BIOCHEMISTRY
including biophysical chemistry & molecular biology

Accelerated Publication

pubs.acs.org/biochemistry

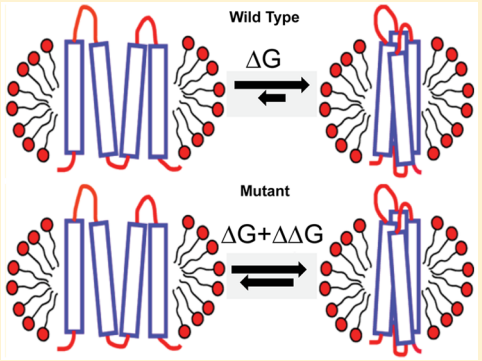
Documentation of an Imperative To Improve Methods for Predicting Membrane Protein Stability

Brett M. Kroncke,[†,‡] Amanda M. Duran,[‡] Jeffrey L. Mendenhall,[‡] Jens Meiler,[*,‡,§] Jeffrey D. Blume,[*,∥] and Charles R. Sanders[*,†,‡]

[†]Department of Biochemistry, [‡]Center for Structural Biology, [§]Departments of Chemistry, Pharmacology, and Bioinformatics, and [∥]Department of Biostatistics, Vanderbilt University, Nashville, Tennessee 37240, United States

Supporting Information

ABSTRACT: There is a compelling and growing need to accurately predict the impact of amino acid mutations on protein stability for problems in personalized medicine and other applications. Here the ability of 10 computational tools to accurately predict mutation-induced perturbation of folding stability ($\Delta\Delta G$) for membrane proteins of known structure was assessed. All methods for predicting $\Delta\Delta G$ values performed significantly worse when applied to membrane proteins than when applied to soluble proteins, yielding estimated concordance, Pearson, and Spearman correlation coefficients of <0.4 for membrane proteins. Rosetta and PROVEAN showed a modest ability to classify mutations as destabilizing ($\Delta\Delta G < -0.5$ kcal/mol), with a 7 in 10 chance of correctly discriminating a randomly chosen destabilizing variant from a randomly chosen stabilizing variant. However, even this performance is significantly worse than for soluble proteins. This study highlights the need for further development of reliable and reproducible methods for predicting thermodynamic folding stability in membrane proteins.



Each individual's genome has, on average, 10000−20000 nonsynonymous single-nucleotide polymorphisms (nsSNPs).[1] Deleterious, loss-of-function nsSNPs constitute the most common cause of monogenic disorders.[2−4] Substantial evidence suggests a majority of disease-promoting nsSNPs act, at least in part, by destabilizing the folded conformation of the encoded protein.[3−7] The resulting loss of thermodynamic stability leads to a reduced population of functional protein available to cells, which in some cases is compounded by the toxicity of the misfolded protein.[8−10] The more accurately mutation-induced changes in protein stability can be determined, the more accurately and specifically we can predict loss-of-function phenotypes for previously uncharacterized point mutations, a growing concern as more genomes are sequenced to unveil variants of unknown significance.[1]

There are many algorithms that predict changes in folded protein stability caused by single- or multiple-amino acid mutations. Some approaches rely on known protein structures using functions that predict the energetic perturbation introduced by the mutation.[11] Other methods train machine learning methods on large data sets to combine selected physical, statistical, and empirical features for stability predictions.[12,13] For water-soluble proteins, several algorithms are able to predict mutation-induced change in stability with a Pearson correlation coefficient near or above 0.7 (Figure 1); however, the performance of these methods on membrane proteins is an open question. Membrane proteins fold and reside in a heterogeneous environment—a lipid bilayer bounded on both sides by water—with distinct forces driving folding and unfolding compared to soluble proteins, and therefore may require treatment separate from that of soluble proteins.[14−17]

Membrane protein structures comprise only ∼1% of the protein structure database (http://www.rcsb.org/pdb/home/ and http://blanco.biomol.uci.edu/mpstruc/), and thermodynamic stability measurements of membrane proteins are grossly underrepresented. This paucity of data dictates that all currently available $\Delta\Delta G$ calculators have been trained and refined from data sets strongly biased toward soluble proteins. Here we evaluate the ability of current methods to predict amino acid mutation-induced free energy changes in membrane protein stability in cases both for which an atomic-resolution structure is available and for which stabilities of wild-type and mutant forms have been measured.

METHODS

Compilation of Experimental $\Delta\Delta G$ Values. We used all available (as of January 2016) experimental $\Delta\Delta G$ data sets for mutant forms of membrane proteins of known structure. The relevant Protein Data Bank (PDB) codes are as follows: 1PY6 for bacteriorhodopsin,[18] 1AFO for glycophorin A,[19] 2XOV for the *Escherichia coli* rhomboid protease (GlpG),[20] 2K73 for

