

Mass balance approximation of unfolding boosts potential-based protein stability predictions

Ivan Rossi¹  | Guido Barducci¹  | Tiziana Sanavia¹  | Paola Turina²  |
Emidio Capriotti^{2,3}  | Piero Fariselli¹ 

¹Department of Medical Sciences, University of Torino, Torino, Italy

²Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy

³Computational Genomics Platform, IRCCS University Hospital of Bologna, Bologna, Italy

Correspondence

Tiziana Sanavia, Department of Medical Sciences, University of Torino, Via Santena 19, 10126 Torino, Italy.
Email: tiziana.sanavia@unito.it

Piero Fariselli, Department of Medical Sciences, University of Torino, Via Santena 19, 10126 Torino, Italy.
Email: piero.fariselli@unito.it

Emidio Capriotti, Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy.
Email: emidio.capriotti@unibo.it

Funding information

Italian Ministry for Education, University and Research; European Union's Horizon 2020, Grant/Award Numbers: 101017598, 101017549

Review Editor: Nir Ben-Tal

Abstract

Predicting protein stability changes upon single-point mutations is crucial in computational biology, with applications in drug design, enzyme engineering, and understanding disease mechanisms. While deep-learning approaches have emerged, many remain inaccessible for routine use. In contrast, potential-like methods, including deep-learning-based ones, are faster, user-friendly, and effective in estimating stability changes. However, most of them approximate Gibbs free-energy differences without accounting for the free-energy changes of the unfolded state, violating mass balance and potentially reducing accuracy. Here, we show that incorporating mass balance as a first approximation of the unfolded state significantly improves potential-like methods. While many machine-learning models implicitly or explicitly use mass balance, our findings suggest that a more accurate unfolded-state representation could further enhance stability change predictions.

KEYWORDS

deep learning models, Gibbs free energy, mass-balance correction, potential-like methods, protein stability prediction, single-point mutations

1 | INTRODUCTION

Predicting protein stability changes upon single-point mutations is a longstanding challenge in computational biology (Benevenuta et al., 2022; Pucci et al., 2022; Sanavia et al., 2020), with significant implications in drug design, enzyme engineering, and understanding disease mechanisms (Thomas et al., 1995). Protein stability is typically quantified by measuring the Gibbs free energy change (ΔG) between the folded and unfolded states

$$\Delta G = G_F - G_U.$$

(1)

However, mutations can dramatically alter this delicate balance. Destabilizing mutations are often linked to diseases (Martelli et al., 2016) such as cancer (Petrosino et al., 2021), while stabilizing mutations can enhance protein function and resilience, especially in industrial and therapeutic settings (Coluzza, 2017; Korendovych & DeGrado, 2020).

From the experimental point of view, the measure of interest is the difference of the unfolding free energy between the mutated and wild-type proteins ($\Delta\Delta G$), calculated as

$$\Delta\Delta G = (G_F(m) - G_F(w)) - (G_U(m) - G_U(w)), \quad (2)$$

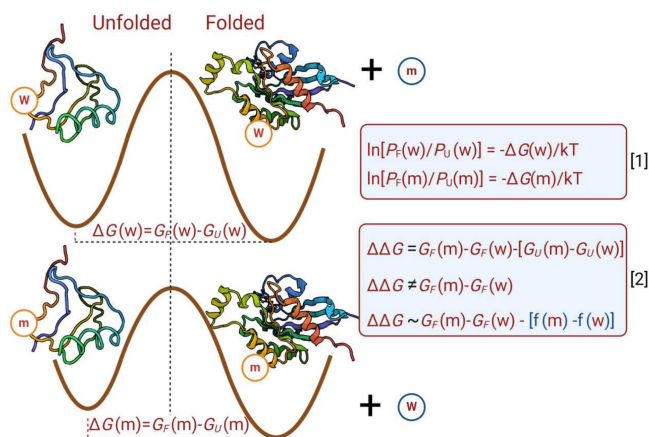


FIGURE 1 Thermodynamics of the variation of the folding free energy upon single point mutation, considering mutated (m) and wild-type (w) states. In box [1] the relation between probability and free energy of folding is reported. In box [2], the correct measure of the difference of the unfolding free energy between the mutated and wild-type proteins, considering the difference between the folded and unfolded states is reported (first equation); however, some potential-like methods approximate it using the difference of the folding state free energy, neglecting the effect of the unfolded states (box [2], second equation). A first approximation can be obtained by adding a mass-balance correction (also a kind of solvation term) to the folding free energy difference (box [2], third equation).

where m and w stand for mutant and wild-type (Figure 1).

$$P_F(w) + P_U(m) = P_F(m) + P_U(w), \quad (3)$$

where P represents the concentration of the protein either in the wild-type (w) or mutant (m) forms, both in the folded (F) or unfolded (U) states. It can be noticed that this kind of “reaction” corresponds to that used in Free-Energy Perturbation (FEP) calculations (York, 2023; Zwanzig, 1954), a widely used method to calculate ΔG differences in molecular modeling and drug design.

