

MuFaDDG: a sequence-based multiscale feature fusion framework for protein stability changes prediction

Jianting Gong^{1,2,†, }, Pengjia Ma^{1,2,†}, Zilin Ren^{3, }, Si Li², Zhiguo Fu², Pingping Sun^{2,*}, Ming Ni^{1,*, }, Xiaochen Bo^{1,*, }

¹Academy of Military Medical Sciences, Beijing, 100850, China

²School of Information Science and Technology, Northeast Normal University, Changchun, 130117, China

³State Key Laboratory of Pathogen and Biosecurity, Key Laboratory of Jilin Province for Zoonosis Prevention and Control, Changchun, 130122, China

*Corresponding authors. Pingping Sun, School of Information Science and Technology, Northeast Normal University, Changchun, 130117, China.

Email: sunpp567@nenu.edu.cn; Ming Ni, Academy of Military Medical Sciences, Beijing, 100850, China. Email: niming@bmi.ac.cn; Xiaochen Bo, Academy of Military Medical Sciences, Beijing, 100850, China. Email: boxc@bmi.ac.cn.

[†]These authors contributed equally.

Associate Editor: Arne Elofsson

Abstract

Motivation: Predicting the thermodynamic stability of proteins upon single-point mutations is a pivotal step in both protein engineering and medicine. In the study of predicting protein thermodynamic stability, various computational methods, whether they extract features at the local-level or global-level, exhibit their respective advantages and limitations. To leverage the advantages of both features, we developed MuFaDDG, a novel sequence-based method that integrated multiscale feature fusion for improved prediction of protein stability changes ($\Delta\Delta G$).

Results: MuFaDDG achieves comparable performance on the S669 benchmark, demonstrating strong capabilities in stabilizing mutations. Notably, it shows a significant advantage in the ACC metric, with values of 0.75, 0.88, and 0.81 on the direct, reverse, and overall datasets of the CAGI5 Challenge's Frataxin, respectively. Furthermore, our method outperforms leading sequence-based approaches including THPLM, DDGemb, DDGun, and INPS-Seq on protein Myoglobin stability prediction. Additionally, MuFaDDG demonstrates exceptional predictive performance with higher PCC and ACC on the protein ThreeFoil, which is uncured by FireProtDB and ProThermDB databases.

Availability and implementation: The source code and data are available at <https://github.com/PengjiaMa23/MuFaDDG>.

1 Introduction

The thermodynamic stability of proteins significantly impacts the function and structure of proteins within organisms, as well as protein engineering, such as industrial biocatalysts (Chapman et al. 2018, Wu et al. 2021, Bell et al. 2021) and pharmaceutical biologics (Gebauer and Skerra 2009, Jay and Lee 2013, Karmakar et al. 2020, Portelli et al. 2021). Decreased thermodynamic stability may lead to increased protein denaturation, thereby affecting biological functions, including enzymatic activity, signal transduction, and structural integrity. On the other hand, proteins with enhanced thermodynamic stability are more amenable to engineering, as they are less prone to unfolding and aggregation (Benevenuta et al. 2022, Rahban et al. 2022, Diaz et al. 2024). To get the more stable candidate protein, computational methods and mutagenesis techniques are proposed to identify advantageous point mutations (Planas-Iglesias et al. 2021). Among them, computational methods for

predicting mutant stability can significantly save costs and time by helping prioritize the most promising mutants for experimental characterization, such as direct evolution (Wang et al. 2021).

Existing computational algorithms for predicting protein thermodynamic stability changes can be roughly divided into two main approaches: structure-based methods and sequence-based methods (Benevenuta et al. 2021, Li et al. 2021, Montanucci et al. 2022). Structure-based methods, such as physics-based methods ABACUS (Xiong et al. 2017), Rosetta (Das and Baker 2008) and FoldX (Schymkowitz et al. 2005), are commonly used in the community. Due to their foundation in physical principles, they tend to offer highly accurate predictions, particularly when high-quality structural data is available. Some studies (Kurniawan and Ishida 2023, Lee et al. 2025, Rossi et al. 2025) have accounted for the free energy changes of the unfolded state, as Rossi et al. (2025) demonstrated that incorporating mass balance as a first approximation of the unfolded state significantly improves some physics-based

Received: 18 August 2025. Revised: 2 November 2025. Accepted: 15 April 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

methods. Recent advancements in machine learning have led to the structure-based predictors' development. Stability Oracle (Diaz et al. 2024) based on graph-transformer, ThermoMPNN (Dieckhaus et al. 2024) based on pre-trained ProteinMPNN (Dauparas et al. 2022) and PremPS (Zhang et al. 2020) based on random forest (RF) regression are proposed and achieved the state-of-the-art performance on one or more testing datasets from protein structures. Structure-based methods typically provide more accurate predictions (Jeffery et al. 2022). However, sequence-based methods have the advantage of not requiring high-quality structural data and being able to efficiently screen large-scale datasets, therefore, they can be used for pre-screening protein variants in early-stage research.

According to reports, following the conclusion of existing prediction approaches and tools on protein stability changes (Table 1, available as supplementary data at *Bioinformatics* online), sequence-based methods occupy only a small portion of these methods. Sequence-based computational approaches for feature extraction emphasize different aspects and can be broadly categorized into residue-level methods (local-level methods) and protein-level methods (global-level methods). Residue-level methods mostly focus on features related to enthalpy changes, primarily relying on local-level features that represent the attractive forces between residues, such as changes in secondary structure and variations in the physicochemical properties of amino acids. The representation works include MU3DSP (Gong et al. 2023), MUpro (Cheng et al. 2006), SAAFEC-SEQ (Li et al. 2021), INPS-Seq (Fariselli et al. 2015), and DDGun (Montanucci et al. 2019). Protein-level methods DDGemb (Savojardo et al. 2024) and THPLM (Gong et al. 2023) represents the protein thermodynamic stability change in view of the global sequence by protein language model, which incorporates knowledge about protein flexibility and disorder (Kabir and Hoque 2024) through its learned representation of protein sequence. Designing a protein with the desired stability is a challenging task. The stability changes of proteins upon single-point mutations are influenced by various factors at different scales. Therefore, extracting and integrating features from both local and global perspectives is crucial for comprehensively assessing the stability changes. Local-level features can provide detailed information about the specific mutation site and its surrounding environment, while global-level features can capture the overall characteristics and context of the protein sequence. By combining these two types of features, we can obtain a more comprehensive and accurate representation of the protein's stability changes.

Here, we developed a thermodynamic stability prediction tool named MuFaDDG, which aimed to be used for pre-screening proteins or protein variants from sequences in early-stage research

in protein engineering. MuFaDDG is a multiscale feature fusion method that integrates both global-level changes and local residue alterations before and after protein single-point mutations. It combines ESM-2's global sequence representations with NetSurfP-3.0 (Høie et al. 2022) predicted local structural features around mutation sites. The fused features' model MuFaDDG outperformed models based on single-scale features and demonstrated comparable and even superior performance compared to other sequence-based methods when tested on commonly used external testing set S669, the CAGI5 Challenge's Frataxin, the protein Myoglobin and the protein ThreeFoil which not curated by both FireProtDB and ProThermDB databases. Moreover, MuFaDDG demonstrated a better performance in stabilizing mutations that are naturally less common.

