JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 1508−1514

pubs.acs.org/jcim

DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks

Huali Cao,[†] Jingxue Wang,[†] Liping He,[†] Yifei Qi,*[,†,‡] and John Z. Zhang*[,†,‡,§]

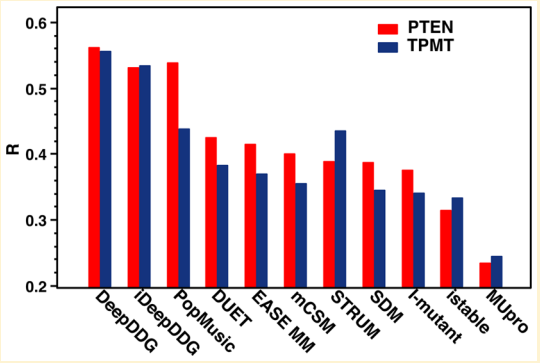
[†]Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China

[‡]NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

[§]Department of Chemistry, New York University, New York, New York 10003, United States

Supporting Information

ABSTRACT: Accurately predicting changes in protein stability due to mutations is important for protein engineering and for understanding the functional consequences of missense mutations in proteins. We have developed DeepDDG, a neural network-based method, for use in the prediction of changes in the stability of proteins due to point mutations. The neural network was trained on more than 5700 manually curated experimental data points and was able to obtain a Pearson correlation coefficient of 0.48−0.56 for three independent test sets, which outperformed 11 other methods. Detailed analysis of the input features shows that the solvent accessible surface area of the mutated residue is the most important feature, which suggests that the buried hydrophobic area is the major determinant of protein stability. We expect this method to be useful for large-scale design and engineering of protein stability. The neural network is freely available to academic users at http://protein.org.cn/ddg.html.



INTRODUCTION

Proteins are essential molecules in most biological processes in cells. Their correct functioning relies on intact three-dimensional structures and, thus, their stability within the cellular environment. Accurate prediction of changes in protein stability resulting from mutations not only provides critical insights into how proteins fold but also has significant applications for bioindustry.[1] Many methods have been developed to this end,[2−17] including machine learning,[3,7,8,11] scoring-function and potentials,[2,6,13,18,19] and physical-based methods.[4,12] Notably, integrated methods that make predictions using outputs from other methods as inputs show better performance and have become increasingly popular in recent years.[10,15,16] Nonetheless, a recent assessment of various methods showed that the accuracy of their predictions is still unsatisfactory.[20]

The deep neural network technique has been advancing rapidly in computational biology and chemistry in recent years due to theoretical advancements, the accumulation of a large amount of data, and the development of new computer hardware, which has pushed the limits of many prediction tasks.[21,22] Successful examples include the scoring of protein−ligand interactions,[23,24] the prediction of protein secondary structure[25,26] and contact maps,[27−29] and prediction of compound toxicity and liver injury,[30−32] among many others. Recently, we have developed a neural network for computational protein design that predicts the probability of all 20 natural amino acids for each residue in a given protein backbone structure (available at http://www.protein.org.cn/cpd.html).[33] In this neural network, the network parameters are shared for each target residue-neighbor residue pair, similarly to a convolutional neural network in which the network for each subset within the input image shares the same set of weights. We named this network structure the "shared residue pair (SRP) network". The encouraging performance of this network structure in protein design makes us believe that it can be applied to the prediction of other residue-specific properties as well.

In addition to the network structure, the key to training an accurate neural network is a large amount of data. Unfortunately, in biology and chemistry, such experimental data is, in many cases, limited and expensive to obtain. So far, a few collections of thermodynamic data in relation to protein mutations have been published. Among them, the Protherm database,[34] which was updated as recently as 2013, remains the most comprehensive and contains more than 25,000 thermodynamic data.[34]

In this study, we have developed a novel method to predict changes in protein stability caused by point mutations using the SRP network and a large set of experimental data. We have manually gathered and curated, to our best knowledge, the

Received: October 7, 2018
Published: February 14, 2019

DOI: 10.1021/acs.jcim.8b00697
J. Chem. Inf. Model. 2019, 59, 1508−1514

largest data set of experimentally determined stability data from various sources and adapted the SRP network to predict stability changes based on the three-dimensional structure of a protein. We examined whether integrating predictions produced using other methods within the neural network would improve its performance. The importance of each input feature to the neural network was also evaluated. Our method outperforms 11 other methods in terms of linear correlation with the experimental data based on results derived from three independent test sets. The network is freely available to academic users at http://protein.org.cn/ddg.html.

METHODS

Data Collection. The thermodynamic data were collected from the Protherm database,[34] two previously published data sets,[14,35] and a literature search. Since we found many errors in the Protherm database (for example, the sign used for the $\Delta\Delta G$ value was incorrect in some items), all data were manually compared to the original reference to ensure their correctness. $\Delta\Delta G$ values from all experimental methods were accepted if a wild-type structure for the protein was available. Each data point consists of the protein name, source, mutation, pH, temperature, reference (PubMed ID, if available), Protein Data Bank (PDB) ID, and UniProt ID. All collected data are available in the Supporting Information.