The folding free energy difference between two protein variants depends on both the folded and unfolded states of each sequence. Studies using molecular dynamics, based on *Alchemical Free Energy Perturbation* (Kurniawan & Ishida, 2023; York, 2023), have demonstrated that accurately modeling the unfolded state is crucial for achieving high predictive performance (Kurniawan & Ishida, 2023), though such approaches require computationally expensive methods. Similar statistical-mechanics approaches describing the contribution of the unfolded state have been presented by Bastolla and coworkers (Arenas et al., 2020; Bastolla, 2014; Minning et al., 2013).

In recent years, deep learning-based approaches have significantly advanced the field of protein stability prediction. Despite their success, these models require substantial computational resources and are

sometimes inaccessible for routine or high-throughput applications (Sanavia et al., 2020).

In contrast, potential-like methods, such as those utilizing empirical energy functions like FoldX (Schymkowitz et al., 2005) structure-based protein-language models such as ProteinMPNN (Dauparas et al., 2022) and ESM-IF1 (Hsu et al., 2022), and methods that directly address the calculation of $\Delta\Delta G$ upon mutation using deep neural networks, such as Pythia (Sun et al., 2025), offer faster and more accessible alternatives. These methods estimate stability changes by calculating either atomistic interactions or the likelihood of an amino acid in a given structural context of the protein. Pythia, for example, employs a self-supervised learning framework to perform zero-shot $\Delta\Delta G$ predictions across a large protein sequence space, offering ultrafast computational performance.

However, one fundamental limitation of the potential-like methods is their simplified approach to Gibbs free energy calculations, where only the folded states $\{G_F(x)\}$ (i.e., the protein structure) are considered. This simplification leads to the following approximation for the mutant (m) and wild-type (w):

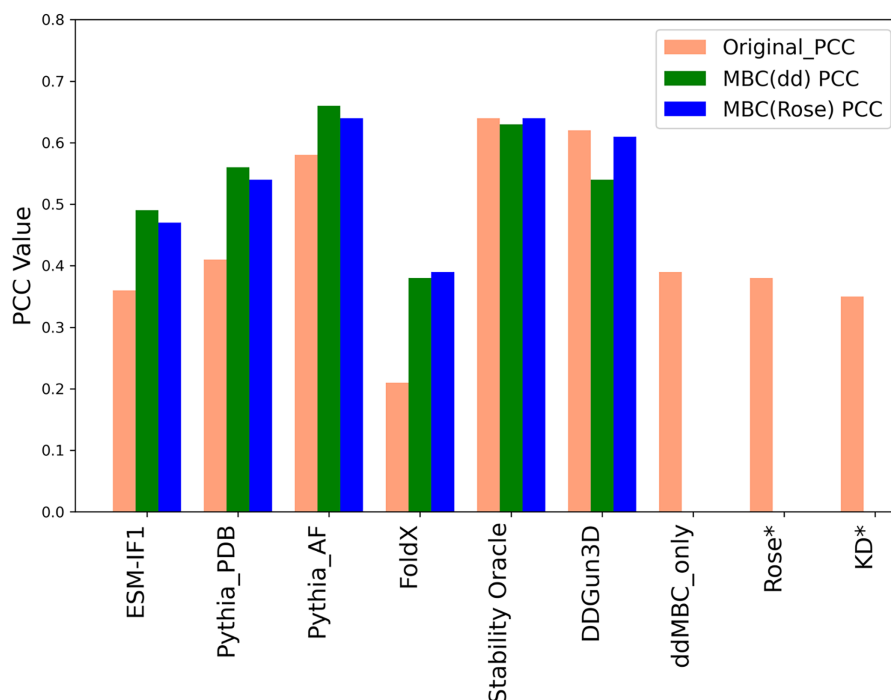
$$\Delta\Delta G_{approx} = (G_F(m) - G_F(w)). \quad (4)$$

Under this approximation, the second term of Equation (2), describing the ΔG between the unfolded states of the two protein sequences, is typically neglected due to the difficulty of properly defining and measuring it. However, this approximation might not always hold, since, for example, different inter-residue interactions and degrees of freedom between wild-type and mutant might persist in the unfolded state. An additional contribution might be the difference in free energy of solvation for the amino acids involved in the mutation (Rose et al., 1985). It should also be observed that the $\Delta\Delta G$ expression is a difference of two terms, and neglecting one of them could lead to significant deviations from the correct solution. Furthermore, neglecting this second term also implicitly means violating the mass conservation for the process, as Gibbs free energy is defined for closed systems where mass is conserved.

