**SUPPORTING INFORMATION**

# Evaluating the predictions of the protein stability change upon single amino acid substitutions for the FXN CAGI5 challenge

Castrense Savojardo[1§], Maria Petrosino[2§], Giulia Babbi[1], Samuele Bovo[1], Carles Corbi-Verge[3], Rita Casadio[1,4], Piero Fariselli[5], Lukas Folkman[6], Aditi Garg[7], Mostafa Karimi[8], Panagiotis Katsonis[9], Philip M. Kim[3,10,11], Olivier Lichtarge[9,12-14], Pier Luigi Martelli[1], Alessandra Pasquo[15], Debnath Pal[7], Yang Shen[8], Alexey V. Strokach[11], Paola Turina[16], Yaoqi Zhou[6,17], Gaia Andreoletti[18], Steven Brenner[18], Roberta Chiaraluce[2*], Valerio Consalvi[2*], Emidio Capriotti[16*].

Affiliations:

1. Biocomputing Group, Department of Pharmacy and Biotechnology, University of Bologna, Italy.
2. Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza University of Roma, Roma, Italy.
3. Donnelly Center for Cellular and Biomolecular Research, University of Toronto, 160 College St, Toronto, ON M5S 3E1, Canada.
4. Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), Italian National Research Council (CNR), Bari, Italy.
5. Department of Medical Sciences University of Torino, 10126 Torino, Italy
6. School of Information and Communication Technology Griffith University, Parklands Dr, Southport, QLD 4222, Australia
7. Department of Computational and Data Sciences. Indian Institute of Science, Bengaluru 560 012, India.
8. Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77840, USA.
9. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.
10. Department of Molecular Genetics, University of Toronto, 1 King's College Cir, Toronto, ON M5S 1A8, Canada.
11. Department of Computer Science, University of Toronto, 214 College St, Toronto, ON M5T 3A1, Canada
12. Department of Biochemistry & Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA.
13. Department of Pharmacology, Baylor College of Medicine, Houston, Texas 77030, USA.
14. Computational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, Texas 77030, USA.
15. ENEA CR Frascati, Diagnostics and Metrology Laboratory,FSN-TECFIS-DIM, Frascati, Italy.
16. Department of Pharmacy and Biotechnology, University of Bologna, 40126 Bologna, Italy.
17. Institute for Glycomics, Griffith University, Parklands Dr, Southport QLD 4222, Australia.
18. Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA.

$ Co-first authors

* Corresponding authors: Roberta Chiaraluce roberta.chiaraluce@uniroma1.it,
Valerio Consalvi valerio.consalvi@uniroma1.it,
Emidio Capriotti emidio.capriotti@unibo.it

## SUPPORTING MATERIALS AND METHODS

### 1 Experimental data analysis

Urea-induced equilibrium unfolding transitions were analysed by fitting baseline and transition region data to a two-state linear extrapolation model (Santoro and Bolen, 1988) according to

$$\Delta G_{\text{unfolding}} = \Delta G^{\text{H}_2\text{O}} + m\left[Urea\right] = -R\text{T}lnK_{\text{unfolding}} \quad [1]$$

where $\Delta G_{\text{unfolding}}$ is the free energy change for unfolding for a given denaturant concentration. $\Delta G^{\text{H}_2\text{O}}$ the free energy change for unfolding in the absence of denaturant and $m$ a slope term which quantifies the change in $\Delta G_{\text{unfolding}}$ per unit concentration of denaturant. $R$ the gas constant. $T$ the temperature and $K_{\text{unfolding}}$ the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_{\text{N}} + s_{\text{N}}[X]_i + \left(y_{\text{U}} + s_{\text{U}}[X]_i\right) * \exp\left[\left(-\Delta G^{H_2O} - m[X]_i\right)/RT\right]}{1 + \exp\left[\frac{\left(-\Delta G^{H_2O} - m[X]_i\right)}{RT}\right]} \quad [2]$$

where $y_i$ is the observed signal, $y_{\text{U}}$ and $y_{\text{N}}$ are the baseline intercepts for unfolded and native protein, $s_{\text{U}}$ and $s_{\text{N}}$ are the baseline slopes for the unfolded and native protein, $[X]_i$ the denaturant concentration after the ith addition, $\Delta G^{\text{H}_2\text{O}}$ the extrapolated free energy of unfolding in the absence of denaturant, $m$ the slope of $\Delta G_{\text{unfolding}}$ versus $[X]$ plot. The denaturant concentration at the midpoint of the transition, $[\text{Urea}]_{0.5}$, according to equation 1, is calculated as:

$$\left[\text{Urea}\right]_{0.5} = \Delta G^{\text{H}_2\text{O}} / m \quad [3]$$

All unfolding transition data were fitted by using Graphpad Prism 5.04.