2 Methods

2.1 Dataset curation and preprocessing

For a fair performance comparison against the methods for predicting $\Delta\Delta G$, we used widely recognized public benchmarking datasets as follows. S2648 (Dehouck et al. 2011, Fariselli et al. 2015) served as the training set for training the MuFaDDG model. Considering the antisymmetric characteristic of the Gibbs free energy difference before and after mutation, S2648 encompasses both forward and reverse mutations. For effective cross-validation, Fariselli et al.'s method (Fariselli et al. 2015) was employed, segmenting S2648 based on sequence homology into five parts for training.

S669 (Pancotti et al. 2022) and S^{sym} (Pucci et al. 2018) were designated as testing sets. S669, sourced from the latest ProTherm database (Nikam et al. 2021), contains variations with less than 25% homology to the sequences in S2648. S^{sym} comprises PDB structures for forward and reverse mutants and has been previously employed to evaluate anti-symmetry and bias across different methods (Mishra 2022).

2.2 MuFaDDG architecture

In this study, we developed a novel model, named MuFaDDG, which combined two feature extractors, an integrated module and a decoder, aimed at accurately predicting $\Delta\Delta G$ upon single-point mutations (Fig. 1A). The overall architecture of MuFaDDG is illustrated in Fig. 1.

Two feature extractors, including local-level and global-level encoders, are designed to characterize changes associated with mutations. On a global scale (sequence-level), embeddings of amino acid sequences before and after mutation are acquired

Table 1 Performance comparison of stability prediction methods on protein ThreeFoil.

Method	RMSE (kcal/mol)	PCC	ACC
MuFaDDG	1.81	0.64	0.60
DDGemb	1.00	0.15	0.40
INPS-Seq	1.22	0.01	0.60
DDGun	1.48	-0.12	0.60
THPLM	1.07	-0.35	0.10

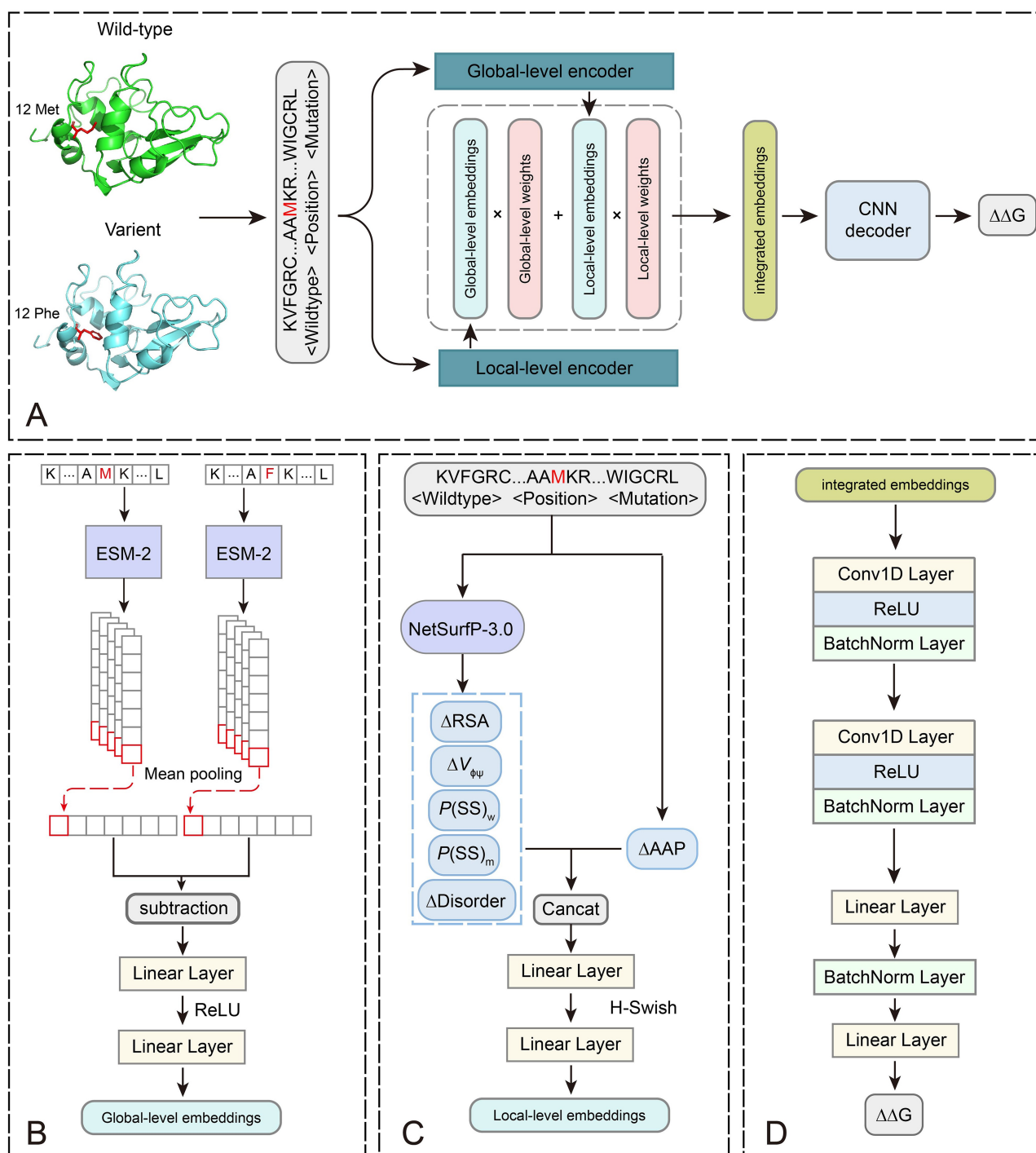


Figure 1 Architecture of the MuFaDDG model. (A) Overview of the MuFaDDG model, comprising two feature extractors (global-level and local-level encoders), an integrated module, and a decoder. (B) Global-level encoder. (C) Local-level encoder. (D) Decoder: Composed of convolutional and fully connected layers, the decoder predicts the $\Delta\Delta G$ value of the mutation.

separately through the pre-trained protein language model ESM-2. These embeddings are then subjected to average pooling across the amino acid residue dimension, and the difference between the pre- and post-mutation embeddings is calculated to obtain a global-level feature representation (Fig. 1B). To represent local-level changes, we computed amino acid properties changes (ΔAAP) of the mutant residue and used NetSurfP (Høie et al. 2022) to obtain a set of structural and physicochemical

features, including secondary structure probabilities ($P(SS)$) for both wild-type and mutant residues, changes in relative solvent accessibility (ΔRSA), disorder propensity ($\Delta Disorder$), and dihedral angle conformations ($\Delta V_{\phi\psi}$). These features are combined to form preliminary local-level features, which are then refined through two linear layers to achieve a representation of local-level characteristics. Next, the integrated module used the linear combination to leverage the local-level and global-level

representations (Fig. 1, available as supplementary data at *Bioinformatics* online). Then, the decoder comprises two convolutional layers and two fully connected layers with three BatchNorm layers for predicting $\Delta\Delta G$.

