Predicting Mutation-Induced Relative Protein-Ligand Binding Affinity Changes via Conformational Sampling and Diversity Integration with Subsampled Alphafold2 in Few-Shot Learning

Liangxu Xie[1], Xiaohua Lu[1], Dawei Zhang[1], Lei Xu[1], Mingjun Yang[2], Young-Seob Jeong[3], Shan Chang[1,*], Xiaojun Xu[1,*]

[1] Institute of Bioinformatics and Medical Engineering, Jiangsu University of Technology, Changzhou, 213001, China

[2] Shenzhen Jingtai Technology Co., Ltd. (XtalPi), Shenzhen 518045, China

[3] Department of Computer Engineering, Chungbuk National University, Cheongju, 28644, South Korea

Correspondence: S.C.: schang@jsut.edu.cn; X.X.: xuxiaojun@jsut.edu.cn

Abstract

Predicting the impact of mutations on protein-ligand binding affinity is crucial in drug discovery, particularly in addressing drug resistance and repurposing existing drugs. Current structure-based methods are constrained by their dependence on known protein-ligand complex structures. The absence of structural data for mutated proteins, along with the potential of mutations to modify protein conformational states, brings challenges for reliable predictions. The downstream application of conformational sampling of Alphafold, such as the prediction of binding free energy changes induced by mutation, has not been fully explored. To tackle the challenge, we propose a few-shot learning approach that integrates AlphaFold2 subsampling, ensemble docking, and Siamese learning to predict the relative binding affinity changes induced by mutations. We validate our approach for predicting binding affinity changes in 31 variants of the mutated Abelson tyrosine kinase (ABL). The conformation ensembles for each ABL variant are generated using AlphaFold2 subsampling. The most likely conformational states are selected by Gaussian fitting to construct the pairwise structural inputs and compute relative binding energies. Then the relative binding affinities are derived by averaging the values predicted by the Siamese learning model across the constructed conformational pairs. By benchmarking on the Tyrosine kinase inhibitors (TKI) dataset and the direct binding affinity using the refined set of PDBbind database, our proposed approach achieves an increased Spearman correlation coefficient for five out of six TKI molecules across 31 mutants when estimating relative binding affinities. Although the proposed method is validated on the ABL, it can be transferred and applied to other drug-target interaction predictions when the conformational flexibility of proteins needs to be considered. We anticipate the approach will be valuable for predicting binding energy changes induced by mutations in proteins or ligands,

facilitating drug discovery efforts.

Keywords: deep learning; relative binding affinity; protein mutation; Siamese learning

Introduction

Proteins are responsible for life-essential specific functions such as cellular structural support, immune protection, enzyme catalysis, cell signaling to transcription, and translation functions [1-5]. These functions are closely related to structural changes in the proteins. The experimentally determined three-dimensional structures remain limited due to the complexity and expense of techniques like X-ray crystallography, nuclear magnetic resonance, and electron cryo-microscopy (cryo-EM). Consequently, prediction of protein structures through computational methods has become crucial [6-8]. Deep learning techniques, notably AlphaFold2 (AF2), have revolutionized the field of protein structure prediction. AF2 provides prediction accuracy close to experimentally determined methods [9]. Following the success of AF2, many predictors have been generated such as RoseTTAFold [10], ColabFold, OpenFold [11]; ESMfold [12], RGN2 [13], and OmegaFold [14]. The success of AF2 brings new possibilities for many downstream explorations, such as analyzing missense variants in genome sequences [15], predicting protein-ligand molecule interaction in drug discovery [16,17], protein engineering in synthetic biology [18,19], and metallic pathway analysis for biological systems [20].

While it effectively predicts static protein structures, the default AF2 has limited capability in capturing the dynamic behavior and conformational changes of proteins. Proteins are not static and undergo conformational changes upon binding of a specific ligand, whereby residues in the pocket adjust their conformation to optimize their interaction with the ligand [21,22]. Significant efforts have been dedicated to refining AF2 to capture different conformations. Recent advancements in protein design have facilitated the detailed characterization of protein dynamics, enabling the identification of both ground and excited states [23]. The AF2 model constructs protein 3D structures by using Multiple Sequence Alignment (MSA) as inputs [24]. Since MSA encodes coevolutionary information, manipulating MSA offers a viable way to alter the conformations by controlling coevolutionary information. Several successful methods have been reported for predicting conformational ensembles of proteins, such as AF-Cluster via sequence clustering [25], subsampling MSA [26], energetic frustration analysis of MSA [27], SPEECH_AF via in silico mutagenesis of MSA [28], randomized alanine sequence scanning [29], etc. The recent successful cases have applied AF2 to directly predict the relative conformation populations by subsampling the comparison of multiple sequences [26,30]. Additionally, analysis of MSA has also been used to identify patterns of relevant mutations within protein families [31,32], predict interacting residues based on sequences [33,34], and construct structure prediction models guided by energy-based processes [35,36].

After predicting the structures of proteins, binding affinity can be predicted for assessing the strength and stability of small molecules binding to protein targets. With advances in databases and computational power, predictions can be made using physics-based methods and deep learning methods [37-40]. However, the binding sites of proteins in different conformational states can exhibit significant variations, which may affect the binding mode. Consequently, the challenge has changed from merely determining whether a protein structure is available to appropriately selecting an optimal structure for accurate modeling. To avoid bias toward a

single protein conformation and to implicitly account for protein flexibility, an ensemble of representative structures is used to dock candidate ligands [41]. Incorporating protein flexibility is essential for accurate binding affinity prediction, as exemplified by studies on Abelson tyrosine kinase (ABL) [42]. Although Tyrosine kinase inhibitors (TKIs) like imatinib have revolutionized the treatment of chronic myeloid leukemia (CML) [43], drug resistance emerges from mutations that alter the conformation, thereby affecting the drug binding mode and binding affinity [44]. The limited number of crystal structures of kinase mutants poses a limitation to the accuracy of both physics-based methods and machine/deep learning methods [45].

Evaluating the changes in binding affinity caused by mutations is crucial for understanding essential biological processes. Despite considerable endeavors, accurately predicting the impact of mutations on binding affinity remains challenging [46-48]. Mutation-induced structural changes bring two major challenges for the traditional computational methods: determining the precise structural changes of proteins, and selecting the most representative structure from conformation ensembles. A real-world scenario in drug development involves understanding the resistance profiles of certain drugs against various mutated structures [49]. For example, mutations in ABL alter the relative population of the active state and inactivate state, which can be selectively targeted by different types of inhibitors [50]. Several states of ABL are structurally similar to known crystal structures capable of binding a variety of inhibitors [51]. Consequently, the conformation ensemble of kinase proteins should be incorporated into the prediction of binding free energy changes induced by mutations. The accuracy of AF2 extends to the resolution of individual amino acids, enabling precise predictions of protein structures at an atomic level [15,52]. AF2 has learned an approximate biophysical energy function that can accurately assess the quality of mutated protein structures [53]. Furthermore, subsampled AF2 is capable of predicting the conformational landscape and the relative populations of conformations at single mutation resolution [26]. In this study, we propose an integrated approach to address the aforementioned limitations by streamlining AF2 subsampling, ensemble docking, and Siamese learning. To incorporate structural diversity, we employ data augmentation by pairing the reference structures with those generated structures from subsampled AF2. We benchmark our proposal by investigating drug resistance in ABL kinase. To efficiently process the structural features, we propose an STGNet (Structure and graph-aware Siamese network) model in the Siamese architecture because Siamese network has been used to predict relative binding affinity for the static structures [54]. In few-shot learning, our approach demonstrates higher ranking accuracy in predicting binding free energy changes induced by mutations.

Results

Construction of augmented pairwise dataset and deep learning architecture

Predicting relative properties has significant applications in drug discovery, for example, changes in absorption, distribution, metabolism, excretion, and toxicity (ADMET) during drug lead optimization [55,56], as well as in assessing relative binding affinities between congeneric ligands [57]. Various models have been proposed for this purpose as illustrated in **Fig. 1**. The

relative properties can be computed using a standard neural network, the delta model [56], or the Siamese model [54]. The Siamese architecture is designed to compare two inputs by generating embeddings, thus capturing subtle distinctions between similar inputs [54]. The pairwise Siamese network further expands on this concept by integrating pairs of representations and calculating average values across various inputs [55].

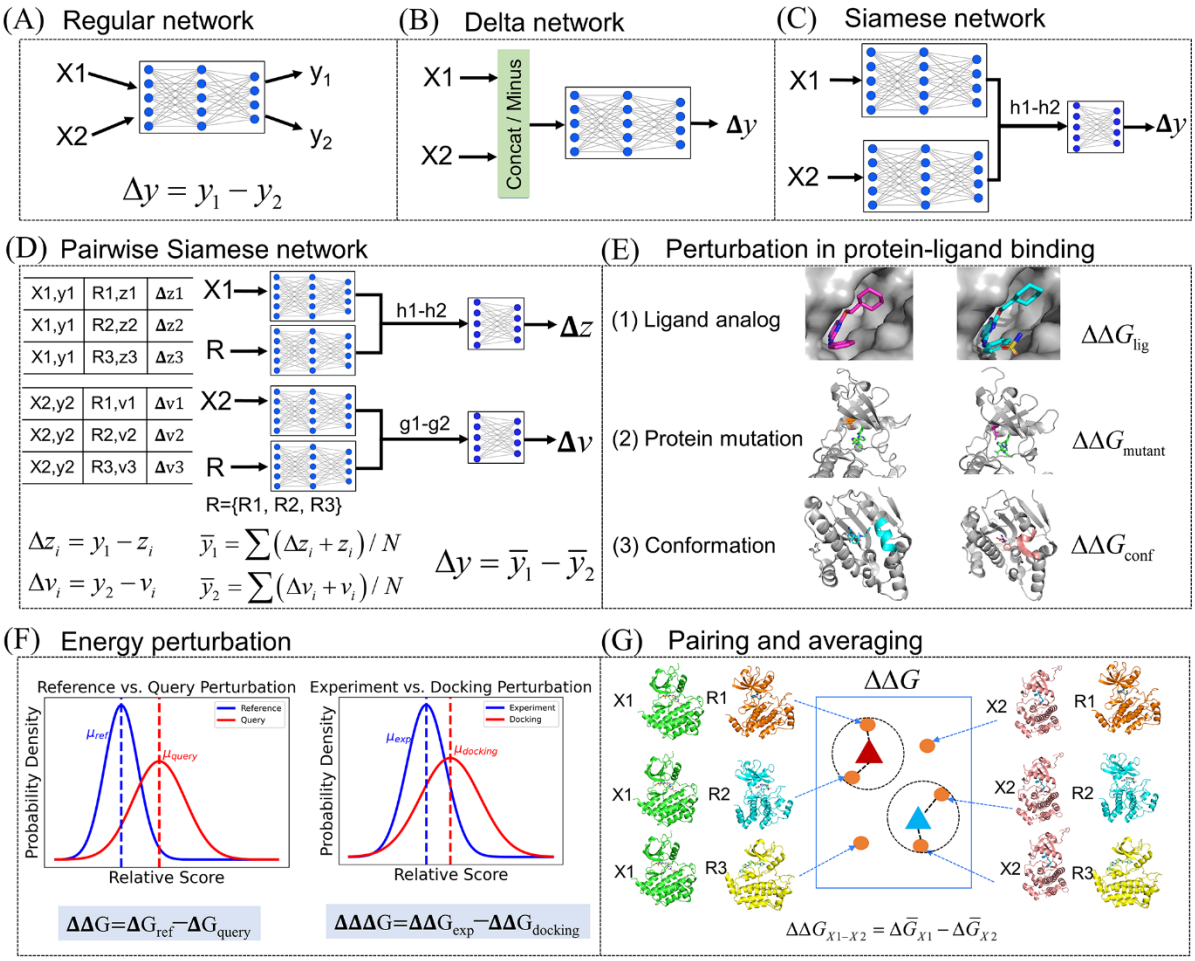

Fig. 1 Models for relative property prediction. (A) Regular neural network. (B) Delta neural network. (C) Siamese network. (D) Pairwise Siamese network. (E) Perturbation in protein-ligand binding includes variations of ligand analogs, protein mutation, and conformations. (F) Two types of energy perturbation. (G) Pairing and averaging with the most populated states when estimating the relative binding free energy. X1 and X2 represent the input, and y_1 and y_2 are the output. R1, R2, and R3 represent the reference states.

