

Predicting protein thermal stability changes upon single and multi-point mutations via restricted attention subgraph neural network

Mohammad Madani ^{a,b}, Anna Tarakanova ^{a,c,*}

^a School of Mechanical, Aerospace, and Manufacturing Engineering, University of Connecticut, Storrs, CT, United States

^b Department of Computer Science & Engineering, University of Connecticut, Storrs, CT, United States

^c Department of Biomedical Engineering, University of Connecticut, Storrs, CT, United States



ARTICLE INFO

Keywords:

Thermal stability
Proteins
Mutations
Graph neural networks
Attention mechanisms

ABSTRACT

Accurate prediction of protein stability changes due to mutations is instrumental for understanding mechanisms of disease and drug failure, as well as for engineering tailored protein-based materials. In recent years, computational tools using machine learning have been developed to predict stability changes and thereby supplement available experimental methods that may be time consuming and costly for bulk mutational analysis. Existing tools are limited, however, to predicting single point mutations and showing antisymmetric bias in direct/reverse mutations. Here, we develop structure and sequence-based models to predict protein stability changes via Gibbs free energy change upon single or multi-point mutations via two parallel augmented gated attention graph neural networks integrated with global attention blocks receiving subgraphs. These subgraphs are related to mutation sites and constructed through predicted protein contact maps to capture spatial structural information. We train our model on direct and reverse mutations obtained from the S5294 dataset and test on one independent test set and eight most commonly used test sets, including S350, P53, S_{sym} , S669, S1925, S250, and myoglobin. Our approach shows considerable improvement in estimating the impacts of stabilizing mutations, and consistently outperforms other methods by at least a 5.2% improvement in root mean square error. This approach can be employed for finding functionally important protein variants, helping to design new proteins with vast untapped potential for broad pharmaceutical applications.

1. Introduction

Proteins are made up of specific combinations of amino acids (Serpente, 2013). A change within a single amino acid through, e.g., a missense mutation can cause significant protein dysfunction by altering the molecule's stability (Casadio et al., 2011; Bromberg and Rost, 2009). Numerous studies document how protein stability changes contribute to a wide spectrum of molecular disorders (Hartl, 2017; Stefl et al., 2013). Conversely, changes in protein stability can offer valuable insights, as exemplified by the identification of precision medicines through an enhanced understanding of how mutations influence thermal stability and impact a patient's drug resistance or sensitivity (Li et al., 2016). Hence, quantifying stability changes in proteins upon mutation is essential for identifying potential functional variants.

* Corresponding author.

E-mail address: anna.tarakanova@uconn.edu (A. Tarakanova).

The prevailing approach for assessing the impact of mutations on protein stability is through quantification of the changes in Gibbs free energy associated with the folding/unfolding process in mutant proteins compared to their wild type counterparts (Stevens, 2000; Chen et al., 2020). Experimental measurement of free energy changes poses challenges in terms of complexity and expense (Siedhoff et al., 2020), limiting its applicability to crystallizable proteins (Stevens, 2000; Chen et al., 2020). Low-cost and high-throughput computational methods have been developed in response to supplement experimental data (Chen et al., 2020; Fang, 2020; Getov et al., 2016; Dehouck et al., 2009; Pires et al., 2014). Initial methods relied on molecular dynamics simulations and physical-based force fields, using principles of statistical physics to predict mutation effects from structure-based folding free energy changes (Guerois et al., 2002; Kollman et al., 2000; Petukh et al., 2015).

Data-driven models, constructed using efficient and accurate machine learning (ML) algorithms, are another promising approach for stability change prediction (Fang, 2020). Different algorithms and feature extraction procedures are used in existing models, including random forest (Montanucci et al., 2019), supportive vector machine (Capriotti et al., 2005), decision tree (Witvliet et al., 2016), gradient boosting machine (Iqbal et al., 2022), and neural networks (Li et al., 2020). Models can use sequence-based or structure-based features for training. Models built with structure-based features generally perform better than those with sequence-based features, as the latter lack critical structural information (Chen et al., 2020). Experimentally characterized datasets are used to evaluate ML-based predictive models, with moderate to low accuracy being reported by most models and better performance shown by a few, such as FoldX (Schymkowitz et al., 2005) and PremPS (Chen et al., 2020).

The use of ML for stability prediction has three outstanding problems: 1) overfitting, where models perform well on training data but poorly on new data; 2) bias towards destabilizing mutations as most datasets are enriched with these, causing the models to learn noise and poor correlations from this data class; 3) lack of consideration for the anti-symmetric nature of thermal stability changes, leading to biased predictions towards the characteristics of the learning dataset (Chen et al., 2020; Samaga et al., 2021). Deep learning models may be a useful to learn important features even from imbalanced data. Li et al. recently presented a 3D CNN-based predictor of thermal stability changes (Li et al., 2020). This model used protein structures as 3D images, extracting features from their 3D voxels. Although it performed well, it was costly to run. An alternative approach augmented data by creating hypothetical stabilizing mutants from destabilizing mutants to address biases and symmetric issues (Chen et al., 2020).

Recently, attention mechanisms and graph neural networks (GNNs) are used to predict protein properties from sequences or structures (Jha et al., 2022; Chen et al., 2021a). To use GNNs protein structures and sequences are converted into graphs: structural graphs are created from protein structures with atoms as nodes and bonds as edges; sequence graphs are made from amino acids and their relationships in contact maps with nodes and edges, respectively. GNNs can accurately and efficiently elicit sequence/structure-property relationships from either structures or sequences (Jha et al., 2022). Attention mechanisms effectively identify key parts of input data determining target properties (Chen et al., 2021b). Integrating GNNs and attention mechanisms in a model effectively captures a residue's contribution to protein stability (ΔG) in its local structure/sequence environment.

This study presents a new model, ThermoAGT-GA, based on augmented gated attention graph neural networks and global attention mechanisms, for predicting $\Delta\Delta G$ in single or multi-point mutations using both protein sequence and structure. The model uses subgraphs of 10 amino acids surrounding mutation sites (five residues surrounding the mutation site on either side), with amino acids represented as nodes and predicted contact maps as edges. The model is constructed by two parallel AGT-GA networks, one for wild type and one for mutant subgraphs, incorporating thermal stability changes between forward and reverse mutations. Our model is trained on the S5294 dataset and tested on 8 commonly used test sets (S350 (Dehouck et al., 2009), P53 (Pires et al., 2013), S_{sym} (Pucci et al., 2018), S669 (Pancotti et al., 2022), S1925 (Masso and Vaisman, 2008), S250 (Savojardo et al., 2019), myoglobin (Kepp, 2015)) and a combined independent test set (S879). It consistently outperformed other methods with at least a 5.2% improvement in root mean square error. The model also demonstrates outstanding antisymmetry properties, performing similarly on direct and reverse mutations and showing substantial improvement in predicting stabilizing mutations. We note that a mutation introduced into a wild type protein to generate a mutant variant is referred to as "direct mutation," whereas the reverse process of introducing a mutation into the mutant variant to revert it back to the wild type is known as "reverse mutation". These results suggest the model is robust against overfitting and capable of predicting informative structure-property relationships based on biochemistry laws.

2. Materials and methods

2.1. Datasets

For training our model, we consider both single and multi-point mutations. The single point mutation dataset contains 2648 mutants for 131 globular proteins obtained from the ProTherm database (Bava et al., 2004). All mutants have been constructed by introducing single point mutations into wild type structures. The nature of this training set is imbalanced because the number of destabilizing and stabilizing mutations are 2080 and 568, respectively. In order to consider the reversibility property of changes in free energy and overcome bias toward destabilizing mutations, we model reverse mutations in our training set. In this way, if the change in free energy for substitution of residue A with residue B is ΔG , we add $-\Delta G$ for replacing residue B with residue A to our training dataset. Thus, the single point mutation dataset includes 2648 direct and 2648 reverse mutations.

For multi-point mutations, we consider the PTmul dataset including 914 mutations for 91 globular proteins (Montanucci et al., 2019; Kumar et al., 2006). The number of simultaneous variations (i.e., the number of residue substitutions) ranges from 2 to 10 for the PTmul dataset (Kumar et al., 2006). We model reverse mutations for the PTmul dataset similarly to the single point mutation dataset. The final multi-point mutation dataset consists of 914 direct and 914 reverse mutations. In total, the final training set has 5296 single point mutations (S5296) and 1500 multi-point mutations (S1500). We use 328 remaining multi-point mutations for testing (S328).

