

## PremPS: Predicting the Effects of Single Mutations on Protein Stability

Yuting Chen<sup>1</sup>, Haoyu Lu<sup>1</sup>, Ning Zhang<sup>1</sup>, Zefeng Zhu<sup>1</sup>, Shuqin Wang<sup>1</sup> and Minghui Li<sup>1,\*</sup>

<sup>1</sup>Center for Systems Biology, Department of Bioinformatics, School of Biology and Basic Medical Sciences, Soochow University, Suzhou, 215123, China

\* - corresponding author, minghui.li@suda.edu.cn

## ABSTRACT

Protein stability is related to its functional activities, and effect on stability or misfolding could be one of the major disease-causing mechanisms of missense mutations. Here we developed a novel machine learning computational method PremPS, which predicts the effects of single mutations on protein stability by calculating the changes in unfolding Gibbs free energy. PremPS uses only ten evolutionary- and structure-based features and is parameterized on five thousand mutations. Our approach outperforms previous methods and shows a considerable improvement in estimating the effects of mutations increasing protein stability. In addition, PremPS presents an outstanding performance in predicting the pathogenicity of missense mutations using an experimental dataset composed of two thousand non-neutral and neutral mutations. PremPS can be applied to many tasks, including finding functionally important variants, revealing the molecular mechanisms of functional influences and protein design. It is freely available at <https://lilab.jysw.suda.edu.cn/research/PremPS/>.

**Key words:** mutation; protein stability; feature selection; random forest regression algorithm

### Key Points:

- Considerable improvement in estimating the effects of mutations increasing protein stability;
- Comprehensive comparison with other 25 computational methods on different test sets;
- An outstanding performance in predicting the pathogenicity of missense mutations;
- PremPS employs only ten distinct features belonging to six categories, and the most important feature describes evolutionary conservation of the site;
- The webserver allows to do large-scale mutational scanning and takes about ten minutes to perform calculations for one thousand mutations from a normal size protein.

## INTRODUCTION

Protein stability is one of the most important factors that characterize protein function, activity and regulation (1). Missense mutations can lead to protein dysfunction by affecting their stabilities and interactions with other biological molecules (2-6). Several studies have shown that the mutations are deleterious due to decreasing/enhancing the stability of the corresponding protein (7-12). To quantify the effects on protein stability requires estimating the changes in folding/unfolding Gibbs free energy induced by mutations. Experimental measurements of protein stability changes are laborious and appropriate only for proteins that can be purified (13). Therefore, computational prediction is urgently required, which would help the prioritization of potentially functionally important variants and become vital to many fields, such as medical applications (14) and protein design (15).

A lot of computational approaches have been developed in the last decades to predict the effects of single mutations on protein stability (16-42). The vast majority of them are machine learning approaches and based on protein 3D structures. They are different in terms of algorithms used for building models, structural optimization procedures or features of energy functions. The prediction performances of these methods have been assessed and compared using several different datasets of experimentally characterized mutants (11,43-46). The results indicate that all methods showed a correct trend in the predictions but with inconsistent performances for different test sets. A majority of methods presented moderate or low accuracies when applied to the independent test sets, and PoPMuSiC (16), FoldX (23) and mCSM (17) showed relatively better performances in comparison with other methods.

The machine learning approaches are prone to have overfitting problems (47), namely their predictions tend to be biased towards the characteristics of learning datasets. The training data sets available so far with experimentally determined protein stability changes are enriched with destabilizing mutations (16,48). Thus, the vast majority of predictors that did not consider the unbalance of training dataset showed a better performance for predicting destabilizing than stabilizing mutations (49,50). A recent study constructed a balanced data set with an equal number of destabilizing and stabilizing mutations and was used to assess the performance of 14 methods (51). The results showed that the 14 widely used predictors present a strong bias toward predicting the destabilizing mutations. Correcting for such bias in method's performance is not a trivial task,

which requires enriching the training set with stabilizing mutations and developing new energy functions.

To address this issue, we developed PremPS that uses a novel scoring function composed of only ten features and trains on a balanced dataset including five thousand mutations, half of which belong to destabilizing mutations and the remaining half are stabilizing mutations. PremPS has been validated to perform significantly better than other methods, especially in predicting the effects of stabilizing mutations. PremPS can be applied to many tasks, including finding functionally important variants, revealing the molecular mechanisms of functional influences and protein design.

## MATERIALS AND METHODS

### Experimental dataset for parameterizing PremPS

ProTherm database is a collection of thermodynamic parameters for wild-type and mutant proteins (48). It contains unfolding Gibbs free energy changes that provide important clues for estimating and interpreting the relationship among structure, stability and function of proteins and their mutants. It is frequently used as training template for developing *in silico* prediction approaches.

S2648 dataset includes 2648 single-point mutations from 131 globular proteins (Figure S1), which was derived from the ProTherm database and compiled by (16). It was used as the training dataset of PoPMuSiC (16), mCSM (17) and DUET (18) methods. Here, we also used the mutations and their unfolding Gibbs free energy changes from the S2648 to parameterize PremPS model. Then, we applied the following criteria to update the protein 3D structures: structure obtained/extracted from monomer or homomer is preferred over heteromer; wild-type protein structure is preferred over mutant; structure with minimal number of ligands is used; crystal structure is preferred over NMR, and higher resolution structure is chosen. The multimeric state of each protein was either assigned by manually checking the corresponding references used to measure protein stability changes or retrieved from the PQS server (52).

Unfolding free energy change ( $\Delta\Delta G$ ) of a system can be characterized as a state function where the  $\Delta\Delta G_{wt \rightarrow mut}$  value of a forward mutation minus  $\Delta\Delta G_{mut \rightarrow wt}$  of its reverse mutation should be equal to zero. Given the unbalanced nature of the S2648 dataset with 2080 destabilizing

(decreasing stability,  $\Delta\Delta G_{exp} \geq 0$ ) and 568 stabilizing (increasing stability,  $\Delta\Delta G_{exp} < 0$ ) mutations, we modelled their reverse mutations in order to establish a more accurate computational method. Therefore, the final training set for parameterizing PremPS model includes 5296 single mutations (it will be referred as S5296) (Table S1). The dataset can be downloaded from <https://lilab.jysw.suda.edu.cn/research/PremPS/download/>.