Considering the extreme flexibility of the neural-networks in implicitly modeling all terms of Equation (1), the approximation of ΔG between the unfolded states of the two protein sequences equal to zero should not affect, in principle, models that explicitly incorporate the protein-sequence composition change among their input features (e.g., I-mutant [Capriotti et al., 2005], ACDC-NN [Pancotti et al., 2021], Stability Oracle [Diaz et al., 2024]). However, as previously mentioned, this approximation might become relevant for models that do not compensate for it, such as most “potential-like” methods.

To address this gap, we propose a novel correction that incorporates “mass balance” back into potential-like scoring methods, improving the accuracy of protein

FIGURE 2 Comparison of Pearson correlation coefficient obtained on S461 dataset between the original method (pink bar) and its adjusted version with mass-balance correction, using both MBC(dd) (green bar) and Rose scale (blue bar). ddMMBC_only represents the prediction made using only the fitted mutation coefficients without incorporating a method. KD* (Kyte & Doolittle, 1982) and Rose* (Rose et al., 1985) are the scale difference values without any fitting.



stability predictions without compromising their usually high computational efficiency. By retrofitting these potential-like models with this extra term, which we call *mass-balance correction* (MBC), our approach adjusts for a key flaw in the evaluation of $\Delta\Delta G$, significantly enhancing the prediction accuracy without any reparameterization of the original model.

Furthermore, the obtained performance for some of these modified methods is comparable, or even better, to those of state-of-the-art models such as Stability Oracle, providing a valuable tool for researchers needing rapid stability assessments.

2 | RESULTS

2.1 | Incorporating mass-balance information as a first approximation of the unfolded state

We first evaluated the performance of three different potential-like methods, representing three different approaches to $\Delta\Delta G$ calculation, with and without the MBC correction, and then compared them to the results of the DDGun3D (Montanucci et al., 2019; Montanucci et al., 2022) “untrained” benchmark model. DDGun3D explicitly incorporates a form of MBC by considering the hydrophobicity difference between mutated and wild-type residues, establishing it as a suitable reference benchmark. We also derived the data-driven MBC term, referred to as ddMMBC_only hereafter, by fitting it to the training set using ridge

regression implemented in Scikit-learn (Pedregosa et al., 2011) with default parameters.

The MBC(dd) term was then compared with the Kyte–Doolittle (Kyte & Doolittle, 1982) and Rose (Rose et al., 1985) scales for scoring the difference of hydrophobicity and solvation, respectively, as first approximations of the unfolded state. Additionally, we included a comparison with the Stability Oracle model, a recent state-of-the-art deep learning-based method. We used the S461 dataset (Hernández et al., 2023) as the test set to perform comparisons (see Table S1 and Figure 2).

The three potential-like methods considered are:

1. **ESM-IF1**, a large protein-language model (PLM) trained to predict a protein sequence likelihood from its backbone atom coordinates.
2. **FoldX**, a widely-used all-atom knowledge-based potential for fast and quantitative estimation of the importance of the interactions contributing to the stability of proteins.
3. **Pythia**, a self-supervised graph neural network tailored for zero-shot $\Delta\Delta G$ predictions, large-scale residue scanning, and missing-residue probability prediction.

On the S461 test set, all methods showed visible performance boosts, with increased Pearson correlation coefficients (PCC) compared to the original methods (for more detail see Table S2) and with Pythia/MBC(dd) being the top performer.