As we have only 3D structures of wild type proteins, protein sequences of mutants and wild type proteins are obtained from Uniprot (Consortium, 2015). To assess the predictive performance of our model, we use several independent test sets widely used in previous studies. These independent test sets are S350 (Dehouck et al., 2009), P53 (Pires et al., 2013), S_{sym} (Pucci et al., 2018), S669²⁹, S1925³⁰, S250³¹, and myoglobin (Kepp, 2015). S_{sym} (Pucci et al., 2018), S669²⁹, and S250³¹ contain equal amounts of direct and reverse mutations to analyze the antisymmetric property of the model on prediction of thermal stability changes upon mutations. We create a novel composite independent test dataset by eliminating redundant mutants or wild types, whose sequence identity exceeds 20% with sequences in the aforementioned datasets, as well as any mutations that overlap with S5296. As a result, the combined independent test set contains 879 single mutations from 52 proteins (S879).

For the forward mutations, we obtained 3D structures of wild type proteins from the Protein Data Bank (PDB). In the case of reverse mutations, we generated initial protein 3D structures using the BuildModel module of FoldX (Schymkowitz et al., 2005), utilizing wild type protein structures as templates. It's important to note that FoldX (Schymkowitz et al., 2005) specifically optimizes neighboring side chains around the mutation site during mutant structure creation. We use FoldX to generate initial mutant structures by introducing the mutation into wild type structures and then optimizing all side chains surrounding the mutation site via energy minimization to produce final mutant structures. Our choice of FoldX (Schymkowitz et al., 2005) was based on its efficiency, especially in comparison to other available models like AlphaFold2 (Jumper et al., 2021). Furthermore, FoldX is a versatile suite encompassing various modules, including a thermal stability predictor. Despite utilizing FoldX for mutant structure relaxation, our approach to feature extraction, model building, etc., is entirely distinct from FoldX's (Schymkowitz et al., 2005) tool for predicting thermal stability changes. Table 1 summarizes the test sets used in this study.

2.2. Graph generation from protein sequences and protein structures

We build both sequence and structure-based ThermoAGT-GA models to predict $\Delta\Delta G$ from sequence and structure if it is available. To this end, we generate required graphs from both sequence and 3D structure as input to the models. To construct the graph, we define a set of nodes and edges. For the structure-based graph, node features consist of a learnable amino-acid-embedding initialized with amino acid properties (SI Table S1). The edge features include inter-residue distances and orientations (SI Figure S1) obtained from C β –C β distances as well as sine and cosine of angles between any pairs of neighboring residues (within an 11-residue amino acid sequence centered around the mutating residue). We use the reference protein's structure to compute the structure-based features for both reference and mutant proteins (even when the mutant structure is not available). For the sequence-based graph, we consider spatial structural information obtained from protein contact maps (Madani et al., 2022). To this end, we construct edge features from the predicted contacts between the residues (C β –C β distances) in the protein sequence. For node features, we use evolutionary features derived from multiple sequence alignments (MSA) such as position specific score matrix, predicted relative solvent accessibility, secondary structure, and other amino acid properties (SI Table S1). By considering the whole protein as a graph, we extract subgraphs related to mutation sites (11-residue amino acid sequence centered around the mutating residue) for both structure- and sequence-based models. After input preparation, we construct our model in which we use a series of augmented gated attention graph neural networks (Ryu et al., 2018) followed by a global attention block (Louis et al., 2020) (AGT-GA) to update initial subgraphs and extract meaningful features maps correlated to stability changes for both wild type and mutant. More details about the feature extraction procedure for both sequence- and structure-based graphs are provided in the SI.

2.3. Model architecture

After construction of subgraphs for mutation sites within mutant and wild type proteins, we create an end-to-end model, ThermoAGT-GA, via attention-based graph neural networks (Fig. 1). Our model consists of two parallel subnets to work on wild type and mutant subgraphs separately. The architecture of each subnet is identical. Each subnet of ThermoAGT-GA employs two graph soft-attention variants to update initial subgraphs and extract meaningful feature maps correlated to stability changes for wild type and mutant from the input subgraphs. The first type of soft-attention consists of additive multi-head attention (four or eight layers) applied to the one-hop neighbors of each residue. These attention layers are named augmented graph attention (AGT) layers because we augment the node feature vectors with the features from their connecting edge. These AGT layers are only used to extract the locally

Table 1

Datasets used in this study for both single and multi-point mutations for protein structures.

Dataset		#Destabilizing mutations	#Stabilizing mutations	#Protein structures
Training dataset	S5296 + S1500	3398	3398	166
Common test set	S350	260	90	67
	S1925	1373	552	55
	S669	335	334	6
	S250	125	125	134
	P53	32	11	1
	S^{sym}	350	334	357
	Myoglobin	38	96	1
Independent combined test set (single-point mutations)	S879	570	309	186
Multi-point mutations	S328	233	95	32

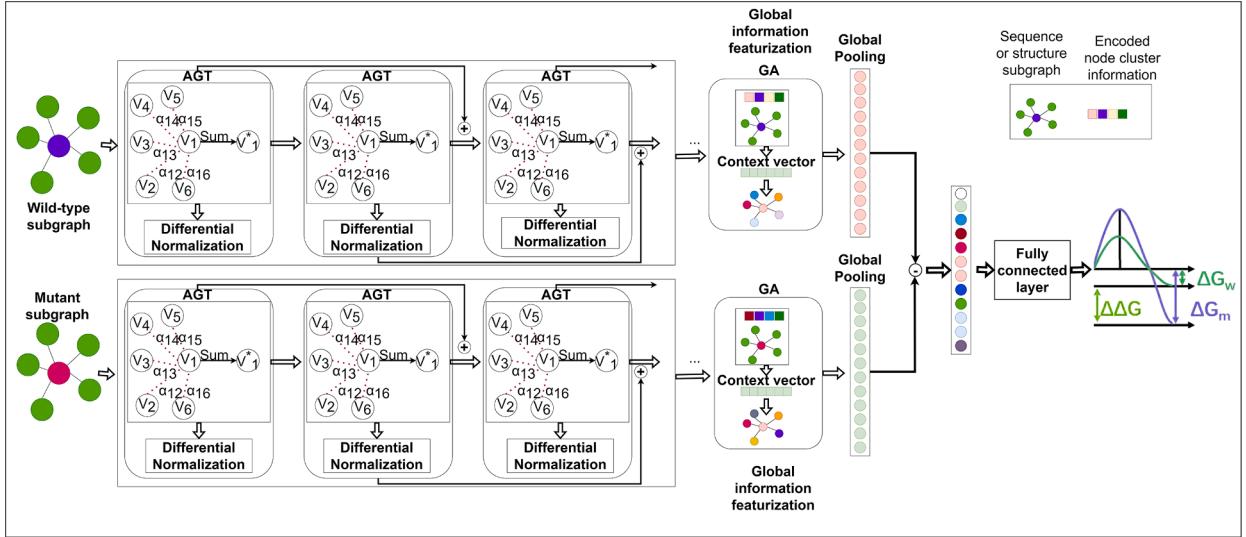


Fig. 1. Model architecture of ThermoAGT-GA.

dependent features between neighboring residues. Upon extracting the local features, our model then uses a unique soft-attention called global attention (GA) at the end to transform neighborhood-dependent information into a global context (concerning all other residues in the protein). The local soft-attention α_{ij} between a node i and a neighbor j can be represented by the following equation:

$$\alpha_{ij} = \frac{\exp(a_{ij})}{\sum_{k \in N_i} \exp(a_{ik})} \quad (1)$$

In Eq. (1), N_i represents the neighborhood of node i, and a_{ij} is the parameterized weight coefficient between nodes i and j, which represents the importance of node j to node i. The global attention, g_i , which is applied immediately before the global pooling, calculates each node's overall importance and can be described as:

$$g_i = \frac{(x_i || E) W}{\sum_{c \in x} (x_c || E) W} \quad (2)$$

In Eq. (2), $x_c \in R^F$ represents a learned embedding, E is a compositional vector of the protein dynamics, $x_c \in R^{1 \times (F+|E|)}$ is a parameterized matrix, and x_c is the learned embedding of any residue c within the protein. By using the combination of eight of these local soft-attentions and one global soft-attention, ThermoAGT-GA can extract comprehensive feature maps from wild type/mutants and also provides interpretable results in terms of each residue's contribution to the change in free energy.

With the expectation that a deeper model should be able to extract even more of these inter-residue-dependent features, we aim to overcome this over-smoothing limitation so that our model can more effectively extract inter-residue-dependent features of the protein. To that end, we devise ThermoAGT-GA which uses additive skip-connections between attention layers that extract the local features and further improve the learning by including differentiable group normalization (DGN) layers. The overall ThermoAGT-GA architecture is shown in Fig. 1. For each subnet, the model consists of 8 AGT layers, followed by DGN operators and skip connections between each of these layers. Then, the last AGT layers are followed by a global attention layer (GA) and a global pooling layer to produce the predicted feature vector with a size of 1×512 . Thus, each subnet generates a 1×512 dimensional feature vector for wild type and mutant contributing to the change in free energy. Finally, by subtracting the wild type feature vector from the mutant feature vector and feeding the produced feature map to the fully connected layer, our model predicts $\Delta\Delta G$.