Neural Network Training and Input Feature Calculation. The neural network was built and trained using the Keras library with the TensorFlow backend. Training was performed for 100 epochs with a learning rate of 0.001 and a batch size of 1000. To reduce overfitting, a dropout value of 0.15 and a L2 kernel regulation value of 8×10^{-4} were used. The loss function used was the mean squared-error, which was minimized using the Adam method. Rectified Linear Unit (ReLu) activation was used for all layers except the output layer, where a softsign function was used. Since the output of a softsign function is between −1 and 1, the experimental $\Delta\Delta G$ values were normalized using a multiplier of 10 during training. During testing and validation, the output of the network was multiplied by 10 to make the final prediction. To enforce the symmetry of the network output, each mutation and the sign of the corresponding $\Delta\Delta G$ value were additionally reversed during training. These reversed mutations were not included during testing and validation.

The solvent accessible surface area, secondary structures, and hydrogen bonds were calculated using Naccess,[36] Stride,[37] and HBPLUS.[38] The position-specific scoring matrix (PSSM) was calculated using PSI-BLAST 2.7.1[39] with 3 iterations based on the rp-seq-55 sequence set. The original PSSM value for amino acid i for a particular residue was transformed using the following softmax function

$$P_{i,\mathrm{new}} = \exp(P_{i,\mathrm{orig}}) \Big/ \sum_i \exp(P_{i,\mathrm{orig}})$$

where the summation is made for all 20 amino acid types. The PSSM value indicates the independent conservation of the amino acid at each residue. To quantify the "fitness" between the target residue and its neighbor residues, we defined a simple feature as the pairwise fitness score (PFS) that is based on a multiple sequence alignment of the target protein

$$\mathrm{PFS} = P_{t,i,n,j} \Big/ P_{t,i} \times P_{n,j}$$

where $P_{t,i,n,j}$ represents the joint probability of amino acid types i and j at the target residue t and neighboring residue n in the alignment, respectively, $P_{t,i}$ is the probability of amino acid type i at the target residue t, and $P_{n,j}$ is the probability of amino acid type j at the neighboring residue n. The PFS was transformed using a softsign function to map the value range to $[0, 1]$. Multiple sequence alignment was performed using HHsuite3.0[40] based on the uniprot20 sequence set. All other features were calculated using in-house scripts.

RESULTS AND DISCUSSION

Collection of Experimental Data. We first set out to collect a reliable data set of experimentally determined folding free-energy changes due to point mutations in proteins that would be as large as possible. Multiple resources were used, including the Protherm database[34] and two published data sets.[14,35] Studies published after 1990 were also searched using Google Scholar to obtain additional data. Since the $\Delta\Delta G$ values from Protherm could not be used directly (some values had the wrong sign, while some represented the value between the folded/unfolded and the intermediate state), we manually compared all collected data to the original references. The final data set contained 5,766 $\Delta\Delta G$ values from 242 proteins collected from ∼600 studies, which was much larger than previously collected data sets such as the S2648 data set.[14] Our data set also included 46 mutations whose $\Delta\Delta G$ values could not be determined based on the original references and were not used in this study. The distribution of the $\Delta\Delta G$ values ranged mostly between −10 and 10 kcal/mol and was biased toward negative values (destabilizing mutations, Figure 1).

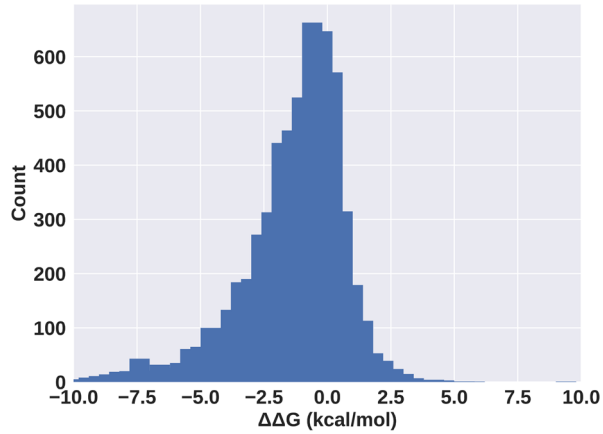


Figure 1. Distribution of the $\Delta\Delta G$ values in our data set. Negative values denote destabilizing mutations.

Thirty-seven of the 242 proteins that are not homologous (sequence identity ≤25%) to proteins in the S2648 set or the remaining proteins in our data set were extracted as an independent test set (Table S1). This test set has 276 mutations and was not used for any training/validation process in this study. The remaining mutations were separated into 10 folds for a 10-fold cross-validation. The separation was residue-based, which ensured that two mutations that corresponded to the same site in two homologous proteins (sequence identity ≥25%) were always in the same fold. The entire data set is available in the Supporting Information.