2.3 Feature representation and fusion

Global-level encoder. Protein language models (PLMs) have seen significant advancements in recent years (Rives et al. 2021, Elnaggar et al. 2022, Madani et al. 2023, Su et al. 2024). Specifically, ESM-2 is a transformer-based model that has shown remarkable performance in capturing the evolutionary information of protein sequences, which is crucial for understanding their functions.

In this study, the global-level encoder consists of an ESM-2 backbone and two linear layers, which generates 128-dimensional embeddings. Considering hardware constraints and model capabilities, a pre-trained ESM-2 model with 650 million parameters (esm2_t36_650M_UR50D) was selected. Protein sequences before and after mutation were input into the ESM-2 model, obtaining high-dimensional embeddings with a dimensionality of $L \times 1280$ for both the wild-type (Seq_w) and mutant (Seq_m) proteins. L denotes the sequence length.

In this study, the global-level encoder consists of an ESM-2 backbone and two linear layers, which generates 128-dimensional embeddings. Considering hardware constraints and model capabilities, a pre-trained ESM-2 model with 650 million parameters (esm2_t36_650M_UR50D) was selected. Protein sequences before and after mutation were input into the ESM-2 model, obtaining high-dimensional embeddings with a dimensionality of $L \times 1280$ for both the wild-type (Seq_w) and mutant (Seq_m) proteins. L denotes the sequence length.

We applied average pooling to reduce the impact of individual amino acid positions while emphasizing global changes, resulting in a unified representation with a dimension of 1×1280 . The global-level feature representation G_{pri} was then obtained by calculating the difference between the wild-type and mutant embeddings. The formulation is as follows:

$$G_{pri} = \text{AvgPool}(ESM(Seq_m)) - \text{AvgPool}(ESM(Seq_w))$$

$$G = w_2(\sigma(w_1 G_{pri} + b_1)) + b_2$$

Next, to effectively integrate and analyze both local-level and global-level representations, enabling the model to capture and utilize the correlations and complementarities between various representations, we adjusted the dimensionality of the global change features to 1×128 through two linear layers.

Local-level encoder. The local-level encoder is composed of two linear layers that transform 25-dimensional manually extracted features into 128-dimensional local embeddings. The manually extracted features are primarily focused on and centered around mutated residues, including 13-dimensional changes of amino acid properties (ΔAAP), 6-dimensional secondary structure probabilities ($P(SS)$), 1-dimensional relative solvent accessibility changes (ΔRSA) at the mutation site, 1-dimensional disorder propensity changes ($\Delta Disorder$), and 4-dimensional dihedral angle conformational changes

($\Delta V_{\Phi\Psi}$), all of them are the decisive factors for stability (Folkman et al. 2016, Liu et al. 2019, Gong et al. 2023).

The $P(SS)$, $\Delta Disorder$, ΔRSA and $\Delta V_{\Phi\Psi}$ features are obtained by NetSurfP-3.0, where $P(SS)$ includes probability values for α -helix, β -sheet, and coil states for both wild-type and mutant residues and $\Delta here $P(SS)$ includes probability values for α -helix, β -sheet, and coil states for both wild-type and mutant residues and CYP119.$

$$RSA = \frac{ASA}{ASA_m}$$

$$\Delta RSA = RSA_m - RSA_w$$

To address angular periodicity artifacts, the backbone dihedral angle changes $\Delta V_{\Phi\Psi}$ are represented using sine/cosine differences:

$$V_{\Phi\Psi} = [\sin(\Phi), \cos(\Phi), \sin(\Psi), \cos(\Psi)]$$

$$\Delta V_{\Phi\Psi} = V_{\Phi\Psi}^m - V_{\Phi\Psi}^w$$

These features are concatenated into a 25-dimensional local residue representation, which is then processed through two linear layers:

$$Res = [\Delta RSA, \Delta V_{\Phi\Psi}, \Delta Disorder, \Delta AAP, P(SS)_w, P(SS)_m]$$

$$R = w_2(\sigma(w_1 Res_x + b_1)) + b_2$$

Res refers to the 25-dimensional manually extracted features. w_1 and w_2 represent the weight matrices of the two linear layers, while b_1 and b_2 are the bias terms. The symbol σ denotes the H-Swish activation function defined by the following expression:

$$H\text{-swish}(x) = x \cdot \text{ReLU6}(x + 3)/6$$

$$\text{ReLU6}(x) = \min(\max(0, x), 6)$$

Finally, Res can be mapped into a 128-dimensional local-level feature representation, denoted as R (Local-level representations around residue).

2.4 Features integration and convolution neural network (CNN) decoder

Following the encoders, we acquired both local-level and global-level representations, each with a dimensionality of 1×128 . Next, we integrated the local feature change embeddings and global feature change embeddings into a unified representation for prediction. After exploring three methods for combining these embeddings, Concatenation, Outer Product and Linear Combination (detailed in Table 2, Fig. 1, available as supplementary data at *Bioinformatics* online), we ultimately selected the linear combination method for integrating features. The integrated new representation is denoted as $f_{integrated}$:

$$f_{\text{Integrated}} = G^{\circ}G_weight + R^{\circ}R_weight$$

Here, G represents the dimensionally adjusted global-level representation, and R denotes the local-level representation. G_weight and R_weight are two learnable weight matrices, each with dimensions of 1×128 . The symbol $\circ$ represents the Hadamard product, facilitating the element-wise multiplication between the two matrices.

Convolution neural network (CNN) Decoder is a regression model for predicting $\Delta\Delta G$ with two CNN layers and two linear layers (Fig. 1D). The CNN layer is composed of a 1-dimensional convolution operation (Conv1d), followed by a Rectified Linear Unit (ReLU) activation function and a batch normalization (BN). Two CNN layers can be represented as follows:

$$E_1 = \text{ReLU}(\text{BN}(\text{Conv1d}(f_{\text{Integrated}})))$$

$$E_2 = \text{ReLU}(\text{BN}(\text{Conv1d}(E_1)))$$

Then, E_2 is passed through two linear layers (*Linear*) to obtain the final predictive result, denoted as $\hat{y}$.

$$\hat{y} = \text{Linear}(\text{BN}(\text{Linear}(\text{flatten}(E_2))))$$

2.5 Evaluation metrics

For the evaluation of our proposed MuFaDDG, we used Root Mean Square Error (RMSE), Accuracy (ACC), Pearson Correlation Coefficient (PCC), antisymmetric property (r_{d-r}) and bias property ($\langle\delta\rangle$) as evaluation metrics. RMSE measures the difference between predicted and experimental $\Delta\Delta G$ s, while PCC assesses the degree of linear correlation between the predicted and experimental $\Delta\Delta G$ s. ACC evaluates the binary classification performance by categorizing mutations as stabilizing ($\Delta\Delta G \geq 0$) or destabilizing ($\Delta\Delta G < 0$). Moreover, r_{d-r} and $\langle\delta\rangle$ between the predicted $\Delta\Delta G$ s of the direct and corresponding reverse variants were used to evaluate the model's ability to anti-symmetric mutation (Mishra 2022).