### 2 Participant groups and submissions

***Lichtarge Lab - Baylor College of Medicine (Group 1)***

The Lichtarge Lab at the Baylor College of Medicine submitted one set of predictions (G1-1) using the Evolutionary Action (EA) method (Katsonis and Lichtarge, 2014) that measures the fitness effect of coding variants analytically from protein evolution data. The participants used the EA scores as predictors of the folding destabilization due to frataxin variations. However, although the variants that destabilize protein structure are more likely to have a strong fitness

effect, the reverse is not always true. Thus, it is expected that EA tends to overestimate the stability effect of some variants. Keeping this in mind, they used 71 homologous sequences of the human frataxin sequence (NP_000135) to predict, on a relative scale from 0 (neutral) to 100 (pathogenic), the fitness effect for each variant. To convert the EA score to $\Delta\Delta G^{H2O}$ values, a linear transformation was performed to match EA scores of 30 (the typical cutoff for benign variants) to 0 kcal/mol and EA scores of 100 to -3 kcal/mol.
In further detail, EA estimates the fitness effect of variations by using an equation that states the phenotype change ($\Delta\varphi$) equals to the product of the sensitivity of the variation site to genotype changes ($\partial\varphi/\partial\gamma$) and the magnitude of the genotype change ($\Delta\gamma$). The sensitivity of the variation site $\partial\varphi/\partial\gamma$ can be approximated with the Evolutionary Trace (ET) algorithm which ranks the relative evolutionary importance of sequence positions in a family of aligned homologs (Lichtarge, et al., 1996) and the genotype change $\Delta\gamma$ can be approximated with inverse amino acid substitution log-odds. The computed fitness change ($\Delta\varphi$), or Evolutionary Action score, has been shown to correlate with experimental loss of function, clinical association, morbidity, and mortality(Katsonis and Lichtarge, 2014; Neskey, et al., 2015). EA is available for non-profit use at http://mammoth.bcm.tmc.edu/EvolutionaryAction.

***Biocomputing Group - University of Bologna (Group 2)***

The Biocomputing Group at the University of Bologna submitted one set of predictions (G2-1) using INPS-3D (Fariselli, et al., 2015; Savojardo, et al., 2016). INPS-3D (Impact of Non-synonymous mutations on Protein Stability) is a method for predicting the impact of non-synonymous Single Nucleotide Variants (nsSNVs) on protein stability, starting from features extracted from the protein 3D structure, namely: relative solvent accessibility and local energy change upon amino acid substitution (computed using pairwise residue contact potentials). INPS-3D adopts a Support Vector Regression (SVR) approach, trained on seven features extracted from the protein primary sequence, including BLOSUM62 substitution score (Henikoff and Henikoff, 1992), hydrophobicity (wild-type and variants) (Kyte and Doolittle, 1982), Dayhoff mutability index (Dayhoff, et al., 1978) of wild-type, molecular weights of wild-type and variant and evolutionary information derived from multiple sequence alignments.
For this challenge the PDB structure 1EKG was used as reference for the frataxin challenge. To each prediction a standard deviation of 0.5 was assigned. INPS-3D is available online at https://inpsmd.biocomp.unibo.it.

***Zhou Lab - Griffith University (Group 3)***

The Zhou Lab at Griffith University submitted three sets of predictions (G3-1, G3-2, G3-3) using the EASE-MM (Evolutionary, Amino acid, and Structural Encodings with Multiple Models) algorithm (Folkman, et al., 2016). EASE-MM is a sequence-based method available as a web-server at http://sparks-lab.org/server/ease. The method predicts protein stability changes using only protein sequence information. The structural properties of the protein are predicted using SPIDER2 (Heffernan, et al., 2015), which correctly predicted seven out of eight residues from the frataxin challenge. EASE-MM comprises five specialized support vector regression (SVR) models to predict $\Delta\Delta G_u$ of amino acid substitutions located in different secondary structure (SS)

elements (helix, sheet, or coil) and with different levels of accessible surface area (ASA) (exposed or buried with a 25% threshold). The final prediction is the average of $\Delta\Delta G_u$ predicted with two models, one selected based on the predicted SS and the other based on the predicted ASA of the variation site. EASE-MM was designed using a dataset of 1676 variations (70 proteins) from the ProTherm database (Kumar, et al., 2006), version February 2013, and tested using a dataset of 236 variations (23 proteins) with a low sequence identity (<25%) to the design dataset. Importantly, these datasets were manually curated to correct erroneous records in ProTherm according to the original publications.