Neural Network Training and Performance. We next adapted the SRP network to predict $\Delta\Delta G$ values for mutations

DOI: 10.1021/acs.jcim.8b00697
J. Chem. Inf. Model. 2019, 59, 1508−1514

(Figure 2). The mutated residue and each of its N closest neighboring residues (based on the C_α−C_α distance) forms a

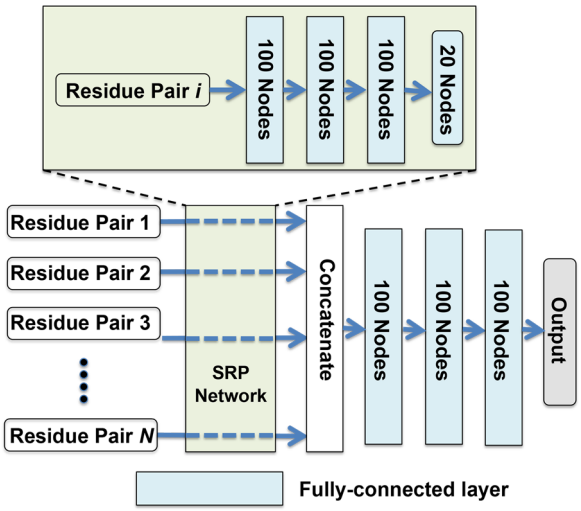


Figure 2. Structure of the full neural network.

residue pair that is the basic input for the SRP network. The number of neighboring residues N was set to 15, which demonstrated the best performance for protein design in our previous study. The weights for the SRP network are shared among all residue pairs, similarly to that in the convolutional network. The features for one residue pair include the following: (1) geometric features such as the backbone dihedrals, solvent accessible surface area, secondary structures, number of hydrogen bonds, distance and orientation between the mutated residue and the neighboring residue; (2) sequence features such as the PSSM value derived from a PSI-BLAST search, the pairwise fitness score (PFS; see Methods) from the multiple sequence alignment, and the probability of the wild-type and mutant amino acids for the target residue based on our protein design neural network; and (3) general features such as the amino acid types of the target and neighboring residues. The details of these features are listed in Table S2.

Since DeepDDG does not rely on outputs from other methods, we first compared it with eight nonintegrated methods (Table 1). DeepDDG shows a linear correlation of 0.681, 0.557, and 0.658 based on 10-fold cross-validation of the training set, the independent test set, and the test set with 5% of the outliers removed, respectively, which is higher than that obtained using other methods ($p < 0.01$, z-test). It also has a comparable or slightly smaller mean absolute error (MAE) than the other methods. The slopes obtained during linear fitting between the experimental and predicted $\Delta\Delta G$ values were 0.99 and 0.67 for the training and test sets, respectively (Figure 3, Table S3). Notably, almost all tested methods demonstrate a higher correlation for the training set than the test set, which indicates some degree of overfitting. For example, STRUM demonstrated a correlation of 0.715 for the training set but only 0.445 for the test set. When the training set is not large enough, such overfitting is inevitable. We have used techniques such as dropout and L2 regulation to reduce overtraining of our network, but there is still a gap between the performance of the training and test sets. With greater accumulation of experimental data, it is hoped that a more generalized neural network could be trained and the performance gap could be reduced.

The integration method uses outputs from other methods as inputs and has become a popular choice for many learning tasks in recently years. To test whether the accuracy of DeepDDG can be improved using the integration method, we fed the outputs of three servers, mCSM, SDM, and DUET, into the concatenation layer of our network. It should be noted that the test proteins were not used for the development of SDM, mCSM, and DUET, which rules out the possibility that the test set was exposed to the network. This integrated network, named iDeepDDG, is compared to three integrated methods, iStable,[10] DynaMut,[15] and DUET[16] in Table 2. iDeepDDG demonstrates R values of 0.718, 0.563, and 0.687 for the training set, test set, and the test set with 5% of the outliers removed, respectively, which is comparable to that obtained from the nonintegrated method DeepDDG but higher than that from other integrated methods. The slope obtained from the linear fitting between the experimental and predicted $\Delta\Delta G$ values is 1.11 for the training and 0.75 for the test set (Figure 4).

Importance of the Input Features. Since our networks use a variety of input features (Table S2), it is necessary to analyze the importance of each input feature, which may provide insights regarding the factors that determine the change in stability due to a mutation. To this end, we used a simple approach that involves the removal of each feature on a one-by-one basis and the re-evaluation of the performance based on the training and test sets (Table 3). Note that the amino acid types of the target and neighboring residues are not analyzed here. Of the eight feature categories that were examined, the residue solvent accessible surface area had the largest effect on the correlation and MAE of the experimental