3 Results

3.1 Cross-validation performance of MuFaDDG

To evaluate the performance and generalizability of the MuFaDDG model, cross-validation is employed by testing it on different subsets of data. The S2648 dataset was partitioned into five cross-validation sets based on homology criteria proposed by Fariselli et al. (Dehouck et al. 2011). We trained five independent cross-validation models and integrated them into a composite model by averaging, which we named MuFaDDG. Specifically, the five-fold cross-validation yielded PCC of 0.74, 0.62, 0.76, 0.61, and 0.74 for the individual models, with an average PCC of 0.69 (Fig. 2, available as supplementary data at Bioinformatics online). The corresponding RMSE of overall variations were 1.25 kcal/mol, 1.33 kcal/mol, 1.25 kcal/mol, 1.46 kcal/mol, and 1.22 kcal/mol, resulting in an average RMSE

of 1.30 kcal/mol (Table 3, available as supplementary data at Bioinformatics online).

3.2 Performance comparison on datasets S669

To validate the robustness of the MuFaDDG model, we compared it with seven advanced sequence-based predictors, THPLM, MU3DSP, DDGun, INPS-Seq, MUpro, SAAFEC-SEQ, and DDGemb on the S669 testing sets. Considering the antisymmetric property of predictors, S669 is divided into direct, reverse, and overall mutation datasets to evaluate these predictors by RMSE and PCC metrics, which are shown in Fig. 2.

On the S669 dataset (Fig. 2A and B), MuFaDDG showed comparable performance with other methods in terms of RMSE and PCC on the S669 direct, reverse, and overall mutation sets. Notably, MuFaDDG achieved the second-best PCC on both direct (PCC=0.45) and reverse (PCC=0.44) mutations. It also obtained the top 3 RMSE on direct (MuFaDDG=1.53 kcal/mol, INPS-Seq=1.52 kcal/mol, DDGemb=1.40 kcal/mol), reverse (MuFaDDG=1.56 kcal/mol, INPS-Seq=1.53 kcal/mol, DDGemb=1.40 kcal/mol), and overall mutations (MuFaDDG=1.55 kcal/mol, INPS-Seq=1.52 kcal/mol, DDGemb=1.40 kcal/mol). In addition, MuFaDDG exhibited strong antisymmetry and low bias, with an antisymmetric property of -0.99 and a bias property ($\langle\delta\rangle$) of 0.04 kcal/mol on dataset S669 (Fig. 2D) as well as an antisymmetric property of -0.99 and a bias property ($\langle\delta\rangle$) of 0.05 kcal/mol on dataset S^{sym} (Fig. 2C).

Notably, stabilizing mutations are naturally less common (Nisthal et al. 2019) and many methods suffer from poor generalization on stabilizing mutations (Diaz et al. 2024). Therefore, we further evaluated the performance of MuFaDDG, DDGemb and INPS-Seq on direct stabilizing mutations of S669. In Fig. 2E and F, MuFaDDG achieved a better performance than INPS-Seq and DDGemb with an ACC of 0.46 (INPS-Seq=0.34, DDGemb=0.41) and a RMSE of 1.54 kcal/mol (INPS-Seq=1.56 kcal/mol, DDGemb=1.73 kcal/mol) on direct stability mutations, respectively. A detailed comparison of MuFaDDG with other sequence-based predictors on stabilizing and reverse destabilizing mutations in the S669 dataset is provided in Table 4, available as supplementary data at Bioinformatics online.

3.3 MuFaDDG prediction preference analysis

To evaluate the applicability spectrum of MuFaDDG, we analyzed its prediction preferences, including the relative solvent accessibility (RSA) and secondary structure of the mutated residues to determine which type of mutated residues that MuFaDDG predicts more effectively. Secondary structure (Ji and Li 2010) refers to the specific structural motifs (like α -helices, β -sheets, and loops) that play key roles in the stability of proteins, while RSA reflects the extent to which residues are exposed to the solvent (typically water) within the protein structure. The secondary structure and RSA of each mutated residue were obtained by the DSSP program (Joosten et al. 2011). In particular, residues with RSA greater than 0.5 are defined as

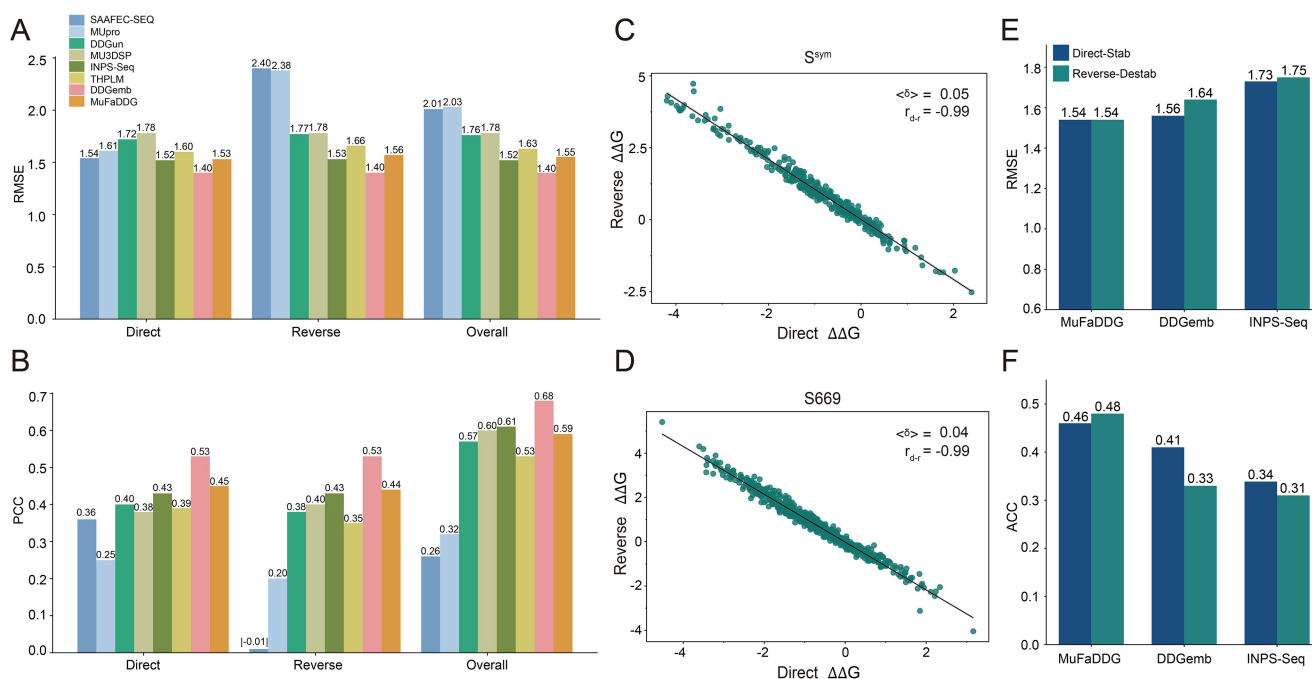


Figure 2 Comparison of MuFaDDG with other sequence-based models on two test sets. (A) RMSE of eight sequence-based methods on the S669 dataset. (B) PCC of eight sequence-based methods on the S669 dataset. (C) Bias and antisymmetry properties of MuFaDDG on the S^{sym} dataset. (D) Bias property and antisymmetry properties of MuFaDDG on the S669 dataset. RMSE (E) and ACC (F) comparison of MuFaDDG, INPS-Seq and DDGemb on the S669 dataset's forward stabilizing mutations and reverse destabilizing mutations.

solvent-exposed residues, while those with RSA less than 0.1 are defined as being buried within the protein.