For the forward mutations ( $\Delta\Delta G_{wt \rightarrow mut}$ ), 3D structures of wild-type proteins were obtained from the Protein Data Bank (PDB) (53). For the reverse mutations ( $\Delta\Delta G_{mut \rightarrow wt}$ ), the 3D structures of mutants were produced by BuildModel module of FoldX (23) using wild-type protein structures as the templates.

## Experimental datasets used for testing

First, we used the following six datasets to assess and compare the predictive performance of PremPS and other 25 computational methods (16-40).

- S350, it is a randomly selected subset from S2648 including 350 mutations from 67 proteins (16). This dataset is widely used to compare the performance of different methods. During the comparison, all methods were retrained after removing S350 from their training sets.
- S605, it was compiled from Protherm database by (21) and contains 605 mutations from 58 proteins, which is the training dataset of Meta-predictor method (21).
- S1925, it includes 1925 mutations from 55 proteins evenly distributed over four major SCOP structural classes, which is the training dataset of AUTOMUTE method (24).
- $S^{sym}$ , a dataset was manually curated by (51). It contains 684 mutations, half of which belong to *forward mutations*, and the remaining half are *reverse mutations* with crystal structures of the corresponding mutant proteins available.
- S134, it consists of experimentally determined stability changes for 134 mutations from sperm-whale myoglobin (43), and six different high-resolution crystal structures of myoglobin were used for the energy calculation.
- p53, it includes 42 mutations within the DNA binding domain of the protein p53 with experimentally determined thermodynamic effects, and the data was obtained from (17).

The number of mutations in each test set is shown in Table S1a, and the number of overlapped mutations between our training set of S5296 and each test set is presented in Table S1b.

Next, we removed the redundant mutations from the above datasets and the overlapped mutations with S5296, then established a combined independent test set. For the proteins with multiple 3D structures, we used the same criteria used in processing the training dataset of S2648 to choose the optimal structure. For the conflicting entries with multiple experimental measurements, if the difference between the maximal and minimal  $\Delta\Delta G_{exp}$  for this mutation is less than 1.0 kcal mol<sup>-1</sup>, we used the average value, otherwise we removed all entries (Figure S2). As a result, the combined independent test set contains 921 single mutations (it will be referred as S921) (Table S1a).

## The model of PremPS

The random forest (RF) regression scoring function of PremPS is composed of ten distinct features belonging to six categories (described below) and parameterized on S5296 dataset. The contribution of each category of features is shown in Table S2.

- *PSSM* score is the Position-Specific Scoring Matrix created by PSI-BLAST. It finds similar protein sequences for the query sequence in which the mutation occurs by searching all protein sequences in NCBI non-redundant database, then builds a PSSM from the resulting alignment (54). The default parameters were applied to construct PSSM profile.
- $\Delta CS$  represents the change of conservation after mutation calculated by PROVEAN method (55). The features of *PSSM* and  $\Delta CS$  illustrate that the evolutionarily conserved sites may play an important role in protein folding.
- $\Delta OMH$  is the difference of hydrophobicity scale between mutant and wild-type residue type. The hydrophobicity scale (OMH) for each type of amino acid residue, obtained from the study of (56), was derived by considering the observed frequency of amino acid replacements among thousands of related structures.
- $SASA_{pro}$  and  $SASA_{sol}$  is the solvent accessible surface area (SASA) of the mutated residue in the protein and in the extended tripeptide respectively. The SASA of a residue in the protein and in the extended tripeptide was calculated by DSSP program (57) and obtained from (58), respectively.
- $P_{FWY}$ ,  $P_{RKDE}$  and  $P_L$  is the fraction of aromatic residues (F, W or Y), charged residues (R, K, D or E) and leucine (L) buried in the protein core, respectively. For instance,  $P_L = \frac{N_L}{N_{All}}$ ,  $N_L$  is

the number of all leucine residues buried in the protein core and  $N_{All}$  is the total number of residues. If the ratio of solvent accessible surface area of a residue in the protein and in the extended tripeptide is less than 0.2 (59), we defined this residue as buried in the core of protein.

- $N_{Hydro}$  and  $N_{Charg}$  is the number of hydrophobic (V, I, L, F, M, W, Y or C) and charged amino acids (R, K, D or E) at 23 sites in the protein sequence, the center of which is at the mutated site, respectively (60).

In addition to Random Forest, we also tried two other popular learning algorithms of Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost), and the results indicate that random forest regression model presents the best performance (Table S3).

PremPS takes about ten minutes to perform calculations for one thousand mutations from a normal size protein ( $\sim 300$  residues). The running time for a single mutation per normal size protein is about four minutes, and it requires  $\sim 0.4$  second for each additional mutation.

### Cross-validation procedures

We performed four types of cross validation (CV1-CV4). For CV1 and CV2, we randomly chose 80% and 50% of mutations from the S5296 set respectively to train the model and used the remaining mutations for blind testing; the procedures were repeated 100 times. The number of mutations is not uniformly distributed over proteins (Figure S1), in order to conquer the bias toward the proteins with large number of mutations, we carried out the third type of cross validation (CV3). Namely, a subset was created by randomly sampling up to 20 mutations for each protein from S5296; the procedure was repeated 10 times and resulted in 1704 mutations in each subset. Then 80% mutations were randomly selected from each subset to train the model and the rest of mutations were used for testing, repeated 10 times. Next, we performed leave-one-protein-out validation (CV4), in which the model was trained on all mutations from 130 protein structures and the rest of protein/mutations were used to evaluate the performance. This procedure was repeated for each protein and its mutations. In all four described cross-validation procedures, during the training/test splits, the forward and their corresponding reverse mutations were retained in the same set, either training or testing.

### Statistical analysis and evaluation of performance

We used two measures of Pearson correlation coefficient (R) and root-mean-square error (RMSE) to verify the agreement between experimental and predicted values of unfolding free energy changes. All correlation coefficients reported in the paper are significantly different from zero with p-value smaller than 0.01 (t-test). RMSE (kcal mol<sup>-1</sup>) is the standard deviation of the residuals (prediction errors). For checking whether the difference in performance between PremPS and other methods is significant, we used Hittner2003 (61) and Fisher1925 (62) tests implemented in package *cocor* from R (63) to compare two correlation coefficients. Receiver operating characteristics (ROC) curves were compared with Delong tests (64).