Received: May 26, 2016
Revised: August 19, 2016
Published: August 26, 2016

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009

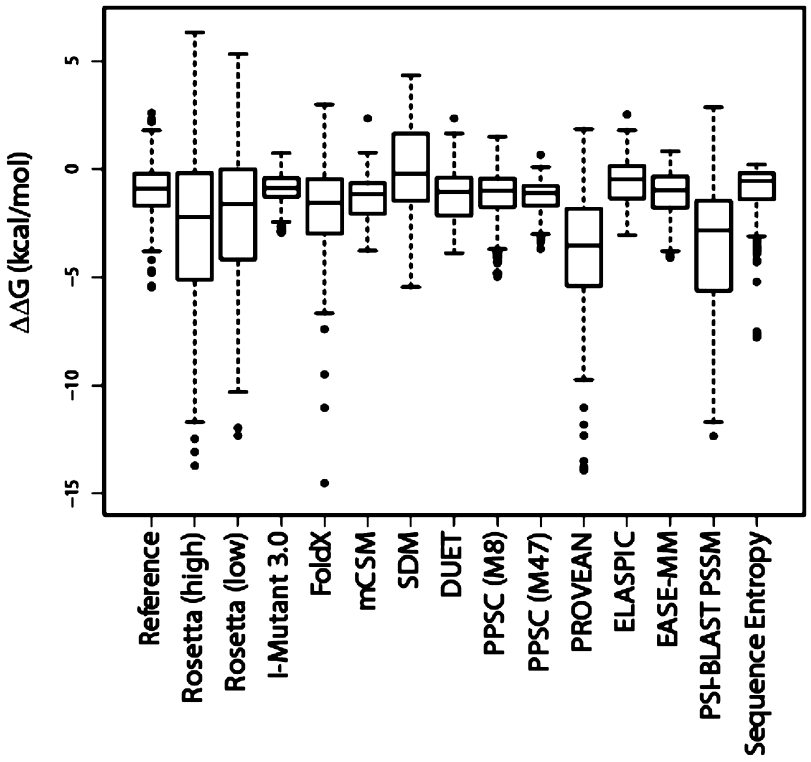


Figure 1. Boxplot of experimental (reference) and predicted value distributions. The middle line in the box is the median, and upper and lower bounds to the boxes are the upper and lower quartiles, respectively. Nonoutlier extrema are bracketed with dashed lines above and below the upper and lower quartiles, respectively. Dots are outliers beyond 1.5 times the upper or lower quartile.

disulfide formation protein B (DsbB),[21] 1QD6 for outer membrane phospholipase A1 (OmpLA),[22] 1QJP for outer membrane protein A (OmpA),[23] and 3GP6 for the lipid A palmitoyltransferase (PagP).[24] The 223 rigorously determined $\Delta\Delta G$ measurements originated from the following studies: bacteriorhodopsin,[18,25−29] glycophorin A,[30,31] GlpG,[32,33] DsbB,[34] OmpLA,[35] OmpA,[16] and PagP.[36]

Protein Stability Programs. We tested available methods for which servers or software were available online and functional as of January 2014 or for which the authors of published algorithms were responsive to our request for software (Table 1). The following programs were used to predict $\Delta\Delta G$ values for each membrane protein mutation in the experimental database mentioned above: Rosetta (revision 58019) with both low-resolution (Rosetta-low) and high-resolution (Rosetta-high) protocols,[37] I Mutant (3.0; http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi),[38] FoldX (3.0, beta 6.1),[11] mCSM,[39] SDM,[40] DUET (http://bleoberis.bioc.cam.ac.uk/duet/stability),[41] PPSC (Prediction of Protein Stability, version 1.0) with the 8 (M8) and 47 (M47) feature sets,[12] PROVEAN (http://provean.jcvi.org/seq_submit.php),[42] ELASPIC (http://elaspic.kimlab.org/),[13] and EASE-MM.[43] We also tested the standard Rosetta ddg_monomer application replacing the minimization score function score12 with membrane_highres_Menv_smooth (RosettaMembrane). In addition we tested the RosettaMP $\Delta\Delta G$ calculating framework, RosettaMPddG. Both attempts failed to improve performance (Figure S1). The membrane protein scoring function adds nothing in accuracy and discrimination for calculating $\Delta\Delta G$ values in Rosetta.

To compare the performance of each $\Delta\Delta G$ calculation method with what could be obtained from sequence information alone, we calculated two parameters. First, the likelihood of a specified amino acid mutation being observed among the wild-type (WT) sequences comprising a particular protein family was assessed according to the position-specific iterative basic local alignment search tool-derived position-specific scoring matrix (PSI-BLAST PSSM). PSI-BLAST PSSM values were calculated, as follows. The PSI-BLAST position-specific scoring matrix value for a given mutant residue amino acid type was subtracted from the value for the native residue (PSI-BLAST employed the UniRef50, nonredundant sequence database, 5-iterations, e-value cutoff of 0.01). This metric gives an estimation of the evolutionary penalty for substituting the WT residue with the specified mutant amino acid. Second, the Shannon (or "sequence entropy") entropy was determined from PSI-BLAST results. Sequence entropy is a description of how often the identity of a particular residue in a protein changes from family member to family member. Shannon/sequence entropy is the PSSM value for amino acids located at a particular position. This parameter is agnostic with regard to the amino acid type of both the mutated-in and native residue. Instead, the Shannon/sequence entropy reports the likelihood that a change in residue identity is evolutionarily tolerated. All numbers were formatted so that negative values indicate destabilization.