In addition to deep learning methods, free energy perturbation (FEP) is a widely employed physics-based method for predicting relative binding affinity [58], which quantifies energy changes between a reference and a target state. In our study, we broaden the scope of perturbation analysis by transitioning from absolute value comparisons to evaluating relative values, as illustrated in **Fig. 1F**. Specifically, we focus on analyzing the relative changes in binding energy induced by mutations. To achieve this, we leverage deep learning techniques to establish a mapping between relative scores and the corresponding relative binding energies associated with mutations. To enhance the reliability of our predictions, we also implement a

pairing and averaging process as shown in **Fig. 1G**. This method will be useful for computing relative binding affinities in protein-ligand complexes that are caused by ligand structure transformation, protein mutations, and conformational variations during interactions.

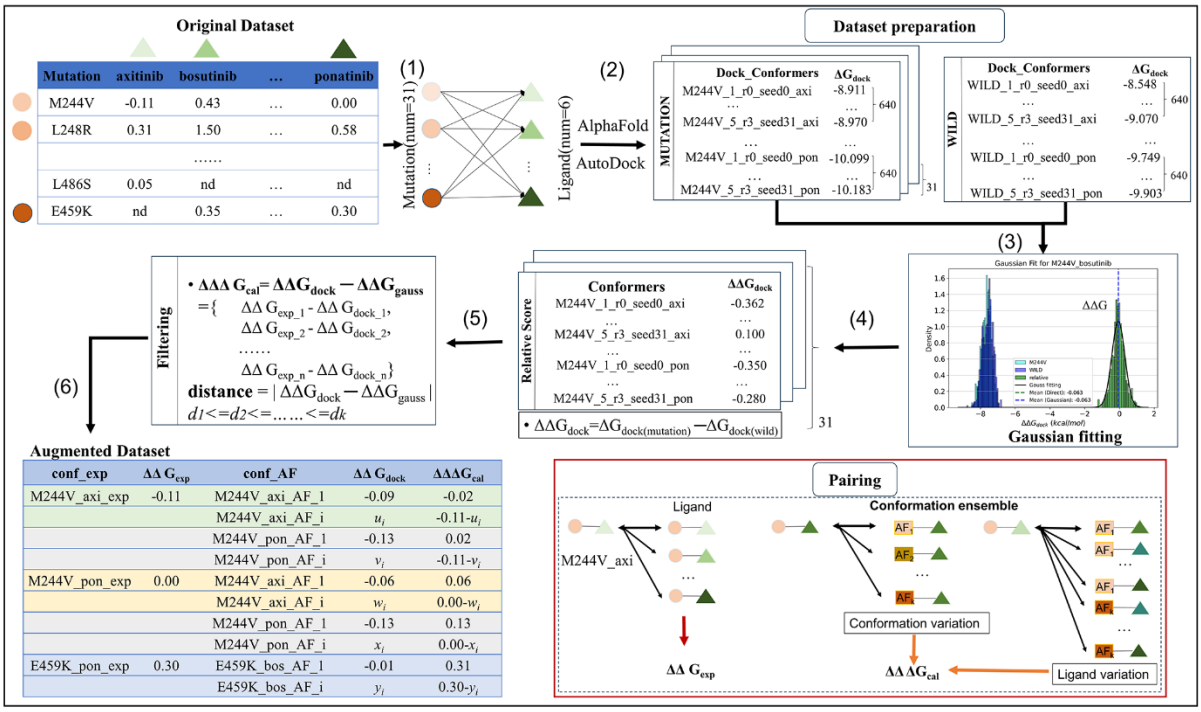

Fig. 2 Pipeline of construction of the augmented pairwise dataset. The workflow includes the creation of initial mutant-ligand data pairs, conformational sampling, ensemble molecular docking, relative binding energy calculation, data filtering, and construction of the final dataset. The red bounding box highlights the pairing between the reference state and the query states, encompassing the conformation variation and ligand variation.

The augmented dataset is constructed by using AF2 subsampling and ensemble docking (refer to the Method section for details in supporting information). In the workflow illustrated in **Fig. 2**, we selected 31 mutants of ABL kinase and the 6 TKI molecules to construct our dataset. The objective is to construct reference-generated conformation pairs for the specified mutants. The reference values are derived from the TKI dataset [59]. The generated values are predicted from subsampling AF2 for 31 mutants and wild-type of ABL kinase. The relative docking score $\Delta\Delta G_{\text{dock}}$ is computed by comparing the docking scores of mutants to those of the wild type.

From the distributions of relative docking scores, we identified the most probable conformations. The selected relative docking scores are then paired with the experimental values. The relative binding affinity for each experimental value will be augmented to a specified number of data points. Then, we can obtain the relative binding free energy difference (referred to as $\Delta\Delta\Delta G_{\text{pred}}$) by subtracting the reference value from the relative docking score.

By specifying the mutation type, this approach emphasizes differences in conformation and variations in bound ligands. The key advantage of using the relative binding affinity difference

is that it allows Siamese learning to focus on the difference between the reference state and the generated conformational states. Experimental-level binding energy accuracy can be achieved through the equation $\Delta\Delta G_{\text{exp}} = \Delta\Delta G_{\text{dock}} + \Delta\Delta\Delta G_{\text{pred}}$. Other computational details can be found in the Methods section.

Data augmentation has been demonstrated to enhance the prediction of binding affinities, offering improved accuracy and robustness [39]. The impact of multi-conformation data on the predictive ability of the model is evaluated by systematically varying the number of conformations, specifically using *k* values of 1, 5, 50, and 100, respectively. This methodology expands the dataset through *k*-fold augmentation by pairing the experimental data with the generated data, incorporating structural diversity for each protein-ligand pair via conformation variation and ligand variation. After obtaining the augmented dataset, the dataset is partitioned into training and test sets according to the names of different TKIs for cross-validation. This splitting method reflects real-world scenarios where we have established knowledge of protein mutations and their associated changes in binding free energy. By leveraging this prior knowledge, we can predict the relative binding free energy of a new drug molecule across all mutations. Further details are described in the Methods section and Fig. S1 in supporting information. All molecules exhibit low similarity as shown in Fig. S2, which ensures that pairing with the experimental values does not result in information leakage. Similar pairwise approaches have been employed to predict relative binding free energies [54,57,60]. Our approach differs from existing work in two key respects. First, previous research, such as pairwise binding comparison network (PBCNet), has focused on predicting changes in binding energy due to variations in ligands. In contrast, our method extends PBCNet by specifically addressing the prediction of binding energy changes caused by protein mutations. Secondly, we introduce the concept of using the difference in relative binding free energy as the prediction target, rather than the relative binding free energies themselves. This modification enables the Siamese learning model to more accurately capture the differences among various conformation states.

To harness the structural information, the input for our Siamese learning includes contact maps of each conformation, the molecular graph of ligand, and the molecular graph of protein-ligand interaction as shown in **Fig. 3**. The molecular graph has been validated in our previous work [61]. The protein contact map with BLOSUM62 matrix scores [62] can reflect the spatial contacts between CA atoms of each residue in the three-dimensional (3D) structure of a protein. The ligand graph contains the chemical properties within the molecule. The protein-ligand interaction graph considers covalent interaction between intramolecular and non-covalent atomic-level interaction information of intermolecular. To learn the structural features, we constructed the STGNet (Structure and graph-aware Siamese network) model to learn the structural information using the Message Passing Neural Networks (MPNN) network [63], and to learn interaction information using the Geometric Interaction Graph Neural Network (GINN) network [64]. The protein contact map and ligand graph are learned using MPNN networks in the Siamese architecture, which helps to capture the 3D folding structures of the protein as well as the internal chemical properties of the ligand. GIGN can learn the binding patterns between proteins and ligands, by combining the properties of the nodes and the 3D coordinate

information. By using MPNN and GIGN together, the model can focus on intramolecular and intermolecular features, respectively. The experimental values and the paired AF2-generated data will be fed into the shared Siamese Network. Therefore, the network can better learn the distinctions between the reference state and the alternative conformational states, thereby enhancing the prediction of relative binding affinity differences.