To build the five models employed by EASE-MM, the design dataset was partitioned according to SS and ASA predicted from the protein sequence. A unique set of predictive features was identified for each of the five SVR models using the sequential forward floating search (Pudil, et al., 1994). As a result, each model included a unique combination of evolutionary conservation features (such as the difference of the wild-type and variant amino acid probabilities in a multiple sequence alignment), amino acid parameters (such as differences in volume, bulkiness, hydrophobicity of the wild-type and variant residues), and predicted structural properties (such as SS elements and ASA). Importantly, these features were selected using "unseen-protein" 10-fold CV, which was devised to avoid over-fitting on specific proteins by splitting the dataset into CV folds so that all variations of a cluster of similar proteins (≥ 25% sequence identity) are always contained within a single fold (Folkman, et al., 2014). The same CV scheme was employed to optimize hyper-parameters of the SVR models and the radial basis function kernel (C, ε, and γ) using grid search. Finally, to maximize the training dataset size, EASE-MM (the final model used in this challenge and available on the web-server) was trained using all available variants (by merging the design and test datasets). Of note, no frataxin variants or variants of related proteins (≥ 25% sequence identity) were included in the dataset.

***Shen Lab - Texas A&M University (Group 4)***

The Shen Lab at the Texas A&M University submitted two sets of predictions (G4-1, G4-2) using iCFN (interconnected Cost Function Network) (Karimi and Shen, 2018). iCFN , an efficient and exact multi-state protein design algorithm, was first applied here to predict changes in folding energy terms upon variation; then a machine learning model was trained with these terms as features to predict unfolding. Specifically, the global flexibility of the wild-type frataxin structure was modeled using an ensemble of substates (PDB IDs: 1EKG, 3S4M, 3S5E, 3S5F, 3T3L, 3T3J, 3T3K and 3T3X); variations of interest were introduced to the structures while keeping the neighboring residues flexible; and the energetically most favorable combinations of substates and conformations were searched for each variant by solving a combinatorial optimization problem exactly. The energy terms of the resulting variant structures (internal energy and continuum electrostatics in an MM) were compared to the wild type. These energy contributions were selected as features for a linear regression model trained over few observed frataxin variant data (D122Y, G130V, I154F and W155R) from (Correia, et al., 2008). This model was used for predicting the $\Delta\Delta G^{H_2O}$ values of the variants in the CAGI5 frataxin challenge.

***Pal Lab - Indian Institute of Science (Group 5)***

The Pal Lab at the Indian Institute of Science in Bangalore submitted two batches of predictions using GROMACS (Van Der Spoel et al 2005). The first step in the process was to run all-atom molecular dynamics (MD) simulations using GROMACS (Version 4.6.5) software on the wild type and all the variant proteins with CHARMM27 (MacKerell, et al., 2000) force field for 1 ns at 300K temperature. In each case was used a cubic box of a specific size with SPC/E (SPC216) water and centered the protein such that it left roughly 10 Å distance to the edge of the box. Thereafter the system was neutralized and subjected to steepest descent energy minimization to remove any overlapping contacts and reduce the maximum force in the system to 1000 kJ/mol/nm. This was followed by NVT equilibration, with 2 fs time step, using modified Berendsen thermostat with total simulation time of 100 ps under a temperature of 300 K. Subsequently the NPT equilibration of 100 ps using 2 fs time step at 1 atm was done using Parinello-Rahman pressure coupling. Structures during unconstrained dynamics simulation were recorded every 10 ps to give a total of 101 frames for the analysis. After completing the MD simulations, the second step of the workflow consisted in clustering the frames on the basis of their Root Mean Square Fluctuation (RMSF) using the g_cluster command of Gromacs utilities. The RMSF threshold was set to obtain only 2 clusters. It was assumed that the cluster which had the highest number of frames is the one having the more stable structures, while the one with a lower number of structures corresponds to less stable. These two states were assumed to represent the folded and unfolded states, respectively for calculation purposes. In the third step, one representative structure from the cluster was selected, that is closest to the cluster centroid and used for free energy calculation using the g_mmpbsa method (Kumari, et al., 2014). This method gives 3 types of energies: molecular mechanics potential energy, apolar and polar energy. The free energy is calculated by summing up all three energies. The unfolding free energy is the difference between the unfolded and the folded state ($\Delta$G). Finally the $\Delta\Delta$G was calculated taking the difference between the variant $\Delta$Gs and the wild-type. The obtained values were not scaled using any experimental reference.