Statistical Analysis of Experimental versus Predicted $\Delta\Delta G$ Values. For each method, the experimental versus predicted $\Delta\Delta G$ data were processed using an in-house R script to calculate correlation coefficients and area-under-the-curve (AUC) values. To analyze the collected data set on the basis of several features, we parsed out and evaluated separately point mutations according to the following classifications: those impacting α-helical versus β-barrel proteins, those with a point mutation site in the aqueous phase, in the aliphatic phase, or at the water−membrane interface, and mutations at positions that

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009

Table 1. Summary of Methods Evaluated

name	brief description	method[a]	calibrated[b]	sequence	Pearson[c]	stability data sets[d]
Rosetta[37]	Structure knowledge-based potential. Score terms considered: van der Waals, electrostatics, solvation, hydrogen bond, rotamer probability. ddG_monomer application	N/A			0.69 (high), 0.68 (low)	ProTherm[46] (test set)
I Mutant 3.0[38]	Support vector machine (SVM)-based predictor; can use sequence information and structure information to predict destabilizing, neutral, and stabilized	SVM	X	X	0.69	Thermodynamic Database for Proteins and Mutants ProTherm (September 2005)
FoldX[11]	Empirical force field calibrated with experimental ddG values. Score terms considered: van der Waals, solvation, hydrogen bonding, water bridges, electrostatic, entropy of backbone and side chain, and atomic clashes	grid search	X		0.8	derived from ProTherm
mCSM[39]	Graph-based structural signatures: distance patterns between atoms to represent the environment. Also considers pharmacophore changes and experimental conditions. Supervised learning machine learning methods trained on regression and classification	ANN	X		0.82	derived from ProTherm
SDM[40]	Statistical potential energy function (structure): evaluates amino acid structural propensities in homologous protein families	N/A		X	0.58	derived from ProTherm
DUET[41]	SVM that combines mCSM and SDM methods	SVM	X	X	0.71	ProTherm (low-redundancy set)
PPSC (M8)[12]	SVM with eight attributes: hydropathy, isotropic surface area, electronic charge, volume, contact energy	SVM	X		0.65	derived from ProTherm
PPSC (M47)[12]	SVM trained with 8 + 40 additional protein features from ref 38 (I Mutant 2)	SVM	X		0.82	derived from ProTherm
PROVEAN[42]	Pairwise sequence alignment scores to predict effects of a mutation, including deletions, insertions, and multiple substitutions	N/A		X	0.71[e]	derived from UniProtKB and Swiss-Prot databases
ELASPIC[13]	Machine learning approach that combines semiempirical force fields, sequence conservation scores, and structural information through stochastic gradient boosting of decision trees	SGBT-DT	X	X	0.77	ProTherm
EASE-MM[43]	Sequence-based SVM model that evaluates the predicted secondary structure and accessible surface area of the region of interest	SVM	X	X	0.56	derived from ProTherm

[a]Type of machine learning method used: artificial neural network (ANN), support vector machine (SVM), and stochastic gradient boosting of decision trees (SGBT-DT). [b]The predictive method is calibrated to experimental $\Delta\Delta G$ values. [c]Reported Pearson correlation coefficient. [d]Used to derive both training and testing sets unless otherwise noted. [e]Activity correlation.

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009

Table 2. Summary of Statistical Methods Used To Evaluate Predictive Methods

quantification method	description
concordance CC[a]	The concordance correlation coefficient measures the degree to which the predicted $\Delta\Delta G$ value equals the actual experimental value (0 indicates no agreement and 1 perfect agreement).
Pearson CC[a]	The Pearson correlation coefficient measures the degree to which a uniform linear transformation of the predicted $\Delta\Delta G$ values (i.e., a shift and scale change) would yield the actual experimental values (0 indicates no agreement after transformation, 1 perfect agreement, and −1 perfect inverse agreement).
Spearman rank CC[a]	The Spearman rank correlation coefficient measures the degree to which the rank ordering of the predicted $\Delta\Delta G$ values matches the rank ordering of the actual experimental values (0 indicates no agreement after transformation, 1 perfect agreement, and −1 perfect inverse agreement).
ROC and AUC	The area-under-the-receiver operating characteristic (ROC) curve tests several cutoff values for binning mutations as neutral or destabilizing between the most negative calculated $\Delta\Delta G$ value and the most positive calculated $\Delta\Delta G$ value, with true positive rates (sensitivity) calculated at each point. As the true positive rate is calculated, the classifier is moved to less extreme values; this yields the ROC curve. The AUC curve is a summary statistic that approximates how well the predictor actually discriminates between the two classifications.

[a]CC indicates correlation coefficient.