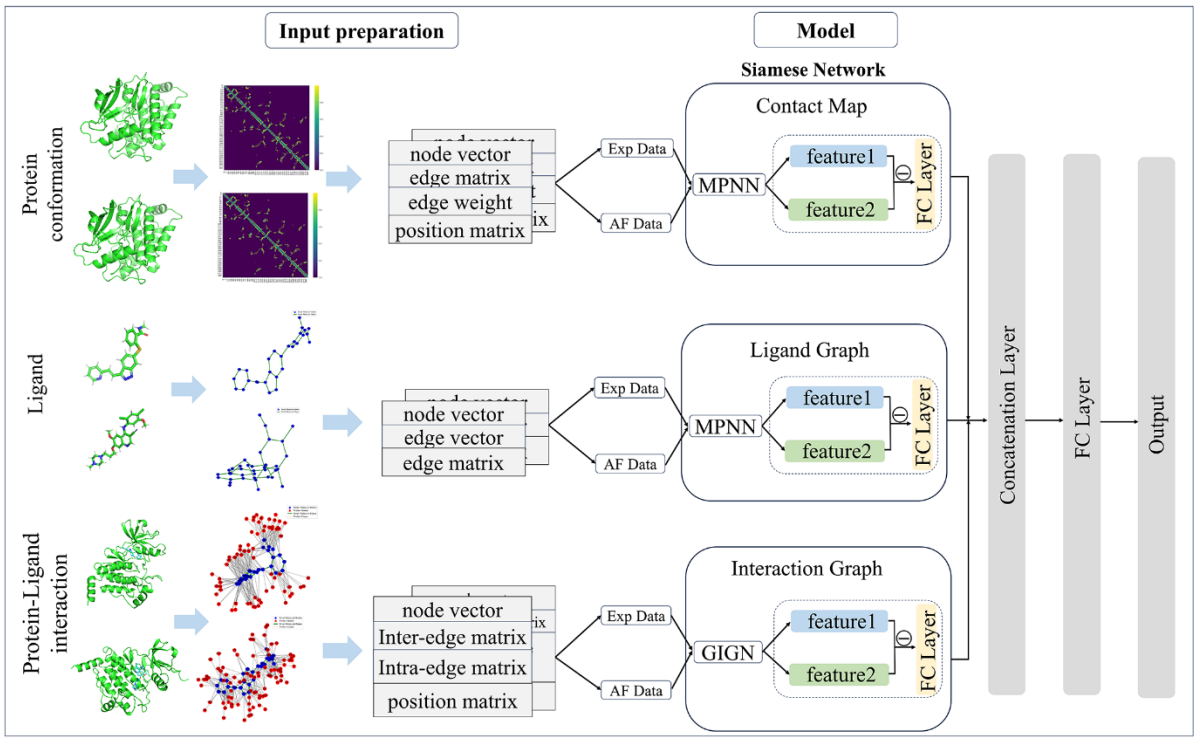

Fig. 3 Architecture of the STGNet model. The reference state and the AF2-generated state are processed by the Siamese network to capture the feature difference. Protein contact maps and ligand graphs are processed by MPNN, and protein-ligand interaction graphs are processed by GIGN. Ligands show as blue nodes and edges, and proteins show as red nodes and edges in molecular graphs.

Conformation sampling for ABL kinase using subsampled AF2

Conformational changes in ABL kinase have important implications for its function and drug resistance [65]. Mutations induce obvious structural changes in the P-loop and A-loop region, thereby modulating the binding affinity of the drug to ABL kinase by interfering with the stability of specific conformations (e.g., DFG-in and DFG-out) [45,66]. Structural analyses comparing type I (DFG-in) and type II (DFG-out) inhibitors reveal that these two classes of drugs exhibit different sensitivities to conformational changes in ABL kinase. **Fig. 4** demonstrates three conformational states of ABL kinase, including the active state, inactive state I_1, and inactive state I_2. In the active state, P-Loop and A-Loop are in the open state and αC-Helix is in an inward position. Inactive state I_2 shows obvious structural changes with active state, with the αC-Helix shifting from the inward position to the outward position, the P-Loop becoming closed, and the A-Loop partially folding. The inactive state I_1 displays relatively minor conformational changes compared to the active state. Mutations may indirectly reduce

the availability of drug-binding sites by stabilizing a certain conformation state and altering the percentage of the active state of the kinase [67].

The previous study also presents a conformational dynamic equilibrium between the active and two inactive states (I_1 and I_2) of 8 mutants of ABL kinase achieved by subsampling MSA[26]. We adopted the same procedure of subsampling MSA and obtained the conformation ensemble for all 31 mutants. The deviation in conformation distributions of the mutants relative to the crystal structure is quantified by computing root mean square deviation (RMSD). **Fig. 4B** depicts the RMSD distribution densities of the loop regions of 31 mutants and the wild-type (WILD). The primary peaks of the RMSD densities for the wild-type are concentrated at RMSD values less than 3.5 Å, which is consistent with previous study [26]. For wild type, the percentage of active state is 83%, which is close to the experimental value of 88%. The different distributions suggest that the mutation alters the conformational balance of the ABL kinase. The relative population of other mutants is summarized in Table S1.

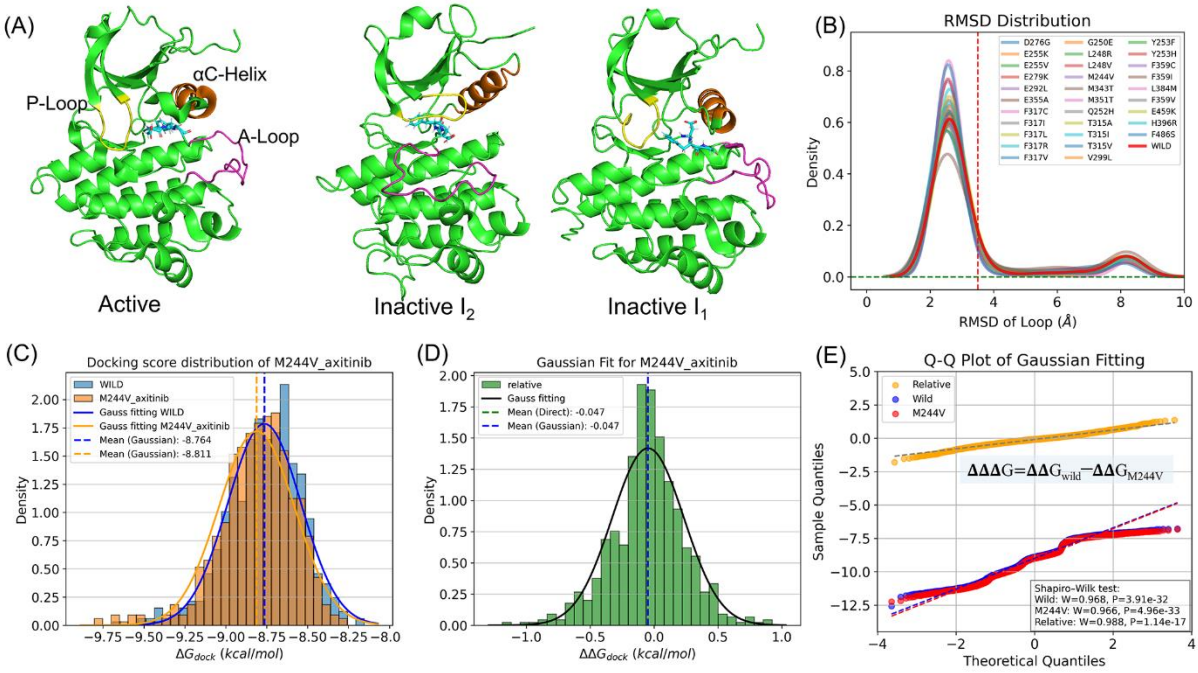

Fig. 4 Mutation effect on the conformation of ABL kinase and the drug binding affinity. (A) ABL kinase conformation in active state (PDB ID: 6XR6) and two inactive conformational states I_1 and I_2 (PDB ID:6XR7, 6XRG). (B) RMSD distribution of the loop regions (A-loop, P-loop, and C-helix) for wild type and 31 mutants. (C) Distribution of AutoDock Vina docking score of the wild and mutant (M244V) of ABL kinase. (D) Density distribution plot based on the relative binding energy obtained from the mutant (M244V) and the wild type. (E) The Quantile-Quantile Plot for the docking score and the relative docking score for the mutant (M244V) and the wild type. More information can be found in Fig. S3-S8.

The distribution of docking score for mutant-ligand (M244V_axitinib) and the wild-type (WILD_axitinib) are shown in **Fig. 4C-D**. The Gaussian fitting indicates the binding free energy distribution follows the principle of maximum likelihood. **Fig. 4E** confirms the Gaussian distribution of relative binding free energy distributions $\Delta\Delta G_{\text{dock}}$ of the M244V

mutant and the wild-type of ABL kinase, as validated by the Sharipo-Wilk test. Similar Gaussian distributions are observed for other muations as shown in Figs. S3-S8. The docking distribution can be interpreted as a perturbation of the experimental distribution. Consequently, our objective is to determine the relative energy changes induced by mutation. Based on the principle of maximum likelihood, we select the *k* conformations and their relative binding affinities, ensuring that the deviation of these affinities from the Gaussian mean value is minimized. The proposed filtering approach provides a robust statistical framework for identifying the most populated states, rather than simply prioritizing the highest-scoring conformation.

Improved predictive performance over physics-based methods (docking and FEP)

Prediction of relative binding free energy delivers import value in drug discovery applications. However, accurate prediction is limited due to technical and practical challenges in physics-based modeling and statistical mechanics methods [68]. The molecular docking and free energy perturbation (FEP+) were compared to our model. The predicted relative binding affinity and the experimental value are shown in **Fig. 5**. Molecular docking can provide a quick ranking, but it fails to reproduce the relative binding free energy between them. FEP+ can reproduce the relative binding affinity at the chemical accuracy, but it requires high computational resources. For our method, we provide the relative binding affinity for all experiment-AF2 pairs, then we compute the statistical average for the required protein-ligand pair, which enhances the stability of the prediction.

The performance of the models was evaluated by calculating RMSE, MAE, Pearson, and Spearman coefficients, which are summarized in **Fig. 6** and Table S2-S5. It can be seen that molecular docking has limited performance for protein-ligand relative binding free energy calculation, as indicated by the low Pearson and Spearman coefficients. The FEP+ results reported by Hauser et al [59], are markedly superior to those obtained from molecular docking, as evidenced by higher Pearson and Spearman coefficients for the axitinib, dasatinib, imatinib, and ponatinib significantly higher than those of molecular docking. The STGNet, employing 50-fold data augmentation (denoted as top50), demonstrates enhanced performance in terms of RMSE, MAE, and Pearson coefficients, obtaining the best values for four types of TKIs: bosutinib, imatinib, nilotinib, and ponatinib. For the computational accuracy, our model cannot competed with The FEP+ results. For the ranking ability, our method outperforms both molecular docking and FEP+ methods, as evidenced by the highest Spearman correlation coefficients obtained for five out of six TKIs: axitinib, bosutinib, dasatinib, imatinib, and nilotinib, highlighting its enhanced ranking performance. By introducing a multi-conformational free energy distribution filtering strategy and applying Siamese learning, our method significantly improves the accuracy in predicting the relative binding free energy changes induced by mutation.