2.4. Training and evaluation

We performed five-fold cross-validation on the training dataset. In each round four folds were employed to train a model that was evaluated on the left one fold. This process was repeated five times, and the performances of five predictions were averaged as the validation performance. The validations were used to optimize all hyper-parameters (SI Table S2). After fine-tuning the optimal hyperparameters, a model was trained on the whole training dataset and independently tested on nine independent test datasets. To predict change in free energy, we use root mean absolute error as loss function (MSE). Given all $\Delta\Delta G$ labels in the training dataset when mutations occur, the training goal is to minimize the MSE loss based on the following equation:

$$l = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

Table 2Model performance on S305, S669, S1925, Myoglobin, P53, S250, and S_{sym} .

Test set	S350		S669		S1925		Myoglobin		P53		S_{sym}		S250	
	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE
ThermoAGT-GA (structure)	0.882	0.899	0.712	1.394	0.738	1.421	0.673	0.776	0.763	1.479	0.848	1.095	0.907	0.989
ThermoAGT-GA (sequence)	0.722	1.109	0.609	1.626	0.658	1.683	0.634	0.834	0.723	1.573	0.766	1.38	0.799	1.210
PremPS	0.794	0.974	0.648	1.541	0.689	1.572	0.657	0.794	0.754	1.478	0.782	1.254	0.823	1.136
SCONES	0.769	1.107	0.599	1.693	0.624	1.704	0.620	0.846	0.727	1.543	0.718	1.555	0.743	1.324
FoldX	0.802	0.937	0.685	1.430	0.716	0.1456	0.678	0.770	0.759	1.432	0.772	1.274	0.867	1.048
PoPMuSiC	0.723	1.112	0.627	1.602	0.662	1.569	0.653	0.789	0.702	1.509	0.743	1.330	0.765	1.294
SAAFEC-SEQ	0.639	1.241	0.483	1.886	0.521	1.808	0.593	0.904	0.649	1.629	0.684	1.623	0.729	1.382
PROST	0.668	1.209	0.509	1.764	0.594	1.722	0.622	0.843	0.674	1.593	0.592	1.883	0.722	1.389
mCSM	0.745	1.050	0.602	1.683	0.642	1.673	0.619	0.857	0.736	1.528	0.664	1.649	0.790	1.220
DeepDDG	0.698	1.182	0.564	1.729	0.658	1.669	0.603	0.872	0.709	1.549	0.689	1.608	0.748	1.343

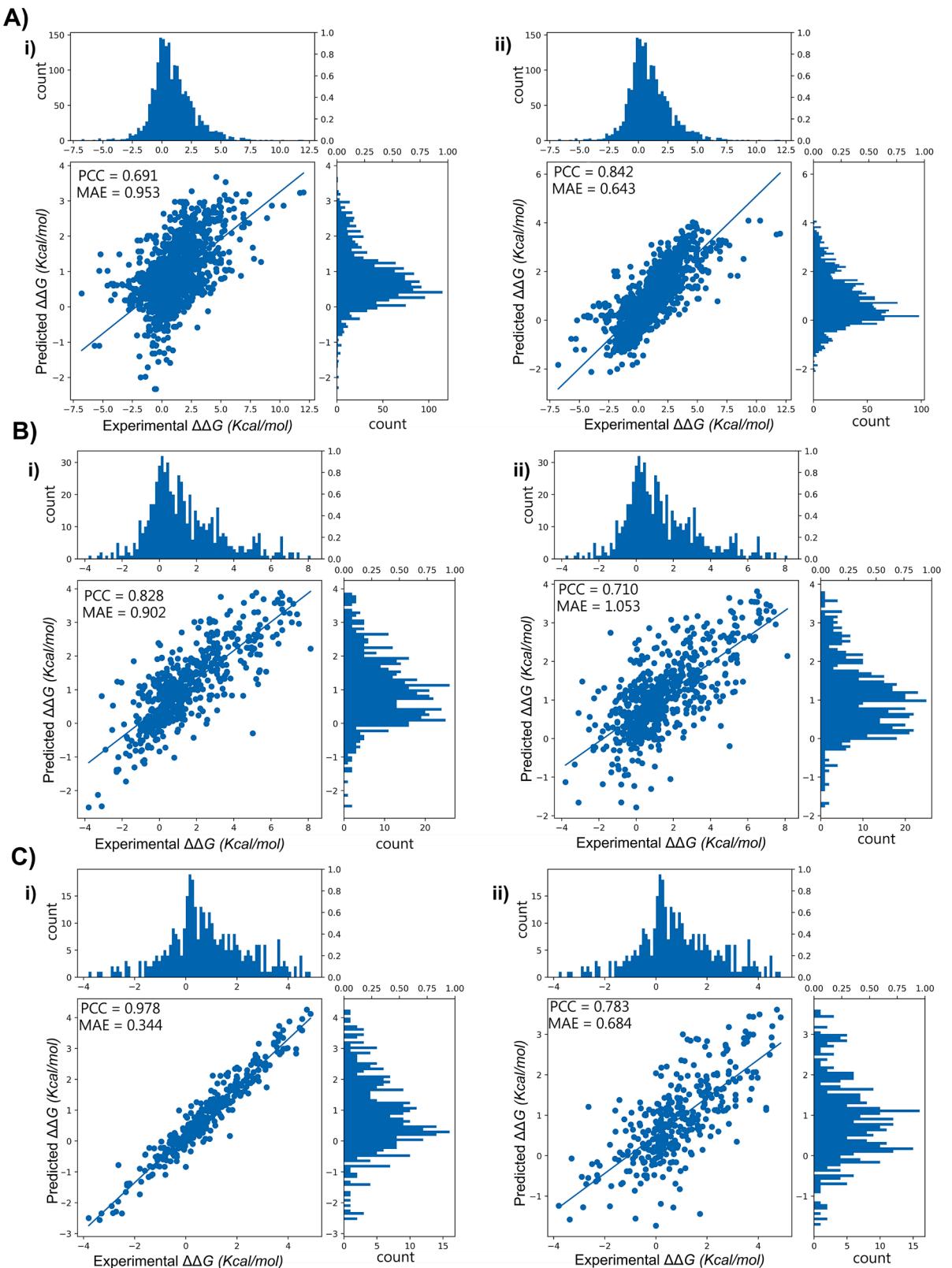


Fig. 2. Performances of both i) structure and ii) sequence-based ThermoAGT-GA models on A) S1925, B)S669, and C) S350 datasets.

where y_i and \hat{y}_i represent the predicted $\Delta\Delta G$ and the experimental $\Delta\Delta G$ for the i th sample, respectively.

The Adam optimizer was used with a weight decay of 0.005 and default parameters for the rest. The initial learning rate was set to 0.005 and decreased by a factor of 10 every time the learning plateaued. The model was trained for 200 epochs with a minibatch size of 4.

To evaluate the predictive performances of ThermoAGT-GA, we employ the root mean square error (RMSE), and Pearson correlation coefficient (PCC). Also, correlation between reverse and direct mutation (r_{fr}) and δ (*bias on prediction*) are used to evaluate the prediction bias between reverse and direct mutation prediction. The equations of r_{fr} and δ are given as follows:

$$r_{fr} = \frac{\text{Covariance } (\Delta GG^{dir}, \Delta GG^{rev})}{\sigma_{dir}\sigma_{rev}} \quad (4)$$

where σ_{dir} and σ_{rev} are the standard deviation for direct and reverse mutations, respectively.

$$\delta = \frac{\sum_i^n (\Delta GG_i^{dir} + \Delta GG_i^{rev})}{2n} \quad (5)$$

We use Receiver Operating Characteristics (ROC) analyses to quantify the performance of our model in distinguishing highly destabilizing ($\Delta\Delta G_{exp} < -1.0$ kcal/mol) and highly stabilizing ($\Delta\Delta G_{exp} > 1.0$ kcal/mol) mutations.