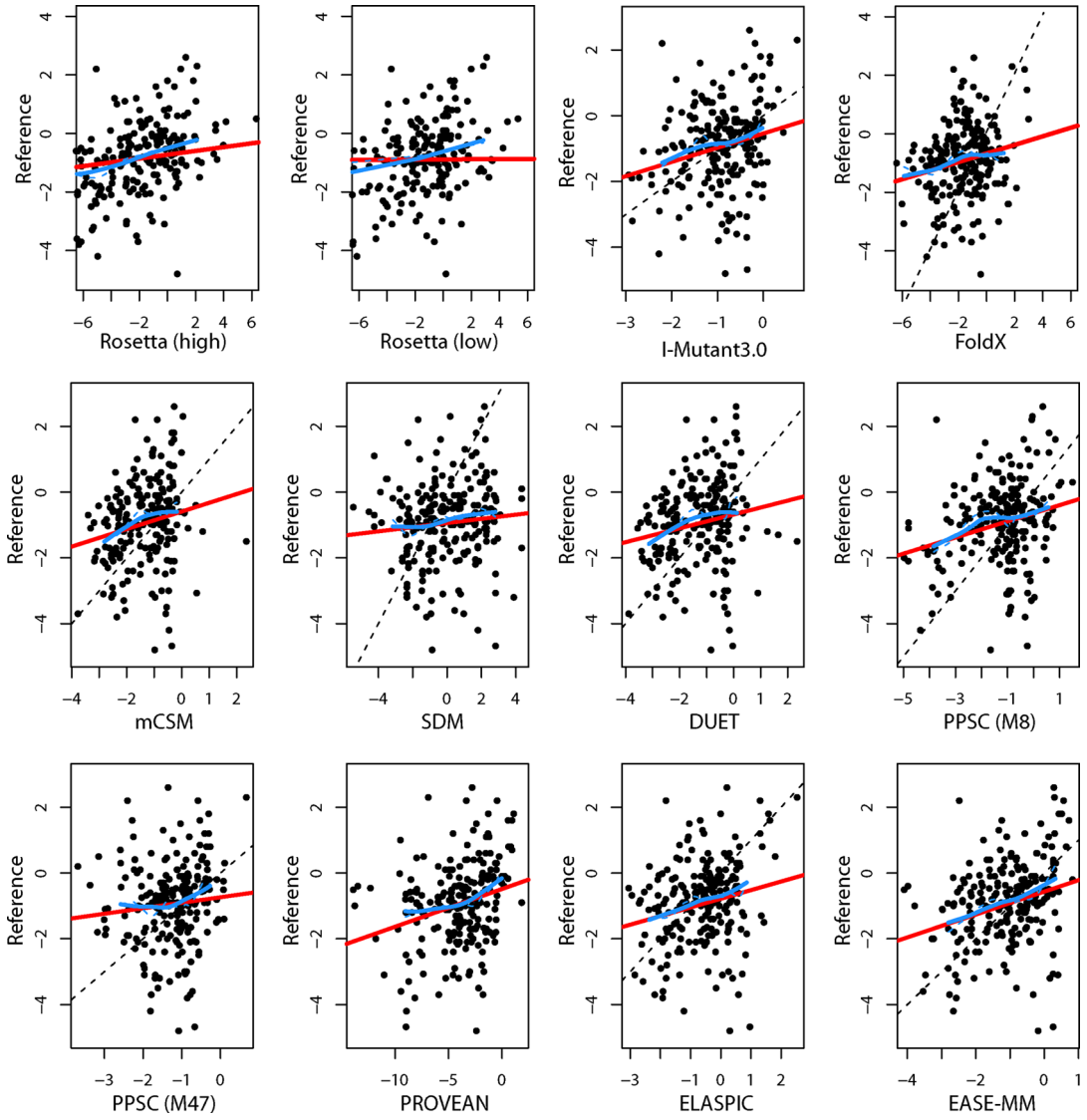


Figure 2. Reference (experimental) $\Delta\Delta G$ values vs calculated ddG values (x-axis) from each method tested (see also Table S1). Red lines are simple linear regressions from which Pearson correlations are derived; blue lines are flexible nonparametric trend lines. For the Rosetta and FoldX plots, a few predicted points were outliers that fall outside of the plotted window. The dashed line is the $y = x$ line measuring perfect agreement between the predicted $\Delta\Delta G$ and the experimental values and is plotted for methods constructed to make direct predictions.

were either buried within the protein or exposed to solvent or lipid (Figures S2−S10). We analyzed the set of predictions for each protein separately and also parsed out point mutations involving proline or glycine (Figures S11−S17). Concordance, Pearson, and Spearman correlations were computed, along with ROC curves (and their AUC values) for predicting a negative $\Delta\Delta G$ of less than −0.5 (see Table 2). The concordance correlation is the proper statistic for assessing agreement among continuous measurements, though the Pearson correlation is more common in the literature. The Spearman

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009

correlation is a rank-based correlation analogue of Pearson that is less reliant on linear assumptions. We used a nonparametric bootstrap (500 replications) to obtain estimates of standard errors and bias-corrected 95% confidence intervals (CIs) for estimates. We used scatter plots with nonparametric trend lines to examine the data. Bland–Altman plots were used to visually examine the agreement between predictions and actual values. As a control for our processing, we also computed correlation coefficients using previous Rosetta $\Delta\Delta G$ prediction results from a large data set containing almost exclusively soluble proteins.[37]

RESULTS AND DISCUSSION

We collected all available experimental $\Delta\Delta G$ data sets for structurally diverse membrane proteins of known structure (which constitutes the vast majority of all $\Delta\Delta G$ measurements made to date for membrane proteins). We acknowledge differences in the cellular folding landscapes of α-helical and β-barrel proteins; however, given the limited number of membrane proteins with known structure and thermodynamic stability measurements, we combined all proteins for analysis and subsequently parsed potentially relevant subsets to evaluate the effect of each. As of early 2016, there were 223 single-amino acid $\Delta\Delta G$ destabilization measurements available for these proteins, with mutated side chains in the following categories: water-exposed, 6% (14); lipid hydrocarbon-exposed, 25% (55); exposed interfacial, 18% (41); or protein-buried, 52% (117).