To quantify the performance of different methods, we performed Receiver Operating Characteristics (ROC) and precision-recall analyses. True positive rate is defined as  $TPR = TP / (TP + FN)$  and false positive rate is defined as  $FPR = FP / (FP + TN)$  (TP: true positive; TN: true negative; FP: false positive; FN: false negative). In addition, Matthews correlation coefficient is calculated for estimating the quality of binary classification and accounting for imbalances in the labeled dataset:

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

To compare across methods, the maximal MCC value was reported for each method through calculating the MCC across a range of thresholds.

## RESULTS and DISCUSSION

PremPS employs only ten features belonging to six categories and is constructed by random forest regression algorithm implemented in R randomForest package (65). The number of trees “ntree” is set to 500 and the number of features, randomly sampled as candidates for splitting at each node, “mtry” value is set to 3. All features have significant contribution to the model (Table S2). The performance of PremPS trained and tested on S5296 is shown in Figure 1a and Table S4. Pearson correlation coefficient between experimental and calculated unfolding free energy changes is 0.82 and the corresponding root-mean-square error and slope is 1.03 kcal mol<sup>-1</sup> and 1.08 respectively.

### Performance on four types of cross-validation

Overfitting is one of the major concerns in machine learning, which may occur when the parameters are over-tuned to minimize the mean square deviations of predicted from experimental values in the training set. To overcome this problem, we performed four types of cross-validation (details were explained in Methods section), which is capable of estimating the performance of a method on previously unseen data. As shown in Figure 1b and Table S4, the correlation coefficient of each round in either CV1 or CV2 is higher than 0.77, and the mean values of R and RMSE for both cross-validation are  $\sim 0.80$  and  $\sim 1.08$  kcal mol $^{-1}$  respectively across the 100 rounds. Taking the bias that the distribution of the number of mutations over proteins is not uniform into account, the CV3 cross-validation was performed. The mean values of R and RMSE is 0.74 and 1.21 kcal mol $^{-1}$  respectively for CV3. Moreover, we evaluated the performance of PremPS on a low redundant set of proteins using leave-one-protein-out validation (CV4), the Pearson correlation coefficient of which reaches to 0.73 (Figure 1c) and RMSE = 1.23 kcal mol $^{-1}$  (Table S4).

### Validation on independent test sets

First, we used six datasets of S350, S605, S1925, S<sup>sym</sup>, S134 and *p53* (details about each set were shown in Methods section) to assess and compare the predictive performance of PremPS and other 25 computational methods that predict protein stability changes upon single mutations. The Pearson correlation coefficient and root-mean-square error for different methods applied on different datasets are presented in Table S5. The results indicate that the PremPS has the best performance, and the next are PoPMuSiC, FoldX and mCSM methods (the detailed descriptions are illustrated in the Table S5). Moreover, we removed the redundant mutations from the above six datasets and the overlapped mutations with the training set of S5296, and established the independent test set of S921. Then, we applied PremPS and other three methods of PoPMuSiC, FoldX and mCSM on this test set. The results reported in Figure 1d, Table 1 and Figure S3 demonstrate that PremPS achieves a high prediction accuracy with R of up to 0.78 and RMSE of 1.48 kcal mol $^{-1}$ , outperforming all three other methods.

One of our goals is to improve the prediction accuracy for mutations increasing stability (stabilizing mutations). Table 1 and Table S4 show the considerable improvement for PremPS estimating the effects of stabilizing mutations. The R is 0.60 for PremPS tested on the stabilizing mutations from S921 set (Table 1). In addition, we evaluated the performance of PremPS when it was trained only on the forward mutation dataset of S2648 and tested on the S921. The correlation

coefficients are 0.73 and 0.29 for destabilizing and stabilizing mutations, respectively (Table S6). The results confirm that the use of reverse mutations improved the performance of our model in estimating the effects of stabilizing mutations without compromising the prediction accuracy for destabilizing mutations.

### Performance on identification of highly destabilizing and stabilizing mutations

We carried out the ROC analysis in order to quantify the performance of PremPS in distinguishing highly destabilizing ( $\Delta\Delta G \geq 1.0 \text{ kcal mol}^{-1}$ ) and highly stabilizing ( $\Delta\Delta G \leq -1.0 \text{ kcal mol}^{-1}$ ) mutations from the others. Figure 2a and Figure S4 show excellent performance of PremPS in evaluating highly destabilizing/stabilizing mutations. We further subdivided mutations into two categories of core and surface according to the location of mutated site in the protein 3D structure (The definition was illustrated in the Methods section). We found that PremPS performed well for both categories and outperforms other methods on the S921 set (Figure 2b, Figure S5 and Table S6). The experimental  $\Delta\Delta G$  values for the majority of surface mutations are distributed near zero (Figure S5a), which might be the reason for the relatively lower correlation coefficient and RMSE compared to mutations in the core (RMSE is 1.81 and 1.11  $\text{kcal mol}^{-1}$  for core and surface mutations, respectively) (Table S6).

### Performance on identification of functionally important mutations

Protein stability is related to its functional activity, and effect on stability could be one of the major disease-causing mechanisms of missense mutations. Therefore, we further checked the capability of PremPS on the identification of functionally important mutations. A previously proposed experimental dataset including 1139 non-neutral/deleterious and 4137 neutral/benign mutations was used here to evaluate the performance of PremPS in distinguishing non-neutral from neutral mutations (66). Among them 1037 non-neutral and 1159 neutral mutations (named as F2196 dataset) could be mapped to the corresponding protein 3D structures and allowed to calculate the unfolding free energy changes (Figure S6). We compared the performance of PremPS with 28 state-of-the-art computational methods developed for predicting the pathogenicity of missense mutations (67,68), such as CHASMplus (69), VEST4 (70), REVEL (71) and B-Score (66) (Figure S6). Table 2 and Figure S6 show that PremPS is capable of distinguishing non-neutral from neutral mutations. Since one protein could be mapped to several PDB structures, the number

of mapped 3D structures for 50 proteins in F2196 dataset is 1554 (Fig. S6a). If the average value of protein stability changes calculated using all mapped structures was used, the performance of PremPS ranks in the middle (Fig. S6b), and if the maximum predicted absolute value was used for each non-neutral mutation and minimum predicted absolute value was used for each neutral mutation, PremPS performs better than all other methods (Table 2 and Fig. S6c).