Table 1. Comparison of DeepDDG with Other Nonintegrated Methods

	training set		test set		test set (5% of outliers removed)	
methods	R	MAE (kcal/mol)	R	MAE (kcal/mol)	R	MAE (kcal/mol)
SCooP[41]	NA[a]	NA[a]	0.078	6.44	0.180	4.35
MUpro1.1[3]	0.551	1.10	0.190	1.06	0.310	0.93
EASE-MM[42]	0.607	1.16	0.402	0.91	0.591	0.76
PopMusic[6]	NA[a]	NA[a]	0.443	0.91	0.619	0.77
STRUM[43]	0.715	1.02	0.447	0.88	0.518	0.77
I-Mutant3.0[11]	0.565	1.17	0.453	0.91	0.538	0.79
mCSM[9]	0.630	1.14	0.467	0.90	0.572	0.78
SDM[2]	0.491	1.41	0.483	1.02	0.604	0.89
DeepDDG	0.681 ± 0.019	1.04 ± 0.02	0.557 ± 0.015	0.86 ± 0.04	0.658 ± 0.014	0.74 ± 0.04

[a]Not evaluated using the training set due to technical issues.

DOI: 10.1021/acs.jcim.8b00697
J. Chem. Inf. Model. 2019, 59, 1508−1514

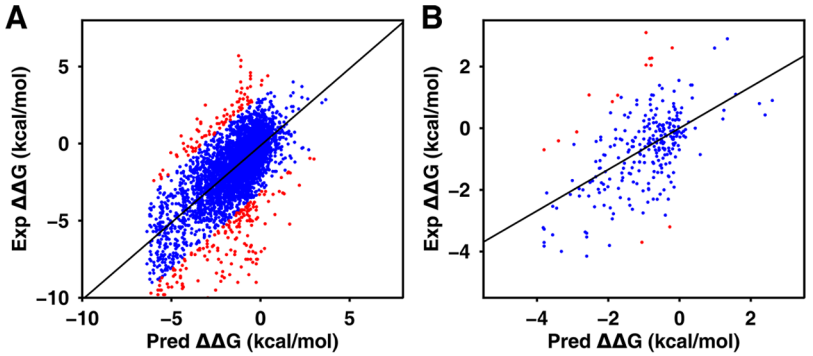


Figure 3. Comparison of the experimental and predicted $\Delta\Delta G$ values obtained from the DeepDDG network. (A) 10-fold cross-validation for the training set. (B) Test set. The outliers (5%) are colored red, and the black lines represent the linear fitting of all data.

Table 2. Comparison of iDeepDDG to Other Integrated Methods

	training set		test set		test set (5% outliers removed)	
methods	*R*	MAE (kcal/mol)	*R*	MAE (kcal/mol)	*R*	MAE (kcal/mol)
DynaMut[15]	NA[a]	NA[a]	0.393	0.97	0.543	0.84
iStable[10]	0.557	1.13	0.427	0.89	0.541	0.78
DUET[16]	0.643	1.11	0.515	0.87	0.636	0.75
iDeepDDG	0.718 ± 0.018	1.00 ± 0.06	0.563 ± 0.015	0.80 ± 0.05	0.687 ± 0.015	0.68 ± 0.06

[a]Not evaluated for the training set due to technical issues.

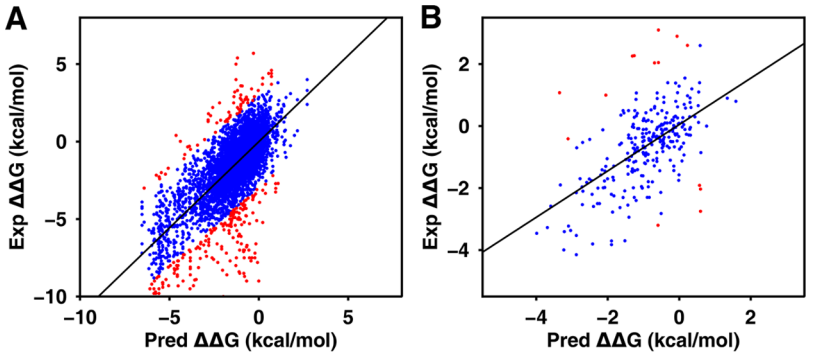


Figure 4. Comparison of experimental and predicted $\Delta\Delta G$ values obtained using the iDeepDDG network. (A) 10-fold cross-validation of the training set. (B) Test set. The 5% outliers are colored red, and the black lines represent the linear fit for all data.