3. Results

3.1. Model performance on training and common test sets

We evaluate the performance of both sequence and structure-based tools first on our training set for both direct and reverse mutations, and second, on common test sets. For each dataset, we compute the performance metrics from the average of all extracted models from five-fold cross-validations. Performance of ThermoAGT-GA tested on S5296 dataset is shown in Table 2 and Fig. 2. PCC and RMSE values between experimental and predicted changes in free energy are 0.865 and 0.977 Kcal/mol for S5296. Our model shows similar performance on reverse (PCC = 0.653) and direct (PCC = 0.668) mutations in S5296. Table 2 represents the performance of the ThermoAGT-GA model on seven common test sets compared to seven existing sequence and structure-based thermal stability predictors including SCONES (Samaga et al., 2021), PremPS (Chen et al., 2020), mCSM (Pires et al., 2013), PoPMuSiC (Dehouck et al., 2009), FoldX (Guerois et al., 2002), SAAFEC-SEQ (Li et al., 2021), PROST (Iqbal et al., 2022), and DeepDDG (Cao et al., 2019). Our structure-based model outperforms existing models for all test sets when performance is evaluated based on PCC and RMSE. For the S350 test set, structure-based ThermoAGT-GA shows at least 0.20 Kcal/mol improvement in RMSE compared to existing models. For the S250 test set, our model achieves lowest RMSE value (0.989 Kcal/mol) and highest PCC value (0.907). Furthermore, on P53, and myoglobin datasets, our structure-based ThermoAGT-GA achieves lowest RMSE (0.776 Kcal/mol and 1.472 Kcal/mol, respectively) and highest PCC values (0.673 and 0.763, respectively) among available models. These results demonstrate that our structure-based thermal stability predictor accurately capture key differences between wild type and mutant proteins related to structural changes in the protein upon point mutations. However, although sequence-based ThermoAGT-GA shows slightly inferior performance on test sets compared to some of available structure-based models, it outperforms existing sequence-based predictors such as SAAFEC-SEQ with a large margin. For example, our sequence-based model shows good performance in the S669 blind test datasets with PCC = 0.609 and MSE = 1.626 Kcal/mol compared with other sequence-based models SAAFEC-SEQ (PCC = 0.483 and MSE = 1.886 kcal/mol) and PROST (PCC = 0.509 and MSE = 1.764 Kcal/mol). Furthermore, both our structure and sequence-based models outperform available models for P53 and myoglobin datasets. For instance, on the myoglobin test set, sequence-based ThermoAGT-GA shows PCC = 0.654 and RMSE = 0.791 Kcal/mol highlighting marginally better performance than all structure-based predictors such as FoldX (PCC = 0.662 and MSE = 0.779 Kcal/mol) and PremPS (PCC = 0.657 and MSE = 0.794 Kcal/mol). Therefore, not only the structure-based model but also the sequence-based ThermoAGT-GA model represent outstanding performance for thermal stability prediction upon mutation. Fig. 2 shows the performance of both models on three different test sets.

3.2. Model performance on S879 test set (combined independent test set)

To evaluate our model, we consider the S879 dataset as it does not have any similarity with the training data, meaning that sequence similarity between the datasets is less than 15 %. We predict the values of stability changes using both ThermoAGT-GA models and seven most commonly tested and reliable predictors including SCONES (Samaga et al., 2021), PremPS (Chen et al., 2020), mCSM (Pires et al., 2013), PoPMuSiC (Dehouck et al., 2009), FoldX (Guerois et al., 2002), SAAFEC-SEQ, PROST (Iqbal et al., 2022), and DeepDDG (Cao et al., 2019). The results reported in Table 3 demonstrate that ThermoAGT-GA achieves the highest prediction accuracy especially for stabilizing mutations with PCC of up to 0.832, 0.792, and 0.677 for all destabilizing and stabilizing mutations, respectively. However, we note that there are 7 mutations with a large difference of more than 4.3 Kcal/mol between experimental and predicted values (Fig. 3A). The experimental values of these mutations, with one exception, exceed 5.83 Kcal/mol, whereas in our training dataset, all mutations have experimental values below 4.6 Kcal/mol. This likely accounts for the significant disparity between the experimental and predicted values for these 7 mutations. To evaluate the effect of adding reverse mutations on performance of the ThermoAGT-GA model, we evaluate the performance of ThermoAGT-GA when trained only on the forward mutation dataset of S2648 and tested on the S879. The PCC is 0.762 and 0.403 for destabilizing and stabilizing mutations, respectively

Table 3

Model performances on S879 datasets for all mutations, highly stabilizing ($\Delta\Delta G_{exp} > 1.0$ Kcal/mol), highly destabilizing ($\Delta\Delta G_{exp} < -1.0$ Kcal/mol), destabilizing, and stabilizing mutations.

Test sets	S879											
	All mutations		Highly stabilizing			Highly destabilizing			Stabilizing		Destabilizing	
	PCC	RMSE	PCC	AUC	RMSE	PCC	AUC	RMSE	PCC	RMSE	PCC	RMSE
ThermoAGT-GA (structure)	0.823	1.405	0.647	0.783	1.582	0.783	0.799	1.349	0.677	1.399	0.792	1.023
ThermoAGT-GA (sequence)	0.763	1.545	0.563	0.718	1.664	0.712	0.725	1.424	0.591	1.583	0.723	1.372
PremPS	0.792	1.443	0.602	0.732	1.588	0.740	0.755	1.372	0.637	1.437	0.765	1.189
SCONES	0.664	1.678	0.589	0.722	1.652	0.622	0.683	1.691	0.642	1.420	0.664	1.475
FoldX	0.593	1.899	0.283	0.583	2.294	0.683	0.722	1.573	0.352	2.274	0.703	1.410
PoPMuSiC	0.702	1.633	0.392	0.665	1.943	0.779	0.809	1.333	0.453	1.984	0.863	0.973
SAAFEC-SEQ	0.479	2.291	0.209	0.532	2.328	0.404	0.593	2.168	0.258	2.573	0.583	1.683
PROST	0.503	2.103	0.399	0.658	2.277	0.438	0.631	2.007	0.471	1.927	0.503	1.903
mCSM	0.559	2.032	0.221	0.587	2.309	0.623	0.694	1.644	0.274	2.473	0.662	1.472
DeepDDG	0.622	1.837	0.317	0.595	2.182	0.684	0.718	1.569	0.348	2.208	0.746	1.274

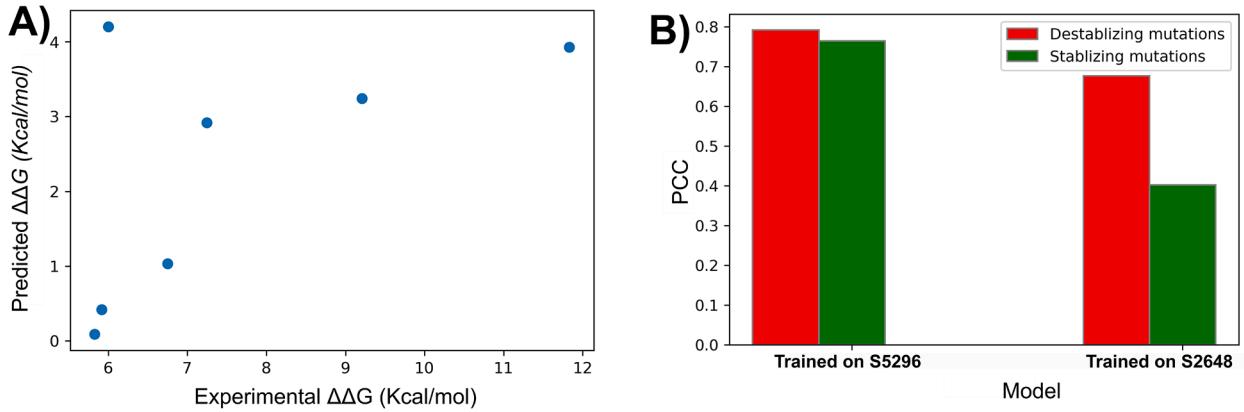


Fig. 3. A) 7 mutations with a large difference of more than 4.3 Kcal/mol between experimental and predicted values. B) Performance of ThermoAGT-GA for prediction of destabilizing and stabilizing mutation after training on S5296 including both direct and reverse mutations and S2648 including only direct mutations.

(Fig. 3B). The results confirm that the usage of reverse mutations improves the performance of our model in estimating the effects of stabilizing mutations without compromising the prediction accuracy.

3.3. Performance of ThermoAGT-GA on highly stabilizing and highly destabilizing mutations

We carry out ROC analysis to quantify the performance of ThermoAGT-GA in distinguishing highly destabilizing ($\Delta\Delta G_{exp} < -1.0$ Kcal/mol) and highly stabilizing ($\Delta\Delta G_{exp} > 1.0$ Kcal/mol) mutations from the others. Fig. 4 and Table 3 show the excellent performance of ThermoAGT-GA in evaluating highly destabilizing/stabilizing mutations, outperforming other methods. The area under curve (AUC) of structure-based ThermoAGT-GA is 0.783 and 0.799 for highly stabilizing and destabilizing mutations, respectively. AUC value for the closest competitor (PremPS (Chen et al., 2020)) is at least 7% less than the ThermoAGT-GA model. Also, the AUC values and ROC curves for both highly stabilizing/destabilizing mutations are almost identical and hint that our method is capable of identifying significantly stabilizing and destabilizing mutations equally well without any bias toward destabilizing mutations. We can conclude that our method is able to learn and explain a substantial portion of the variance in the experimental data. It shows balanced performance for both stabilizing and destabilizing mutations, classifying significant mutations ($|\Delta\Delta G| \geq 1$ kcal/mol) fairly well with nearly all the samples in S879 classified correctly.