Fig. 3A and Fig. 3B show the predictive performance of MuFaDDG on residues of different secondary structures located in different positions within the proteins in the S^{sym} testing sets. Under different secondary structures, MuFaDDG performs best in predicting mutations in residues with β -sheet structures (PCC=0.85) and shows slightly poorer performance for α -helix (PCC=0.79) and random coil (PCC=0.72) structures (Fig. 3A). In terms of different mutation residue positions, Fig. 3B demonstrates that MuFaDDG predictive performance is better for buried residues (PCC=0.81) than for surface residues (0.58). A possible reason is that mutations in surface residues might also need to consider environmental factors such as solvents, and the model lacks this type of information in its input features, leading to poorer performance in predicting thermodynamic stability changes caused by mutations in protein surface residues.

3.4 Ablation study

To determine whether combining global-level features with local-level features contributes to MuFaDDG's performance, we conducted an ablation study by removing one of them. We found MuFaDDG significantly outperforms the other two models (Fig. 4A and B), which were trained using a single type of feature. Specifically, the ablated model 'wo Global' (excluding global features) retained only the local residue feature extraction module, whereas 'wo Local' (excluding local features) kept only the global-level encoder module. Importantly, MuFaDDG demonstrates significant improvements over the ablated models.

Compared to 'wo Local', it achieves 3.75%, 12.16%, and 7.79% higher ACC on the direct, reverse, and overall S^{sym} datasets, respectively. Compared to 'wo Global', it shows 18.57%, 15.27%, and 16.90% ACC improvements on the corresponding datasets. Detailed comparative data from the ablation studies are provided in Table 5, available as supplementary data at Bioinformatics online.

Since 'wo Local' outperformed 'wo Global' in terms of PCC and RMSE on the direct, reverse, and overall datasets (Fig. 4A and B), we further analyzed the contribution of these two feature types in the MuFaDDG model. Fig. 4C demonstrates the absolute weights on 128-dimensional embeddings when integrating both feature types, revealing which features ('wo Global' or 'wo Local') contribute more significantly to predicting protein stability changes. It shows that the model attends to both local structural features and global sequence features, while assigning slightly greater importance to local structural features. This weighting pattern indicates that the model effectively captures meaningful information from protein structural features while maintaining a balanced consideration of global sequence characteristics.

3.5 The effect of mutations on a three-fold symmetric globular protein ThreeFoil

ThreeFoil is a three-fold symmetric globular protein by computational reconstructed and repeated the β -trefoil subdomain module (Broom et al. 2012). ThreeFoil is not recorded in both FireProtDB (Stourac et al. 2021) and ProThermDB (Nikam et al. 2021) databases, and it displays lower 30% identity similarity with training dataset S2648. Therefore, we applied MuFaDDG on

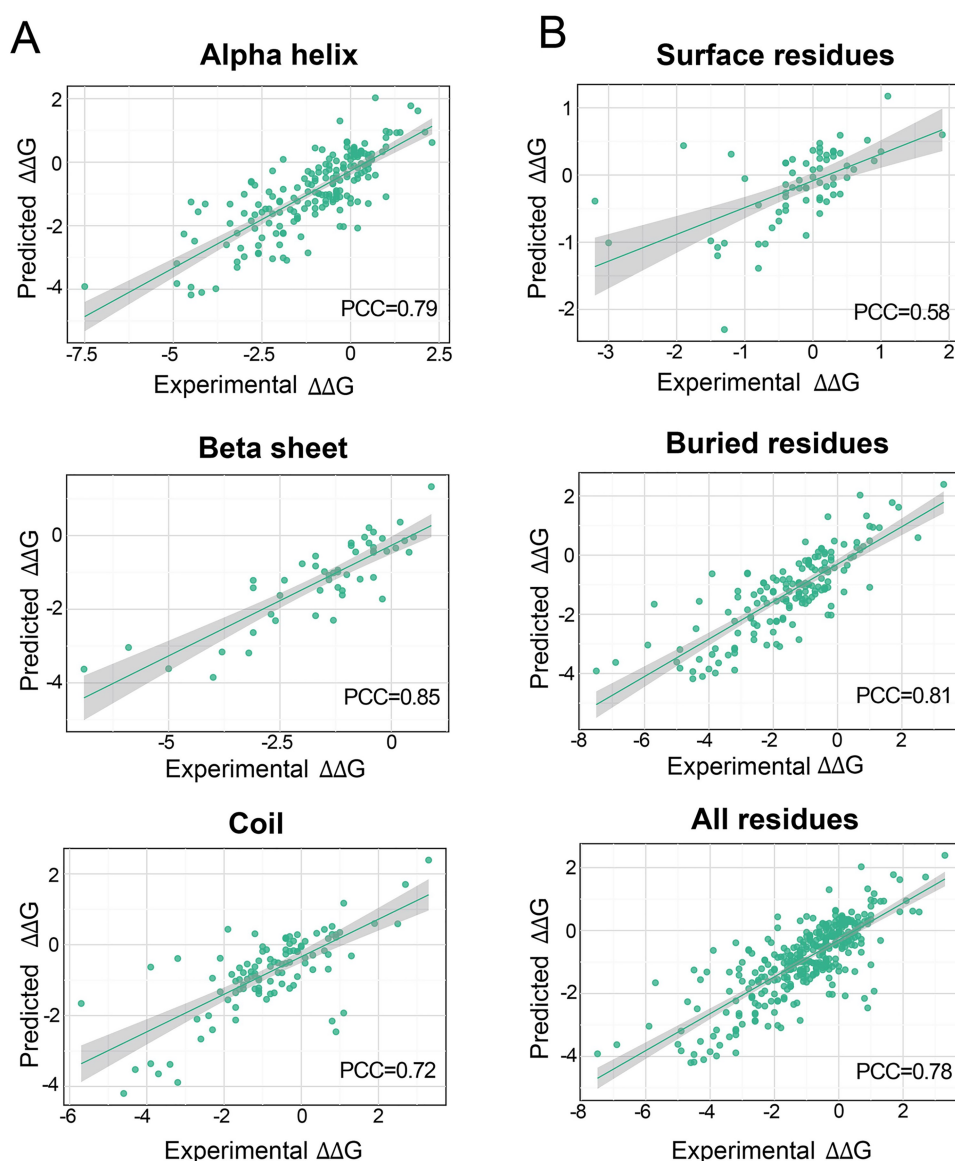


Figure 3 Analysis of secondary structure and mutation residue position. (A) Performance of MuFaDDG on different secondary structures in the S^{sym} dataset. (B) Performance of MuFaDDG on different mutation residue positions in the S^{sym} dataset.

protein ThreeFold that contained ten mutations consisting of three no significant effect mutations, four stabilizing mutations, and three destabilizing mutations (Broom et al. 2017). As Table 1 demonstrated that our model outperformed other predictors including THPLM, INPS-Seq, DDGemb and DDGun reaching PCC of 0.64 and ACC of 0.60, further confirming the model's generalization ability.