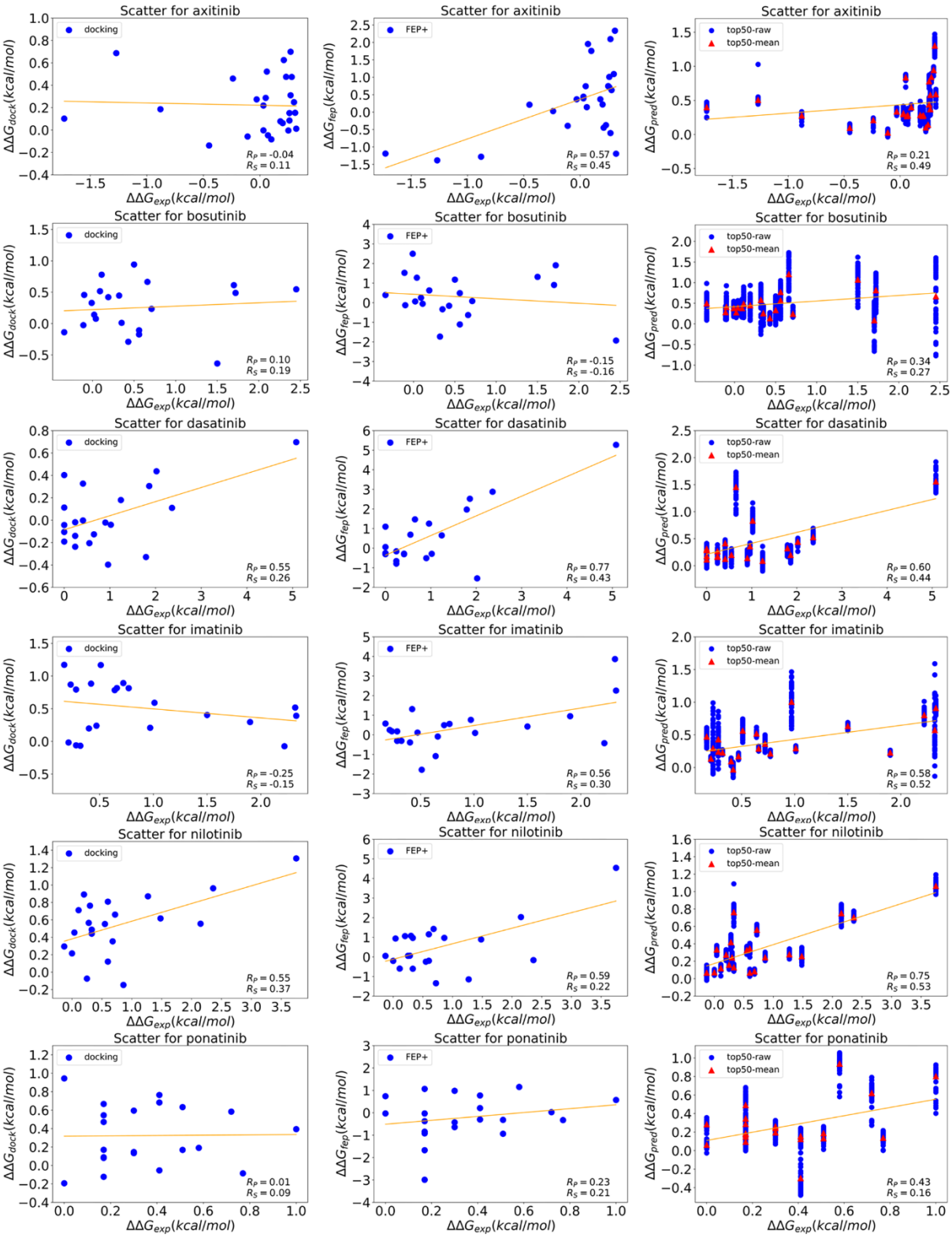

Fig. 5 Comparison of experimental and predicted relative binding energies for six TKI molecules. The predictions were generated using molecular docking, FEP+, and STGNet (denoted as top50-raw for the raw prediction and top50-mean for the average over the 50 data points derived from pairs sharing the same reference state). Each blue sphere represents one prediction and the red triangle represents the average value.

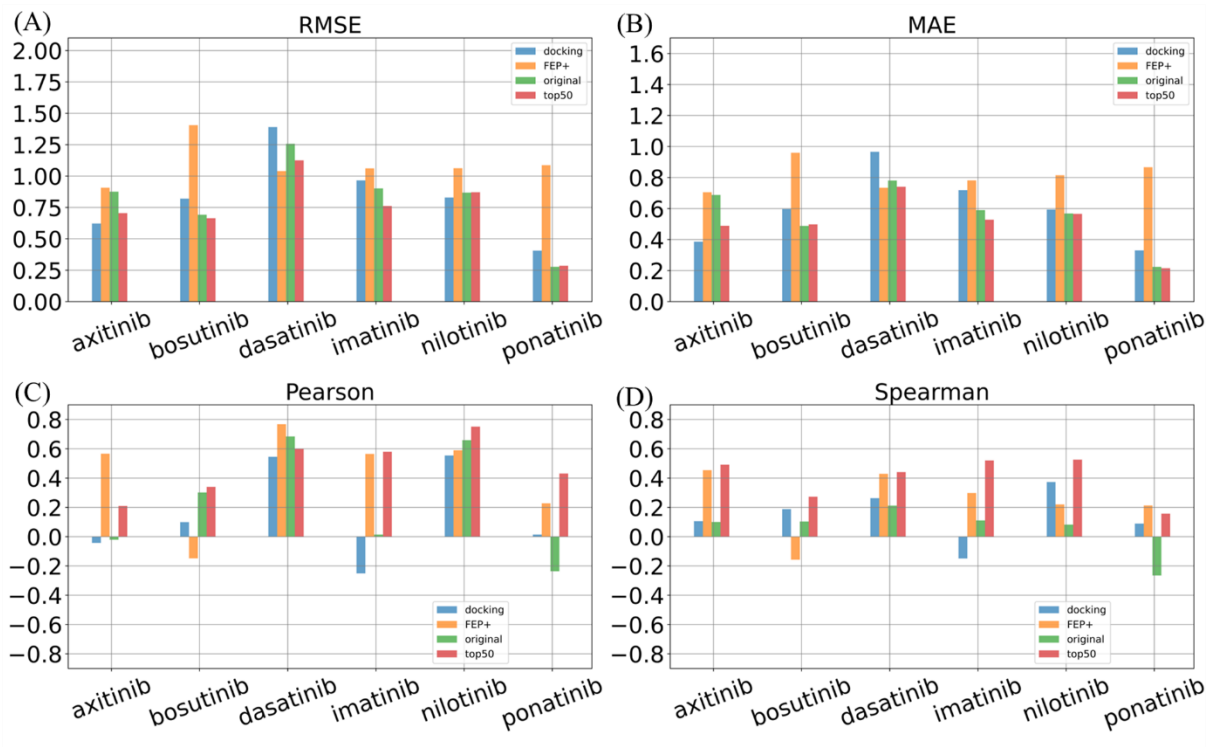

Fig. 6 Evaluation metrics of four different methods. Performance is evaluated by RMSE, MAE, Pearson coefficients, and Spearman coefficients for docking, FEP+, STGNet using the original TKI dataset (denoted as original, which represents that no conformational flexibity is included), and STGNet using the top 50 samples (denoted as top50).

Relative value prediction outperforms absolute value prediction

We compare the data augmentation method with direct absolute prediction of binding energy using the refined set of PDBbind after excluding the TKI-related data. The model for the regular neural network, illustrated in Fig. S9, was trained using a comparable network architecture except for omitting the Siamese network and the pairwise input. The model predicts the absolute value of the binding free energy. As shown in **Fig. 7A**, the regular model achieves satisfactory performance in predicting absolute binding affinities, with strong correlations (Pearson and Spearman coefficients are 0.90, and 0.90 for the hold-out set, and for the CASF-2016 set). However, the model exhibits limitations in predicting both absolute and relative binding energies for ABL mutants, as illustrated in **Fig. 7B and 7C**. Specifically, it fails to accurately capture the effects of single mutations in proteins. The Spearman coefficients in **Fig. 7D** support this observation, indicating the regular model's inability to reliably predict relative binding energy changes induced by single mutation in proteins.

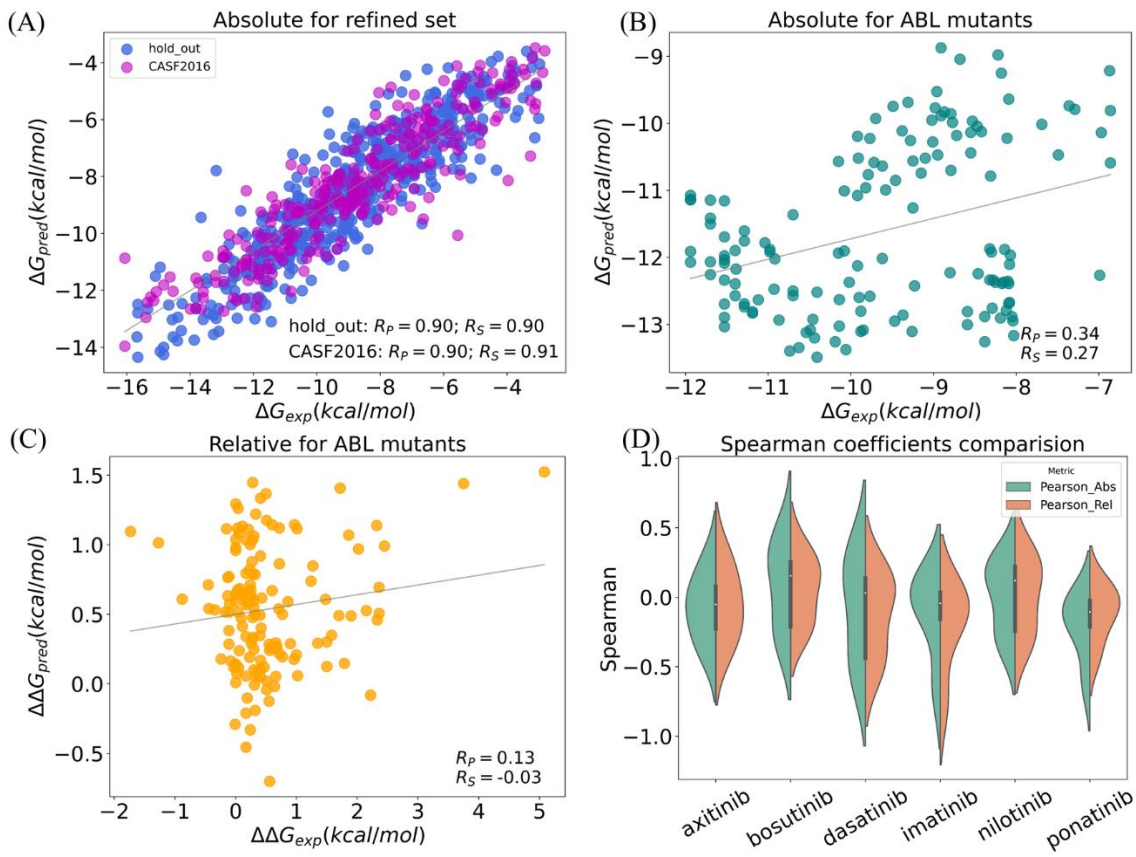

Fig. 7 Performance of regular network in the prediction of absolute binding energy. (A) Correlation between the absolute values of the prediction and experimental values in the CASF-2016 set. (B) Correlation between the absolute values of prediction and experimental values in the TKI set. (C) Correlation between the relative values of prediction and experimental values in the TKI set. (D) Spearman correlation coefficients for both absolute values and the relative values of predictions compared to experimental values in five-fold validation.