***Kim Lab - University of Toronto (Group 6)***

The Kim Lab at the University of Toronto submitted three sets of predictions (G6-1, G6-2, G6-3) using the ELAPSIC algorithm (Berliner, et al., 2014; Witvliet, et al., 2016). In brief, a set of structural features describing the thermodynamics of the wild-type and variant proteins, and a set of sequential features, describing the impact that the variation is likely to have on the viability of the organism were considered, and the gradient-boosting decision tree (GBDT) algorithm (Friedman, 2002) was adopted to combine those features into a final score that correlates well with the variant-induced change in the Gibbs free energy of folding ($\Delta\Delta G_{fold}$) and binding ($\Delta\Delta G_{bind}$). FoldX (Guerois, et al., 2002), Stide (Frishman and Argos, 1995), MSMS (Sanner, et al., 1996), and a number of in-house developed routines were used to construct the set of structural features, and Provean (Choi, et al., 2012), chemical similarity of the wild-type and variant amino acids, and metrics evaluating the alignment between the protein in question and its closest structural template were used to construct the set of sequential features. The $\Delta\Delta G_{fold}$

predictor is trained using the experimental data in the Protherm (Kumar, et al., 2006) database, while the $\Delta\Delta G_{bind}$ predictor is trained using variations from the Skempi dataset (Moal and Fernandez-Recio, 2012). In both cases, the hyperparameters of the GBDT algorithm are selected by maximizing a metric which includes the accuracy of the predictor on the training dataset, evaluated using 6-fold cross-validation, as well as the accuracy of the predictor on a validation dataset containing benign variants as well as variants involved in disease. The purpose of this submission to the CAGI5 frataxin challenge was to validate the accuracy of the ELASPIC webserver (Witvliet, et al., 2016) and, more specifically, to make sure that no overfit on the features extracted from FoldX, Provean, and other tools was present. To that end, the first submission (G6-1) contains variant ΔΔG values calculated using the ELASPIC webserver, the second submission (G6-2) contains ΔΔG values calculated using FoldX, and the third submission (G6-3) contains the deleteriousness scores calculated using Provean.

## 3 Measures of performance

***Performance in regression mode***

For evaluating the performance of the methods in the regression task we compared the predicted and experimental values of the variation of free energy change upon amino acid substitution ($\Delta\Delta G^{H_20}$).

The standard scoring values calculated in our assessment are the Pearson, Spearman and Kendall-Tau correlation coefficients ($r_P$, $r_S$, *and* $r_{KT}$ respectively), the root mean square error (RMSE) and the mean absolute error (MAE). They are defined as follows:

$$r_P = \frac{n\sum_{i=1}^{n} y_i \bar{y}_i - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} \bar{y}_i\right)}{\sqrt{\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}\sqrt{\left[n\sum_{i=1}^{n} \bar{y}_i^2 - \left(\sum_{i=1}^{n} \bar{y}_i\right)^2\right]}} \quad [4]$$

$$r_S = \frac{n\sum_{i=1}^{n} r(y_i) r(\bar{y}_i) - \left(\sum_{i=1}^{n} r(y_i)\right)\left(\sum_{i=1}^{n} r(\bar{y}_i)\right)}{\sqrt{\left[n\sum_{i=1}^{n} r(y_i)^2 - \left(\sum_{i=1}^{n} r(y_i)\right)^2\right]}\sqrt{\left[n\sum_{i=1}^{n} r(\bar{y}_i)^2 - \left(\sum_{i=1}^{n} r(\bar{y}_i)\right)^2\right]}} \quad [5]$$

$$r_{KT} = \frac{2}{n(n-1)} = \sum_{i=1}^{n}\sum_{j=1}^{i-1} sign(y_i - y_j)\ sign(\bar{y}_i - \bar{y}_j) \quad [6]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}{n}} \quad [7]$$

$$MAE = \frac{\sum_{i=1}^{n} \left| y_i - \bar{y}_i \right|}{n} \quad [8]$$

where $y_i$ and $\bar{y}_i$ are the predicted and experimental $\Delta\Delta G^{H_2O}$ values respectively and $r(y_i)$, $r(\bar{y}_i)$ their ranks.