Table 3. Correlation and MAE after Removing Input Features from the DeepDDG Network

	training set		test set	
removed feature category	*R*	MAE (kcal/mol)	*R*	MAE (kcal/mol)
none	0.681	1.04	0.557	0.86
distance and orientation of the neighboring residue	0.670	1.04	0.555	0.86
backbone dihedral	0.671	1.04	0.553	0.84
PSSM	0.661	1.07	0.551	0.88
hydrogen bond	0.677	1.05	0.550	0.86
secondary structure	0.673	1.06	0.550	0.86
protein design probability	0.658	1.08	0.543	0.86
PFS	0.664	1.06	0.539	0.87
residue solvent accessible surface area	0.656	1.08	0.518	0.88

data. This is understandable because mutation of buried residues usually reduces the stability of a protein, and therefore, the surface area is strongly correlated with the change in stability. In addition to surface area, the probability calculated by our protein design network and PFS are the most important features. These three features are derived from large-scale structural and sequence alignment analysis, suggesting that the inclusion of evolutionary information for the protein structure and sequence is crucial to determining the change in stability. Surprisingly, simple structural properties such as the backbone dihedrals and distance and orientation of the neighboring residue show minimal effects on accuracy. This is likely because our network uses a static protein structure and does not take the structural rearrangement caused by the mutation into account. Nonetheless, to ensure compatibility with future developments, these features are still included in the networks.

Independent Testing of the Effects of Mutations on Melting Temperature. To facilitate additional independent testing of our networks, we collected 173 experimental melting temperature changes (ΔT_m) caused by mutations in six proteins in the Protherm database and the literature (available in the Supporting Information). These proteins were selected based on the number of mutations (>10) and sequence identity when compared to proteins in our training set. Five of them are not homologous (sequence identity <30%) to the

DOI: 10.1021/acs.jcim.8b00697
J. Chem. Inf. Model. 2019, 59, 1508−1514

Table 4. Linear Correlation Coefficients Obtained for Comparisons between Experimental ΔT_m and Predicted $\Delta\Delta G$ Values Using Various Methods for Six Test Proteins

method	1AQH	1H8V	1OSI	1XAS	2FJF	GK[a]	average
DeepDDG	**0.688**	0.163	0.253	**0.562**	**0.759**	**0.500**	**0.488**
iDeepDDG	−0.140	0.131	0.347	**0.476**	**0.754**	0.456	**0.337**
STRUM	0.240	0.091	0.306	0.344	0.620	0.392	0.332
EASE_MM	0.172	**0.309**	**0.375**	−0.065	0.528	**0.551**	0.312
SDM	−0.058	0.301	0.328	0.099	0.579	0.328	0.263
PopMusic	0.081	0.118	**0.442**	0.032	0.621	*b*	0.259
DUET	−0.434	**0.329**	0.360	0.410	0.393	0.374	0.239
iStable	**0.342**	−0.148	0.289	0.368	0.237	*b*	0.218
mCSM	−0.549	0.227	0.325	0.335	0.340	0.344	0.170
I-Mutant3.0	−0.289	0.031	0.321	0.300	0.248	0.304	0.153
MUpro1.1	0.313	0.123	0.183	−0.473	0.302	0.334	0.130
DynaMut	−0.634	0.217	0.235	0.122	0.289	0.332	0.094
no of. exp data	11	16	24	12	14	96	

[a]Guanylate kinase. [b]Not determined because the GK structure could not be uploaded.

training proteins (Table S4); one of them, 2FJF, has one chain that is homologous to eight of the training proteins. We performed multiple sequence alignment between this chain and the eight proteins to ensure that the included mutations do not have homologous mutations in the training set (solid boxes in Figure S1). In addition, the structure of guanylate kinase (GK) from *Branchiostoma floridae* (UniProt: C3YEM4_BRAFL) was modeled by Modeler 9.20[44] using the mouse GK structure (PDB ID 1LVG[45]) as a template. A protein-dependent linear relationship between T_m and ΔG has been observed,[46] which allows the correlation between the predicted $\Delta\Delta G$ value and the experimental ΔT_m to serve as an indicator of the accuracy of computational methods. The results of the comparison between DeepDDG and iDeepDDG and ten other methods are shown in Table 4. DeepDDG produces the highest correlation for three of the six proteins and yields an average correlation of 0.488, which is much better than that obtained using the other methods. iDeepDDG yields an average correlation of 0.337, which is comparable to that of STRUM. The reduction in correlation is likely because of the great effects of the input from SDM, DUET, and mCSM. Moreover, the correlation produced by all methods varies greatly depending upon the protein. Such large variation is probably due to bias stemming from the limited number of proteins used in the training set and highlights the challenges of accurately predicting the change in stability.

Independent Testing of the PTEN and TPMT Data Sets from CAGI. For the third independent test, we obtained two data sets for the phosphatase and tensin homologue (PTEN) and thiopurine S-methyl transferase (TPMT) proteins from the Critical Assessment of Genome Interpretation 5 challenge (CAGI 5, https://genomeinterpretation.org/content/predict-effect-missense-mutations-pten-and-tpmt-protein-stability). In this challenge, a deep mutational scan[47] was performed on both proteins during which thousands of mutations were assessed in parallel, and the steady state abundance of each mutation was transformed into a stability score, for which 0 denoted instability, 1 represented stability similar to that of the wild-type, and >1 denoted greater stability than the wild-type. Since the full-length PTEN structure is not available in the PDB, we used the crystal structure 5BUG, which contains coordinates for residues 14−351 and has a highest sequence identity of 23.3% to the training set proteins. For TPMT, we used the structure 2BZG, which has a maximum sequence identity of 23.1% to the proteins in the training set. As we were not able to obtain predictions from the DynaMut server in a reasonable amount of time, it was not included in the comparison. Figure 5 shows the correlation