## ONLINE WEB SERVER

### Input

The 3D structure of a protein is required by the webserver, and the user can provide the Protein Data Bank (PDB) code or upload the coordinate file. When the user provides the PDB code, biological assemblies or asymmetric unit can be retrieved from the Protein Data Bank (Fig. 3). After the structure is retrieved correctly, the server will display a 3D view colored by protein chains and list the corresponding protein name (Fig. S7). At the second step one or multiple chains that must belong to one protein can be assigned to the following energy calculation. The third step is to select mutations and three options are provided: “Upload Mutation List”, “Alanine Scanning for Each Chain” and “Specify One or More Mutations Manually” (Fig. 3 and Fig. S7). “Upload Mutation List” allows users to upload a list of mutations for large-scale mutational scans. “Alanine Scanning for Each Chain” allows users to perform alanine scanning for each chain. In the option of “Specify One or More Mutations Manually”, users can not only perform calculations for specified mutations but also allow to view the mutated residues in the protein structure.

### Output

For each mutation of a protein, the PremPS server provides the following results (Figure 3):  $\Delta\Delta G$  (kcal mol<sup>-1</sup>), predicted unfolding free energy change induced by a single mutation (positive and negative sign corresponds to destabilizing and stabilizing mutations, respectively); location of the mutation occurred (COR: core or SUR: surface), a residue is defined as buried in the protein core if the ratio of solvent accessible surface area of this residue in the protein and in the extended tripeptide is less than 0.2, otherwise it is located on the surface of protein. In addition, for each mutation, the PremPS outputs the contribution of each feature in the target function and provides an interactive 3D viewer showing the non-covalent interactions between the mutated site and its adjacent residues, generated by Arpeggio (72) (Fig. S8).

## ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China (Grant No. 31701136), Natural Science Foundation of Jiangsu Province, China (Grant No. BK20170335), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

## Competing interests

The authors declare no competing interests.

## References

1. Tanford, C. (1968) Protein denaturation. *Advances in protein chemistry*, **23**, 121-282.
2. Shoichet, B.K., Baase, W.A., Kuroki, R. and Matthews, B.W. (1995) A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, **92**, 452-456.
3. Bromberg, Y. and Rost, B. (2009) Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinformatics*, **10 Suppl 8**, S8.
4. 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. *Human mutation*, **32**, 1161-1170.
5. Zhang, N., Chen, Y., Zhao, F., Yang, Q., Simonetti, F.L. and Li, M. (2018) PremPDI estimates and interprets the effects of missense mutations on protein-DNA interactions. *PLoS Comput Biol*, **14**, e1006615-e1006615.
6. Li, M., Simonetti, F.L., Gonçarenc, A. and Panchenko, A.R. (2016) MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. *Nucleic Acids Res*, **44**, W494-W501.
7. Hashimoto, K., Rogozin, I.B. and Panchenko, A.R. (2012) Oncogenic potential is related to activating effect of cancer single and double somatic mutations in receptor tyrosine kinases. *Human mutation*, **33**, 1566-1575.
8. Peng, Y., Norris, J., Schwartz, C. and Alexov, E. (2016) Revealing the Effects of Missense Mutations Causing Snyder-Robinson Syndrome on the Stability and Dimerization of Spermine Synthase. *Journal of biomolecular structure & dynamics*, **17**.
9. Smith, I.N., Thacker, S., Jaini, R. and Eng, C. (2019) Dynamics and structural stability effects of germline PTEN mutations associated with cancer versus autism phenotypes. *Journal of biomolecular structure & dynamics*, **37**, 1766-1782.
10. Chiang, C.H., Grauffel, C., Wu, L.S., Kuo, P.H., Doudeva, L.G., Lim, C., Shen, C.K. and Yuan, H.S. (2016) Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Scientific reports*, **6**, 21581.
11. Kumar, V., Rahman, S., Choudhry, H., Zamzami, M.A., Sarwar Jamal, M., Islam, A., Ahmad, F. and Hassan, M.I. (2017) Computing disease-linked SOD1 mutations: deciphering protein stability and patient-phenotype relations. *Scientific reports*, **7**, 4678.
12. Peng, Y. and Alexov, E. (2016) Investigating the linkage between disease-causing amino acid variants and their effect on protein stability and binding. *Proteins*, **84**, 232-239.
13. Stevens, R.C. (2000) High-throughput protein crystallization. *Current opinion in structural biology*, **10**, 558-563.
14. Kiel, C. and Serrano, L. (2014) Structure-energy-based predictions and network modelling of RASopathy and cancer missense mutations. *Molecular systems biology*, **10**, 727.
15. Goldenzweig, A., Goldsmith, M., Hill, S.E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J. et al. (2016) Automated Structure- and Sequence-

Based Design of Proteins for High Bacterial Expression and Stability. *Molecular cell*, **63**, 337-346.

16. Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P. and Roonan, M. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics (Oxford, England)*, **25**, 2537-2543.
17. Pires, D.E.V., Ascher, D.B. and Blundell, T.L. (2013) mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics (Oxford, England)*, **30**, 335-342.
18. 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 Research*, **42**, W314-W319.
19. 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 Research*, **39**, W215-W222.
20. Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S. and Lackner, P. (2015) MAESTRO--multi agent stability prediction upon point mutations. *BMC Bioinformatics*, **16**, 116.
21. Broom, A., Jacobi, Z., Trainor, K. and Meiering, E.M. (2017) Computational tools help improve protein stability but with a solubility tradeoff. *Journal of Biological Chemistry*, **292**, 14349-14361.
22. Kwasigroch, J.M., Gilis, D., Dehouck, Y. and Roonan, M. (2002) PoPMuSiC, rationally designing point mutations in protein structures. *Bioinformatics (Oxford, England)*, **18**, 1701-1702.
23. 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.
24. Masso, M. and Vaisman, II. (2008) Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics (Oxford, England)*, **24**, 2002-2009.
25. Zhou, H. and Zhou, Y. (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein science : a publication of the Protein Society*, **11**, 2714-2726.
26. Yin, S., Ding, F. and Dokholyan, N.V. (2007) Eris: an automated estimator of protein stability. *Nature methods*, **4**, 466-467.
27. Pandurangan, A.P., Ochoa-Montano, B., Ascher, D.B. and Blundell, T.L. (2017) SDM: a server for predicting effects of mutations on protein stability. *Nucleic Acids Res*, **45**, W229-w235.
28. Chen, C.W., Lin, J. and Chu, Y.W. (2013) iStable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics*, **14 Suppl 2**, S5.