The distribution of experimental $\Delta\Delta G$ values is consistent with a random sampling of residue point mutation stabilities (Figure 1): 65% of point mutations resulted in $\Delta\Delta G$ values of less than −0.5 kcal/mol, considered destabilizing; 24% between −0.5 and 0.5 kcal/mol, considered neutral; and 11% greater than 0.5 kcal/mol, considered stabilizing, as suggested previously.[44] All programs except Rosetta, PROVEAN, SDM, and FoldX have a narrow, slightly negative distribution of predicted $\Delta\Delta G$ values (Figures 1 and 2). The PSI-BLAST PSSM scores were also more dispersed than results for the majority of the programs tested. Interestingly, SDM tended to classify nearly as many mutations as stabilizing as destabilizing, which perhaps is a consequence of restricting mutant classification to neutral or destabilizing only if $|\Delta\Delta G| > 2$ kcal/mol. Most methods tended to underestimate $\Delta\Delta G$ for destabilizing mutations and overestimate $\Delta\Delta G$ for neutral to stabilizing mutations.

To evaluate the predictive ability of each method tested, we compared concordance, Pearson, and Spearman rank correlation coefficients (Figure 2A; a glossary for statistical parameters is provided in Table 2). Note that we distinguish methods that were calibrated to predict $\Delta\Delta G$ values from methods that compute metrics that are expected to linearly correlate with $\Delta\Delta G$ values, such as ROSETTA. This distinction is important, as for optimal performance in the former group we expect a regression line that passes through the coordinate origin and has a slope of 1. In such a case, concordance, Pearson, and Spearman correlation coefficients would be equal to 1. In the latter group, for optimal performance, Pearson and Spearman correlation coefficients, but not the concordance, would be equal to 1.

None of the programs tested performed well in calculating $\Delta\Delta G$ values for membrane proteins compared to their performance in previous studies of soluble protein data sets (Figure 3A). The concordance correlation coefficients for the various methods are all relatively low, the highest being ~0.2

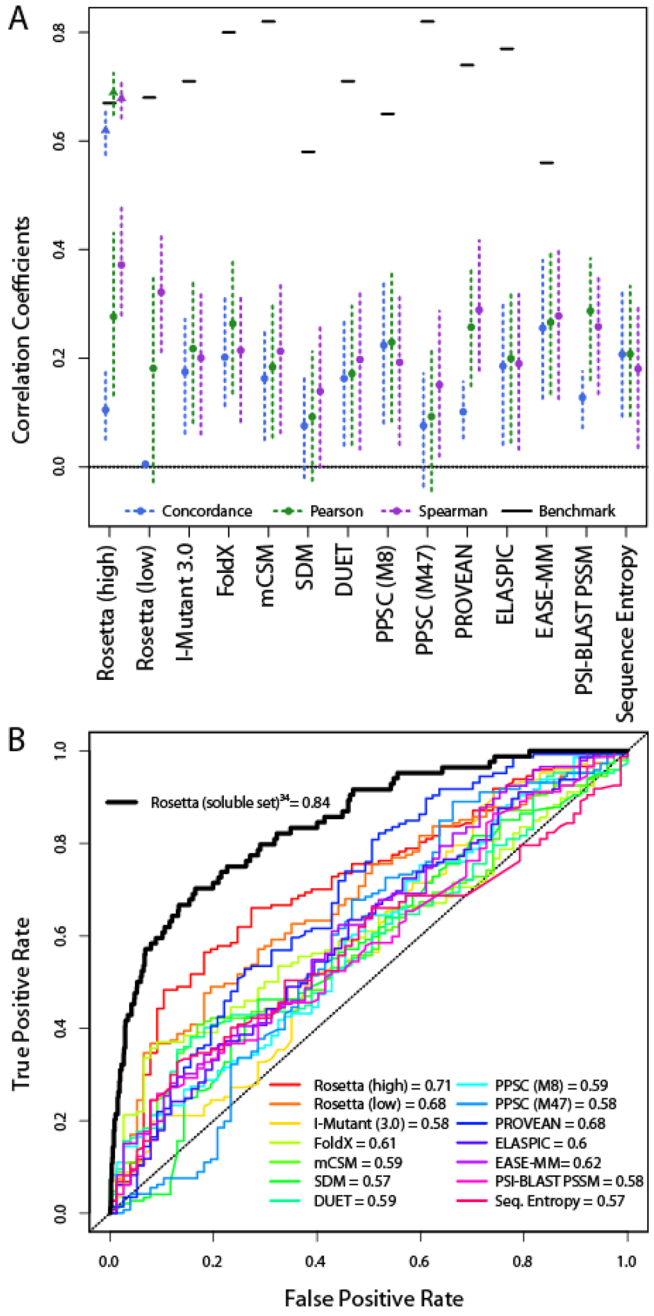


Figure 3. (A) Performance of each evaluated method in predicting true $\Delta\Delta G$ values (concordance correlation coefficient), linearly correlated ddG values (Pearson correlation coefficient), and rank order (Spearman rank order correlation coefficient). The hash marks in the upper portions of this plot indicate the published results for each method. We also evaluated the concordance, Pearson, and Spearman correlation coefficients using the calculated and experimental data previously reported[37] for a mostly water-soluble protein data set to control for processing differences, shown as triangles. (B) Receiver operating characteristic curves of the classification of variants that are more destabilized or less destabilized than 0.5 kcal/mol. We generated the black bold trace using data from a previous $\Delta\Delta G$ calculation effort[37] involving mostly soluble proteins.