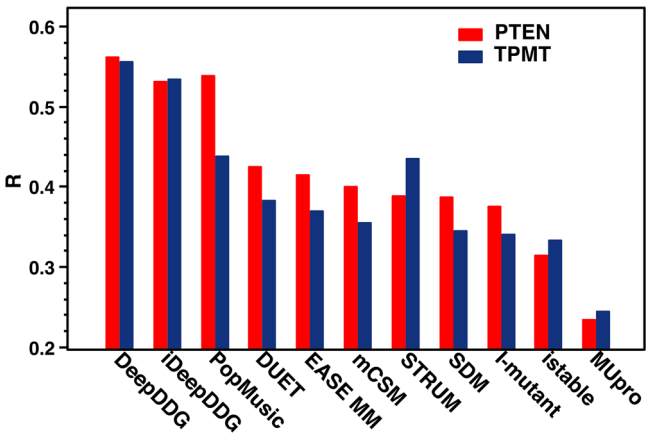


Figure 5. Correlations between predicted $\Delta\Delta G$ values obtained using various methods and the stability scores for PTEN and TPMT.

between the $\Delta\Delta G$ values obtained using various methods and the stability scores for the two proteins. DeepDDG produced a correlation of ∼0.55 for the two data sets, and iDeepDDG produced a correlation of ∼0.53. For the other methods, only PopMusic produced a comparable correlation for the PTEN data set; all other correlations were below 0.45. In addition, we have calculated the Pearson' correlation on the stabilizing and destabilizing mutations from the two proteins (Table S5). Unfortunately, all methods demonstrate much lower correlation on the stabilizing mutations than that on the destabilizing ones, which is likely caused by the unbalanced data (many more destabilizing mutations than stabilizing ones) in the training set of these methods. Nonetheless, considering the sizes of the two data sets (∼2900 mutations for PTEN and ∼3300 for TPMT) and the low sequence identity of the two test proteins to our training set proteins, this independent test inarguably demonstrates that our methods are more accurate than the other tested methods.

DOI: 10.1021/acs.jcim.8b00697
J. Chem. Inf. Model. 2019, 59, 1508−1514

CONCLUSIONS

In this study, we have developed a neural network-based method that can be used to predict changes in protein stability due to mutations. The improved accuracy of our method can be attributed mostly to the use of a large collection of experimental data and the power of the neural network. Although our method shows higher accuracy than other methods for the three test sets, the correlation values obtained upon comparison with experimental data suggest that there is room for improvement. Since our method (as well as many other methods) only takes into account the local environment of the mutation, the incorporation of the global sequence and structural information for the protein is a possible strategy for improvement. Considering the batch-processing capability of neural networks, we expect that the DeepDDG/iDeepDDG method will be a useful tool for large-scale engineering to improve protein stability.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00697.

Figure S1, multiple sequence alignment of L chain of 2FJF and homologous chains within training set; Table S1, highest sequence identity for each test protein upon comparison to training proteins; Table S2, features used by neural network for target and neighboring residues; Table S3, correlation, slope, and intercept obtained during linear fitting of experimental and calculated $\Delta\Delta G$ values for test set; Table S4, highest sequence identity of each protein in ΔT_m set to training proteins; Table S5, Pearson's correlation between predicted $\Delta\Delta G$ values and experimental stabilities on stabilizing and destabilizing mutations from PTEN and TPMT data sets (PDF)

Data set of collected experimental $\Delta\Delta G$ values; Protein name, source, mutation, pH, temperature, reference (PubMed ID, if available), Protein Data Bank (PDB) ID, and UniProt ID for each data point (XLSX)

Data set of collected ΔT_m values (XLSX)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: yfqi@chem.ecnu.edu.cn.

*E-mail: John.Zhang@nyu.edu.

ORCID

Yifei Qi: 0000-0003-2853-7910

John Z. Zhang: 0000-0003-4612-1863

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China (grant no. 2016YFA0501700), the National Natural Science Foundation of China (grant nos. 31700646, 21433004, 91753103), Shanghai Putuo District (grant 2014-A-02), the Innovation Program of the Shanghai Municipal Education Commission (201701070005E00020), and the NYU Global Seed Grant. We thank the Supercomputer Center of East China Normal University for providing us with computer time.

REFERENCES

(1) Magliery, T. J. Protein Stability: Computation, Sequence Statistics, and New Experimental Methods. *Curr. Opin. Struct. Biol.* **2015**, *33*, 161−168.

(2) Worth, C. L.; Preissner, R.; Blundell, T. L. Sdm–a Server for Predicting Effects of Mutations on Protein Stability and Malfunction. *Nucleic Acids Res.* **2011**, *39*, W215−222.

(3) Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins: Struct., Funct., Genet.* **2006**, *62*, 1125−1132.