3.4. ThermoAGT-GA captures antisymmetric thermodynamic properties of forward/reverse mutations

We investigate the antisymmetric thermodynamic features of our model by assessing performance on S_{sym} (Pucci et al., 2018), S_{669}^{29} , and S_{250}^{31} datasets, including forward mutations and their corresponding reverse mutations in identical experimental conditions. The symmetric consistency has been traditionally measured using a correlation coefficient R_{fr} and mean prediction bias δ , the correlation coefficient between the predictions for forward mutations and their corresponding reverse mutations. The δ for a sample is defined as $\Delta\Delta G_{forward} + \Delta\Delta G_{reverse}$, and the average δ over all samples (denoted by $\langle \delta \rangle$) and R_{fr} are used as a measure of overall consistency of the method. A method with perfect symmetric consistency will have an R_{fr} value of -1, $\langle \delta \rangle$ a value of 0. Our method

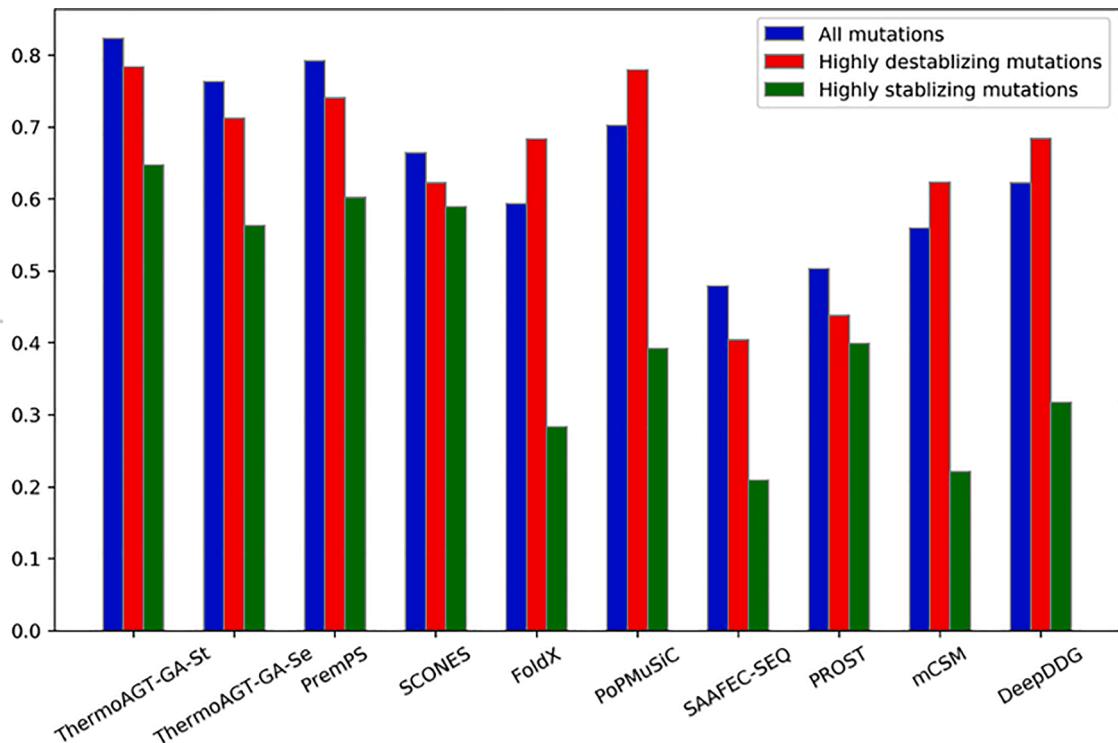


Fig. 4. Models' performances on S879 datasets for all mutations, highly stabilizing, and highly destabilizing mutations.

performs well among all test cases, has the highest prediction accuracy for forward and reverse mutations, a very low antisymmetric prediction bias, and still shows robust performance with each test set (Fig. 5 and Table 4). We also compare the performance of our method on Ssym, S669, and S250 with existing methods. Our model predicts data points within these three datasets with R_{fr} and δ values of -0.98 and -0.04 , respectively (Table 4). Furthermore, our predictions for both forward and reverse mutations, along with the distribution of the prediction bias for S250, Ssym, and S669, confirm that our method exhibits a near-perfect R_{fr} value of -0.98 (Fig. 6).

3.5. Model performance for multi-point mutations

One of the notable contributions of our proposed model is its capability for predicting changes in free energy upon multi-point mutations; to our knowledge, this is the first such machine learning-based tool available. We train our model on the S1500 dataset including both reverse and direct multi-point mutations. To evaluate the performance of our model, we test on the S328 dataset randomly selected from the PTmul dataset. ThermoAGT-GA achieve PCC of up to 0.518 which is 18 % higher than DDGun model (Montanucci et al., 2019) (Table 5). This excellent performance in prediction of change in free energy for multi-point mutations highlights the outstanding capability of ThermoAGT-GA in capturing key structural changes upon mutations in different sites of a protein by extracting non-linear relationships between various location of mutations through subgraphs.

3.6. Effect of resolution and experimental method on prediction of thermal stability changes upon mutations

Recent investigations have undertaken a critical analysis of stability change predictors, particularly examining their reliance on the 3D structures of proteins. This scrutiny has considered factors such as protein resolution and the methods employed for structure determination, revealing intriguing insights into their interplay. To systematically evaluate the impact of experimental strategies for protein structure acquisition on thermal stability predictors, we partitioned the proteins within the S879 dataset into two distinct groups: those acquired through nuclear magnetic resonance (NMR) spectroscopy and those determined via X-ray diffraction. As delineated in Fig. 6, our comprehensive analysis unearthed discernible performance variations across these two subsets. The accompanying error bars underscore the sensitivity of most methods to the chosen experimental strategy, with an overarching trend suggesting slightly improved performance for NMR-derived structures. However, it is noteworthy that FoldX (Schymkowitz et al., 2005) exhibited a pronounced predilection for X-ray-derived structures. Significantly, our ThermoAGT-GA model emerged as highly sensitive to NMR-based structures and demonstrated exceptional performance across both NMR and X-ray-derived structures. This substantial performance divergence between NMR and X-ray-derived structures likely stems from differences in the number of training structures drawn from these methodologies rather than any inherent bias towards a particular experimental technique, as an