29. 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.
30. 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 Suppl 2**, S6.
31. 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.
32. Giollo, M., Martin, A.J., Walsh, I., Ferrari, C. and Tosatto, S.C. (2014) NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation. *BMC genomics*, **15 Suppl 4**, S7.
33. Quan, L., Lv, Q. and Zhang, Y. (2016) STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics (Oxford, England)*, **32**, 2936-2946.
34. 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.
35. Parthiban, V., Gromiha, M.M. and Schomburg, D. (2006) CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res*, **34**, W239-242.
36. 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.
37. 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 science : a publication of the Protein Society*, **17**, 1212-1219.
38. Cohen, M., Potapov, V. and Schreiber, G. (2009) Four distances between pairs of amino acids provide a precise description of their interaction. *PLoS computational biology*, **5**, e1000470.
39. Deutsch, C. and Krishnamoorthy, B. (2007) Four-body scoring function for mutagenesis. *Bioinformatics (Oxford, England)*, **23**, 3009-3015.
40. Dehouck, Y., Kwasigroch, J.M., Gilis, D. and Roonan, M. (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, **12**, 151.
41. Getov, I., Petukh, M. and Alexov, E. (2016) SAAFEC: Predicting the Effect of Single Point Mutations on Protein Folding Free Energy Using a Knowledge-Modified MM/PBSA Approach. *Int J Mol Sci*, **17**, 512.
42. Zhang, Z., Wang, L., Gao, Y., Zhang, J., Zhenirovskyy, M. and Alexov, E. (2012) Predicting folding free energy changes upon single point mutations. *Bioinformatics (Oxford, England)*, **28**, 664-671.
43. Kepp, K.P. (2015) Towards a “Golden Standard” for computing globin stability: Stability and structure sensitivity of myoglobin mutants. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1854**, 1239-1248.

44. Kepp, K.P. (2014) Computing Stability Effects of Mutations in Human Superoxide Dismutase 1. *The Journal of Physical Chemistry B*, **118**, 1799-1812.
45. Khan, S. and Vihinen, M. (2010) Performance of protein stability predictors. *Human mutation*, **31**, 675-684.
46. 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 : PEDS*, **22**, 553-560.
47. Hawkins, D.M. (2004) The problem of overfitting. *Journal of chemical information and computer sciences*, **44**, 1-12.
48. 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 Research*, **32**, D120-D121.
49. Montanucci, L., Savojardo, C., Martelli, P.L., Casadio, R. and Fariselli, P. (2019) On the biases in predictions of protein stability changes upon variations: the INPS test case. *Bioinformatics (Oxford, England)*, **35**, 2525-2527.
50. Usmanova, D.R., Bogatyreva, N.S., Arino Bernad, J., Eremina, A.A., Gorshkova, A.A., Kanevskiy, G.M., Lonishin, L.R., Meister, A.V., Yakupova, A.G., Kondrashov, F.A. *et al.* (2018) Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. *Bioinformatics (Oxford, England)*, **34**, 3653-3658.
51. Pucci, F., Bernaerts, K.V., Kwasigroch, J.M. and Roonan, M. (2018) Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics (Oxford, England)*, **34**, 3659-3665.
52. Henrick, K. and Thornton, J.M. (1998) PQS: a protein quaternary structure file server. *Trends in Biochemical Sciences*, **23**, 358-361.
53. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235-242.
54. Bhagwat, M. and Aravind, L. (2007) PSI-BLAST tutorial. *Methods in molecular biology (Clifton, N.J.)*, **395**, 177-186.
55. 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.
56. Sweet, R.M. and Eisenberg, D. (1983) Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. *Journal of Molecular Biology*, **171**, 479-488.
57. Joosten, R.P., te Beek, T.A., Krieger, E., Hekkelman, M.L., Hooft, R.W., Schneider, R., Sander, C. and Vriend, G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res*, **39**, D411-419.
58. Rose, G.D., Geselowitz, A.R., Lesser, G.J., Lee, R.H. and Zehfus, M.H. (1985) Hydrophobicity of amino acid residues in globular proteins. *Science (New York, N.Y.)*, **229**, 834-838.

59. Hou, Q., Kwasigroch, J.M., Rooman, M. and Pucci, F. (2019) SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics (Oxford, England)*.
60. Yang, Y., Urolagin, S., Niroula, A., Ding, X., Shen, B. and Vihinen, M. (2018) PON-tstab: Protein Variant Stability Predictor. Importance of Training Data Quality. *Int J Mol Sci*, **19**, 1009.
61. Hittner, J.B., May, K. and Silver, N.C. (2003) A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of general psychology*, **130**, 149-168.
62. Edwards, A.W.F. (2005) R.A. Fischer, statistical methods for research workers, first edition (1925). *Landmark Writings in Western Mathematics 1640-1940*, 856-870.
63. Diedenhofen, B. and Musch, J. (2015) cocor: a comprehensive solution for the statistical comparison of correlations. *PloS one*, **10**, e0121945.
64. DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, **44**, 837-845.
65. Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32.
66. Brown, A.L., Li, M., Goncearenco, A. and Panchenko, A.R. (2019) Finding driver mutations in cancer: Elucidating the role of background mutational processes. *PLoS computational biology*, **15**, e1006981.
67. Li, J., Zhao, T., Zhang, Y., Zhang, K., Shi, L., Chen, Y., Wang, X. and Sun, Z. (2018) Performance evaluation of pathogenicity-computation methods for missense variants. *Nucleic Acids Res*, **46**, 7793-7804.
68. Liu, X., Jian, X. and Boerwinkle, E. (2011) dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Human mutation*, **32**, 894-899.
69. Tokheim, C. and Karchin, R. (2019) CHASMplus Reveals the Scope of Somatic Missense Mutations Driving Human Cancers. *Cell systems*, **9**, 9-23.e28.
70. Carter, H., Douville, C., Stenson, P.D., Cooper, D.N. and Karchin, R. (2013) Identifying Mendelian disease genes with the variant effect scoring tool. *BMC genomics*, **14 Suppl 3**, S3.
71. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D. *et al.* (2016) REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American journal of human genetics*, **99**, 877-885.
72. Jubb, H.C., Higueruelo, A.P., Ochoa-Montano, B., Pitt, W.R., Ascher, D.B. and Blundell, T.L. (2017) Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol*, **429**, 365-371.