We compare the result of the regular model with our STGNet model. As can be seen in **Fig. 8**, the RMSE and MAE of STGNet using the different numbers of samples (5, 50, and 100 are denoted as top5, top50, and top100) were lower than those of the refined dataset for five out of six molecules, including axitinib, bosutinib, dasatinib, nilotinib, and ponatinib. STGNet presents higher Spearman coefficients than that of the refined set. Combining experimental and AF2 data, the STGNet method is better able to capture the dynamic changes of proteins in different conformations and the effects of mutations on these changes. By combining the results from **Fig. 7** and **Fig. 8**, we demonstrate that our approach significantly enhances the ability of the STGNet model to recognize and interpret structural and energetic perturbations induced by single amino acid substitutions. Overall, our method enhances mutation effect prediction by integrating multiple conformations, applying a rational filtering strategy, and Siamese learning with the STGNet model.

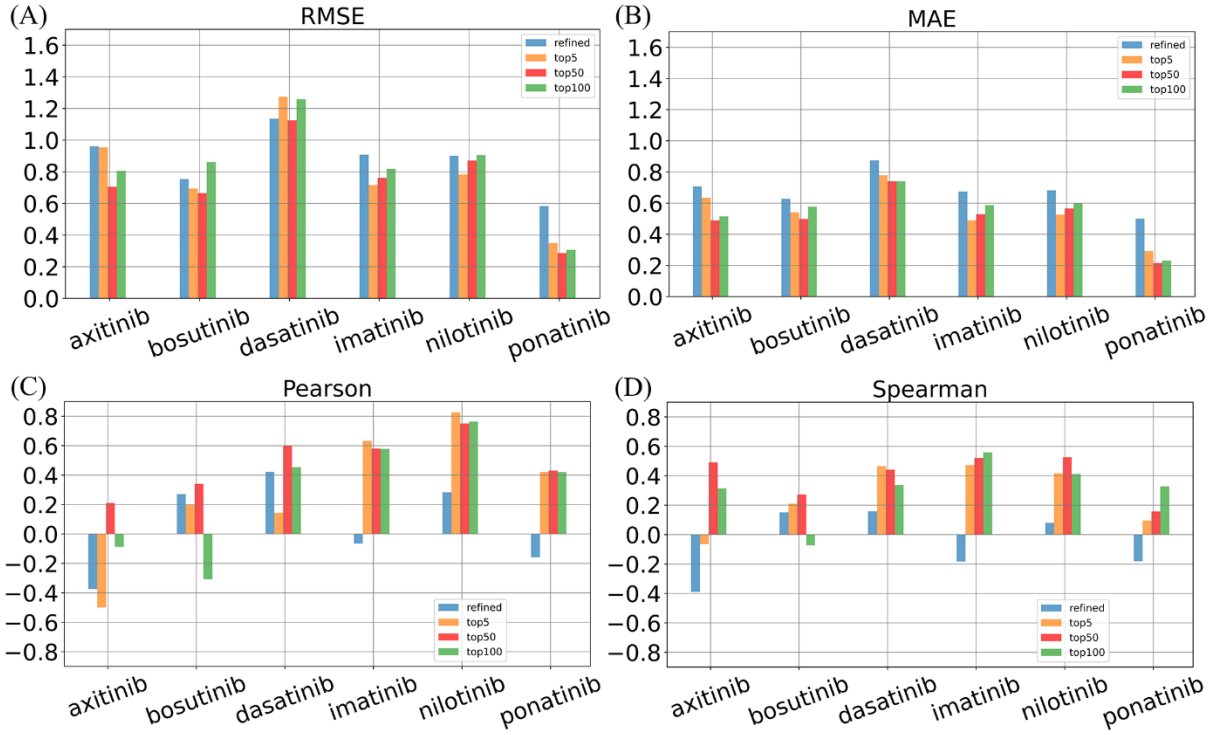

Fig. 8 Evaluation metrics for using different number of generated conformational samples. Performance is evaluated using RMSE, MAE, Pearson coefficients, and Spearman coefficients for a refined set of PDBbind, and our method is evaluated by using 5, 50, and 100 samples.

Data augmentation improves prediction accuracy

We assess the impact of sample size on prediction performance by expanding the dataset to 5, 50, and 100 times the original dataset. As can be seen from **Fig. 8**, the prediction based on the original dataset results in higher errors and lower correlations. The original dataset fails to reflect the impact of mutations and conformation on free energy changes. The best performance is achieved when the number of data is expanded by 50-fold (denoted as top50) on the original dataset. The top50 dataset reduced the RMSE and MAE by the data enhancement strategy, especially in the case of bosutinib, dasatinib, and ponatinib. In contrast, top5 and top100 performed slightly lower than top50. This indicates that the proper choice in the number of selected conformations can further improve the accuracy of the model prediction. From Pearson and Spearman coefficients, both top5 and top50 showed higher correlations on these datasets, especially top50 showed the best correlation on almost all drugs. The top50 also shows less variance over other datasets as shown in Fig. S10. The top50 may effectively achieve the balance between capturing the structural diversity and minimizing noise, while the top100 dataset may include excessive conformational changes, which also introduces variability in prediction. Due to the diverse conformational distribution among TKIs, both excessive and insufficient data can hinder improvements in model performance. The result also indicates that data augmentation strategies play an important role in improving model performance. Optimal performance will be achieved when conformational diversity is adequately represented without introducing redundancy or noise. Further optimizing the ratio of the number of samples can be

conducted in the future.

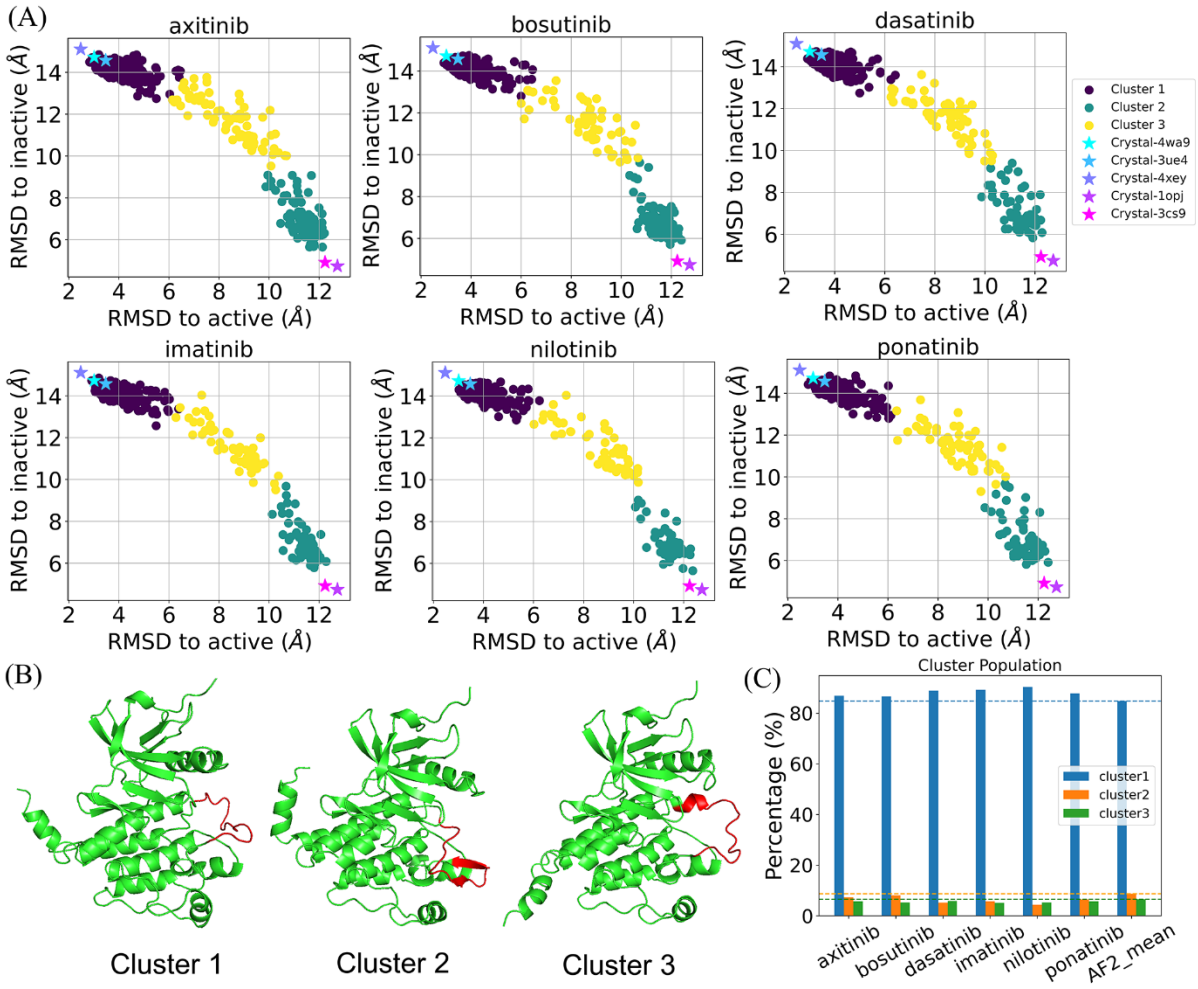

Fig. 9 RMSD clustering of top50 dataset and the representative conformations. (A) RMSD distributions of conformations docked for six TKIs. (B) Representative conformations in each cluster center. (C) Percentage changes in each cluster after filtering. Crystal structures are labeled as the stars. The A-Loop region is highlighted in red in the structures.

Conformations distribution indicates rational choice of conformations

The RMSD distribution characterizes the relative distribution of proteins between the active and inactive states after mutation. We characterize the filtered conformations of mutants in our model by clustering conformations based on their RMSD from the reference crystal structures. The clustering result in **Fig. 9A** reflects the preference of TKIs for kinase conformations. Cluster 1 is near the active state region, cluster 2 is near the inactive state region, and cluster 3 is in the middle transition region, which is validated by the extracted structures of each cluster center. Similar binding modes are observed for six TKI molecules as shown in **Fig. 9B** and Fig. S11. We also observe the changes in conformational distributions in the three clusters as shown in **Fig. 9C**. The selection process guided by the most probable distribution during data augmentation partially addresses the question of how to appropriately select a proper structure from ensembles. We demonstrate the application of subsampled AF2 in the downstream

prediction of drug resistance. This approach offers new insights into studying the dynamic properties of protein-ligand interactions in drug discovery.

Conclusion

We present a method for predicting relative free energy in protein mutations by generating the relative distribution of conformations through AF2 subsampling. The data enhancement approach significantly improves the prediction performance of mutation effects by integrating the structural diversity of conformational ensembles, implementing a filtering strategy, and Siamese learning. Benchmarked using the TKI dataset and a refined set of PDBbind, the results demonstrate that the dataset augmented by 50-fold (top50), generated by using AF2 subsampling and Gaussian fitting, not only increases the structural diversity of mutation-ligand pairs but also effectively captures potential interaction patterns between different mutations and ligands. The STGNet model proposed in this study successfully captures the key features of protein-ligand interactions, thereby significantly improving the accuracy and stability of mutation effect predictions. We extend the application of integrating subsampled AF2 with Siamese learning to predict relative binding affinity changes, which holds promise for investigating drug resistance. This approach provides a promising tool for predicting the relative binding energy of mutation-ligand interactions without relying on the crystal structures of protein mutants.