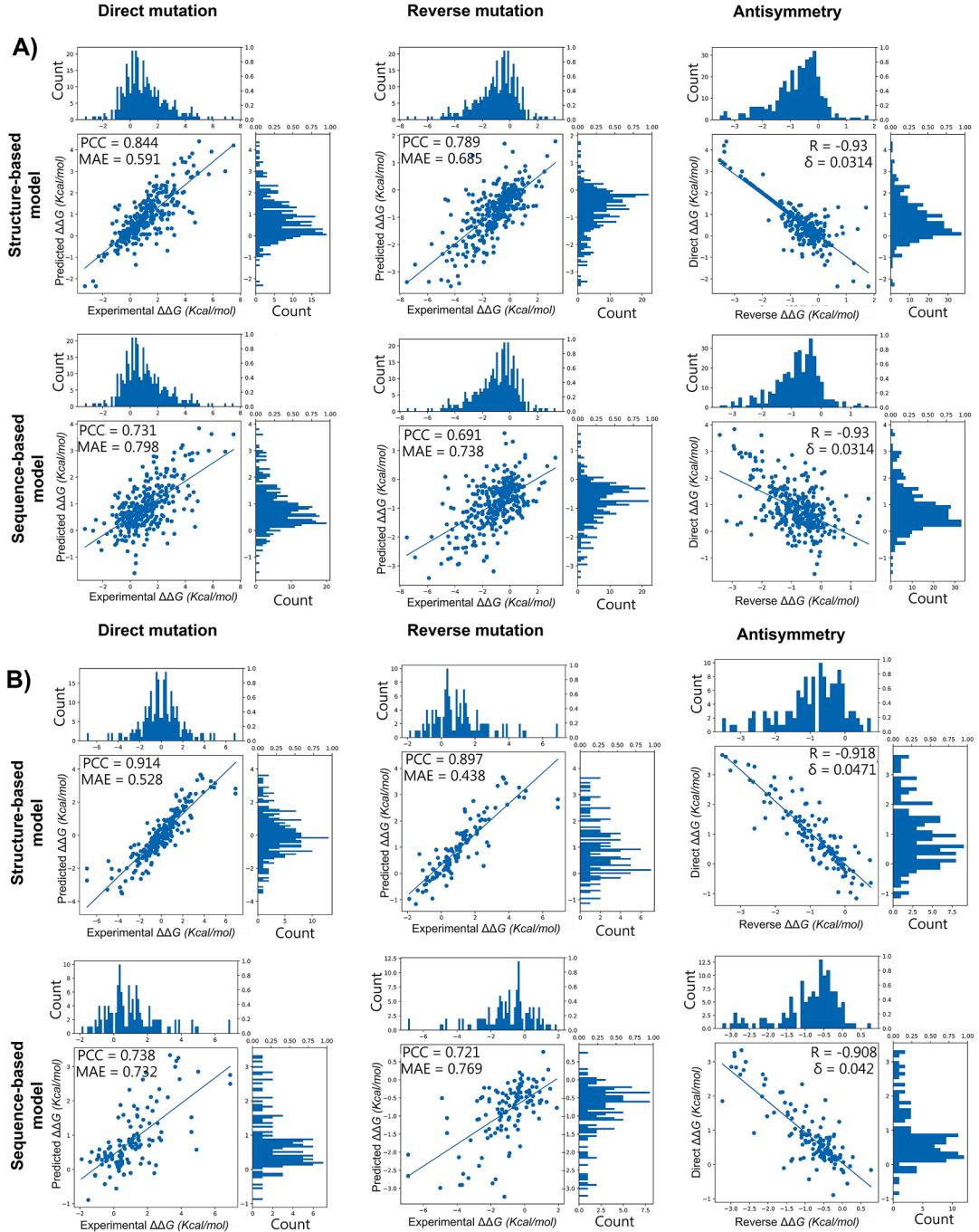


Fig. 5. Performance of both structure- and sequence-based ThermoAGT-GA models on direct and reverse mutations to evaluate antisymmetric properties of all models for thermal stability change predictions for A) S_{sym} and B) S250 datasets.

overwhelming 69 % of the training data originated from NMR-based sources.

Additionally, we explored the influence of structural resolution on model performance. To do this, we segregated the X-ray-derived structures into two distinct categories based on their resolution, distinguishing between structures with resolutions higher and lower than 3 Å. As depicted in Fig. 6, our findings underscore that structural resolution exerts minimal impact on the majority of the models, with a notable exception being the sequence-based models, SAAFEC-SEQ (Li et al., 2021) and PROST (Iqbal et al., 2022), which exhibited higher sensitivity to structural resolution.

Table 4

Model performance on direct and reverse mutations to evaluate antisymmetric thermodynamic properties of all models for thermal stability change predictions.

Test sets	S669						S250						S_{sym}					
	Reverse		Direct		Antisymmetry		Reverse		Direct		Antisymmetry		Reverse		Direct		Antisymmetry	
	PCC	RMSE	PCC	RMSE	R _{fr}	δ	PCC	RMSE	PCC	RMSE	R _{fr}	δ	PCC	RMSE	PCC	RMSE	R _{fr}	δ
Metrics																		
ThermoAGT-GA (structure)	0.652	1.419	0.641	1.388	-0.97	-0.04	0.814	1.228	0.820	1.189	-0.98	-0.03	0.797	1.128	0.802	1.107	-0.98	-0.03
ThermoAGT-GA (sequence)	0.573	1.662	0.588	1.635	-0.94	-0.07	0.702	1.442	0.727	1.428	-0.93	-0.09	0.724	1.377	0.739	1.362	-0.95	-0.07
PremPS	0.618	1.602	0.593	1.618	-0.95	-0.06	0.729	1.403	0.768	1.219	-0.88	-0.14	0.741	1.332	0.764	1.291	-0.91	-0.10
SCONES	0.473	1.883	0.504	1.793	-0.90	-0.11	0.639	1.573	0.663	1.473	-0.90	-0.12	0.653	1.563	0.693	1.443	-0.89	-0.16
FoldX	0.439	1.893	0.692	1.334	-0.78	-0.32	0.653	1.539	0.841	1.100	-0.65	-0.48	0.632	1.644	0.734	1.389	-0.68	-0.46
PoPMuSiC	0.407	1.934	0.563	1.745	-0.74	-0.38	0.620	1.567	0.686	1.447	-0.79	-0.31	0.619	1.723	0.709	1.430	-0.71	-0.29
SAAFEC-SEQ	0.284	2.092	0.462	1.867	-0.72	-0.43	0.583	1.683	0.673	1.499	-0.72	-0.40	0.493	1.921	0.673	1.622	-0.73	-0.27
PROST	0.362	1.908	0.509	1.820	-0.86	-0.18	0.603	1.564	0.653	1.542	-0.89	-0.12	0.509	1.883	0.528	1.899	-0.88	-0.16
mCSM	0.433	1.876	0.593	1.705	-0.85	-0.2	0.778	1.283	0.723	1.427	-0.84	-0.22	0.617	1.342	0.649	1.399	-0.86	-0.2

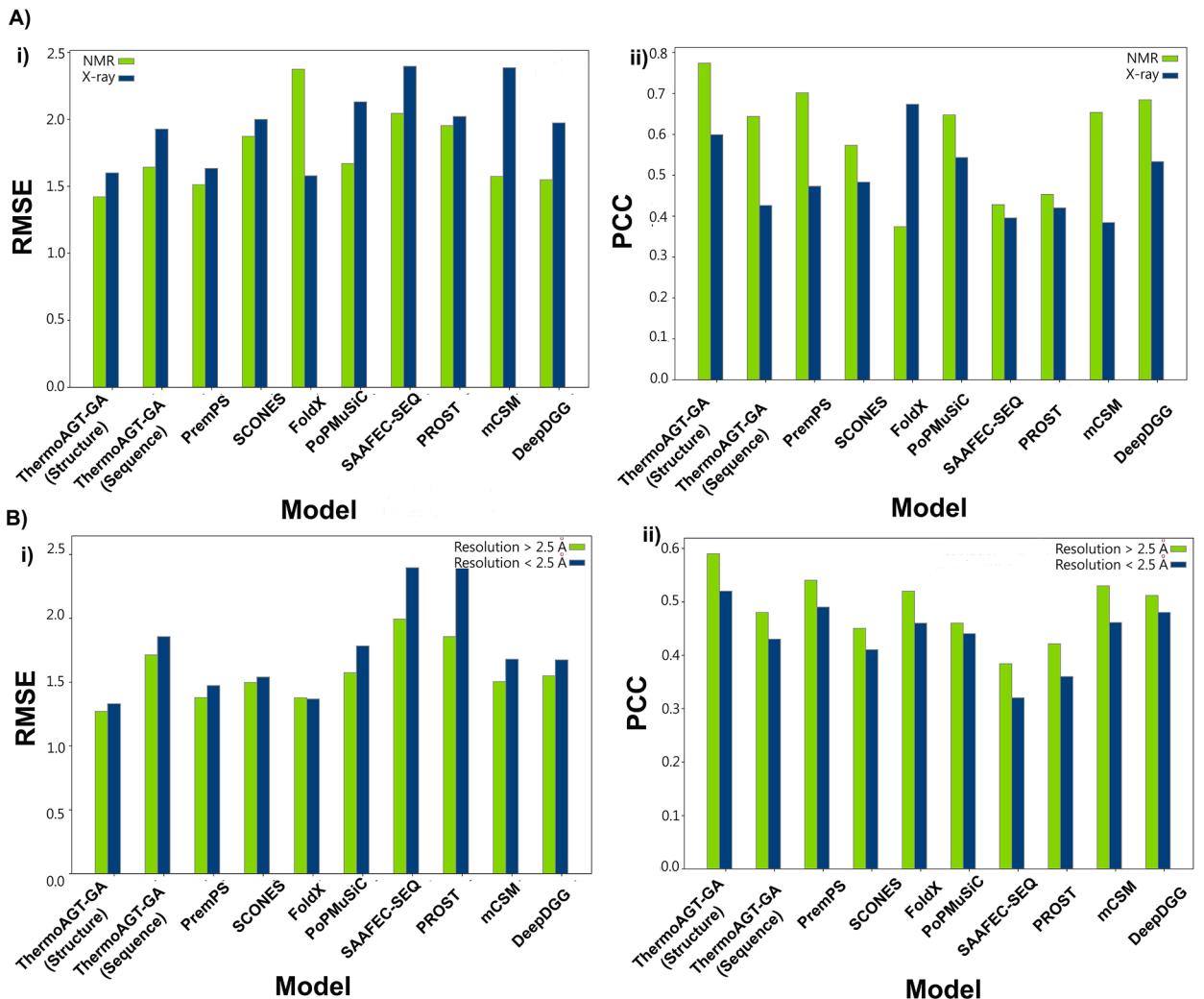


Fig. 6. Effect of A) NMR and X-ray crystallographic experimental methods for 3D structure generations i) RMSE ii) PCC and B) protein resolutions of performances of thermal stability predictors i) RMSE ii) PCC.

Table 5

Performance of ThermoAGT-GA models and DDGun model (Montanucci et al., 2019) on multi-point mutation dataset (S328 dataset).

Test sets	S328	
Metrics	PCC	RMSE
ThermoAGT-GA (structure)	0.518	1.899
ThermoAGT-GA (sequence)	0.395	2.139
DDGun	0.322	2.374

3.7. Case study for validation of ThermoAGT-GA performance

Our model operates hierarchically, enabling the prediction of ΔG contributions arising from individual residue-residue interactions. These predictions hold the potential to offer valuable insights into the factors driving protein stabilization or destabilization. We utilize an interaction map to consolidate these intermediate predictions, facilitating a more comprehensive assessment.

