sum_j^{\text{types}} w_{i,j}} \quad (\text{Eq. 2})$$

where $w_{i,j}$ is the weight for the i th tool on the j th mutation type.

Weights for each tool and each type of mutation are given in [supplemental Table S4](#). In the case of the work presented here, all 11 tools were used, but as shown in Fig. 1, benefit can still be derived from combining a subset of those tools if not all are readily available. The web-server for the meta-predictor (meieringlab.uwaterloo.ca/stabilitypredict/) does not include the PoPMuSiC predictor, but it can be used individually from the authors' website (dezye.com).

Meta-prediction of ThreeFoil point mutations

The $\Delta\Delta G$ of point mutations to ThreeFoil were predicted with each of the 11 tools and then combined into a meta- $\Delta\Delta G$ as above (see Equation 1). To preserve function, we excluded 18 residues in the three symmetric carbohydrate-binding sites (Asp-17/64/111, Ile-30/77/124, Tyr-32/79/126, Ser-35/82/129, Asn-39/86/133, and Gln-40/87/134) and three residues (Asn-28/75/122) in the single sodium-binding site, 21 residues in total. This left 120 residues to test. ThreeFoil is well-behaved in solution, and to preserve this, we chose to avoid incorporation of cysteine, which could cause aberrant cross-linking and aggregation (ThreeFoil has no cysteine residues). Given the remaining set of 18 natural amino acids (not including the WT amino acid) and 120 positions, we tested a total of 2160 mutations with each tool (except for EGAD, which cannot predict mutations to or from glycine or proline (14)).

Because ThreeFoil has a 3-fold symmetric sequence and structure, we reasoned that the quality of the predictions could be improved by averaging the results across symmetric positions, because the mutation impact ought to be very similar at

each position. Experimental testing of point mutations was performed using the symmetric position in the second/middle sub-domain module.

Expression and purification of ThreeFoil mutants

ThreeFoil was expressed from a pET-28a plasmid (Novagen) in BL21 DE3 *Escherichia coli* cells, as described previously (43, 56). Expression was induced with 1 mM isopropyl 1-thio- β -D-galactopyranoside, and cells were harvested after 24 h at 37 °C. Inclusion bodies were solubilized in buffered urea (6 M urea, 100 mM sodium phosphate, 10 mM Tris, pH 8.1), bound to a nickel-nitrilotriacetic acid column, and eluted with the same buffer as above but at pH 4.5. The protein was then refolded by dialysis (1:10, 10 times) in lyophilization buffer in 50 mM ammonium acetate and lyophilized.

Kinetic analysis of ThreeFoil mutants

All measurements were performed at 27 °C. Folding and unfolding were monitored by fluorescence using a SpectraMax M5 plate reader (Molecular Devices) with excitation at 274 nm and emission at 317 nm. Fluorescence was measured from the bottom of 96-well UV Star[®] (Greiner Bio-One) black-well plates with clear, UV-transparent bottoms and the top sealed with HD Clear sealing tape (Hampton Research) to prevent evaporation. To folded or unfolded protein (10 mg/ml ThreeFoil ($\sim 550 \mu\text{M}$) in standard buffer (50 mM Tris, 150 mM NaCl, pH 8.1) with or without 4 M GuSCN), varying concentrations of GuSCN in standard buffer were added, such that the final volume in each well was 250 μl at a protein concentration of 3.3 μM . For mutants with much reduced solubility under folded conditions (Q78I, MMut1, and MMut2), the final concentration was kept at 3.3 μM , but the initial stock was of lower concentration; thus, the range of denaturant concentrations that could be explored was less.

The shortest experiments ran for 30 min, whereas the longest ran for 4 days, and the interval for taking plate readings (10 s to 30 min) was chosen such that each run would have ~ 200 measurements. Each kinetic trace was fit to a single exponential equation,

$$S = B + Ce^{-kt} \quad (\text{Eq. 3})$$

where S is the fluorescence signal, B is the offset, C is the amplitude, k is the rate constant, and t is the time in seconds. The chevron of observed rate constants, k_{obs} , as a function of denaturant activity, A , was fit to the equation for a two-state transition between folded (f) and unfolded (u) states of the protein (97),