Methods

Mathematical formulation

The energy difference between two states can be analyzed using perturbation theory. The perturbation depends on the selection of the reference state. In cases where the two states differ in the bound ligands, the perturbation corresponds to the energy difference in binding affinity induced by variations in the ligand. If two states differ in the conformational states of the protein, the perturbation reflects the energy difference in binding affinity resulting from conformational changes, which can be expressed as follows.

$$\Delta G_{i,j} = \Delta G_{\text{ref}} + \Delta\Delta G_{\text{conf}} \tag{1}$$

Transitioning from absolute binding energy $\Delta G_{i,j}$ to relative binding energy $\Delta\Delta G_{i,j}$, we can define experimental value as the reference state, and the perturbed state is represented by the relative binding energy obtained from molecular docking. Thus, in our case, we can interpret the distribution of relative docking scores as the perturbations applied to the experimental distribution. The energy difference is computed from the reference states plus the energy variation as follows.

$$\Delta\Delta G_{i,j} = \Delta\Delta G_{\text{ref}} + \Delta\Delta\Delta G_{\text{pertb}} \tag{2}$$

Therefore, the purpose of employing deep learning is to compute the differences in relative binding energy $\Delta\Delta\Delta G_{\text{pertb}}$. After generating the conformation ensemble, we conduct the molecular docking. We obtain the relative docking score difference between mutants and the wild-type protein.

$$\Delta\Delta G_{\text{dock}} = \Delta G_{\text{dock(mutation)}} - \Delta G_{\text{dock(wild)}} \tag{3}$$

The relative docking score is used to select the most probable distribution based on the deviation from the Gaussian fitting mean value. The selected docking score and their corresponding conformation will then be paired with the experimental values. In the current study, we select the experimental value as the reference state. The difference in relative binding energy is computed from the following equation:

$$\Delta\Delta\Delta G_{\text{pred}} = \Delta\Delta G_{\text{exp}} - \Delta\Delta G_{\text{dock}} \tag{4}$$

Here, $\Delta\Delta G_{\text{exp}}$ is the energy of reference state, $\Delta\Delta\Delta G_{\text{pred}}$ represents the energy difference caused by the conformation change and ligand variety. We aim to include conformational diversity in the prediction. Siamese networks are particularly well-suited for learning the differences between the two states, so we constructed a relative difference based on the experimental values and generated values, which serves as the regression target for our network learning.

The molecular representations and the difference binding energies are input into STGNet. After training the model, we predicted the binding energy difference $\Delta\Delta\Delta G_{\text{pred}}$ for the test dataset. Siamese network was employed to learn the energy differences resulting from both conformation changes and mutations in protein sequences. The predicted energy differences $\Delta\Delta\Delta G_{\text{pred}}$ are recovered to the predicted relative binding energy $\Delta\Delta G_{\text{pred}}$ from the following equation:

$$\Delta\Delta G_{\text{pred}} = \Delta\Delta\Delta G_{\text{pred}} + \Delta\Delta G_{\text{dock}} \tag{5}$$

Each experimental value has been paired with N-generated samples, thereby we obtained N predicted relative binding energy. The final predicted relative binding energy is the average of the predicted relative binding energy:

$$\overline{\Delta\Delta G_{\text{pred}}} = \frac{\Delta\Delta G_{\text{pred}}}{N} \tag{6}$$

From the above workflow, we introduce the conformation ensemble in the prediction of relative binding energy. The relative binding energy can be recalibrated to the experimental level $\overline{\Delta\Delta G_{\text{pred}}}$ by computing the statistical average of STGNet predictions for the specified ligand molecule.

Workflow for constructing the augmented dataset

As shown in **Fig. 2**, the workflow outlines a systematic approach to enhancing drug docking studies through the integration of reference data, conformational generation, and statistical analysis. By utilizing AlphaFold2 for conformational modeling and implementing ensemble docking procedures, we obtain relative docking scores that are subsequently analyzed using Gaussian fitting. The workflow emphasizes ranking these scores, filtering conformations based on probability proximity to the average, and ultimately constructing an augmented dataset.

Step 1: Construct reference data. Compile the original dataset, which includes 31 types of mutations and 6 associated compounds. Prepare mutation-ligand pairs and the corresponding relative binding energies.

Step 2: Generate conformations and conduct the docking. Utilize subsampled AlphaFold2 to generate protein conformations for wild and 31 types of ABL mutants. We generate 640 conformations for each type of mutation. Conduct molecular docking simulations on these generated conformations to assess binding affinities.

Step 3: Obtain docking scores and perform Gaussian fitting. Calculate the docking scores from the docking using AutoDock Vina. Check Gaussian fitting on the distribution of these docking scores to identify statistical averages.

Step 4: Prepare relative docking scores. Obtain relative docking scores between the wild and mutants to evaluate the binding efficacy of each conformation.

Step 5: Filter conformations. Filter the conformations and their corresponding relative docking scores to identify those that are close to the average score with the highest probability density.

Step 6: Construct an augmented dataset. Create an augmented dataset that includes the filtered conformations, their associated relative docking scores, and the difference in the relative docking scores. Ensure that the augmented dataset captures variability and distinct conformational states for further analysis.

Datasets and preprocessing

TKI dataset. We validated the proposed approach on the TKI dataset collected by Hauser et al., which includes binding affinity data for mutation-ligand pairs ($\Delta\Delta G$ in Kcal/mol). The 'nd' values were excluded based on the original dataset, resulting in the selection of 144 valid docking energy data for the mutants. In total, 31 mutant and 1 wild-type proteins were selected. The conformations from each protein were then adopted as the target to dock with different TKI molecules using AutoDock Vina, thus obtaining docking energy data for both mutant and wild-type conformations. The investigated TKI molecules include axitinib, bosutinib, dasatinib, imatinib, nilotinib, and ponatinib. The data is partitioned into training and test sets for cross-validation based on the name of TKI.

Conformational ensemble prediction from AF2 subsampling

Monteiro et al. developed subsampled AlphaFold2 to predict the relative populations of protein

conformations by subsampling MSA. Their results show that the subsampling method can accurately predict the relative conformational populations of proteins when compared to the data from NMR experiments. They also obtained the accurate relative state population changes for the eight mutants for ABL kinase. We employed the same implementation to sample conformation ensembles of 31 mutants of ABL kinase. The residue length indices from 248 to 534 according to the crystal structure (PDB ID: 6XR6), following the same prediction procedure using ColabFold. ColabFold utilizes the fast homology search of MMseqs2 with AF2 to facilitate rapid prediction of protein structures and complexes. The value of max_seq:extra_seq was set to 256:512, which would result in the most diverse results concerning the activation loop conformation. We utilized 5 models of AF2 per prediction, performed 4 rounds of optimization, and set the number of seeds as 32. In total, we generated 640 conformations for each mutation.

Molecular docking and preprocessing of the relative binding energy

Different conformations of the proteins were obtained through subsampling, resulting in a relative conformational distribution. Molecular docking for each sampled conformation was performed using AutoDock Vina to determine the binding energy of the ligand across different mutations as shown in Step 2 in **Fig. 2**. The docking center is set as the coordinate center of four residues with residue index 267, 288, 378, and 399. The exhaustiveness of AutoDock Vina is set as 32. The docking energies for both the mutant and wild type were derived from ensemble docking. We computed the difference in docking score between the mutant and the wild type in Step 3 and Step 4. Then we filtered the conformational data with the smallest deviation from the Gaussian mean of the binding energy distribution of each mutant-ligand combination in Step 5. We selected the top k (eg. 1, 5, 50, and 100) values with the smallest absolute differences value from the Gaussian mean value for pairing. The k equals 1, which means no flexibiligy is incuded. We select the k values near the average because the Gaussian fitting indicates that values close to the average have the highest probability distribution. We calculate $\Delta\Delta\Delta G_{\mathrm{cal}}$ different TKIs by calculating the differences between the experimental value and the generated value in Step 6. Finally, all composite confirmation data are grouped to construct the required data-augmented dataset. The total number of raw predictions is equal to the number of conformations multiplied by the number of ligands, represented as $N=k\times 6$. As shown in **Fig. 2**, conf_exp consists of a single conformation, while conf_AF comprises k conformations generated from subsampled AF2 corresponding to the same mutation type.

Refined set of PDBbind for direct absolute value prediction

The refined set of PDBbind version 2020 was selected to compare with our approach. After the removal of ABL samples, the refined set was divided into the training and validation sets in a 9:1 ratio. The CASF-2016 is used to benchmark the performance of the regular network model. The test set data comprises the 144 mutation data in the original TKI dataset alongside the six TKIs for wild type. The training was conducted using a similar network architecture, except

for removing the Siamese network. After training the model, we predict the absolute binding free energy separately for both the wild type and mutations. Subsequently, we computed the relative binding affinity by subtracting the absolute binding energy of mutation from that of the wild type, respectively.

Input features of the Siamese learning

We adopt both the structural information and molecular graph as the molecular representation to capture the contribution from the conformation and the interaction between proteins and ligands. The molecular representations are summarized in Table S6.

Protein contact maps. Predicting geometrical relationships between residues (including distances and orientations) has demonstrated its importance in structure predictions. Protein contact maps use the adjacency matrix $A_{i,j}$ to represent the graph structure of a protein, which contains information about whether or not amino acids are found in the protein chain, specifically, a protein map $G=\left(N_p, M_p\right)$ is a contact map generated by taking an input protein containing L_p residues. A pair of residues is considered to be in contact when the Euclidean distance $d_{i,j}$ between the CA atoms of the two pairs of amino acids is less than a threshold d_c. The residues in a pair are considered to be in contact with each other. These connections are determined from the three-dimensional structure of the proteins. We set a threshold of 10 Å.

$$A_{i,j}=\begin{cases} d_{i,j} & d_{i,j}\le d_c \\ 0 & d_{i,j}>d_c \end{cases} \tag{7}$$

The protein contact map of the MPNN structure can be represented as $G_{contactMap}=(V,E)$, where V denotes a set of n atomic nodes, and each node feature vector denotes (x_v), including the one-hot coding of the protein sequence, the 12 amino acid nature features, and the BLOSUM62 matrix features[62]. E denotes the relationship of the edges of the neighbor matrix in the molecular graph, and the eigenvector of each edge is denoted (e_{vw}) as the Euclidean distance between each pair of amino acids.