To evaluate the practical applicability of our model, we performed a case study involving Pyrazinamide (PZA). PZA is a primary drug used to combat latent Mycobacterium tuberculosis isolates, but resistance can arise due to mutations in the pncA and rpsA genes, responsible for encoding pyrazinamidase (PZase) and ribosomal protein S1 (RpsA), respectively (Khan et al., 2019). Specifically, we introduced two mutations, D343N and I351F, into RpsA (Fig. 7). For our analysis, we leveraged the 3D structure of the RpsA protein

(PDB ID: 4NNI) obtained from the Protein Data Bank (Berman et al., 2000). To generate mutant structures, we followed the same procedure utilized in dataset preprocessing.

Our structure-based model successfully predicted a change in free energy of 3.06 Kcal/mol for 4NNI-D343N and 3.11 Kcal/mol for 4NNI-I351F. Remarkably, these predictions closely aligned with experimental findings recently confirmed by Khan et al., who reported free energy landscape changes of 3.2 Kcal/mol and 2.9 Kcal/mol for the 4NNIA-D343N and 4NNIA-I351F mutations in ribosomal protein S1 (rpsA), associated with pyrazinamide resistance (Khan et al., 2019). This congruence underscores the potential of ThermoAGT-GA as a rapid estimator of free energy changes resulting from protein mutations, showcasing its viability in medical contexts for swift and accurate assessments.

4. Discussion

Our model addresses two significant gaps in thermal stability change prediction. Firstly, existing thermodynamic datasets used for predicting thermal stability changes often display a skew toward destabilizing mutations, resulting in an imbalance among different mutation types. Secondly, many machine learning-based models suffer from a lack of symmetric consistency due to their reliance on biased training data. To overcome these challenges, our model takes a dual-pronged approach. It incorporates hypothetical reverse mutations and hierarchical symmetry within its architecture while considering transitive properties to enhance performance. In our model, each residue is treated as a node capable of estimating inter-residue interaction energies. ThermoAGT-GA calculates the contribution of neighboring residues at the mutation site toward protein stability by summing all interaction energies with their neighbors. This approach enables our model to learn key interactions between residues through message passing on graphs during training. We intentionally avoid using features spanning more than two residues or involving both reference and mutant proteins, ensuring robustness against overfitting. The network's exposure to all possible inter-residue interactions during training equips it to handle previously unseen samples at test time, potentially generalizing to mutations with limited representation in the training set.

Also, our method effectively captures a significant portion of the variance in experimental data. It exhibits balanced performance for both stabilizing and destabilizing mutations across all test sets. Notably, ThermoAGT-GA stands as the first tool capable of utilizing both protein sequence and structure for thermal stability prediction, enhancing its robustness and comprehensiveness.

To gauge method performance accuracy, we conducted a comprehensive evaluation using multiple test sets, demonstrating that ThermoAGT-GA outperforms commonly used predictors across eight test sets. It exhibits excellent symmetric and transitive consistency, aligning with state-of-the-art methods. While surface mutations typically pose greater prediction challenges, our method's performance in this regard may relate to its omission of explicit residue-solvent interactions. We anticipate that improved sampling

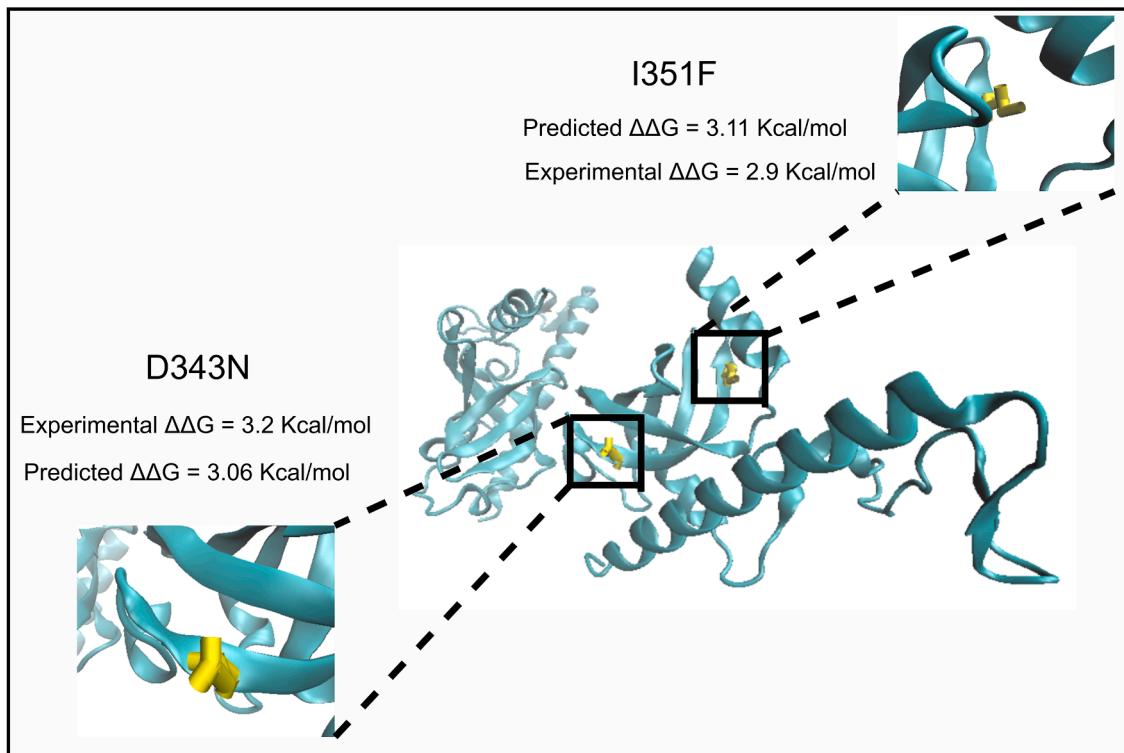


Fig. 7. 3D structure of RpsA with 2 introduced mutations (D343N and I351F). For D343N mutation, experimental and predicted values of change in free energy are 3.2 and 3.06 Kcal/mol, respectively. For I351F mutation, experimental and predicted values of change in free energy are 2.9 and 3.11 Kcal/mol, respectively.

procedures, transfer learning, and auxiliary losses can mitigate intraensemble variance.

Our hierarchical structure, similar to how force field calculations work, allows us to easily add extra guidance from force fields or statistical information to help with fine-tuning and optimizing. Moreover, our method displays low variance in ensemble performance across diverse training datasets, indicating its relative resilience to overfitting. The architecture further grants insights into intermediate computations, allowing for the incorporation of explainable machine learning methods that shed light on the model's predictive mechanisms, crucial in light of limited and biased training data because of capabilities of our model to work on wild type and mutant separately in parallel architecture.

We note that our model employs an attention-based graph neural network (AGT-GA). Attention-based models, by their nature, excel at capturing key aspects of inputs to predict specific properties. The AGT-GA model specifically incorporates both local and global attention networks. The local attention network captures the influence of neighboring residues on the mutation site, providing a nuanced understanding of local interactions. Meanwhile, the global attention block extracts the most significant features from all nodes (residues) in the protein on the mutation site, ensuring a comprehensive and interpretable model. The dual attention mechanism enhances the overall interpretability of our model by capturing both local and global context in predicting the desired property.

Another pivotal aspect of our model lies in its ability to predict thermal stability changes resulting from multi-point mutations—a feature lacking in many existing models. ThermoAGT-GA's architecture enables it to effectively assess the impact of various mutation sites on changes in free energy. Additionally, our model exhibits low dependence on the experimental techniques employed for protein production and the 3D resolution of proteins obtained from X-ray crystallography.

5. Conclusion

Prediction of protein stability can be utilized for a myriad of applications in biochemistry and protein engineering, from molecular structure prediction, to vaccine design and computationally-guided protein-based material synthesis. Despite recent advancements, access to suitable, diverse, and unbiased data remains a challenge, necessitating innovative model architectures and datasets. In this study, we propose graph-based models with a hierarchical architecture that integrates domain knowledge to overcome these data limitations. Our model is trained on both reverse and direct mutations, mitigating bias towards destabilizing mutations and considering antisymmetric and transitive properties of $\Delta\Delta G$. ThermoAGT-GA accurately predicts thermal stability changes resulting from single and multi-point mutations using either protein sequence or structure. It calculates $\Delta\Delta G$ by independently predicting ΔG contributions for reference and mutant residues in missense mutations, reducing reliance on data for self-consistency properties and enhancing resilience to unbalanced datasets. ThermoAGT-GA surpasses existing models by a substantial margin on commonly used datasets and a consolidated independent test set. Notably, it excels at predicting stabilizing mutations, facilitating the identification of optimal mutation sites to enhance protein stability. In summary, ThermoAGT-GA serves as a powerful tool for the rapid, reliable, and cost-effective prediction of protein stability changes, guiding more efficient protein engineering endeavors.