$$\ln(k_{\text{obs}}) = \ln(e^{f+m_f A} + e^{u+m_u A}) \quad (\text{Eq. 4})$$

where m_f and m_u are the linear denaturant dependence of folding and unfolding, respectively, and $f = \ln(k_f^{\text{H}_2\text{O}})$ and $u = \ln(k_u^{\text{H}_2\text{O}})$ are the natural logarithms of the respective folding and unfolding rate constants in water (measured in s^{-1}). The denaturant activity, A , was calculated as by Cota and Clarke (98),

$$A = [\text{GuSCN}] \times (C_{0.5}/(C_{0.5} + [\text{GuSCN}])) \quad (\text{Eq. 5})$$

where $C_{0.5}$ is the [GuSCN] at half activity, equal to 6.47 M. The Gibbs free energy of unfolding, also referred to as the thermo-

dynamic stability of the protein, was calculated from the ratio of the folding and unfolding rate constants in the absence of denaturant,

$$\Delta G_{U-F} = -RT \ln(k_u^{H_2O}/k_f^{H_2O}) \quad (\text{Eq. 6})$$

where R is the universal gas constant, and T , the temperature, is 300 K.

Solubility measurements and predictions

A saturated protein solution was obtained by adding 250 μl of buffer (150 mM NaCl, 50 mM Tris, pH 8.1) to 10 mg of lyophilized protein. After incubating for 2 h at room temperature, the sample was centrifuged at $13,300 \times g$ for 10 min to pellet undissolved protein. The protein concentration of the supernatant was measured spectrophotometrically after a 1:20 dilution in buffer, using a molar absorption coefficient of $33,600 \mu\text{l mol}^{-1} \text{cm}^{-1}$ (56).

Prediction of mutant solubility was performed using the web servers for Aggrescan3D (99) (biocomp.chem.uw.edu.pl/A3D/),³ CamSol (100), PASTA (101) (protein.bio.unipd.it/pasta2/),³ TANGO (102) (tango.crg.es/),³ ZipperDB (103) (services.mbi.ucla.edu/zipperdb/intro),³ and Zyggregator (104). The hydrophobicity scale for side chains based on solvation free energy was taken from Table 2 of Wimley *et al.* (65). All values were normalized such that the WT score was 1 and predictions of reduced solubility would score lower than 1. For predictors that represent reduced solubility as lower scores, this was done by dividing the WT and mutant scores by the WT score. For predictors that represent reduced solubility as higher scores (*i.e.* higher aggregation propensity), this was done by dividing the inverse of each score by the inverse of the WT score.

Author contributions—A. B. and E. M. M. conceived of and coordinated the study and wrote the paper. A. B. and Z. J. designed, performed, and analyzed the computational predictions/modeling and wet laboratory experiments. A. B. and K. T. curated and analyzed the Protherm database. All authors reviewed the results and approved the final version of the manuscript.