3.6 Performance comparison on CAGI5 challenge's Frataxin

We further evaluated MuFaDDG using the Frataxin dataset from the CAGI5 challenge (Fowler and Fields 2014). MuFaDDG was compared with four leading sequence-based predictors (THPLM, DDGemb, INPS-Seq, and DDGun) that considered the $\Delta\Delta G$ prediction of direct and reverse variations across the direct, reverse, and overall datasets.

The results (Table 6, available as supplementary data at Bioinformatics online, Fig. 5) demonstrate that although MuFaDDG slightly underperforms in the PCC metric compared to some methods, it exhibits a significant advantage in the ACC by achieving values of 0.75, 0.88, and 0.81 on the direct, reverse, and overall datasets, respectively.

3.7 Achieves the best performance on stability prediction of myoglobin

Myoglobin, a crucial oxygen-binding protein in muscle tissues, serves as an excellent model for studying protein stability. We evaluated MuFaDDG on a myoglobin dataset derived from Umerenkov et al.'s work (Umerenkov et al. 2023). The $\Delta\Delta G$ values of duplicate mutations from myoglobin dataset were averaged, adhering to the methodology of the training dataset (Dehouck et al. 2011), to ensure data uniformity and yield a final

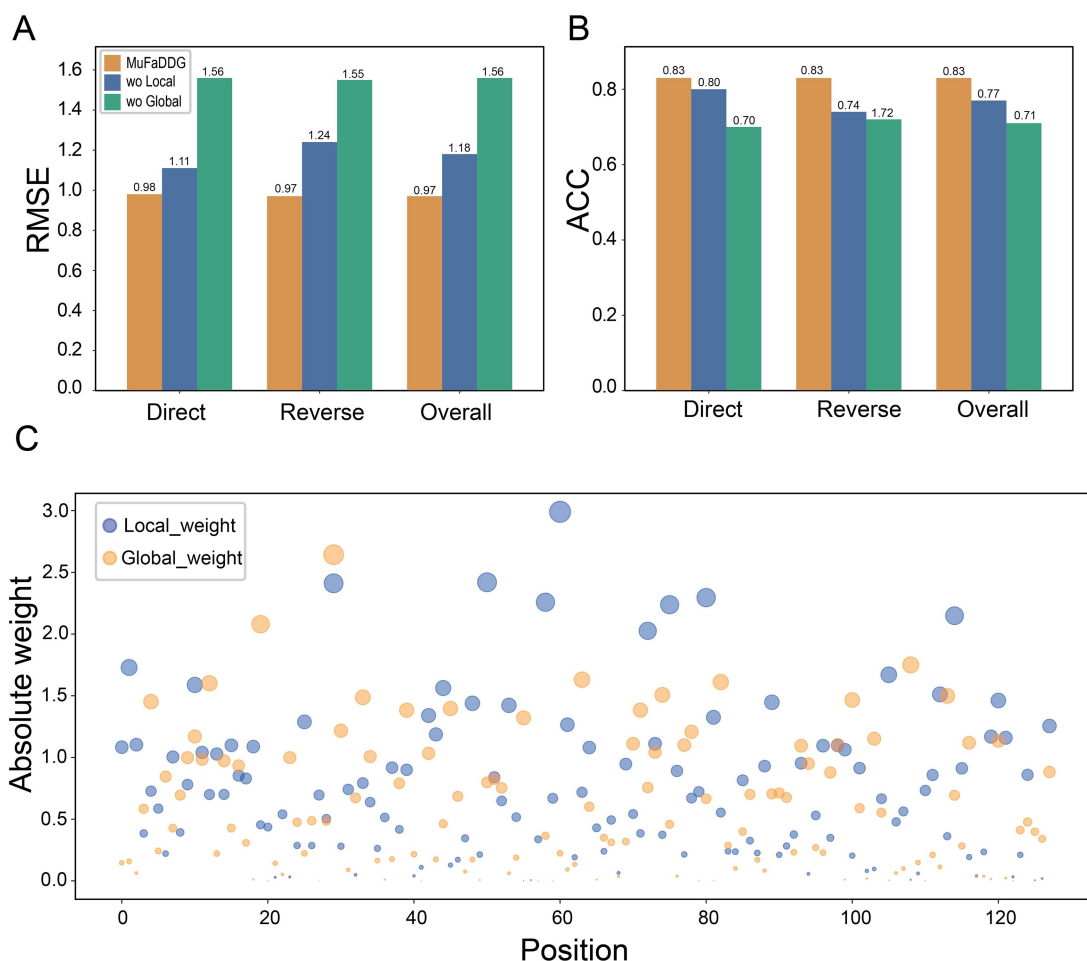


Figure 4 Ablation experiments, and feature matrix visualization. (A) Ablation study results of MuFaDDG on the S^{sym} test set, showing the RMSE metric. (B) Ablation study results of MuFaDDG on the S^{sym} test set, showing the ACC metric. (C) Visualization of global and local feature matrix weights in MuFaDDG.

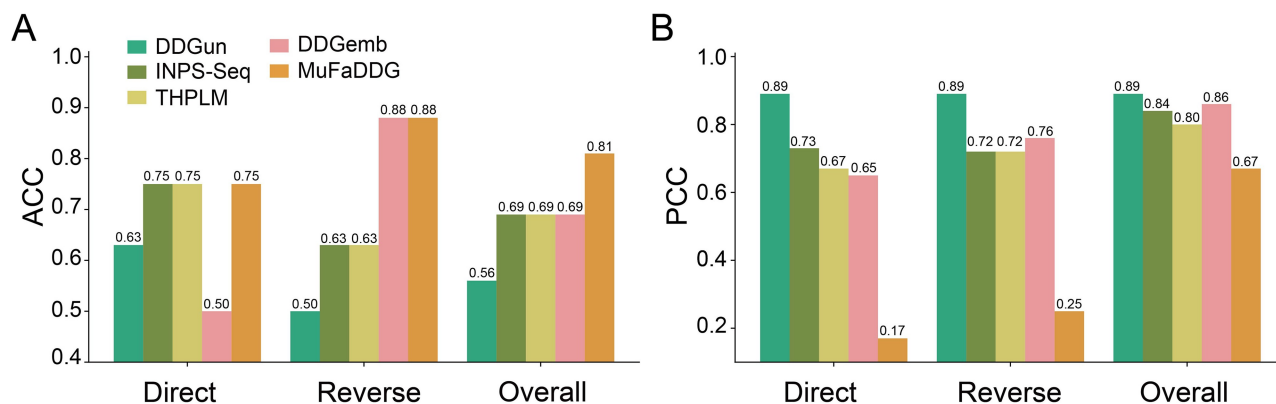


Figure 5 Performance comparison of sequence-based stability changes prediction methods on CAGI5 Challenge's Frataxin.

set of 113 unique variants. As shown in Fig. 6, MuFaDDG outperformed all sequence-based methods including THPLM, DDGun, DDGemb and INPS-Seq across all metrics, achieved an RMSE of 0.87 Kcal/mol, PCC of 0.64, and ACC of 0.76, respectively. These results demonstrated MuFaDDG's robust capability in predicting Myoglobin stability changes, confirming its utility for practical applications in protein engineering.