(4) Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability-Large-Scale Validation of Md-Based Relative Free Energy Calculations. *J. Mol. Biol.* **2017**, *429*, 948−963.

(5) Zhang, Z.; Wang, L.; Gao, Y.; Zhang, J.; Zhenirovskyy, M.; Alexov, E. Predicting Folding Free Energy Changes Upon Single Point Mutations. *Bioinformatics* **2012**, *28*, 664−671.

(6) Dehouck, Y.; Kwasigroch, J. M.; Gilis, D.; Rooman, M. Popmusic 2.1: A Web Server for the Estimation of Protein Stability Changes Upon Mutation and Sequence Optimality. *BMC Bioinf.* **2011**, *12*, 151.

(7) Capriotti, E.; Fariselli, P.; Casadio, R. A Neural-Network-Based Method for Predicting Protein Stability Changes Upon Single Point Mutations. *Bioinformatics* **2004**, *20* (Suppl 1), i63−i68.

(8) Jokinen, E.; Heinonen, M.; Lahdesmaki, H. Mgpfusion: Predicting Protein Stability Changes with Gaussian Process Kernel Learning and Data Fusion. *Bioinformatics* **2018**, *34*, i274−i283.

(9) Pires, D. E.; Ascher, D. B.; Blundell, T. L. Mcsm: Predicting the Effects of Mutations in Proteins Using Graph-Based Signatures. *Bioinformatics* **2014**, *30*, 335−342.

(10) Chen, C. W.; Lin, J.; Chu, Y. W. Istable: Off-the-Shelf Predictor Integration for Predicting Protein Stability Changes. *BMC Bioinformatics* **2013**, *14* (Suppl 2), S5.

(11) Capriotti, E.; Fariselli, P.; Casadio, R. I-Mutant2.0: Predicting Stability Changes Upon Mutation from the Protein Sequence or Structure. *Nucleic Acids Res.* **2005**, *33*, W306−310.

(12) Steinbrecher, T.; Abel, R.; Clark, A.; Friesner, R. Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations. *J. Mol. Biol.* **2017**, *429*, 923−929.

(13) Schymkowitz, J.; Borg, J.; Stricher, F.; Nys, R.; Rousseau, F.; Serrano, L. The Foldx Web Server: An Online Force Field. *Nucleic Acids Res.* **2005**, *33*, W382−388.

(14) Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and Accurate Predictions of Protein Stability Changes Upon Mutations Using Statistical Potentials and Neural Networks: Popmusic-2.0. *Bioinformatics* **2009**, *25*, 2537−2543.

(15) Rodrigues, C. H.; Pires, D. E.; Ascher, D. B. Dynamut: Predicting the Impact of Mutations on Protein Conformation, Flexibility and Stability. *Nucleic Acids Res.* **2018**, *46*, W350−W355.

(16) Pires, D. E.; Ascher, D. B.; Blundell, T. L. Duet: A Server for Predicting Effects of Mutations on Protein Stability Using an Integrated Computational Approach. *Nucleic Acids Res.* **2014**, *42*, W314−319.

(17) Masso, M.; Vaisman, I. I. Auto-Mute: Web-Based Tools for Predicting Stability Changes in Proteins Due to Single Amino Acid Replacements. *Protein Eng., Des. Sel.* **2010**, *23*, 683−687.

(18) Parthiban, V.; Gromiha, M. M.; Schomburg, D. Cupsat: Prediction of Protein Stability Upon Point Mutations. *Nucleic Acids Res.* **2006**, *34*, W239−242.

(19) Zhou, H.; Zhou, Y. Distance-Scaled, Finite Ideal-Gas Reference State Improves Structure-Derived Potentials of Mean Force for Structure Selection and Stability Prediction. *Protein Sci.* **2002**, *11*, 2714−2726.

(20) Pucci, F.; Bernaerts, K.; Kwasigroch, J. M.; Rooman, M. Quantification of Biases in Predictions of Protein Stability Changes Upon Mutations. *Bioinformatics* **2018**, *34*, 3659−3665.

(21) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436−444.