References

1. Rockah-Shmuel, L., Tóth-Petróczy, Á., and Tawfik, D. S. (2015) Systematic mapping of protein mutational space by prolonged drift reveals the deleterious effects of seemingly neutral mutations. *PLoS Comput. Biol.* **11**, e1004421
2. Kwon, W. S., Da Silva, N. A., and Kellis, J. T. (1996) Relationship between thermal stability, degradation rate and expression yield of barnase variants in the periplasm of *Escherichia coli*. *Protein Eng.* **9**, 1197–1202
3. McLendon, G., and Radany, E. (1978) Is protein turnover thermodynamically controlled? *J. Biol. Chem.* **253**, 6335–6337
4. Sauerborn, M., Brinks, V., Jiskoot, W., and Schellekens, H. (2010) Immunological mechanism underlying the immune response to recombinant human protein therapeutics. *Trends Pharmacol. Sci.* **31**, 53–59
5. Manning, M. C., Chou, D. K., Murphy, B. M., Payne, R. W., and Katayama, D. S. (2010) Stability of protein pharmaceuticals: an update. *Pharm. Res.* **27**, 544–575
6. Boulet-Audet, M., Byrne, B., and Kazarian, S. G. (2014) High-throughput thermal stability analysis of a monoclonal antibody by attenuated total reflection ft-ir spectroscopic imaging. *Anal. Chem.* **86**, 9786–9793
7. Bommarius, A. S., and Paye, M. F. (2013) Stabilizing biocatalysts. *Chem. Soc. Rev.* **42**, 6534–6565
8. Redington, J. M., Breydo, L., and Uversky, V. N. (2017) When good goes awry: the aggregation of protein therapeutics. *Protein Pept. Lett.* **24**, 340–347
9. Magliery, T. J. (2015) Protein stability: computation, sequence statistics, and new experimental methods. *Curr. Opin. Struct. Biol.* **33**, 161–168
10. Romero, P. A., and Arnold, F. H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876
11. Tripathi, A., and Varadarajan, R. (2014) Residue specific contributions to stability and activity inferred from saturation mutagenesis and deep sequencing. *Curr. Opin. Struct. Biol.* **24**, 63–71
12. Foit, L., Morgan, G. J., Kern, M. J., Steimer, L. R., von Hacht, A. A., Titchmarsh, J., Warriner, S. L., Radford, S. E., and Bardwell, J. C. A. (2009) Optimizing protein stability *in vivo*. *Mol. Cell* **36**, 861–871
13. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., and Robins, K. (2012) Engineering the third wave of biocatalysis. *Nature* **485**, 185–194
14. Pokala, N., and Handel, T. M. (2005) Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J. Mol. Biol.* **347**, 203–227
15. Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., and Rooman, M. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: Popmusic-2.0. *Bioinformatics* **25**, 2537–2543
16. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005) The foldx web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388
17. Kellogg, E. H., Leaver-Fay, A., and Baker, D. (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830–838
18. Capriotti, E., Fariselli, P., Rossi, I., and Casadio, R. (2008) A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics* **9**, S6
19. Gilis, D., McLennan, H. R., Dehouck, Y., Cabrita, L. D., Rooman, M., and Bottomley, S. P. (2003) *In vitro* and *in silico* design of α 1-antitrypsin mutants with different conformational stabilities. *J. Mol. Biol.* **325**, 581–589
20. Cabrita, L. D., Gilis, D., Robertson, A. L., Dehouck, Y., Rooman, M., and Bottomley, S. P. (2007) Enhancing the stability and solubility of TEV protease using *in silico* design. *Protein Sci.* **16**, 2360–2367
21. Yang, D.-F., Wei, Y.-T., and Huang, R.-B. (2007) Computer-aided design of the stability of pyruvate formate-lyase from *Escherichia coli* by site-directed mutagenesis. *Biosci. Biotechnol. Biochem.* **71**, 746–753
22. Zhang, S.-B., and Wu, Z.-L. (2011) Identification of amino acid residues responsible for increased thermostability of feruloyl esterase A from *Aspergillus niger* using the popmusic algorithm. *Bioresour. Technol.* **102**, 2093–2096
23. Komor, R. S., Romero, P. A., Xie, C. B., and Arnold, F. H. (2012) Highly thermostable fungal cellobiohydrolase i (cel7a) engineered using predictive methods. *Protein Eng. Des. Sel.* **25**, 827–833
24. Silva, I. R., Larsen, D. M., Jers, C., Derkx, P., Meyer, A. S., and Mikkelsen, J. D. (2013) Enhancing RGI lyase thermostability by targeted single point mutations. *Appl. Microbiol. Biotechnol.* **97**, 9727–9735
25. Song, X., Wang, Y., Shu, Z., Hong, J., Li, T., and Yao, L. (2013) Engineering a more thermostable blue light photo receptor *Bacillus subtilis* YtvA LOV domain by a computer aided rational design method. *PLoS Comput. Biol.* **9**, e1003129
26. Wijma, H. J., Floor, R. J., Jekel, P. A., Baker, D., Marrink, S. J., and Janssen, D. B. (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng. Des. Sel.* **27**, 49–58
27. Floor, R. J., Wijma, H. J., Colpa, D. I., Ramos-Silva, A., Jekel, P. A., Szymański, W., Feringa, B. L., Marrink, S. J., and Janssen, D. B. (2014) Computational library design for increasing haloalkane dehalogenase stability. *ChemBiochem* **15**, 1660–1672
28. Deng, Z., Yang, H., Li, J., Shin, H.-D., Du, G., Liu, L., and Chen, J. (2014) Structure-based engineering of alkaline α -amylase from alkaliphilic