[EASE-MM, FoldX, and PPSC (M8)]. This is compared to a concordance correlation coefficient in the range of 0.6 for the Rosetta-based method applied to an almost exclusively water-soluble protein data set. The performance of the different methods at predicting the rank order is improved compared to

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002–5009

their ability to predict absolute $\Delta\Delta G$ values (Figure 3A), but all Spearman correlation coefficients are below 0.4, compared to 0.7 for the Rosetta-based method applied to a largely water-soluble protein data set. This means the majority of predicted rankings are still incorrect. Rosetta (high and low) and PROVEAN have the highest Spearman rank order correlation coefficients overall (0.37, 0.32, and 0.29, respectively) but still significantly underperform compared to results for soluble proteins. The general failure of these methods to reliably rank order the impact of membrane protein point mutations on stability is disappointing, as one of the anticipated applications for these methods is to aid researchers in identifying the most or least destabilizing mutations out of a hypothetical set, which then would be experimentally tested for the purpose of protein engineering.

Another application that can be envisioned is predicting the stability class for a given variant. For example, one might seek to identify mutants that have a $\Delta\Delta G$ value above or below −0.5 kcal (−0.5 is the typical uncertainty in experimentally determined stabilities[45]). To compare the discriminating power of these methods, we plotted receiver operating characteristic curves [ROC (Figure 3B)], which show the ability to correctly classify point mutations as destabilizing ($\Delta\Delta G < -0.5$) or neutral/stabilizing ($\Delta\Delta G > -0.5$). ROC curves that are skewed toward a higher true positive rate (sensitivity) classify mutations more accurately, as quantified by AUC (ranging between 1.0 and 0.5 for perfect and chance classification, respectively). Rosetta and PROVEAN had the largest areas under the curve (95% CIs of 0.65−0.79 and 0.61−0.76, respectively). This is surprising because neither method was constructed or calibrated to predict $\Delta\Delta G$ values but is consistent with their better Spearman correlation performance. PROVEAN is designed to estimate the probability that a variant will be functionally compromised without accounting for structure, while Rosetta is optimized to incorporate protein structural features. The AUC of ∼0.8 for the soluble protein set calculated here, similar to previously reported values for these methods, further emphasizes the conclusion that the unique properties of membrane proteins require separate treatments in constructing stability prediction methods.

A priori, there are several potential explanations for the observed disparity in calculating $\Delta\Delta G$ values for soluble versus membrane proteins. One confounding factor could be the persistence of α-helical structure in the unfolded states of helical membrane proteins, which is typically not the case for unfolded states of soluble proteins. In an effort to test this hypothesis, we separately evaluated β-barrels, expected to have no persistent secondary structure in the unfolded state, and α-helical membrane proteins. The correlation coefficients for the β-barrel protein set have considerably larger 95% confidence intervals but suggest that several programs perform somewhat better for β-barrel proteins (Spearman correlation coefficient of 0.29) than for α-helical membrane proteins (average Spearman correlation coefficient of 0.22) (Figures S2 and S3), although the poor performance for both groups of proteins proves no method is reliable at this task. Interestingly, differences in correlation and ranking ability were not uniform between the methods evaluated: FoldX performed better on α-helical proteins (second-highest Spearman correlation coefficient) than on β-barrels (lowest Spearman correlation coefficient), with estimated Spearman correlations of 0.35 and 0.01, respectively. We also evaluated the effect of parsing out the secondary structure-disrupting residues, glycine and proline. Surprisingly, even removing proline and glycine residues did not improve Spearman correlation coefficients appreciably; 95% confidence intervals narrowed, and estimated values increased from 0.23 to 0.29 (Figure 3A and Figure S4).

Another potential cause of the disparity between soluble and membrane proteins may be the unique solvent environment of the membrane. We parsed $\Delta\Delta G$ values based on residue position: water-exposed (Figure S6), at the membrane interface (Figure S7), membrane-exposed (Figure S8), solvent-facing (Figure S9), or buried in the protein (Figure S10). Given the small number of water-exposed variants assessed, the 95% confidence interval is extremely wide, precluding any real assessment. In any case, no parsing of residue position yielded significant improvements in Spearman correlations. Indeed, to our surprise, all methods tended toward worse predictive ranking for protein-buried residues (average Spearman correlation coefficient of 0.19) than for solvent-exposed residues (Spearman correlation coefficient of 0.25).

Finally, it should be acknowledged that the methods used for experimentally measuring membrane protein ddG values are not yet highly standardized, reflecting use of denaturants as different as sodium dodecyl sulfate and urea, as well as model membranes as different as micelles and bilayer vesicles. The degree to which the stability of a single membrane protein is similar when measured using different methods has yet to be extensively tested.

An open question is whether more computationally intensive strategies, such as molecular dynamics-based approaches, will improve predictive power for membrane proteins. We did not investigate this kind of approach here because of the limiting throughput that can be achieved at present.