4 Conclusion

In this work, we present MuFaDDG, a novel sequence-based method trained by integrating manually extracted features around mutant residues (local-level features) and ESM2's representations of variants (global-level features). Our model reports second-best performance with other methods in direct, reverse

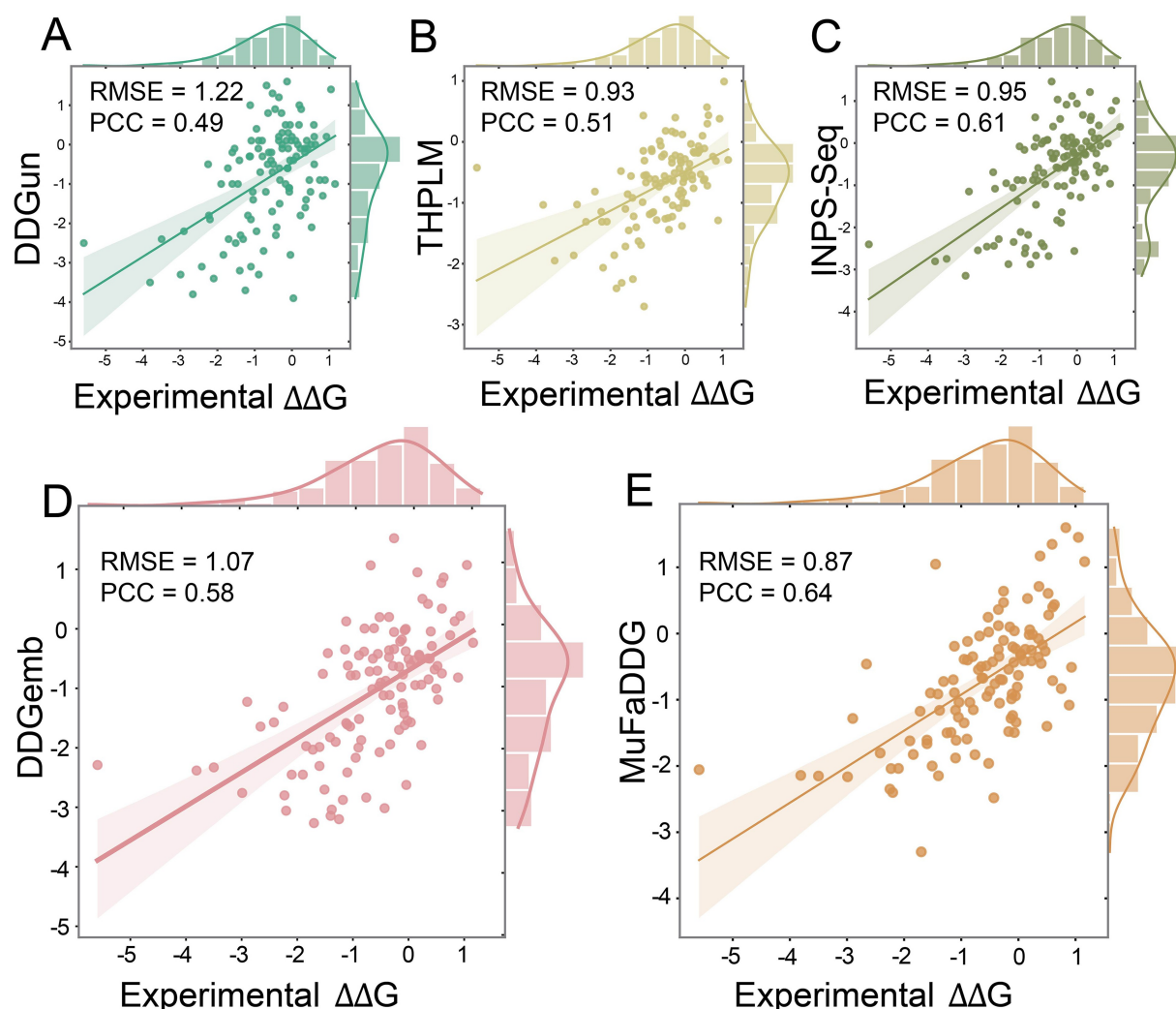


Figure 6 MuFaDDG and other sequence-based methods including THPLM, DDGun, DDGemb and INPS-Seq prediction performance on stability changes of Myoglobin.

and overall variations of S669 datasets and superior performance in direct stabilizing variations of S669 (Fig. 2A and B). For further analysis, mutations located in β -sheets and buried residues demonstrated a higher performance, which is easier captured by the feature extraction methods of MuFaDDG. In addition, to validate MuFaDDG, we applied it to three experimentally validated proteins (case studies) that were excluded from the training dataset. The first case study focuses on the CAG15 Challenge's Frataxin. MuFaDDG demonstrated a significant advantage to classify stabilizing and destabilizing mutations with 0.81 of ACC on overall variations. The second protein is the Myoglobin. Our model outperformed leading sequence-based methods including THPLM, DDGun, DDGemb, and INPS-Seq across PCC and RMSE on Myoglobin with an improvement from 0.61 (INPS-Seq) to 0.64. The rest is a three-fold symmetric globular protein ThreeFoil which is uncured in FireProtDB and ProThermDB databases. Our model outperformed other predictors including THPLM, INPS-Seq, DDGemb, and DDGun reaching ACC of 0.64 and PCC of 0.60. All in all, MuFaDDG fusing both local and global features demonstrated a comparable robustness and generalization ability for predicting thermodynamic stability changes on low-homology sequences.

Author contributions

Jianting Gong (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Methodology [equal], Validation [equal], Writing—original draft [lead]), Pengjia Ma (Data curation [equal], Formal analysis [equal], Software [lead], Visualization [equal], Writing—review & editing [equal]), Zilin Ren (Formal analysis [equal], Validation [equal]), Si Li (Validation [equal], Visualization [equal]), Zhiguo Fu (Validation [equal], Writing—review & editing [equal]), Pingping Sun (Methodology [equal], Validation [equal]), Ming Ni (Methodology [equal], Supervision [equal], Writing—review & editing [equal]), and Xiaochen Bo (Project administration [lead], Resources [lead])

Supplementary material

Supplementary material is available at *Bioinformatics* online.

Conflicts of interest

All authors disclosed no relevant relationships.

Funding

This work has been supported by the Project.

Data availability

The source code and data are available at <https://github.com/PengjiaMa23/MuFaDDG>.