Novelty statement

To the best of our knowledge, the authors declare that no existing work has significant overlap with the submitted manuscript.

CRediT authorship contribution statement

Mohammad Madani: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Tarakanova:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All datasets and source codes are available in <https://github.com/mahan-fcb/ThermoAGT>.

Acknowledgments

This work used Stampede 2 and Ranch at the Texas Advanced Computing Center and Bridges at the Pittsburg Supercomputing Center through allocation MCB180008 from the Extreme Science and Engineering Discovery Environment (XSEDE) (Towns et al., 2014), which was supported by National Science Foundation grant number #1548562, as well as from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program (Boerner et al., 2023), which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmps.2023.105531.

References

- Bava, K.A., Gromiha, M.M., Uedaira, H., Kitajima, K., Sarai, A., 2004. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.* 32 (suppl_1), D120–D121.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. *Nucleic Acids Res.* 28 (1), 235–242.
- Boerner, T.J., Deems, S., Furlani, T.R., Knuth, S.L., Towns, J., 2023. ACCESS: advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: services & support. Practice and Experience in Advanced Research Computing, pp. 173–176.
- Bromberg, Y., Rost, B., 2009. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinf.* 10 (8), S8.
- Cao, H., Wang, J., He, L., Qi, Y., Zhang, J.Z., 2019. DeepDDG: predicting the stability change of protein point mutations using neural networks. *J. Chem. Inf. Model.* 59 (4), 1508–1514.
- Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* 33 (suppl_2), W306–W310.
- Casadio, R., Vascura, M., Tiwari, S., Fariselli, P., Luigi Martelli, P., 2011. Correlating disease-related mutations to their effect on protein stability: a large-scale analysis of the human proteome. *Hum. Mutat.* 32 (10), 1161–1170.
- Chen, Y., Lu, H., Zhang, N., Zhu, Z., Wang, S., Li, M., 2020. PremPS: predicting the impact of missense mutations on protein stability. *PLoS Comput. Biol.* 16 (12), e1008543.
- Chen, J., Zheng, S., Zhao, H., Yang, Y., 2021a. Structure-aware protein solubility prediction from sequence through graph convolutional network and predicted contact map. *J. Cheminform.* 13 (1), 1–10.
- Chen, C., Wu, T., Guo, Z., Cheng, J., 2021b. Combination of deep neural network with attention mechanism enhances the explainability of protein contact prediction. *Proteins Struct. Funct. Bioinf.* 89 (6), 697–707.
- Consortium, U., 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43 (D1), D204–D212.
- Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., Rooman, M., 2009. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* 25 (19), 2537–2543.
- Fang, J., 2020. A critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. *Brief. Bioinformatics* 21 (4), 1285–1292.
- Getov, I., Petukh, M., Alexov, E., 2016. SAAFEC: predicting the effect of single point mutations on protein folding free energy using a knowledge-modified MM/PBSA approach. *Int. J. Mol. Sci.* 17 (4), 512.
- Guerois, R., Nielsen, J.E., 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), 369–387.
- Hartl, F.U., 2017. Protein misfolding diseases. *Annu. Rev. Biochem.* 86 (1), 21–26.
- Iqbal, S., Ge, F., Li, F., Akutsu, T., Zheng, Y., Gasser, R.B., Yu, D.J., Webb, G.I., Song, J., 2022. PROST: alphaFold2-aware sequence-based predictor to estimate protein stability changes upon missense mutations. *J. Chem. Inf. Model.* 62 (17), 4270–4282.
- Jha, K., Saha, S., Singh, H., 2022. Prediction of protein–protein interaction using graph neural networks. *Sci. Rep.* 12 (1), 1–12.
- Juniper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589.
- Kepp, K.P., 2015. Towards a "Golden Standard" for computing globin stability: stability and structure sensitivity of myoglobin mutants. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomes* 1854 (10), 1239–1248. Part A.
- Khan, M.T., Khan, A., Rehman, A.U., Wang, Y., Akhtar, K., Malik, S.I., Wei, D.Q., 2019. Structural and free energy landscape of novel mutations in ribosomal protein S1 (rpsA) associated with pyrazinamide resistance. *Sci. Rep.* 9 (1), 7482.
- Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33 (12), 889–897.
- Kumar, M.S., Bava, K.A., Gromiha, M.M., 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 (suppl_1), D204–D206.
- Li, M., Kales, S.C., Ma, K., Shoemaker, B.A., Crespo-Barreto, J., Cangelosi, A.L., Lipkowitz, S., Panchenko, A.R., 2016. Balancing protein stability and activity in cancer: a new approach for identifying driver mutations affecting CBL ubiquitin ligase activation. Mutations affect multiple stages of CBL activation cycle. *Cancer Res.* 76 (3), 561–571.
- Li, B., Yang, Y.T., Capra, J.A., Gerstein, M.B., 2020. Predicting changes in protein thermodynamic stability upon point mutation with deep 3D convolutional neural networks. *PLoS Comput. Biol.* 16 (11), e1008291.
- Li, G., Panday, S.K., Alexov, E., 2021. SAAFEC-SEQ: a sequence-based method for predicting the effect of single point mutations on protein thermodynamic stability. *Int. J. Mol. Sci.* 22 (2), 606.
- Louis, S.Y., Zhao, Y., Nasiri, A., Wang, X., Song, Y., Liu, F., Hu, J., 2020. Graph convolutional neural networks with global attention for improved materials property prediction. *Phys. Chem. Chem. Phys.* 22 (32), 18141–18148.
- Madani, M., Behzadi, M.M., Song, D., Ilies, H.T., Tarakanova, A., 2022. Improved inter-residue contact prediction via a hybrid generative model and dynamic loss function. *Comput. Struct. Biotechnol. J.* 20, 6138–6148.
- Masso, M., Vaisman, I.I., 2008. Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics* 24 (18), 2002–2009.
- Montanucci, L., Capriotti, E., Frank, Y., Ben-Tal, N., Fariselli, P., 2019. DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC Bioinf.* 20 (14), 1–10.
- Pancotti, C., Benevenuta, S., Birolo, G., Alberini, V., Repetto, V., Sanavia, T., Capriotti, E., Fariselli, P., 2022. Predicting protein stability changes upon single-point mutation: a thorough comparison of the available tools on a new dataset. *Brief. Bioinformatics* 23 (2).
- Petukh, M., Li, M., Alexov, E., 2015. Predicting binding free energy change caused by point mutations with knowledge-modified MM/PBSA method. *PLoS Comput. Biol.* 11 (7), e1004276.
- Pires, D.E.V., Ascher, D.B., Blundell, T.L., 2013. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30 (3), 335–342.
- Pires, D.E.V., Ascher, D.B., Blundell, T.L., 2014. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res.* 42 (W1), W314–W319.
- Pucci, F., Bernaerts, K.V., Kwasigroch, J.M., Rooman, M., 2018. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* 34 (21), 3659–3665.
- Ryu, S., Lim, J., Hong, S.H., Kim, W.Y., 2018. Deeply learning molecular structure-property relationships using attention-and gate-augmented graph convolutional network. *ArXiv. arXiv:1805.10988.*

- Samaga, Y.B., Raghunathan, S., Priyakumar, U.D., 2021. SCONES: self-consistent neural network for protein stability prediction upon mutation. *J. Phys. Chem. B* 125 (38), 10657–10671.
- Savojardo, C., Martelli, P.L., Casadio, R., Fariselli, P., 2019. On the critical review of five machine learning-based algorithms for predicting protein stability changes upon mutation. *Brief. Bioinformatics* 22 (1), 601–603.
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., 2005. The FoldX web server: an online force field. *Nucleic. Acids. Res.* 33 (suppl_2), W382–W388.
- Serpente, N., 2013. Beyond a pedagogical tool: 30 years of molecular biology of the cell. *Nat. Rev. Mol. Cell Biol.* 14 (2), 120–125.
- Siedhoff, N.E., Schwaneberg, U., Davari, M.D., 2020. Machine learning-assisted enzyme engineering. *Meth. Enzymol.* 643, 281–315.
- Stefl, S., Nishi, H., Petukh, M., Panchenko, A.R., Alexov, E., 2013. Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* 425 (21), 3919–3936.
- Stevens, R.C., 2000. High-throughput protein crystallization. *Curr. Opin. Struct. Biol.* 10 (5), 558–563.
- Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., 2014. XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.* 16 (5), 62–74.
- Witvliet, D.K., Strokach, A., Giraldo-Forero, A.F., Teyra, J., Colak, R., Kim, P.M., 2016. ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity. *Bioinformatics* 32 (10), 1589–1591.