Although we used PDB structures to train our model, we observed that the performance of both the baseline ESM-IF1 and Pythia models noticeably

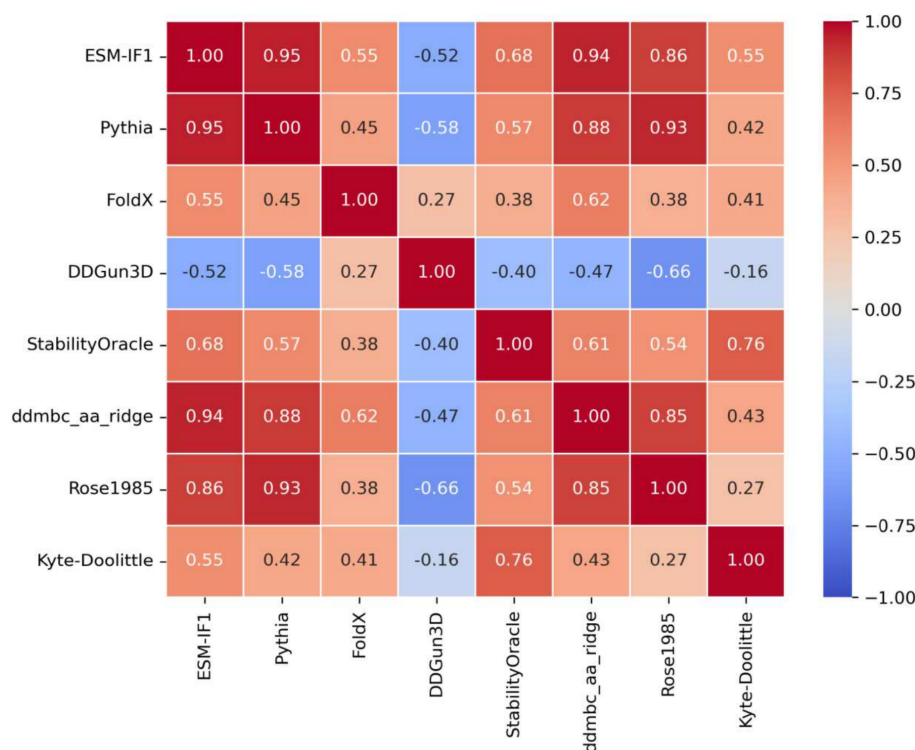


FIGURE 3 Correlation among the residue coefficients of the different methods and two hydrophobicity scales (Kyte–Doolittle and Rose). DDGun3D contains explicitly the difference of the Kyte–Doolittle values. ddmcb_aa_ridge is highly correlated with the Rose scale.

depends on the type of structure used. Namely, the performance of both of these methods is higher if AlphaFold (Jumper et al., 2021) models are used instead of experimental x-ray structures from PDB. This is probably due to the way these methods have been parameterized: for both ESM-IF1 and Pythia training sets, the percentage of AlphaFold structure exceeds 90%, thus any bias that may be introduced by using models instead of experimental structures is captured by the methods. Nonetheless, the MBC(dd) validity is not affected by the choice of model origin (Table S2): using the MBC(dd) correction derived from the PDB structures on the same test sets, but giving in input the AlphaFold structures instead of those from PDB, results in models that are even better performing. Both Stability Oracle and Pythia/MBC(dd)-AF achieve a PCC higher than the one obtained by the benchmark DDGun3D method (PCC: 0.62), whose performance on the S461 data set is very strong (Figure 2 and Table S2). We also computed the MBC(dd) correction for Stability Oracle and DDGun3D benchmarks, and, as expected, the result is worse for both methods (Figure 2). This supports our expectation that these methods, which already account for descriptors of the unfolded state in their input, such as the stoichiometry of the mutation process, are effectively capturing the correct underlying physics without requiring any posterior corrections. See Figure 2 for a summary of the results, while detailed reports are included within the Supporting Information.

2.2 | Comparison between residue specific-coefficients and experimental solvation scales

We performed a Pearson correlation analysis among the residue-specific parameters fitted using the VBS3322 dataset (see Methods section) to assess their consistency across different methods. Additionally, we included solvation and hydrophobicity scale values in the correlation comparison to evaluate their relationship with the fitted parameters. As shown in Figure 3, the amino acid-specific parameters (a_1 to a_{20}) exhibit strong correlations across the potential-like methods. Furthermore, these fitted parameters show a notable correlation with the Kyte and Doolittle hydrophobicity scale and an even stronger correlation with the experimentally derived Rose scale, which was specifically designed to predict the average change in solvent accessible surface area of amino acids upon folding.

In agreement with these observations, we then computed a new MBC based on the Rose scale, referred to as MBC(Rose). This correction was derived using a two-parameter linear combination between the original-method delta and the Rose-scale delta (see Equation 8), with results summarized in Figure 2. The performance of MBC(Rose) is consistent with, or in some cases superior to, that obtained by the MBC(dd) approach.