Ligand graph. We choose the molecular graph as the representation of the ligands. The molecular graph of the ligand for the MPNN structure can be represented as $G_{ligand}=(V,E)$. where V denotes a set of n atomic nodes, each represented by an eigenvector of 75 values (x_v), including 44 atomic types, 11 atomic degrees, 7 implicit valence states, charge, number

of free electrons, 5 hybridization types, aromaticity, and 5 hydrogen atoms. E represents the relationship of the edges of the neighbor matrix in the molecular graph. Each edge is represented by 6 values called eigenvectors (e_{vw}), including 4 bond types, conjugacy, and intra-ring bonds. The molecular graph has been validated in our previous work.

Protein-ligand interaction graph. We adopt the 3D interaction graph structure data used by GIGN. The complex is represented as a 3D interaction graph $G_{inter} = (V, \varepsilon, R) = (V_l \cup V_p, \varepsilon_l \cup \varepsilon_{lp}, R_l \cup R_p)$, where V_l and V_p are sets of ligand and protein nodes (atoms), ε_l and ε_p are sets of ligand and protein covalent edges (covalent bonds/covalent interactions). R_l and R_p are sets of ligand and protein 3D coordinates. Each node $v_i \in V$ has initial features $x_i \in \mathrm{R}^n$ and 3D coordinates $r_i \in \mathrm{R}^3$. When atoms v_i and v_j are connected by a covalent edge, the covalent edge $e_{ij} \in \varepsilon_l \cup \varepsilon_p$. A noncovalent edge $e_{ik} \in \varepsilon_{lp}$ exists when $d_{ik} < 15$ Å, where $d_{ik} = \|r_i - r_k\|^2$ is the distance between the ligand atom $v_i \in V_l$ and the protein atom $v_k \in V_p$.

Siamese learning

Message Passing Neural Networks. We used message-passing networks to process contact maps for proteins and graph data for ligands. Gilmer et al. designed an MPNN framework for processing chemical graphs. Forward propagation consists of two phases, the message-passing phase and the readout phase. The message passing phase runs for T time steps and is defined according to the message function M_t and vertex update function U_t . In the message passing phase, the hidden state h_v^t of each node is updated according to the message m_v^{t+1} as follows.

$$m_v^{t+1} = \sum_{w \in N(v)} M_t\left(h_v^t, h_w^t, e_{vw}\right) \tag{8}$$

$$h_v^{t+1} = U_t\left(h_v^t, m_v^{t+1}\right) \tag{9}$$

Where, $N(v)$ denotes the neighbor of node v in graph G , M_t denotes the message function, U_t denotes the vertex update function. The readout phase uses the readout function R to compute the feature vectors of the entire graph according to the following equation.

$$\hat{y} = R\left(\left\{h_v^T \mid v \in G\right\}\right) \tag{10}$$

where R denotes the readout function.

Geometric interaction graph neural network. Yang et al. designed an ensemble interaction graph neural network, a Geometric interaction graph neural network (GIGN) that combines 3D structural and physical interactions to learn node representations more efficiently through a heterogeneous interaction layer that unifies covalent and non-covalent interactions into a message-passing phase.

In the Geometric Interaction Graph Neural Network (GIGN), the process of integrating covalent and noncovalent interactions into node representations is handled through a message-passing mechanism that accounts for the heterogeneity of these interactions. Each node in the 3D interaction graph, representing an atom, updates its representation by aggregating messages from its neighboring nodes, which are categorized as either covalent or noncovalent interactions. The messages from covalent neighbors are calculated using a function $M^{(\mathrm{cov})}$ that incorporates the node features x and their 3D coordinates r, while messages from noncovalent neighbors are computed using a similar function $M^{(\mathrm{ncov})}$. These messages are then aggregated separately, respecting their distinct nature.

The aggregated messages, which differ in scale due to the significant difference in the number of noncovalent versus covalent interactions, are normalized to the same scale by multilayer perceptrons $U^{(\mathrm{cov})}$ and $U^{(\mathrm{ncov})}$ before integration. This integration step ensures that the information from both interaction types is effectively combined to refine the node's representation. The final representation update for each node is given as follows.

$$h_i^{(t+1)} = U^{(\mathrm{cov})}\left(h_i^{(t)}, m_i^{(\mathrm{cov})}\right) + U^{(\mathrm{ncov})}\left(h_i^{(t)}, m_i^{(\mathrm{ncov})}\right) \tag{11}$$

Where, m_i^{cov} and m_i^{ncov} are the aggregated messages from covalent and noncovalent neighbors, respectively, and $h_i^{(t)}$ is the current representation of node v_i at layer t. This approach maintains the invariance property concerning input translations and rotations, avoiding the need for computationally expensive data augmentation and leading to a more efficient and biologically meaningful prediction of protein-ligand binding affinities.

STGNet model. We propose the SGTNet model for predicting the relative binding energy for protein-ligand complexes. The SGTNet model uses dual inputs of experimental data and generated data as shown in **Fig. 3**. The pairings are made between the reference conformation and the generated conformations. This enables the fusion of ligand-binding features with information on the structural changes triggered by the mutation. For the learning with protein contact maps and small molecules, we take the graph information as inputs and perform feature extraction through the MPNN model with shared parameters to get the feature differences between the two inputs, followed by further extraction of the deep features through the fully

connected layer. For the learning of protein-ligand interactions, we take the graph information of the docked protein-ligand structure as inputs perform feature extraction via the GIGN model with shared parameters, and further extract deep features via the fully-connected layer after calculating the feature differences. Then, all the learned graph features are concatenated together to form a joint feature, which generates the final prediction $\Delta\Delta\Delta G_{\text{pred}}$ through a fully connected layer.

To obtain the target prediction value $\overline{\Delta\Delta G_{\text{pred}}}$, the calculation method is shown in Fig. S1. The results of the raw predictions for a given test set include specific ligands, and each mutation corresponds to the predicted binding energies for these ligands. Consequently, the target prediction value needs to be averaged from the raw predicted results. We evaluated the STGNet model using the original dataset, and three pairwise augmented datasets with *k* chosen as 5, 50, and 100, respectively.

References

1. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., Bridgland, A., Cowie, A., Meyer, C., and Laydon, A. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590.
2. Watson, J.L., Juergens, D., Bennett, N.R., Trippe, B.L., Yim, J., Eisenach, H.E., Ahern, W., Borst, A.J., Ragotte, R.J., Milles, L.F., et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature *620*, 1089-1100.
3. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493-500.
4. Pillai, A., Idris, A., Philomin, A., Weidle, C., Skotheim, R., Leung, P.J., Broerman, A., Demakis, C., Borst, A.J., and Praetorius, F. (2024). De novo design of allosterically switchable protein assemblies. Nature *632*, 911-920.
5. Krishna, R., Wang, J., Ahern, W., Sturmfels, P., Venkatesh, P., Kalvet, I., Lee, G.R., Morey-Burrows, F.S., Anishchenko, I., Humphreys, I.R., et al. (2024). Generalized biomolecular modeling and design with RoseTTAFold All-Atom. Science, eadl2528.
6. Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W., and Bridgland, A. (2020). Improved protein structure prediction using potentials from deep learning. Nature *577*, 706-710.
7. AlQuraishi, M. (2021). Protein-structure prediction revolutionized. Nature *596*, 487-488.
8. Marx, V. (2022). Method of the Year: protein structure prediction. Nat Methods *19*, 5-10.
9. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., and Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583-589.
10. Humphreys, I.R., Pei, J., Baek, M., Krishnakumar, A., Anishchenko, I.V., Ovchinnikov,

S., Zhang, J., Ness, T.J., Banjade, S., Bagde, S.R., et al. (2021). Computed structures of core eukaryotic protein complexes. Science *374*, eabm4805.

11. Ahdritz, G., Bouatta, N., Floristean, C., Kadyan, S., Xia, Q., Gerecke, W., O'Donnell, T.J., Berenberg, D., Fisk, I., and Zanichelli, N. (2024). OpenFold: Retraining AlphaFold2 yields new insights into its learning mechanisms and capacity for generalization. Nat. Methods *21*, 1514-1524.
12. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science *379*, 1123-1130.
13. Chowdhury, R., Bouatta, N., Biswas, S., Floristean, C., Kharkar, A., Roy, K., Rochereau, C., Ahdritz, G., Zhang, J., and Church, G.M. (2022). Single-sequence protein structure prediction using a language model and deep learning. Nat. Biotechnol. *40*, 1617-1623.
14. Wu, R., Ding, F., Wang, R., Shen, R., Zhang, X., Luo, S., Su, C., Wu, Z., Xie, Q., and Berger, B. (2022). High-resolution de novo structure prediction from primary sequence. BioRxiv.
15. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492.
16. Bryant, P., Pozzati, G., and Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. Nat. Commun. *13*, 1265.
17. Yang, Z., Zeng, X., Zhao, Y., and Chen, R. (2023). AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduct. Target Ther. *8*, 115.
18. Kortemme, T. (2024). De novo protein design-From new structures to programmable functions. Cell *187*, 526-544.
19. Ma, W., Zhang, S., Li, Z., Jiang, M., Wang, S., Lu, W., Bi, X., Jiang, H., Zhang, H., and Wei, Z. (2022). Enhancing Protein Function Prediction Performance by Utilizing AlphaFold-Predicted Protein Structures. J. Chem. Inf. Model. *62*, 4008-4017.
20. Grinter, R., Kropp, A., Venugopal, H., Senger, M., Badley, J., Cabotaje, P.R., Jia, R., Duan, Z., Huang, P., Stripp, S.T., et al. (2023). Structural basis for bacterial energy extraction from atmospheric hydrogen. Nature *615*, 541-547.
21. Amaro, R.E. (2019). Will the real cryptic pocket please stand out? Biophys. J *116*, 753-754.
22. Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., and Zarrinkar, P.P. (2011). Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol. *29*, 1046-1051.
23. Broerman, A.J., Pollmann, C., Lichtenstein, M.A., Jackson, M.D., Tessmer, M.H., Ryu, W.H., Abedi, M.H., Sahtoe, D.D., Allen, A., Kang, A., et al. (2024). Design of facilitated dissociation enables control over cytokine signaling duration. bioRxiv, 2024.2011.2015.623900.
24. Jänes, J., and Beltrao, P. (2024). Deep learning for protein structure prediction and design—progress and applications. Mol. Syst. Biol. *20*, 162-169.
25. Wayment-Steele, H.K., Ojoawo, A., Otten, R., Apitz, J.M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., and Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. Nature *625*, 832-839.