In this study, a series of diverse statistical criteria are in uniform agreement that current methods for predicting $\Delta\Delta G$ values of point mutations in membrane proteins will need to be improved or superseded to be reliable and useful. According to our evaluation, the predictive ability of the 10 methods assessed was not greatly improved from that of the PSI-BLAST PSSM and sequence entropy scores, i.e., what one could infer on the basis of mutated site evolutionary sequence conservation. We did not find any method to be robust at predicting either the rank order of mutations or absolute $\Delta\Delta G$ values. This study highlights the need to separately evaluate the performance of $\Delta\Delta G$ calculators on membrane proteins in the future, as well as the need for a much larger training database of experimentally measured stabilities for wild-type and mutant membrane proteins.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.6b00537.

Figures S1−S17 contain a comparison of concordance, Pearson, and Spearman correlation coefficients from different parsings of the $\Delta\Delta G$ data. Figure S1 compares membrane protein-specific scoring in Rosetta to the standard scoring used for membrane proteins. Figures S2 and S3 compare β-barrel proteins and α-helical proteins, respectively. Figures S4 and S5 compare only mutations that involve a proline or glycine and point mutations that do not involve a proline or glycine. Figures S6−S8 compare results for residues in the aqueous phase,

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009

residues at the interface between membrane and aqueous phases, and residues in the aliphatic phase of the membrane. Figures S9 and 10 compare solvent-exposed residues and buried residues. Figures S11−17 compare bacteriorhodopsin, glycophorin A, GlpG, DsbB, OmpLA, OmpA, and PagP (PDF)

Excel file containing all compiled experimental $\Delta\Delta G$ and calculated $\Delta\Delta G$ values (ZIP)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: chuck.sanders@vanderbilt.edu. Phone: +1-615-936-3756. Fax: +1-615-936-2211.

*E-mail: jens.meiler@vanderbilt.edu.

*E-mail: j.blume@vanderbilt.edu.

Author Contributions

C.R.S. and B.M.K. designed the research and compiled the experimental database. B.M.K., A.M.D., and J.L.M. analyzed the data under the guidance of J.D.B. B.M.K., J.M., J.D.B., and C.R.S. wrote the manuscript.

Funding

This project was supported by National Institutes of Health (NIH) Grant R01 HL122010. A.M.D. was supported by the National Science Foundation Graduate Research Fellowship Program under Grants 0909667 and 1445197. B.M.K. was supported by NIH Grant F32 GM113355.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Jonathan Schlebach and Sirui Ma for critical feedback on this manuscript, Shane O'Connor for providing the data set used in ref 37, and Lukas Folkman for assistance with EASE-MM.

ABBREVIATIONS

ANN, artificial neural network; AUC, area under the curve; CIs, confidence intervals; nsSNP, nonsynonymous single-nucleotide polymorphism; PSSM, position-specific scoring matrices; ROC, receiver operating characteristic; SVV, support vector machine; SGBT-DT, stochastic gradient boosting of decision trees.

REFERENCES

(1) Kroncke, B. M., Vanoye, C. G., Meiler, J., George, A. L., Jr., and Sanders, C. R. (2015) Personalized biochemistry and biophysics. *Biochemistry 54*, 2551−2559.

(2) Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shaw, K., and Cooper, D. N. (2012) The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Current Protocols in Bioinformatics*, Chapter 1, Unit 1, 13, Wiley, New York.

(3) Wang, Z., and Moult, J. (2001) SNPs, protein structure, and disease. *Hum. Mutat. 17*, 263−270.

(4) Yue, P., Li, Z., and Moult, J. (2005) Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol. 353*, 459−473.

(5) Casadio, R., Vassura, M., Tiwari, S., Fariselli, P., and Luigi Martelli, P. (2011) Correlating disease-related mutations to their effect on protein stability: a large-scale analysis of the human proteome. *Hum. Mutat. 32*, 1161−1170.

(6) Shi, Z., and Moult, J. (2011) Structural and functional impact of cancer-related missense somatic mutations. *J. Mol. Biol. 413*, 495−512.

(7) Stefl, S., Nishi, H., Petukh, M., Panchenko, A. R., and Alexov, E. (2013) Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol. 425*, 3919−3936.

(8) Calamini, B., and Morimoto, R. I. (2013) Protein homeostasis as a therapeutic target for diseases of protein conformation. *Curr. Top. Med. Chem. 12*, 2623−2640.

(9) Knowles, T. P., Vendruscolo, M., and Dobson, C. M. (2014) The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol. 15*, 384−396.

(10) Valastyan, J. S., and Lindquist, S. (2014) Mechanisms of protein-folding diseases at a glance. *Dis. Models & Mech. 7*, 9−14.

(11) Guerois, R., Nielsen, J. E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol. 320*, 369−387.

(12) Yang, Y., Chen, B., Tan, G., Vihinen, M., and Shen, B. (2013) Structure-based prediction of the effects of a missense variant on protein stability. *Amino Acids 44*, 847−855.

(13) 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.

(14) Neumann, J., Klein, N., Otzen, D. E., and Schneider, D. (2014) Folding energetics and oligomerization of polytopic alpha-helical transmembrane proteins. *Arch. Biochem. Biophys. 564*, 281−296.

(15) Cymer, F., von Heijne, G., and White, S. H. (2015) Mechanisms of integral membrane protein insertion and folding. *J. Mol. Biol. 427*, 999−1022.