Tradeoff between protein stability and solubility

- alkalimonas amylolytica for improved thermostability. *Appl. Microbiol. Biotechnol.* **98**, 3997–4007
29. Larsen, D. M., Nyffenegger, C., Swiniarska, M. M., Thygesen, A., Strube, M. L., Meyer, A. S., and Mikkelsen, J. D. (2015) Thermostability enhancement of an endo-1,4- β -galactanase from *talaromyces stipitatus* by site-directed mutagenesis. *Appl. Microbiol. Biotechnol.* **99**, 4245–4253
 30. Heselpoth, R. D., Yin, Y., Moulton, J., and Nelson, D. C. (2015) Increasing the stability of the bacteriophage endolysin pIyc using rationale-based foldx computational modeling. *Protein Eng. Des. Sel.* **28**, 85–92
 31. Bednar, D., Beerens, K., Sebestova, E., Bendl, J., Khare, S., Chaloupkova, R., Prokop, Z., Brezovsky, J., Baker, D., and Damborsky, J. (2015) Fireprot: energy- and evolution-based computational design of thermostable multiple-point mutants. *PLoS Comput. Biol.* **11**, e1004556
 32. Risso, V. A., Gavira, J. A., and Sanchez-Ruiz, J. M. (2014) Thermostable and promiscuous precambrian proteins. *Environ. Microbiol.* **16**, 1485–1489
 33. Porebski, B. T., and Buckle, A. M. (2016) Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251
 34. Potapov, V., Cohen, M., and Schreiber, G. (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein engineering, design & selection* **22**, 553–560
 35. Li, Z., Yang, Y., Zhan, J., Dai, L., and Zhou, Y. (2013) Energy functions in de novo protein design: current challenges and future prospects. *Annu. Rev. Biophys.* **42**, 315–335
 36. Wan, J., Kang, S., Tang, C., Yan, J., Ren, Y., Liu, J., Gao, X., Banerjee, A., Ellis, L. B. M., and Li, T. (2008) Meta-prediction of phosphorylation sites with weighted voting and restricted grid search parameter selection. *Nucleic Acids Res.* **36**, e22
 37. Yang, J., Roy, A., and Zhang, Y. (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* **29**, 2588–2595
 38. Qin, S., and Zhou, H.-X. (2007) meta-ppisp: a meta web server for protein-protein interaction site prediction. *Bioinformatics* **23**, 3386–3387
 39. Ishida, T., and Kinoshita, K. (2008) Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* **24**, 1344–1348
 40. Kozlowski, L. P., and Bujnicki, J. M. (2012) Metadisorder: a meta-server for the prediction of intrinsic disorder in proteins. *BMC Bioinformatics* **13**, 111
 41. Emily, M., Talvas, A., and Delamarche, C. (2013) Metamyl: a meta-predictor for amyloid proteins. *PLoS One* **8**, e79722
 42. Bava, K. A., Gromiha, M. M., Uedaira, H., Kitajima, K., and Sarai, A. (2004) Protherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.* **32**, D120–D121
 43. Broom, A., Ma, S. M., Xia, K., Rafalia, H., Trainor, K., Colón, W., Gosavi, S., and Meiering, E. M. (2015) Designed protein reveals structural determinants of extreme kinetic stability. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14605–14610
 44. Bloom, J. D., Labthavikul, S. T., Otey, C. R., and Arnold, F. H. (2006) Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5869–5874
 45. Tokuriki, N., and Tawfik, D. S. (2009) Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604
 46. Parthiban, V., Gromiha, M. M., and Schomburg, D. (2006) Cupsat: prediction of protein stability upon point mutations. *Nucleic Acids Res.* **34**, W239–W242
 47. Worth, C. L., Preissner, R., and Blundell, T. L. (2011) Sdm—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **39**, W215–W222
 48. Cohen, M., Potapov, V., and Schreiber, G. (2009) Four distances between pairs of amino acids provide a precise description of their interaction. *PLoS Comput. Biol.* **5**, e1000470
 49. Cheng, J., Randall, A., and Baldi, P. (2006) Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* **62**, 1125–1132
 50. Deutsch, C., and Krishnamoorthy, B. (2007) Four-body scoring function for mutagenesis. *Bioinformatics* **23**, 3009–3015