26. Monteiro da Silva, G., Cui, J.Y., Dalgarno, D.C., Lisi, G.P., and Rubenstein, B.M. (2024). High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. Nat Commun *15*, 2464.
27. Guan, X., Tang, Q.-Y., Ren, W., Chen, M., Wang, W., Wolynes, P.G., and Li, W. (2024). Predicting protein conformational motions using energetic frustration analysis and AlphaFold2. Proc. Natl. Acad. Sci. U.S.A. *121*, e2410662121.
28. Stein, R.A., and McHaourab, H.S. (2022). SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with Alphafold2. PLoS Comput. Biol. *18*, e1010483.
29. Raisinghani, N., Alshahrani, M., Gupta, G., Tian, H., Xiao, S., Tao, P., and Verkhivker, G.M. (2024). Integration of a Randomized Sequence Scanning Approach in AlphaFold2 and Local Frustration Profiling of Conformational States Enable Interpretable Atomistic Characterization of Conformational Ensembles and Detection of Hidden Allosteric States in the ABL1 Protein Kinase. J. Chem. Theory Comput. *20*, 5317-5336.
30. Del Alamo, D., Sala, D., Mchaourab, H.S., and Meiler, J. (2022). Sampling alternative conformational states of transporters and receptors with AlphaFold2. Elife *11*, e75751.
31. Freiberger, M.I., Ruiz-Serra, V., Pontes, C., Romero-Durana, M., Galaz-Davison, P., Ramírez-Sarmiento, C.A., Schuster, C.D., Marti, M.A., Wolynes, P.G., and Ferreiro, D.U. (2023). Local energetic frustration conservation in protein families and superfamilies. Nat. Commun. *14*, 8379.
32. Tessa, A., Battini, R., Rubegni, A., Storti, E., Marini, C., Galatolo, D., Pasquariello, R., and Santorelli, F. (2016). Identification of mutations in AP4S1/SPG52 through next generation sequencing in three families. Eur. J. Neurol. *23*, 1580-1587.
33. Jones, D.T., Singh, T., Kosciolek, T., and Tetchner, S. (2015). MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. Bioinformatics *31*, 999-1006.
34. Wang, S., Sun, S., Li, Z., Zhang, R., and Xu, J. (2017). Accurate de novo prediction of protein contact map by ultra-deep learning model. PLoS Comput. Biol. *13*, e1005324.
35. Heo, L., and Feig, M. (2018). Experimental accuracy in protein structure refinement via molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. *115*, 13276-13281.
36. Park, H., Ovchinnikov, S., Kim, D.E., DiMaio, F., and Baker, D. (2018). Protein homology model refinement by large-scale energy optimization. Proc. Natl. Acad. Sci. U.S.A. *115*, 3054-3059.
37. Steinbrecher, T.B., Dahlgren, M., Cappel, D., Lin, T., Wang, L., Krilov, G., Abel, R., Friesner, R., and Sherman, W. (2015). Accurate binding free energy predictions in fragment optimization. J. Chem. Inf. Model. *55*, 2411-2420.
38. Moore, J.H., Margreitter, C., Janet, J.P., Engkvist, O., de Groot, B.L., and Gapsys, V. (2023). Automated relative binding free energy calculations from SMILES to ΔΔG. Chem. Commun. *6*, 82.
39. Cao, D., Chen, G., Jiang, J., Yu, J., Zhang, R., Chen, M., Zhang, W., Chen, L., Zhong, F., Zhang, Y., et al. (2024). Generic protein–ligand interaction scoring by integrating physical prior knowledge and data augmentation modelling. Nat. Mach. Intell. *6*, 688-700.
40. Gilson, M.K., and Zhou, H.-X. (2007). Calculation of Protein-Ligand Binding Affinities. Annu. Rev. Biophys. *36*, 21-42.

41. Amaro, R.E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J.A., Miao, Y., and Smith, J.C. (2018). Ensemble Docking in Drug Discovery. Biophys. J *114*, 2271-2278.
42. Wang, J.Y. (2014). The capable ABL: what is its biological function? Mol. Cell. Biol. *34*, 1188-1197.
43. O'Brien, S.G., Guilhot, F., Larson, R.A., Gathmann, I., Baccarani, M., Cervantes, F., Cornelissen, J.J., Fischer, T., Hochhaus, A., and Hughes, T. (2003). Imatinib compared with interferon and low-dose cytarabine for newly diagnosed chronic-phase chronic myeloid leukemia. N. Engl. J. Med. *348*, 994-1004.
44. Weisberg, E., Manley, P.W., Cowan-Jacob, S.W., Hochhaus, A., and Griffin, J.D. (2007). Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. Nat. Rev. Cancer *7*, 345-356.
45. Zhou, Y., Portelli, S., Pat, M., Rodrigues, C.H., Nguyen, T.-B., Pires, D.E., and Ascher, D.B. (2021). Structure-guided machine learning prediction of drug resistance mutations in Abelson 1 kinase. Comput. Struct. Biotechnol. J. *19*, 5381-5391.
46. Yu, Y., Wang, Z., Wang, L., Tian, S., Hou, T., and Sun, H. (2022). Predicting the mutation effects of protein–ligand interactions via end-point binding free energy calculations: strategies and analyses. J. Cheminf. *14*, 56.
47. Yang, Z., Ye, Z., Qiu, J., Feng, R., Li, D., Hsieh, C., Allcock, J., and Zhang, S. (2023). A mutation-induced drug resistance database (MdrDB). Chem. Commun. *6*, 123.
48. Li, P., and Liu, Z.-P. (2024). MuToN Quantifies Binding Affinity Changes upon Protein Mutations by Geometric Deep Learning. Adv. Sci. *11*, 2402918.
49. Friedman, R. (2022). Computational studies of protein–drug binding affinity changes upon mutations in the drug target. Wiley Interdiscip. Rev.: Comput. Mol. Sci. *12*, e1563.
50. Roskoski, R. (2016). Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacol. Res. *103*, 26-48.
51. Paul, F., Meng, Y., and Roux, B. (2020). Identification of Druggable Kinase Target Conformations Using Markov Model Metastable States Analysis of apo-Abl. J. Chem. Theory Comput. *16*, 1896-1912.
52. McBride, J.M., Polev, K., Abdirasulov, A., Reinharz, V., Grzybowski, B.A., and Tlusty, T. (2023). AlphaFold2 Can Predict Single-Mutation Effects. Phys. Rev. Lett. *131*, 218401.
53. Roney, J.P., and Ovchinnikov, S. (2022). State-of-the-art estimation of protein model accuracy using AlphaFold. Phys Rev Lett *129*, 238101.
54. McNutt, A.T., and Koes, D.R. (2022). Improving ΔΔG Predictions with a Multitask Convolutional Siamese Network. J. Chem. Inf. Model. *62*, 1819-1829.
55. Feng, B., Liu, Z., Huang, N., Xiao, Z., Zhang, H., Mirzoyan, S., Xu, H., Hao, J., Xu, Y., Zhang, M., and Wang, S. (2024). A bioactivity foundation model using pairwise meta-learning. Nat. Mach. Intell. *6*, 962-974.
56. Fralish, Z., Chen, A., Skaluba, P., and Reker, D. (2023). DeepDelta: predicting ADMET improvements of molecular derivatives with deep learning. J. Cheminf. *15*, 101.
57. Yu, J., Li, Z., Chen, G., Kong, X., Hu, J., Wang, D., Cao, D., Li, Y., Huo, R., Wang, G., et al. (2023). Computing the relative binding affinity of ligands based on a pairwise binding comparison network. Nat. Comput. Sci. *3*, 860-872.
58. Jorgensen, W.L., and Thomas, L.L. (2008). Perspective on Free-Energy Perturbation

Calculations for Chemical Equilibria. J. Chem. Theory Comput. *4*, 869-876.
59. Hauser, K., Negron, C., Albanese, S.K., Ray, S., Steinbrecher, T., Abel, R., Chodera, J.D., and Wang, L. (2018). Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. Commun. Biol. *1*, 70.
60. Jiménez-Luna, J., Pérez-Benito, L., Martínez-Rosell, G., Sciabola, S., Torella, R., Tresadern, G., and De Fabritiis, G. (2019). DeltaDelta neural networks for lead optimization of small molecule potency. Chem. Sci. *10*, 10911-10918.
61. Lu, X., Xie, L., Xu, L., Mao, R., Xu, X., and Chang, S. (2024). Multimodal fused deep learning for drug property prediction: Integrating chemical language and molecular graph. Comput. Struct. Biotechnol. J. *23*, 1666-1679.
62. Eddy, S.R. (2004). Where did the BLOSUM62 alignment score matrix come from? Nat. Biotechnol. *22*, 1035-1036.
63. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., and Dahl, G.E. (2017). Neural message passing for quantum chemistry. (JMLR.org), pp. 1263-1272.
64. Yang, Z., Zhong, W., Lv, Q., Dong, T., and Yu-Chian Chen, C. (2023). Geometric Interaction Graph Neural Network for Predicting Protein–Ligand Binding Affinities from 3D Structures (GIGN). J. Phys. Chem. Lett. *14*, 2020-2033.
65. Lyczek, A., Berger, B.-T., Rangwala, A.M., Paung, Y., Tom, J., Philipose, H., Guo, J., Albanese, S.K., Robers, M.B., and Knapp, S. (2021). Mutation in Abl kinase with altered drug-binding kinetics indicates a novel mechanism of imatinib resistance. Proc. Natl. Acad. Sci. U.S.A. *118*, e2111451118.
66. Fernandes, P.O., Martins, D.M., de Souza Bozzi, A., Martins, J.P.A., de Moraes, A.H., and Maltarollo, V.G. (2021). Molecular insights on ABL kinase activation using tree-based machine learning models and molecular docking. Mol. Divers. *25*, 1301-1314.
67. Xie, T., Saleh, T., Rossi, P., and Kalodimos, C.G. (2020). Conformational states dynamically populated by a kinase determine its function. Science *370*, eabc2754.
68. Cournia, Z., Allen, B., and Sherman, W. (2017). Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J. Chem. Inf. Model. *57*, 2911-2937.

Funding

This work was supported by the National Natural Science Foundation of China (62373172, 22003020 and 12074151), the Natural Science Foundation of Jiangsu Province (BK20191032), and Changzhou Sci. & Tech. Program (CJ20200045).

Author contributions

L.Xie.: Initiated the project and proposed the method. X.L.: Performed modeling. L.Xie. and X.L.: Conducted data analysis. D.Z. Y.J. and L.Xu.: Contributed the data analysis. S.C. and X.X.: Supervised the work.

Competing interests

The authors declare no competing interests.

Data availability

AlphaFold2 subsampling prediction data generated in this study have been deposited in the Zenodo: https://zenodo.org/records/14607966.

Code availability

The code used to generate the results shown in this study is available under an MIT License via GitHub at https://github.com/AIMedDrug/STGNet.

Additional information

The detailed analysis can be found in the supplementary material.

Key Points

- Predicting the impact of mutations on protein-ligand binding affinity is vital for addressing drug resistance and repurposing drugs. Current methods face challenges due to reliance on known protein-ligand structures.
- We proposed a novel few-shot learning framework that integrates AlphaFold2 subsampling, ensemble docking, and Siamese learning to predict relative binding affinity changes caused by mutations.
- We augmented the dataset by pairing the reference state with the structures generated by AlphaFold2. By incorporating the generated data, we enhance the model's prediction accuracy and robustness, particularly for targets with high structural variability.
- The proposed method achieves higher prediction accuracy for relative binding affinities by incorporating structural diversity from AlphaFold.