As a further validation, we computed the Pythia/MBC(dd) and Pythia/MBC(Rose) scores using the

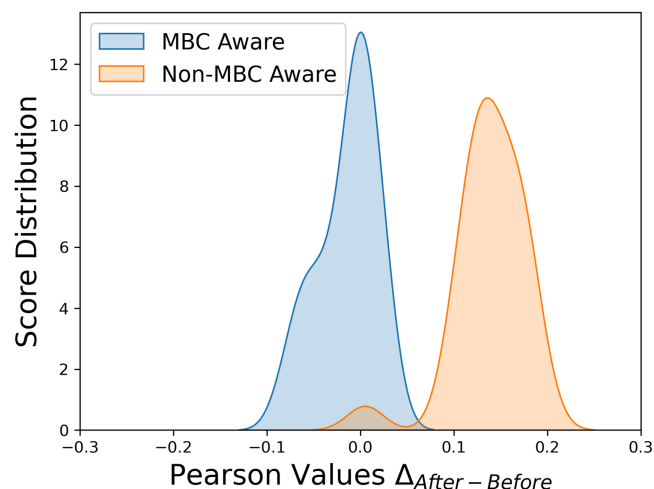


FIGURE 4 Comparison with methods that directly include a mass-balance correction (MBC aware) with those that compute only a difference between the folding states (non-MBC aware). The plot reports the distribution of the difference between the Pearson's correlation after and before the mass-balance term is added. Source: The data are from Reeves and Kalyaanamoorthy (2024).

parameters derived from our VBS3322 training set and tested them on the independent mega-scale dataset (Tsuboyama et al., 2023), which was not used in the parameter derivation. The results show an improvement (PCC: +0.07) over the original Pythia score, achieving a PCC close to 0.70 and an RMSE of 1.43 kcal/mol. Further details are provided in Table S5.

2.3 | Generalization of the mass-balance correction across different methods

Reeves and Kalyaanamoorthy (2024) recently highlighted that structure-based and sequence-based PLMs can be linearly combined to improve the performance, indicating that these two methodological classes provide complementary information. They further noted that “... PSLMs can be reliably augmented with physicochemical properties to exceed the median performance of the benchmark stability predictor ...”. This aligns with our model, since.

$$\begin{aligned}\Delta\Delta G &= (G_F(m) - G_F(w)) - (G_U(m) - G_U(w)) \\ &= \Delta\Delta H_F - T\Delta\Delta S_F.\end{aligned}\quad (5)$$

Thus, it is reasonable to think that both sequence and structure-based terms correspond to the $\Delta\Delta G$ term for the unfolded and folded states, respectively.

Additionally, the molecular volume and the solvent-accessible surface area (SASA) play a crucial role in estimating the solvation energy changes (a large part of $\Delta\Delta S_F$) when a molecule interacts with a solvent. This

concept has been widely applied in different implicit solvation models, such as the GBSA family of models (Godschalk et al., 2013; Onufriev & Case, 2019).

From this perspective, the MBC can be seen as a proxy of this information. Our model provides a simple, yet effective, way to estimate the Gibbs free energy difference between wild-type and mutated proteins in their unfolded states. Alternatively, it can be interpreted as describing the differences in the entropy of folding (which is largely dictated by solvation effects), while the potential-like methods primarily approximate the enthalpic contribution to the folding.

We thus tested whether our approach is able to generalize across different methods, considering the predictions of 48 methods on the S461 dataset taken from Reeves and Kalyaanamoorthy (2024) and supplemented by the Pythia data. To fit the two scale values related to the method and to the Rose scale (see Methods equation 8) we used the prediction reported by the same authors on the Ssym dataset (Pucci et al., 2018).

Figure 4 reports the obtained results, grouping the methods into MBC-aware (i.e. trained with some mass-balance correction) and non-MBC-aware approaches (such as PLMs, which does not account for the mass balance). As expected, the MBC approach notably improved the performance of non-MBC-aware methods.

More specifically, Table S3 reports consistent improvements across most of the potential-like methods, with substantial increases in PCC for Pythia (+0.19), ProteinMPNN (+0.18), and FoldX (+0.18), leading to performance close to the current state-of-the-art for Pythia/MBC(Rose). In contrast, methods that already incorporate MBC information exhibited only minor improvements (e.g., Stability Oracle: +0.01 PCC) or even decreasing performance (e.g., PoPMuSiC (Dehouck et al., 2009): −0.04 PCC).

3 | CONCLUSIONS

The mass-balance correction (MBC), whether data-driven or based on an experimentally derived scale, demonstrates broad applicability, enhancing the performance of various potential-like methods developed through different approaches. These include knowledge-based potentials, sequence- and structure-based protein language models (PLMs), and a self-supervised deep graph-neural network. Notably, MBC achieves these improvements without requiring any reparameterization of the base methods and with negligible additional computational cost.