 51. Yang, Y., and Zhou, Y. (2008) Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci.* **17**, 1212–1219
 52. Davey, J. A., and Chica, R. A. (2014) Improving the accuracy of protein stability predictions with multistate design using a variety of backbone ensembles. *Proteins* **82**, 771–784
 53. Meiering, E. M., Serrano, L., and Fersht, A. R. (1992) Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* **225**, 585–589
 54. Shoichet, B. K., Baase, W. A., Kuroki, R., and Matthews, B. W. (1995) A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 452–456
 55. Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. (2008) How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, e1000002
 56. Broom, A., Doxey, A. C., Lobsanov, Y. D., Berthin, L. G., Rose, D. R., Howell, P. L., McConkey, B. J., and Meiering, E. M. (2012) Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. *Structure* **20**, 161–171
 57. Deng, Z., Huang, W., Bakkalbasi, E., Brown, N. G., Adamski, C. J., Rice, K., Muzny, D., Gibbs, R. A., and Palzkill, T. (2012) Deep sequencing of systematic combinatorial libraries reveals β -lactamase sequence constraints at high resolution. *J. Mol. Biol.* **424**, 150–167
 58. Araya, C. L., Fowler, D. M., Chen, W., Muniez, I., Kelly, J. W., and Fields, S. (2012) A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16858–16863
 59. Klesmith, J. R., Bacik, J.-P., Wrenbeck, E. E., Michalczyk, R., and Whitehead, T. A. (2017) Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2265–2270
 60. Voynov, V., Chennamsetty, N., Kayser, V., Helk, B., and Trout, B. L. (2009) Predictive tools for stabilization of therapeutic proteins. *mAbs* **1**, 580–582
 61. Thompson, M. J., Sievers, S. A., Karanicolas, J., Ivanova, M. I., Baker, D., and Eisenberg, D. (2006) The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4074–4078
 62. Lawrence, M. S., Phillips, K. J., and Liu, D. R. (2007) Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–10112
 63. Warwicker, J., Charonis, S., and Curtis, R. A. (2014) Lysine and arginine content of proteins: computational analysis suggests a new tool for solubility design. *Mol. Pharm.* **11**, 294–303
 64. Rauscher, S., Baud, S., Miao, M., Keeley, F. W., and Pomès, R. (2006) Proline and glycine control protein self-organization into elastomeric or amyloid fibrils. *Structure* **14**, 1667–1676
 65. Wimley, W. C., Creamer, T. P., and White, S. H. (1996) Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35**, 5109–5124
 66. Cordes, M. H., and Sauer, R. T. (1999) Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Sci.* **8**, 318–325
 67. Poso, D., Sessions, R. B., Lorch, M., and Clarke, A. R. (2000) Progressive stabilization of intermediate and transition states in protein folding reactions by introducing surface hydrophobic residues. *J. Biol. Chem.* **275**, 35723–35726
 68. Funahashi, J., Takano, K., Yamagata, Y., and Yutani, K. (2002) Positive contribution of hydration structure on the surface of human lysozyme to the conformational stability. *J. Biol. Chem.* **277**, 21792–21800
 69. Machius, M., Declerck, N., Huber, R., and Wiegand, G. (2003) Kinetic stabilization of *Bacillus licheniformis* α -amylase through introduction of hydrophobic residues at the surface. *J. Biol. Chem.* **278**, 11546–11553
 70. Ayuso-Tejedor, S., Abián, O., and Sancho, J. (2011) Underexposed polar residues and protein stabilization. *Protein Eng. Des. Sel.* **24**, 171–177
 71. Dantas, G., Kuhlman, B., Callender, D., Wong, M., and Baker, D. (2003) A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332**, 449–460
 72. Jacak, R., Leaver-Fay, A., and Kuhlman, B. (2012) Computational protein design with explicit consideration of surface hydrophobic patches. *Proteins* **80**, 825–838