## Figure legends

**Figure 1.** Pearson correlation coefficients between experimentally-determined and calculated values of changes in protein stability ( $\Delta\Delta G$ ) for PremPS trained and tested on S5296 (a), applying three types of cross-validation (b) and leave-one-protein-out validation on S5296 (c), and tested on the independent dataset of S921 (d). See also Table S1 and Table S4.

**Figure 2.** Comparative performance of PremPS and three other methods of PoPMuSiC, FoldX and mCSM on the independent test set of S921. (a) ROC curves for predicting highly destabilizing ( $\Delta\Delta G_{exp} \geq 1 \text{ kcal mol}^{-1}$ ) and highly stabilizing mutations ( $\Delta\Delta G_{exp} \leq -1 \text{ kcal mol}^{-1}$ ). PremPS has substantially higher AUC-ROC than other methods (p-value  $\ll 0.01$ , DeLong test, Figure S4b). (b) Pearson correlation coefficients between predicted and experimental  $\Delta\Delta G$  for mutations located in protein core and surface. The difference in R between PremPS and other methods is significant (p-value  $< 0.01$ , Hittner2003). More details are shown in the Figure S4 and S5.

**Figure 3.** Left corner: the entry page of PremPS server; right corner: the third step for selecting mutations and three options are provided: “Specify One or More Mutations Manually”, “Upload Mutation List” and “Alanine Scanning for Each Chain”, see also Figure S7; and bottom: final results, see also Figure S8.