In several cases, the enhancement of the performance due to MBC is substantial (Table S2). Specifically, in the case of Pythia, the results are particularly notable, bringing Pythia/MBC close to state-of-the-art

performance (from 0.41 to 0.56 PCC; Table S2) while also addressing the method's poor antisymmetry (Table S6). More generally, MBC preserves the antisymmetry of the improved methods whenever the original methods exhibit this property.

This finding strongly supports our hypothesis that a better description of the unfolded state of the proteins might be a necessary step to improve the current state-of-the-art protein stability-change predictions. The MBC correction is just a simple, yet effective, zero-order correction. Thus, it is clearly possible to envision more sophisticated and, eventually, better-performing methods. Nonetheless, we believe that the simplicity of our approach has its own merits per se, since it allows the retrofitting of several existing approaches, achieving good performance and avoiding extra computational costs.

4 | MATERIALS AND METHODS

4.1 | Datasets composition

The main training set used in this work, namely VBS3322, consists of 3322 mutations obtained by combining the VariBench (Sasidharan Nair & Vihinen, 2013) and the S2648 (Dehouck et al., 2009) data sets. In the cases where the same mutation is reported in both data sets, the VariBench value is considered. We also augmented the dataset by including the antisymmetric complement of each mutation, as suggested in a previous work (Capriotti et al., 2008).

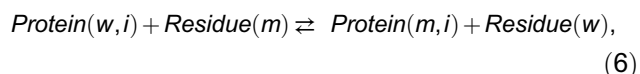
The test set used for the benchmarking is the S461. For all the structures that showed missing backbone atoms, we preprocessed the structure using the PDBFixer utility (Eastman et al., 2017).

The FoldX results used for both training and evaluation have already been published (Pancotti et al., 2022), while Stability Oracle results for the S461 dataset have been computed from the data provided by its authors on Github (<https://github.com/danny305/StabilityOracle>).

4.2 | Mass balance correction

The simplest approach to calculate the $\Delta\Delta G$ for the sequence-mutation process is to assume that the second term of Equation (1) depends only on the amino acids involved in the mutation.

This simplification leads to the following reaction, considering the wild-type (w) and mutated (m) residues:



where $\text{Protein}(x, i)$ represents a protein with residue x in position i , while $\text{Residue}(x)$ refers to a single amino

acid. Conceptually, this corresponds to estimating the difference in the (effective) Gibbs free energy of solvation for two amino acids in solution and within the field of the protein. From another perspective, this approach approximates the free energy of the unfolded state as the sum of independent contributions from each amino acid. Physically, these contributions may arise from the conformational entropy of both the side chain and main chain, as well as their interactions with the solvent. Under this approximation, all terms disappear except for the contributions of the wild-type and mutated amino acids, significantly simplifying the calculation.

4.3 | Input encoding

We encode the mutation in the sequence as a 20-element array, one element for each of the natural amino acids, and we encode their occurrence (O) as -1 for the wild-type amino acid and $+1$ for the substitution.

The modified expression to calculate $\Delta\Delta G$ is then expressed as a linear combination of the original-method score (S) for the wild-type and the mutated protein:

$$\Delta\Delta G = a_0(S(m) - S(w)) + \sum_{i=1}^{20} a_i O_i, \quad (7)$$

where the first term is the (scaled) output of the original method and the second term represents the pseudo- $\Delta\Delta G$ of solvation for the amino acids involved in the mutation (the data-driven MBC).

The first term $S(x)$ thus corresponds to the $\Delta\Delta G$ predicted by the original method, while the second term depends on amino acid-related parameters.

It should also be observed that Equation (7), being antisymmetric by definition, preserves the antisymmetry in the prediction of the original methods, if present.

The 21 coefficients for the linear model above can be easily derived via a simple linear regression with respect to the training set.

Similarly, the MBC(Rose) correction is computed as a two-parameter linear combination of the original-method score (S) and Rose-scale delta.

$$\Delta\Delta G = a_0(S(m) - S(w)) + a_1(R(m) - R(w)), \quad (8)$$

where $R(m)$ and $R(w)$ are the values of the Rose scale for the mutated- and wild-type amino acid, respectively.