(16) Hong, H., Park, S., Flores Jiménez, R. H., Rinehart, D., and Tamm, L. K. (2007) Role of aromatic side chains in the folding and thermodynamic stability of integral membrane proteins. *J. Am. Chem. Soc. 129*, 8320−8327.

(17) Popot, J. L., and Engelman, D. M. (2000) Helical membrane protein folding, stability, and evolution. *Annu. Rev. Biochem. 69*, 881−922.

(18) Faham, S., Yang, D., Bare, E., Yohannan, S., Whitelegge, J. P., and Bowie, J. U. (2004) Side-chain contributions to membrane protein structure and stability. *J. Mol. Biol. 335*, 297−305.

(19) MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) A transmembrane helix dimer: structure and implications. *Science 276*, 131−133.

(20) Vinothkumar, K. R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S., and Freeman, M. (2010) The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J. 29*, 3797−3809.

(21) Zhou, Y., Cierpicki, T., Jimenez, R. H., Lukasik, S. M., Ellena, J. F., Cafiso, D. S., Kadokura, H., Beckwith, J., and Bushweller, J. H. (2008) NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol. Cell 31*, 896−908.

(22) Snijder, H. J., Ubarretxena-Belandia, I., Blaauw, M., Kalk, K. H., Verheij, H. M., Egmond, M. R., Dekker, N., and Dijkstra, B. W. (1999) Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature 401*, 717−721.

(23) Pautsch, A., and Schulz, G. E. (2000) High-resolution structure of the OmpA membrane domain. *J. Mol. Biol. 298*, 273−282.

(24) Cuesta-Seijo, J. A., Neale, C., Khan, M. A., Moktar, J., Tran, C. D., Bishop, R. E., Pomes, R., and Prive, G. G. (2010) PagP crystallized from SDS/cosolvent reveals the route for phospholipid access to the hydrocarbon ruler. *Structure 18*, 1210−1219.

(25) Schlebach, J. P., Woodall, N. B., Bowie, J. U., and Park, C. (2014) Bacteriorhodopsin folds through a poorly organized transition state. *J. Am. Chem. Soc. 136*, 16574−16581.

(26) Yohannan, S., Yang, D., Faham, S., Boulting, G., Whitelegge, J., and Bowie, J. U. (2004) Proline substitutions are not easily accommodated in a membrane protein. *J. Mol. Biol. 341*, 1−6.

(27) Joh, N. H., Oberai, A., Yang, D., Whitelegge, J. P., and Bowie, J. U. (2009) Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *J. Am. Chem. Soc. 131*, 10846−10847.

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009

(28) Joh, N. H., Min, A., Faham, S., Whitelegge, J. P., Yang, D., Woods, V. L., and Bowie, J. U. (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature 453*, 1266−1270.
(29) Cao, Z., Schlebach, J. P., Park, C., and Bowie, J. U. (2012) Thermodynamic stability of bacteriorhodopsin mutants measured relative to the bacterioopsin unfolded state. *Biochim. Biophys. Acta, Biomembr. 1818*, 1049−1054.
(30) Fleming, K. G., and Engelman, D. M. (2001) Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc. Natl. Acad. Sci. U. S. A. 98*, 14340−14344.
(31) Fleming, K. G., Ackerman, A. L., and Engelman, D. M. (1997) The effect of point mutations on the free energy of transmembrane alpha-helix dimerization. *J. Mol. Biol. 272*, 266−275.
(32) Paslawski, W., Lillelund, O. K., Kristensen, J. V., Schafer, N. P., Baker, R. P., Urban, S., and Otzen, D. E. (2015) Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc. Natl. Acad. Sci. U. S. A. 112*, 7978−7983.
(33) Baker, R. P., and Urban, S. (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol. 8*, 759−768.
(34) Otzen, D. E. (2011) Mapping the folding pathway of the transmembrane protein DsbB by protein engineering. *Protein Eng., Des. Sel. 24*, 139−149.
(35) Moon, C. P., and Fleming, K. G. (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A. 108*, 10174−10177.
(36) Huysmans, G. H., Baldwin, S. A., Brockwell, D. J., and Radford, S. E. (2010) The transition state for folding of an outer membrane protein. *Proc. Natl. Acad. Sci. U. S. A. 107*, 4099−4104.
(37) 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: Struct., Funct., Genet. 79*, 830−838.
(38) Capriotti, E., Fariselli, P., and Casadio, R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res. 33*, W306−310.
(39) Pires, D. E. V., Ascher, D. B., and Blundell, T. L. (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics 30*, 335−342.
(40) 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.
(41) Pires, D. E. V., Ascher, D. B., and Blundell, T. L. (2014) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res. 42*, W314−W319.
(42) Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P. (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One 7*, e46688.
(43) Folkman, L., Stantic, B., Sattar, A., and Zhou, Y. (2016) EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. *J. Mol. Biol. 428*, 1394−1405.
(44) Zhou, Y., and Bowie, J. U. (2000) Building a thermostable membrane protein. *J. Biol. Chem. 275*, 6975−6979.
(45) Khatun, J., Khare, S. D., and Dokholyan, N. V. (2004) Can contact potentials reliably predict stability of proteins? *J. Mol. Biol. 336*, 1223−1238.
(46) Kumar, M. D., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H., and Sarai, A. (2006) ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res. 34*, D204−D206.

DOI: 10.1021/acs.biochem.6b00537
Biochemistry 2016, 55, 5002−5009