***Performance of the binary classifier***

For the assessment of the frataxin challenge we transformed the predictions in a binary classification task considering a threshold of -1.0 kcal/mol for discriminating between destabilizing variants ($\Delta\Delta G^{H_2O}$ < -1.0 kcal/mol), indicated as negative cases, and not destabilizing variant ($\Delta\Delta G^{H_2O}$ ≥ -1.0 kcal/mol), indicated as positive cases. A representation of the classification of the 8 points in the frataxin dataset is plotted in Fig. S1.

According to the previous classification scheme, for each submission we assessed the performance using the following metrics: true and false positive rates (*TPR, FPR*) and balanced accuracy ($BQ_2$)

$$FPR = \frac{FP}{FP + TN} \quad TPR = \frac{TP}{TP + FN}$$
$$BQ_2 = \frac{TP}{2P} + \frac{TN}{2N} \quad [9]$$

where N and P are the number of negative and positive cases, TN and FN are the true and false negative, and TP and FP are the true and false positive respectively. In our assessment we computed the Matthew's correlation coefficient *MCC* as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)\ (TP + FN)\ (TN + FP)\ (TN + FN)}} \quad [10]$$

and we also calculated the area under the Receiver Operating Characteristic (ROC) curve (AUC), by plotting the True Positive Rate (TPR) as a function of the False Positive Rate (FPR) at different classification thresholds.

## SUPPORTING REFERENCES


Berliner N, Teyra J, Colak R, Garcia Lopez S, Kim PM. 2014. Combining structural modeling with ensemble machine learning to accurately predict protein fold stability and binding affinity effects upon mutation. PLoS One 9(9):e107353.

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS One 7(10):e46688.

Correia AR, Pastore C, Adinolfi S, Pastore A, Gomes CM. 2008. Dynamics, stability and iron-binding activity of frataxin clinical mutants. FEBS J 275(14):3680-90.

Dayhoff MO, Schwartz RM, Orcutt BC. 1978. A model of evolutionary change in proteins. . Atlas Protein Sequence Struct. 5(3):345- 352.

Fariselli P, Martelli PL, Savojardo C, Casadio R. 2015. INPS: predicting the impact of non-synonymous variations on protein stability from sequence. Bioinformatics 31(17):2816-21.

Folkman L, Stantic B, Sattar A. 2014. Towards sequence-based prediction of mutation-induced stability changes in unseen non-homologous proteins. BMC Genomics 15 Suppl 1:S4.

Folkman L, Stantic B, Sattar A, Zhou Y. 2016. EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. J Mol Biol 428(6):1394-1405.

Friedman JH. 2002.

Stochastic gradient boosting. Computational Statistics & Data Analysis - Nonlinear methods and data mining 38(4):367-378

Frishman D, Argos P. 1995. Knowledge-based protein secondary structure assignment. Proteins 23(4):566-79.

Guerois R, Nielsen JE, Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320(2):369-87.

Heffernan R, Paliwal K, Lyons J, Dehzangi A, Sharma A, Wang J, Sattar A, Yang Y, Zhou Y. 2015. Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning. Sci Rep 5:11476.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. Proc Natl Acad Sci U S A 89(22):10915-9.

Karimi M, Shen Y. 2018. iCFN: an efficient exact algorithm for multistate protein design. Bioinformatics 34(17):i811-i820.

Katsonis P, Lichtarge O. 2014. A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. Genome Res 24(12):2050-8.

Kumar MD, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A. 2006. ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. Nucleic Acids Res 34(Database issue):D204-6.

Kumari R, Kumar R, Open Source Drug Discovery C, Lynn A. 2014. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54(7):1951-62.

Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol 157(1):105-32.

Lichtarge O, Bourne HR, Cohen FE. 1996. An evolutionary trace method defines binding surfaces common to protein families. J Mol Biol 257(2):342-58.

MacKerell AD, Jr., Banavali N, Foloppe N. 2000. Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56(4):257-65.

Moal IH, Fernandez-Recio J. 2012. SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. Bioinformatics 28(20):2600-7.

Neskey DM, Osman AA, Ow TJ, Katsonis P, McDonald T, Hicks SC, Hsu TK, Pickering CR, Ward A, Patel A, Yordy JS, Skinner HD and others. 2015. Evolutionary Action Score of TP53 Identifies High-Risk Mutations Associated with Decreased Survival and Increased Distant Metastases in Head and Neck Cancer. Cancer Res 75(7):1527-36.