**Table 1. Comparison of methods' performance on the independent test set of S921.**

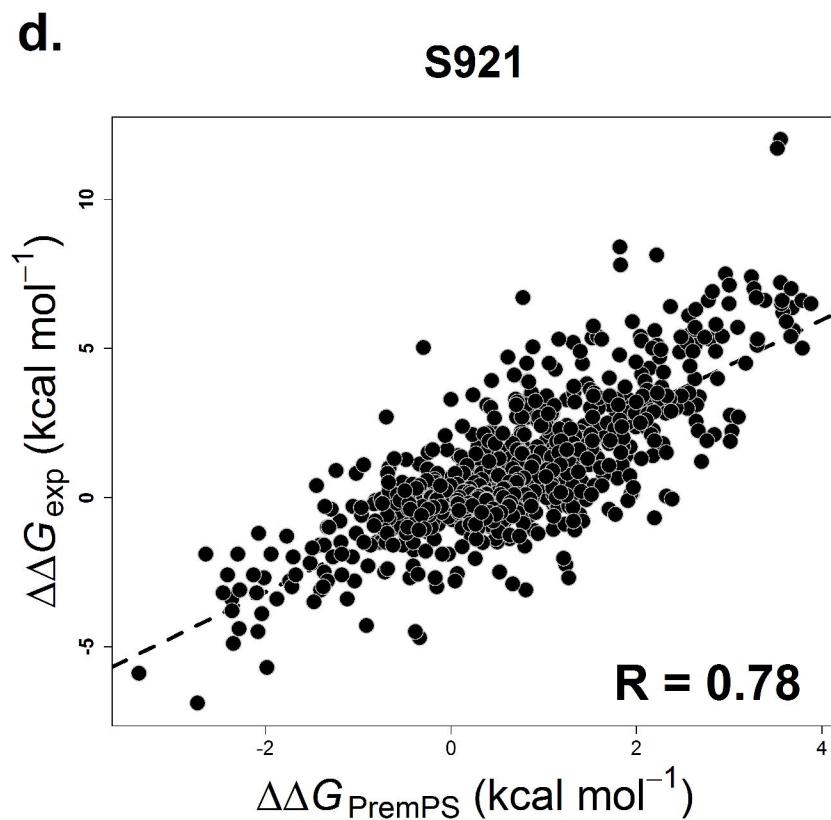
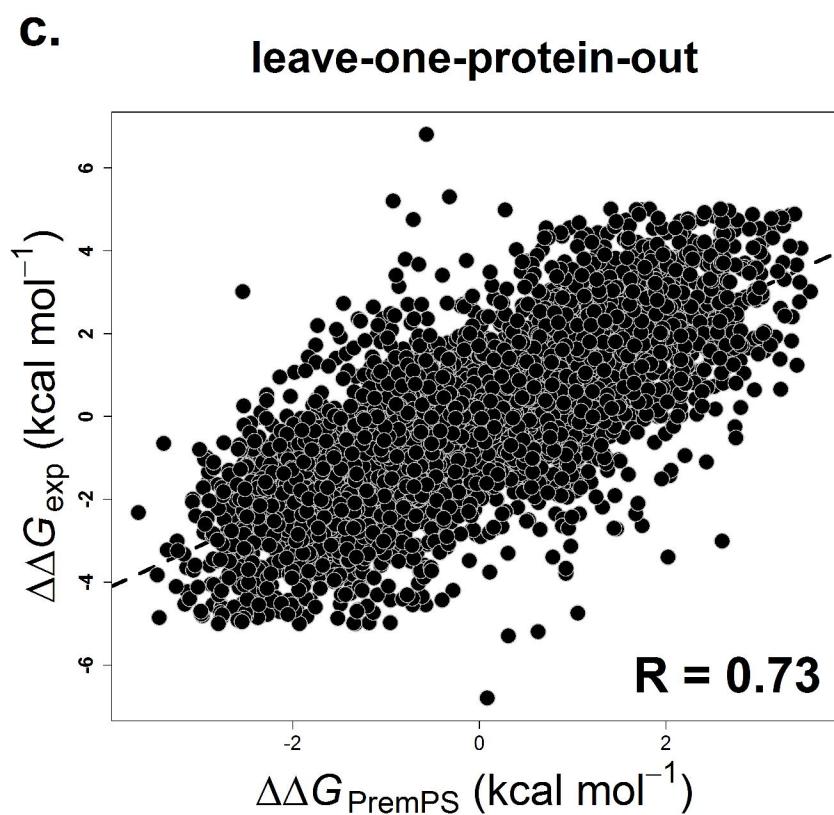
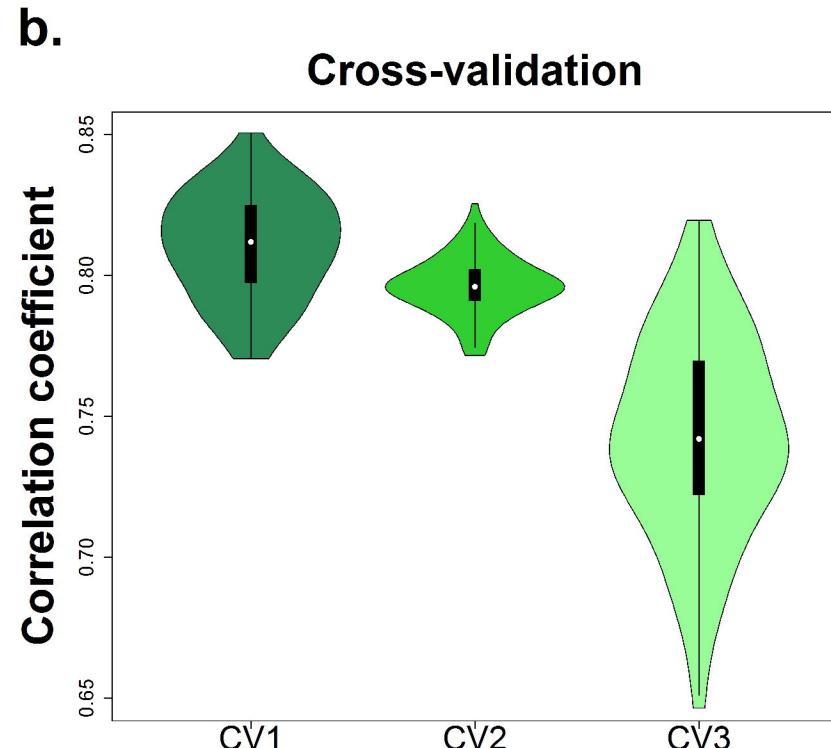
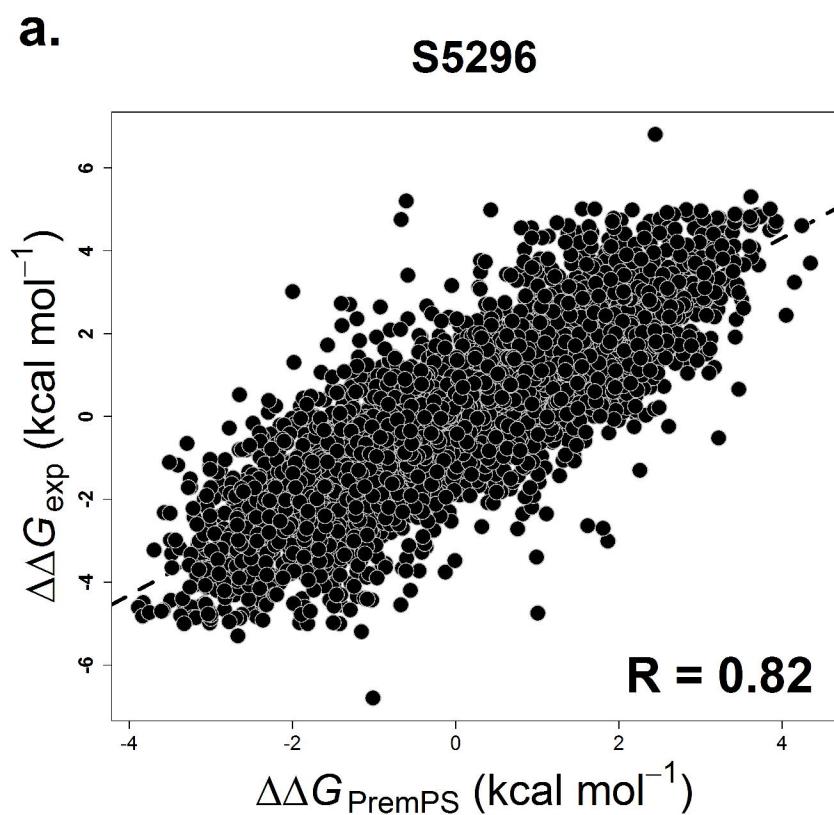
Method	All mutations		Destabilizing		Stabilizing	
	R	RMSE	R	RMSE	R	RMSE
PremPS	0.78	1.48	0.72	1.54	0.60	1.33
PoPMuSiC	0.64	1.68	0.68 <sup>#</sup>	1.48	-	2.06
FoldX	0.57	2.06	0.56	1.99	0.22	2.21
mCSM	0.52	1.85	0.57	1.63	-	2.25

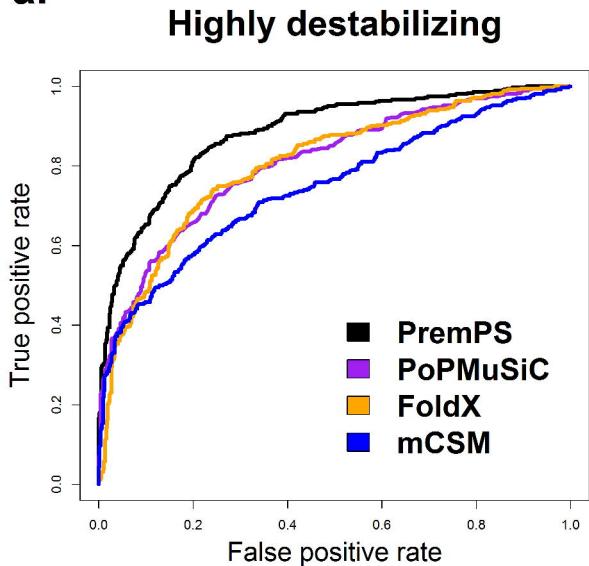
R: Pearson correlation coefficient between experimental and predicted  $\Delta\Delta G$  values. RMSE (kcal mol<sup>-1</sup>): root-mean square error. Only correlation coefficients with statistically significantly different from zero (p-value < 0.01) are shown. The difference in correlation coefficients between PremPS and other methods is significant (all p-values << 0.01 compared to PremPS except <sup>#</sup>p-value = 0.04, Hittner2003 test). The number of destabilizing and stabilizing mutations in S921 is 634 and 287, respectively (Table S1).