(22) Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep Learning for Computational Chemistry. *J. Comput. Chem.* **2017**, *38*, 1291−1307.
(23) Ragoza, M.; Hochuli, J.; Idrobo, E.; Sunseri, J.; Koes, D. R. Protein-Ligand Scoring with Convolutional Neural Networks. *J. Chem. Inf. Model.* **2017**, *57*, 942−957.
(24) Gomes, J.; Ramsundar, B.; Feinberg, E. N.; Pande, V. S. Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity. 2017, arXiv:1703.10603, ArXiv e-prints, https://arxiv.org/abs/1703.10603 (accessed Feb 19, 2019).
(25) Wang, S.; Peng, J.; Ma, J.; Xu, J. Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Sci. Rep.* **2016**, *6*, 18962.
(26) Heffernan, R.; Paliwal, K.; Lyons, J.; Dehzangi, A.; Sharma, A.; Wang, J.; Sattar, A.; Yang, Y.; Zhou, Y. Improving Prediction of Secondary Structure, Local Backbone Angles, and Solvent Accessible Surface Area of Proteins by Iterative Deep Learning. *Sci. Rep.* **2015**, *5*, 11476.
(27) Eickholt, J.; Cheng, J. Predicting Protein Residue-Residue Contacts Using Deep Networks and Boosting. *Bioinformatics* **2012**, *28*, 3066−3072.
(28) Di Lena, P.; Nagata, K.; Baldi, P. Deep Architectures for Protein Contact Map Prediction. *Bioinformatics* **2012**, *28*, 2449−2457.
(29) Wang, S.; Sun, S.; Li, Z.; Zhang, R.; Xu, J. Accurate De Novo Prediction of Protein Contact Map by Ultra-Deep Learning Model. *PLoS Comput. Biol.* **2017**, *13*, No. e1005324.
(30) Xu, Y.; Dai, Z.; Chen, F.; Gao, S.; Pei, J.; Lai, L. Deep Learning for Drug-Induced Liver Injury. *J. Chem. Inf. Model.* **2015**, *55*, 2085−2093.
(31) Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. Deeptox: Toxicity Prediction Using Deep Learning. *Front. Environ. Sci.* **2016**, *3*, 80.
(32) Unterthiner, T.; Mayr, A.; Klambauer, G.; Hochreiter, S. Toxicity Prediction Using Deep Learning. 2015, arXiv:1503.01445, ArXiv e-prints, https://arxiv.org/abs/1503.01445 (accessed Feb 19, 2019).
(33) Wang, J.; Cao, H.; Zhang, J. Z. H.; Qi, Y. Computational Protein Design with Deep Learning Neural Networks. *Sci. Rep.* **2018**, *8*, 6349.
(34) Bava, K. A.; Gromiha, M. M.; Uedaira, H.; Kitajima, K.; Sarai, A. Protherm, Version 4.0: Thermodynamic Database for Proteins and Mutants. *Nucleic Acids Res.* **2004**, *32*, D120−D121.
(35) Pucci, F.; Bourgeas, R.; Rooman, M. High-Quality Thermodynamic Data on the Stability Changes of Proteins Upon Single-Site Mutations. *J. Phys. Chem. Ref. Data* **2016**, *45*, 023104.
(36) Hubbard, S. J.; Thornton, J. M. *Naccess*, Computer Program; Department of Biochemistry and Molecular Biology, University College London, 1993.
(37) Frishman, D.; Argos, P. Knowledge-Based Protein Secondary Structure Assignment. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 566−579.
(38) McDonald, I. K.; Thornton, J. M. Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol.* **1994**, *238*, 777−793.
(39) Schaffer, A. A.; Aravind, L.; Madden, T. L.; Shavirin, S.; Spouge, J. L.; Wolf, Y. I.; Koonin, E. V.; Altschul, S. F. Improving the Accuracy of Psi-Blast Protein Database Searches with Composition-Based Statistics and Other Refinements. *Nucleic Acids Res.* **2001**, *29*, 2994−3005.
(40) Remmert, M.; Biegert, A.; Hauser, A.; Soding, J. Hhblits: Lightning-Fast Iterative Protein Sequence Searching by Hmm-Hmm Alignment. *Nat. Methods* **2012**, *9*, 173−175.
(41) Pucci, F.; Kwasigroch, J. M.; Rooman, M. Scoop: An Accurate and Fast Predictor of Protein Stability Curves as a Function of Temperature. *Bioinformatics* **2017**, *33*, 3415−3422.
(42) Folkman, L.; Stantic, B.; Sattar, A.; Zhou, Y. Ease-Mm: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. *J. Mol. Biol.* **2016**, *428*, 1394−1405.
(43) Quan, L.; Lv, Q.; Zhang, Y. Strum: Structure-Based Prediction of Protein Stability Changes Upon Single-Point Mutation. *Bioinformatics* **2016**, *32*, 2936−2946.
(44) Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.
(45) Sekulic, N.; Shuvalova, L.; Spangenberg, O.; Konrad, M.; Lavie, A. Structural Characterization of the Closed Conformation of Mouse Guanylate Kinase. *J. Biol. Chem.* **2002**, *277*, 30236−30243.
(46) Watson, M. D.; Monroe, J.; Raleigh, D. P. Size-Dependent Relationships between Protein Stability and Thermal Unfolding Temperature Have Important Implications for Analysis of Protein Energetics and High-Throughput Assays of Protein-Ligand Interactions. *J. Phys. Chem. B* **2018**, *122*, 5278−5285.
(47) Fowler, D. M.; Fields, S. Deep Mutational Scanning: A New Style of Protein Science. *Nat. Methods* **2014**, *11*, 801−807.

DOI: 10.1021/acs.jcim.8b00697
J. Chem. Inf. Model. 2019, 59, 1508−1514