Petrosino M, Pasquo A, Novak L, Toto A, Gianni S, Mantuano E, Veneziano L, Minicozzi V, Pastore A, Puglisi R, Capriotti E, Chiaraluce R and others. 2019. Characterization of human frataxin missense variants in cancer tissues. Human Mutation In press:10.1002/humu.23789.

Pudil P, Novovičová J, Kittler J. 1994. Floating search methods in feature selection. Pattern Recognition Letters. Pattern Recognition Letters 15(11):1119-1125.

Sanner MF, Olson AJ, Spehner JC. 1996. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 38(3):305-20.

Santoro MM, Bolen DW. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry 27(21):8063-8.

Savojardo C, Fariselli P, Martelli PL, Casadio R. 2016. INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics 32(16):2542-4.

Witvliet DK, Strokach A, Giraldo-Forero AF, Teyra J, Colak R, Kim PM. 2016. ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity. Bioinformatics 32(10):1589-91.

## SUPPORTING TABLES

**Supporting Table S1.** Experimental $\Delta\Delta G^{H_20}$ values from CD and Fluorescence

| DNA (hg38) | mRNA (NM_000144.4) | Protein (NP_000135.2) | $\Delta\Delta G_{CD}$ kcal/mol | $\Delta\Delta G_{Flu}$ kcal/mol |
|---|---|---|---|---|
| chr9:g.69053187A>G | c.311A>G | p.D104G | 0.2±0.7* | 0.6±0.5* |
| chr9:g.69053196C>T | c.320C>T | p.A107V | 0.8±1.0* | -0.7±0.5* |
| chr9:g.69053201T>C | c.325T>C | p.F109L | -2.1±0.6* | -3.6±0.4* |
| chr9:g.69053244A>C | c.368A>C | p.Y123S | -4.9±0.5* | -5.4±0.4* |
| chr9:g.69065035G>T | c.482G>T | p.S161I | -3.3±0.7* | -2.9±0.5* |
| chr9:g.69072648G>T | c.519G>T | p.W173C | -9.2±0.5* | -9.8±0.4* |
| chr9:g.69072671C>T | c.542C>T | p.S181F | -3.1±0.5* | -2.9±0.5* |
| chr9:g.69072734C>T | c.605C>T | p.S202F | -0.2±0.6* | -0.2±0.5* |

Variation of unfolding free energy change at zero solvent concentration ($\Delta\Delta G^{H_2O}$) obtained by circular dichroism ($\Delta\Delta G_{CD}$) and Fluorescence ($\Delta\Delta G_{Flu}$). These values are calculated using the experimental unfolding free energy of wild-type and variant protein recently published (Petrosino, et al., 2019). The variant p.W173C does not fold into a three-dimensional. Thus, for calculating the $\Delta\Delta G^{H_2O}$ of p.W173C we assumed that its $\Delta G^{H_2O}$ = 0 kcal/mol. It follows that $\Delta\Delta G^{H_2O}$ is equal to $-\Delta G^{H_2O}$ of the wild-type, which is -9.50 kcal/mol.

**Supporting Table S2.** Summary of the prediction submitted by the participants

| Group | Institution | Main Method | Submissions |
|---|---|---|---|
| Lichtarge Lab | Baylor College of Medicine, Houston (USA) | EA | G1-1 |
| Biocomputing Group | University of Bologna (Italy) | INPS-3D | G2-1 |
| Zhou Lab | Griffith University, Brisbane (Australia) | EASE-MM | G3-1,G3-2, G3-3 |
| Shen Lab | Texas A&M University, College Station (USA) | iCFN | G4-1, G4-2 |
| Pal Lab | Indian Institute of Science, Bangalore (India) | GROMACS | G5-1, G5-2 |
| Kim Lab* | University of Toronto (Canada) | ELAPSIC | G6-1, G6-2, G6-3 |

Predictions submitted by the participants are indicated with the number of the group and the number of the submission. * The submissions from Kim's Lab labelled with G6-R1, G6-R2 and G6-R3 are the reverse submissions of G6-1, G6-2, G6-3 respectively.