References

- Bell EL, Finnigan W, France SP *et al.* Biocatalysis. *Nat Rev Methods Primers* 2021;**1**:46.
- Benevenuta S, Birolo G, Sanavia T *et al.* Challenges in predicting stabilizing variations: an exploration. *Front Mol Biosci* 2022;**9**:1075570.
- Benevenuta S, Pancotti C, Fariselli P *et al.* An antisymmetric neural network to predict free energy changes in protein variants. *J Phys D: Appl Phys* 2021;**54**:245403.
- Broom A, Doxey AC, Lobsanov YD *et al.* Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. *Structure* 2012;**20**:161–71.
- Broom A, Jacobi Z, Trainor K *et al.* Computational tools help improve protein stability but with a solubility tradeoff. *J Biol Chem* 2017;**292**:14349–61.
- Chapman J, Ismail AE, Dinu CZ. Industrial applications of enzymes: recent advances, techniques, and outlooks. *Catalysts* 2018;**8**:238.
- Cheng JL, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins-Struct Funct Bioinform* 2006;**62**:1125–32.
- Das R, Baker D. Macromolecular modeling with rosetta. *Annu Rev Biochem* 2008;**77**:363–82.
- Dauparas J, Anishchenko I, Bennett N *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 2022;**378**:49–56.
- Dehouck Y, Kwasigroch JM, Gilis D *et al.* PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinform* 2011;**12**:151.
- Diaz DJ, Gong C, Ouyang-Zhang J *et al.* Stability oracle: a structure-based graph-transformer framework for identifying stabilizing mutations. *Nat Commun* 2024;**15**:6170.
- Dieckhaus H *et al.* Transfer learning to leverage larger datasets for improved prediction of protein stability changes. *Proc Natl Acad Sci USA* 2024;**121**:e2314853121.
- Elnaggar A, Heinzinger M, Dallago C *et al.* ProtTrans: toward understanding the language of life through self-supervised learning. *IEEE Trans Pattern Anal Mach Intell* 2022;**44**:7112–27.
- Fariselli P, Martelli PL, Savojardo C *et al.* INPS: predicting the impact of non-synonymous variations on protein stability from sequence. *Bioinformatics* 2015;**31**:2816–21.
- Folkman L, Stantic B, Sattar A *et al.* EASE-MM: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. *J Mol Biol* 2016;**428**:1394–405.
- Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. *Nat Methods* 2014;**11**:801–7.
- Gebauer M, Skerra A. Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Chem Biol* 2009;**13**:245–55.
- Gong J, Wang J, Zong X *et al.* Prediction of protein stability changes upon single-point variant using 3D structure profile. *Comput Struct Biotechnol J* 2023;**21**:354–64.
- Gong J, Jiang L, Chen Y *et al.* THPLM: a sequence-based deep learning framework for protein stability changes prediction upon point variations using pretrained protein language model. *Bioinformatics* 2023;**39**:btad646.
- Høie MH, Kiehl EN, Petersen B *et al.* NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res* 2022;**50**:W510–W515.
- Jay SM, Lee RT. Protein engineering for cardiovascular therapeutics: untapped potential for cardiac repair. *Circ Res* 2013;**113**:933–43.
- Jeffery CJ, Dorgan KM, Pysh L. Promoting a more integrated approach to structure and function. *Integr Comp Biol* 2022;**61**:2020–30.
- Ji Y-Y, Li Y-Q. The role of secondary structure in protein structure selection. *Eur Phys J E Soft Matter* 2010;**32**:103–7.
- Joosten RP, Te Beek TAH, Krieger E *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res* 2011;**39**:D411–419.
- Kabir MWU, Hoque MT. DisPredict3.0: prediction of intrinsically disordered regions/proteins using protein language model. *Appl Math Comput* 2024;**472**:128630.
- Karmakar M, Rodrigues CHM, Horan K *et al.* Structure guided prediction of pyrazinamide resistance mutations in pncA. *Sci Rep* 2020;**10**:1875.
- 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.
- Lee H *et al.* Protein folding stability estimation with explicit consideration of unfolded states. *bioRxiv*, 2025. 2025.2002.2010.637420.
- Li G, Panday SK, Alexov E. SAAFEC-SEQ: a sequence-based method for predicting the effect of single point mutations on protein thermodynamic stability. *Int J Mol Sci* 2021;**22**:606.
- Liu Z *et al.* Entropic contribution to enhanced thermal stability in the thermostable P450 CYP119. *Proceedings of the National Academy of Sciences* 2018;**115**:E10049–E10058.
- Madani A, Krause B, Greene ER *et al.* Large language models generate functional protein sequences across diverse families. *Nat Biotechnol* 2023;**41**:1099–106.
- Mishra SK. PSP-GNM: predicting protein stability changes upon point mutations with a Gaussian network model. *Int J Mol Sci* 2022;**23**.
- Montanucci L, Capriotti E, Birolo G *et al.* DDGun: an untrained predictor of protein stability changes upon amino acid variants. *Nucleic Acids Res* 2022;**50**:W222–W227.
- Montanucci L, Capriotti E, Frank Y *et al.* DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC Bioinformatics* 2019;**20**:335.
- Nikam R, Kulandaisamy A, Harini K *et al.* ProThermDB: thermodynamic database for proteins and mutants revisited after 15 years. *Nucleic Acids Res* 2021;**49**:D420–D424.

- Nisthal A, Wang CY, Ary ML *et al.* Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proc Natl Acad Sci USA* 2019;**116**:16367–77.
- Pancotti C, Benevenuta S, Birolo G *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.
- Planas-Iglesias J, Marques SM, Pinto GP *et al.* Computational design of enzymes for biotechnological applications. *Biotechnol Adv* 2021;**47**:107696.
- Portelli S *et al.* Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource (vol 52, pg 999, 2020). *Nat Genet* 2021;**53**:254.
- Pucci F, Bernaerts KV, Kwasigroch JM *et al.* Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* 2018;**34**:3659–65.
- Rahban M, Zolghadri S, Salehi N *et al.* Thermal stability enhancement: fundamental concepts of protein engineering strategies to manipulate the flexible structure. *Int J Biol Macromol* 2022;**214**:642–54.
- Rives A *et al.* Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acad Sci USA* 2021;**118**:e2016239118.
- Rossi I, Barducci G, Sanavia T *et al.* Mass balance approximation of unfolding boosts potential-based protein stability predictions. *Protein Sci* 2025;**34**:e70134.
- Savojardo C, Manfredi M, Martelli PL *et al.* DDGemb: predicting protein stability change upon single-and multi-point variations with embeddings and deep learning. *Bioinformatics* 2024;**41**:btaf019.
- Schymkowitz J, Borg J, Stricher F *et al.* The FoldX web server: an online force field. *Nucleic Acids Res* 2005;**33**:W382–388.
- Stourac J, Dubrava J, Musil M *et al.* FireProtDB: database of manually curated protein stability data. *Nucleic Acids Res* 2021;**49**:D319–d324.
- Su J *et al.* SaProt: protein language modeling with structure-aware vocabulary. bioRxiv. 2024.
- Umerenkov D, Nikolaev F, Shashkova TI *et al.* PROSTATA: a framework for protein stability assessment using transformers. *Bioinformatics* 2023;**39**:btad671.
- Wang Y, Xue P, Cao M *et al.* Directed evolution: methodologies and applications. *Chem Rev* 2021;**121**:12384–444.
- Wu S, Snajdrova R, Moore JC *et al.* Biocatalysis: enzymatic synthesis for industrial applications. *Angew Chem Int Ed* 2021;**60**:88–119.
- Xiong P, Chen Q, Liu H. Computational protein design under a given backbone structure with the ABACUS statistical energy function. *Methods Mol Biol* 2017;**1529**:217–26.
- Zhang R-R, Meng N-N, Liu C *et al.* PDB-1 from potentilla discolor bunge induces apoptosis and autophagy by downregulating the PI3K/Akt/mTOR signaling pathway in A549 cells. *Biomed Pharmacother* 2020;**129**:110378.