4.4 | Measures of performance

To evaluate the performance of the methods in the regression task, we compared the predicted (p) and experimental (e) values of the variation of unfolding free

energy change upon mutation ($\Delta\Delta G$). The standard scoring values calculated in our assessment are the Pearson correlation coefficients (PCC) and the root mean square error (RMSE), defined as follows:

$$\text{PCC} = \frac{\sum_{i=1}^N (\Delta\Delta G_e - \overline{\Delta\Delta G_e}) (\Delta\Delta G_p - \overline{\Delta\Delta G_p})}{\sqrt{\sum_{i=1}^N (\Delta\Delta G_e - \overline{\Delta\Delta G_e})^2} \sqrt{\sum_{i=1}^N (\Delta\Delta G_p - \overline{\Delta\Delta G_p})^2}}, \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\Delta\Delta G_p - \overline{\Delta\Delta G_e})^2}{N}}, \quad (10)$$

where $\overline{\Delta\Delta G_p}$ and $\overline{\Delta\Delta G_e}$ are the average predicted and experimental $\Delta\Delta G$ values, respectively.

AUTHOR CONTRIBUTIONS

Ivan Rossi: Conceptualization; software; methodology; investigation; validation; formal analysis; writing – review & editing; writing – original draft; visualization. **Guido Barducci:** Methodology; software; investigation; validation; writing – review & editing; formal analysis. **Tiziana Sanavia:** Conceptualization; investigation; writing – review & editing. **Paola Turina:** Data curation; methodology; investigation; validation; writing – review & editing; visualization. **Emidio Capriotti:** Conceptualization; data curation; investigation; validation; formal analysis; writing – original draft; writing – review & editing; visualization. **Piero Fariselli:** Conceptualization; methodology; software; investigation; validation; supervision; project administration; writing – review & editing; resources; funding acquisition.

ACKNOWLEDGMENTS

The authors thank the Italian Ministry for Education, University and Research under the program “Ricerca Locale ex-60%” and PNRR M4C2 HPC—1.4 “CENTRI NAZIONALI”—Spoke 8 for fellowship support. In addition, the authors thank the European Union's Horizon 2020 projects Brainteaser (Grant Agreement ID: 101017598) and GenoMed4All (Grant Agreement ID: 101017549). Dr. Tiziana Sanavia thanks PRIN project “Investigating the role of NF-YA isoform/lncRNA axis in mesoderm specification” of the Italian Ministry for Education, University and Research (Grant ID: 20224TWKNJ). Ivan Rossi would like to dedicate this work to his past supervisor Prof. Donald G. Truhlar. Open access publishing facilitated by Università degli Studi di Torino, as part of the Wiley - CRUI-CARE agreement.

DATA AVAILABILITY STATEMENT

The python codes and the data used in this study can be downloaded from Github at <https://github.com/combiomed-unito/ddMBC>.

ORCID

Ivan Rossi  <https://orcid.org/0000-0002-2077-7496>
 Guido Barducci  <https://orcid.org/0009-0005-1052-8495>
 Tiziana Sanavia  <https://orcid.org/0000-0003-3288-0631>
 Paola Turina  <https://orcid.org/0000-0001-9174-511X>
 Emidio Capriotti  <https://orcid.org/0000-0002-2323-0963>
 Piero Fariselli  <https://orcid.org/0000-0003-1811-4762>

REFERENCES

- Arenas M, Weber CC, Liberles DA, Bastolla U. ProtASR: an evolutionary framework for ancestral protein reconstruction with selection on folding stability. *Syst Biol*. 2017;66:1054–64.
- Bastolla U. Detecting selection on protein stability through statistical mechanical models of folding and evolution. *Biomolecules*. 2014;4:291–314.
- Benevenuta S, Birolo G, Sanavia T, Capriotti E, Fariselli P. Challenges in predicting stabilizing variations: an exploration. *Front Mol Biosci*. 2022;9:1075570.
- 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:W306–10.
- Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. *BMC Bioinform*. 2008;9(Suppl 2):S6.
- Coluzza I. Computational protein design: a review. *J Phys Condens Matter*. 2017;29:143001.
- Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*. 2022;378:49–56.
- 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;25:2537–43.
- Diaz DJ, Gong C, Ouyang-Zhang J, Loy JM, Wells J, Yang D, et al. Stability Oracle: a structure-based graph-transformer framework for identifying stabilizing mutations. *Nat Commun*. 2024;15:6170.
- Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*. 2017;13(7):e1005659. <https://doi.org/10.1371/journal.pcbi.1005659>
- Godschalk F, Genheden S, Söderhjelm P, Ryde U. Comparison of MM/GBSA calculations based on explicit and implicit solvent simulations. *Phys Chem Chem Phys*. 2013;15:7731–9.
- Hernández IM, Dehouck Y, Bastolla U, López-Blanco JR, Chacón P. Predicting protein stability changes upon mutation using a simple orientational potential. *Bioinformatics*. 2023;39(1):btad011. <https://doi.org/10.1093/bioinformatics/btad011>
- Hsu C, Verkuil R, Liu J, Lin Z, Hie B, Sercu T, et al. Learning inverse folding from millions of predicted structures. *Proceedings of the 39th international conference on machine learning*. Westminster: PMLR; 2022. p. 8946–70 [cited 10 Dec 2024]. Available from: <https://proceedings.mlr.press/v162/hsu22a.html>
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–9.
- Korendovych IV, DeGrado WF. De novo protein design, a retrospective. *Q Rev Biophys*. 2020;53:e3.
- Kurniawan J, Ishida T. Comparing supervised learning and rigorous approach for predicting protein stability upon point mutations in difficult targets. *J Chem Inf Model*. 2023;63:6778–88.

- Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157:105–32.
- Martelli PL, Fariselli P, Savojardo C, Babbi G, Aggazio F, Casadio R. Large scale analysis of protein stability in OMIM disease related human protein variants. *BMC Genomics.* 2016;17(Suppl 2):397.
- Minning J, Porto M, Bastolla U. Detecting selection for negative design in proteins through an improved model of the misfolded state. *Proteins.* 2013;81:1102–12.
- Montanucci L, Capriotti E, Birolo G, Benevenuta S, Pancotti C, Lal D, et al. DDGun: an untrained predictor of protein stability changes upon amino acid variants. *Nucleic Acids Res.* 2022;50(W1):gkac325. <https://doi.org/10.1093/nar/gkac325>
- Montanucci L, Capriotti E, Frank Y, Ben-Tal N, Fariselli P. DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC Bioinform.* 2019;20(Suppl 14):335.
- Onufriev AV, Case DA. Generalized born implicit solvent models for biomolecules. *Annu Rev Biophys.* 2019;48:275–96.
- Pancotti C, Benevenuta S, Birolo G, Alberini V, Repetto V, Sanavia T, et al. Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. *Brief Bioinform.* 2022;23:bbab555.
- Pancotti C, Benevenuta S, Repetto V, Birolo G, Capriotti E, Sanavia T, et al. A deep-learning sequence-based method to predict protein stability changes upon genetic variations. *Genes (Basel).* 2021;12:911.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–30.
- Petrosino M, Novak L, Pasquo A, Chiaraluce R, Turina P, Capriotti E, et al. Analysis and interpretation of the impact of missense variants in cancer. *Int J Mol Sci.* 2021;22:5416.
- Pucci F, Bernaerts KV, Kwasigroch JM, Rooman M. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics.* 2018;34:3659–65.
- Pucci F, Schwersensky M, Rooman M. Artificial intelligence challenges for predicting the impact of mutations on protein stability. *Curr Opin Struct Biol.* 2022;72:161–8.
- Reeves S, Kalyanamoorthy S. Zero-shot transfer of protein sequence likelihood models to thermostability prediction. *Nat Mach Intell.* 2024;6:1063–76.
- Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. Hydrophobicity of amino acid residues in globular proteins. *Science.* 1985;229:834–8.
- Sanavia T, Birolo G, Montanucci L, Turina P, Capriotti E, Fariselli P. Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine. *Comput Struct Biotechnol J.* 2020;18:1968–79.
- Sasidharan Nair P, Vihinen M. VariBench: a benchmark database for variations. *Hum Mutat.* 2013;34:42–9.
- 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:W382–8.
- Sun J, Zhu T, Cui Y, Wu B. Structure-based self-supervised learning enables ultrafast protein stability prediction upon mutation. *Innovation (Camb).* 2025;6:100750.
- Thomas PJ, Qu BH, Pedersen PL. Defective protein folding as a basis of human disease. *Trends Biochem Sci.* 1995;20:456–9.
- Tsuboyama K, Dauparas J, Chen J, Laine E, Mohseni Behbahani Y, Weinstein JJ, et al. Mega-scale experimental analysis of protein folding stability in biology and design. *Nature.* 2023;620:434–44.
- York DM. Modern alchemical free energy methods for drug discovery explained. *ACS Phys Chem Au.* 2023;3:478–91.
- Zwanzig RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys.* 1954;22:1420–6.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rossi I, Barducci G, Sanavia T, Turina P, Capriotti E, Fariselli P. Mass balance approximation of unfolding boosts potential-based protein stability predictions. *Protein Science.* 2025;34(5):e70134. <https://doi.org/10.1002/pro.70134>