**Table 2. Comparison of the performance of different computational methods in distinguishing non-neutral from neutral mutations using the experimental dataset of F2196.**

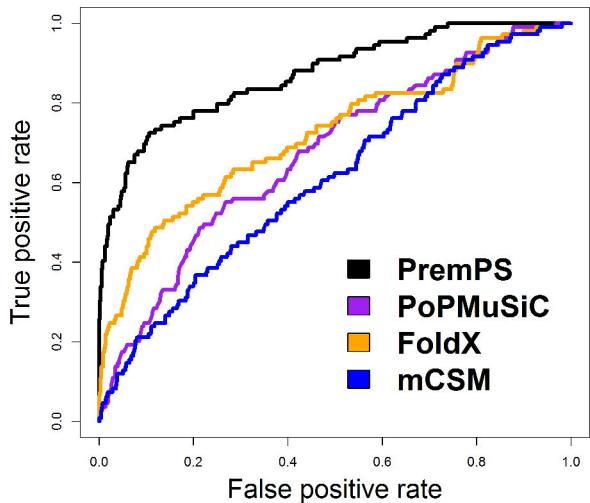
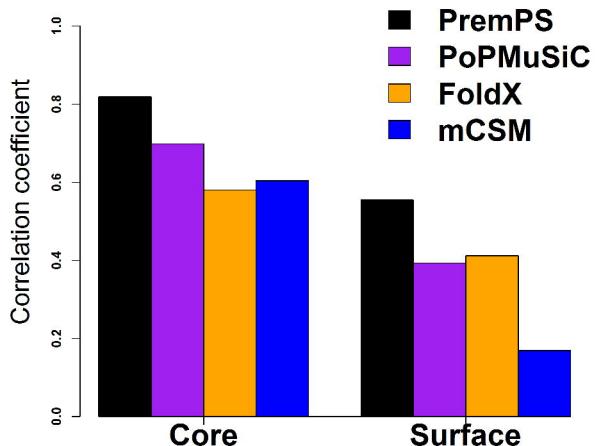
Method	AUC-ROC	AUC-PR	MCC	Sensitivity at 10% FPR
PremPS	0.88	0.84	0.61	0.67
VEST4	0.81	0.79	0.51	0.56
CHASMplus	0.78	0.80	0.47	0.54
B-Score	0.74	0.78	0.47	0.53

AUC-ROC: Area under the receiver operator curve, AUC-PR: Area under the precision recall curve, MCC: maximum Matthews correlation coefficient, and FPR: False positive rate. PremPS: the maximum predicted absolute value of stability change was used for each non-neutral mutation and minimum predicted absolute value was used for each neutral mutation. The difference of AUC-ROC between PremPS and other methods is significant (p-value << 0.01, DeLong test). See Figure S6 for more details.



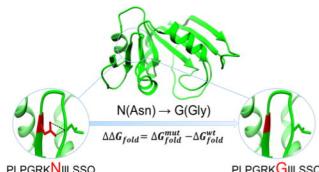
**a.**

### Highly stabilizing

**b.**

## PremPS - Predicting the Effects of Mutations on Protein Stability

PremPS evaluates the effects of single mutations on protein stability by calculating the changes in unfolding Gibbs free energy. It can be applied to a large number of tasks, including finding functionally important variants, understanding their molecular mechanisms and protein design. 3D structure of a protein is required for this method.



### Step1 - Select Protein

Input PDB code:  Bioassembly [?](#)  1st   
 Example: [1YU5](#) Asymmetric Unit

Crystal structure of Native Sperm Whale myoglobin from low ionic strength environment (Form 1)

Upload PDB file:  No file chosen

Format description [?](#) for uploaded file

Next

School of Biology & Basic Medical Sciences, Sichuan University  
199 Ren-Ai Road, Sichuan, Chengdu, 2015123 PR. China

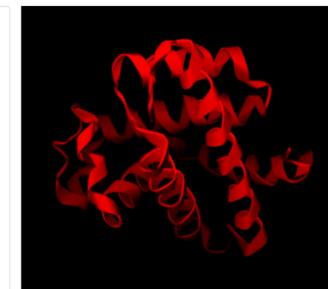


### Step 3 - Select Mutations

PDB id: [1U7S](#)

Chains

Chain A



Reset Zoom

[Manually select](#) [Upload file](#) [Alanine Scanning](#)

Specify One or More Mutations:

Chain to Mutate	Residue	Mutant Residue	View in Structure <a href="#">?</a>
Chain A	Q 26 (GLN)	A (ALA)	<input type="button" value="View"/>
Chain A	L 104 (LEU)	D (ASP)	<input type="button" value="View"/>
Chain A	R 118 (ARG)	C (CYS)	<input type="button" value="View"/>

Submit Job



Add or Remove Mutations [?](#)

Job id: 2019121211341159727269857

### • Summary

PDB ID	Chains	Number of mutations	Start time (EST)	Process time	Results <a href="#">?</a>
1U7S	A	3	2019-12-12 06:41	3 min	<a href="#">Download</a>

### • Results

#	Mutated Chain	Mutation	ΔΔG <a href="#">?</a>	Location <a href="#">?</a>	Structure <a href="#">?</a>
1	A	Q26A	-0.64	COR	<a href="#">Explore</a>
2	A	L104D	1.82	COR	<a href="#">Explore</a>
3	A	R118C	0.27	SUR	<a href="#">Explore</a>