**Supporting Table S3.** Classification performance on the frataxin challenge dataset

| Group | Submission | $BQ_2$ | MCC | AUC | TN | FP | FN | TP |
|---|---|---|---|---|---|---|---|---|
| Kim Lab | G6-R1 | 0.80 | 0.60 | 0.93 | 3 | 2 | 0 | 3 |
| Biocomp | G2-1 | 0.80 | 0.60 | 0.80 | 3 | 2 | 0 | 3 |
| FoldX | - | 0.73 | 0.47 | 0.87 | 4 | 1 | 1 | 2 |
| Zhou Lab | G3-1 | 0.70 | 0.45 | 0.80 | 2 | 3 | 0 | 3 |
| I-Mutant2.0 | - | 0.70 | 0.45 | 0.73 | 2 | 3 | 0 | 3 |
| Lichtarge Lab | G1-1 | 0.63 | 0.26 | 0.87 | 3 | 2 | 1 | 2 |
| Shen Lab | G4-2 | 0.70 | 0.45 | 0.60 | 2 | 3 | 0 | 3 |
| Kim Lab | G6-R2 | 0.63 | 0.26 | 0.80 | 3 | 2 | 1 | 2 |
| Pal Lab | G5-1 | 0.63 | 0.26 | 0.67 | 3 | 2 | 1 | 2 |
| Shen Lab | G4-1 | 0.60 | 0.29 | 0.60 | 1 | 4 | 0 | 3 |
| Kim Lab | G6-R3 | 0.50 | 0.00 | 0.80 | 5 | 0 | 3 | 0 |
| Pal Lab | G5-2 | 0.50 | 0.00 | 0.27 | 5 | 0 | 3 | 0 |
| Kim Lab | G6-3 | 0.50 | 0.00 | 0.20 | 0 | 5 | 0 | 3 |
| Kim Lab | G6-2 | 0.50 | 0.00 | 0.20 | 0 | 5 | 0 | 3 |
| Kim Lab | G6-1 | 0.50 | 0.00 | 0.07 | 0 | 5 | 0 | 3 |

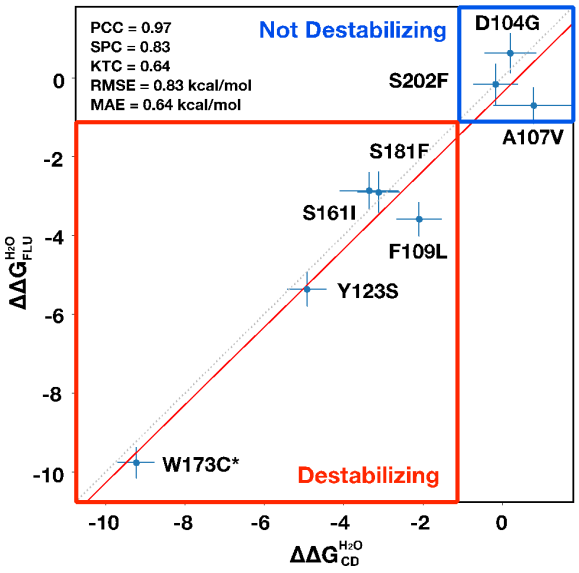
The measures of performance are defined above (section 3). $BQ_2$ = Balanced accuracy. MCC = Matthews correlation coefficient. AUC = Area Under the ROC Curve. TN = True Negative. FP = False Positive. FN = False Negative.TP = True Positive.