73. Huang, P.-S., Boyken, S. E., and Baker, D. (2016) The coming of age of *de novo* protein design. *Nature* **537**, 320–327
74. Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., *et al.* (2014) Contribution of hydrogen bonds to protein stability. *Protein Sci.* **23**, 652–661
75. Bazzoli, A., Kelow, S. P., and Karanicolas, J. (2015) Enhancements to the rosetta energy function enable improved identification of small molecules that inhibit protein-protein interactions. *PLoS One* **10**, e0140359
76. Rykunov, D., and Fiser, A. (2007) Effects of amino acid composition, finite size of proteins, and sparse statistics on distance-dependent statistical pair potentials. *Proteins* **67**, 559–568
77. Rooman, M., and Gilis, D. (1998) Different derivations of knowledge-based potentials and analysis of their robustness and context-dependent predictive power. *Eur. J. Biochem.* **254**, 135–143
78. Thomas, P. D., and Dill, K. A. (1996) Statistical potentials extracted from protein structures: how accurate are they? *J. Mol. Biol.* **257**, 457–469
79. Broom, A., Trainor, K., MacKenzie, D. W., and Meiering, E. M. (2016) Using natural sequences and modularity to design common and novel protein topologies. *Curr. Opin. Struct. Biol.* **38**, 26–36
80. Boersma, Y. L., and Plückthun, A. (2011) Darpins and other repeat protein scaffolds: advances in engineering and applications. *Curr. Opin. Biotechnol.* **22**, 849–857
81. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K., and Plückthun, A. (2008) Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J. Mol. Biol.* **376**, 241–257
82. Balaji, S. (2015) Internal symmetry in protein structures: prevalence, functional relevance and evolution. *Curr. Opin. Struct. Biol.* **32**, 156–166
83. Tzul, F. O., Schweiker, K. L., and Makhatadze, G. I. (2015) Modulation of folding energy landscape by charge-charge interactions: linking experiments with computational modeling. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E259–E266
84. Worth, C. L., and Blundell, T. L. (2010) On the evolutionary conservation of hydrogen bonds made by buried polar amino acids: the hidden joists, braces and trusses of protein architecture. *BMC Evol. Biol.* **10**, 161
85. Bolen, D. W., and Rose, G. D. (2008) Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **77**, 339–362
86. Bolon, D. N., Marcus, J. S., Ross, S. A., and Mayo, S. L. (2003) Prudent modeling of core polar residues in computational protein design. *J. Mol. Biol.* **329**, 611–622
87. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., and Tawfik, D. S. (2007) The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369**, 1318–1332
88. Berliner, N., Teyra, J., Colak, R., Garcia Lopez, S., and Kim, P. M. (2014) Combining structural modeling with ensemble machine learning to accurately predict protein fold stability and binding affinity effects upon mutation. *PLoS One* **9**, e107353
89. Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., and Fleishman, S. J. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* **63**, 337–346
90. Bendl, J., Stourac, J., Sebestova, E., Vavra, O., Musil, M., Brezovsky, J., and Damborsky, J. (2016) Hotspot wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic Acids Res.* **44**, W479–W487
91. Davey, J. A., Damry, A. M., Euler, C. K., Goto, N. K., and Chica, R. A. (2015) Prediction of stable globular proteins using negative design with non-native backbone ensembles. *Structure* **23**, 2011–2021
92. Krivov, G. G., Shapovalov, M. V., and Dunbrack, R. L. (2009) Improved prediction of protein side-chain conformations with scwrl4. *Proteins* **77**, 778–795
93. Creighton, T. (1993) *Proteins: Structure and molecular properties*, p. 4, W. H. Freeman & Co., NY
94. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38
95. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637
96. Matthews, B. W. (1975) Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Biophys. Acta* **405**, 442–451
97. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J.-C., Bona, D., Miller, E. J., Vallée-Bélisle, A., Main, E. R. G., Bemporad, F., Qiu, L., Teilum, K., *et al.* (2005) Protein folding: defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci.* **14**, 602–616
98. Cota, E., and Clarke, J. (2000) Folding of β -sandwich proteins: three-state transition of a fibronectin type iii module. *Protein Sci.* **9**, 112–120
99. Zambrano, R., Jamroz, M., Szczasiuk, A., Pujols, J., Kmiecik, S., and Ventura, S. (2015) Aggrescan3d (a3d): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res.* **43**, W306–W313
100. Sormanni, P., Aprile, F. A., and Vendruscolo, M. (2015) The camsol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **427**, 478–490
101. Walsh, I., Seno, F., Tosatto, S. C. E., and Trovato, A. (2014) Pasta 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* **42**, W301–W307
102. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306
103. Goldschmidt, L., Teng, P. K., Riek, R., and Eisenberg, D. (2010) Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3487–3492
104. Tartaglia, G. G., and Vendruscolo, M. (2008) The zyggregator method for predicting protein aggregation propensities. *Chem. Soc. Rev.* **37**, 1395–1401