**Supporting Table S4.** Classification performance on the frataxin challenge dataset excluding p.W173C

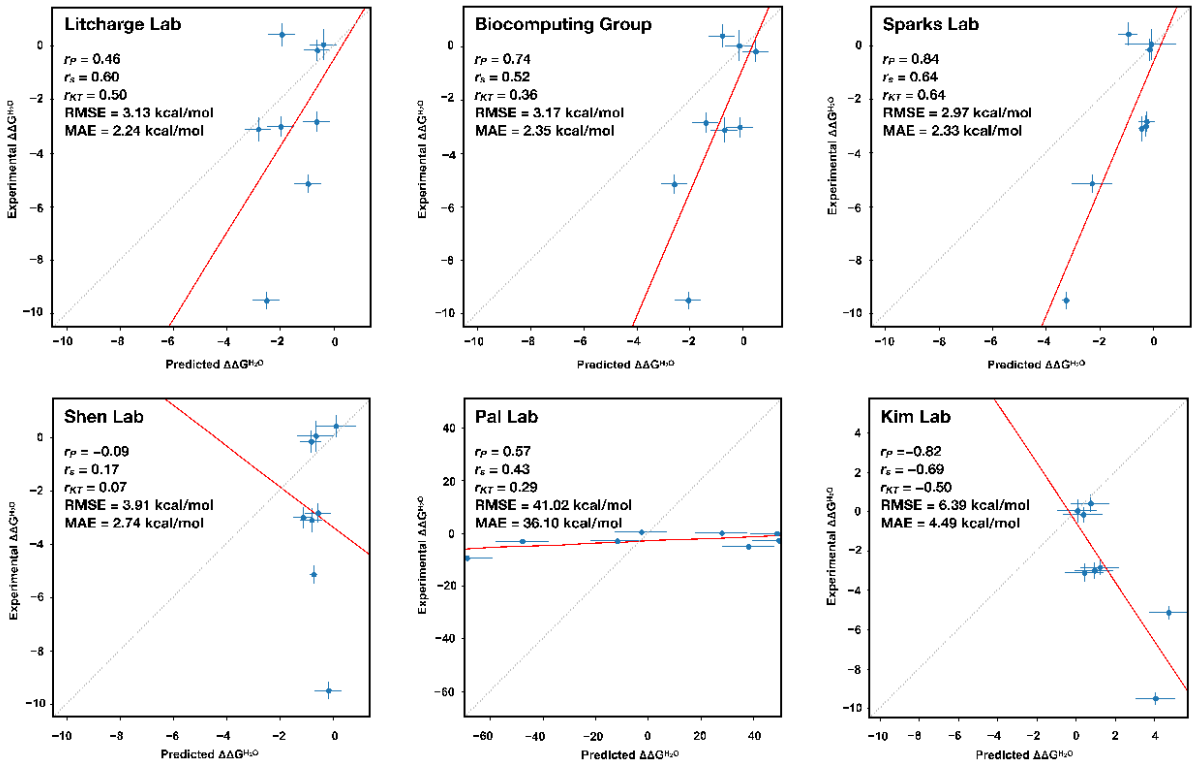
| Group | Submission | $BQ_2$ | MCC | AUC | TN | FP | FN | TP |
|---|---|---|---|---|---|---|---|---|
| Kim Lab | G6-R1 | 0.75 | 0.55 | 0.92 | 2 | 2 | 0 | 3 |
| Biocomp | G2-1 | 0.75 | 0.55 | 0.75 | 2 | 2 | 0 | 3 |
| FoldX | - | 0.71 | 0.42 | 0.83 | 3 | 1 | 1 | 2 |
| Shen Lab | G4-2 | 0.75 | 0.55 | 0.58 | 2 | 2 | 0 | 3 |
| Zhou Lab | G3-1 | 0.63 | 0.35 | 0.75 | 1 | 3 | 0 | 3 |
| I-Mutant2.0 | - | 0.63 | 0.35 | 0.67 | 1 | 3 | 0 | 3 |
| Shen Lab | G4-1 | 0.63 | 0.35 | 0.67 | 1 | 3 | 0 | 3 |
| Lichtarge Lab | G1-1 | 0.58 | 0.17 | 0.83 | 2 | 2 | 1 | 2 |
| Kim Lab | G6-R2 | 0.58 | 0.17 | 0.75 | 2 | 2 | 1 | 2 |
| Pal Lab | G5-1 | 0.58 | 0.17 | 0.58 | 2 | 2 | 1 | 2 |
| Kim Lab | G6-R3 | 0.50 | 0.00 | 0.75 | 4 | 0 | 3 | 0 |
| Pal Lab | G5-2 | 0.50 | 0.00 | 0.33 | 4 | 0 | 3 | 0 |
| Kim Lab | G6-3 | 0.50 | 0.00 | 0.25 | 0 | 4 | 0 | 3 |
| Kim Lab | G6-2 | 0.50 | 0.00 | 0.25 | 0 | 4 | 0 | 3 |
| Kim Lab | G6-1 | 0.50 | 0.00 | 0.08 | 0 | 4 | 0 | 3 |

The measures of performance are defined above (section 3). $BQ_2$ = Balanced accuracy. MCC = Matthews correlation coefficient. AUC = Area Under the ROC Curve. TN = True Negative. FP = False Positive. FN = False Negative.TP = True Positive.

## SUPPORTING FIGURES




**Supporting Figure S1**. Comparison of the experimental $\Delta\Delta G^{H_2O}$ values from CD and Fluorescence. Classification of the variants in destabilizing (in the red box) and not destabilizing (in the blue box) using a $\Delta\Delta G^{H_2O}$ discrimination threshold of -1.0 kcal/mol.




**Supporting Figure S2**. Regression analysis for the best prediction submitted by each group. The measures of performance $r_P$, $r_S$, $r_{KT}$, *RMSE and MAE* are defined above.