

PremPLI: Predicting the Effects of Missense Mutations on Protein-Ligand Interactions

Tingting Sun

Soochow University

Yuting Chen

Soochow University

Yuhao Wen

Soochow University

Zefeng Zhu

Soochow University <https://orcid.org/0000-0002-2761-3291>

Minghui Li (✉ minghui.li@suda.edu.cn)

Soochow University

Article

Keywords: PremPLI, protein-ligand interactions, therapeutic drugs

Posted Date: May 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-417047/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Communications Biology on November 19th, 2021. See the published version at <https://doi.org/10.1038/s42003-021-02826-3>.

PremPLI: Predicting the Effects of Missense Mutations on Protein-Ligand Interactions

ABSTRACT

Protein-ligand interactions trigger a multitude of signal transduction processes and resistance to small-molecule drugs is the main cause of the failure of therapeutic drugs in clinical practice. Missense mutations altering the binding of ligands to proteins are one of the critical mechanisms that result in genetic disease and drug resistance. Computational methods have made a lot of progress for predicting binding affinity changes and identifying resistance mutations, but they are still not satisfied and need to be further improved in both accuracy and speed. To address these issues, we introduced PremPLI, a structure-based machine learning method for quantitatively estimating the effects of single mutations on ligand binding affinity changes. A comprehensive comparison of the predictive performance of PremPLI with other available methods on two benchmark datasets confirms that our approach performs robustly and presents similar or even higher predictive accuracy than the approaches relying on first-principle statistical mechanics and mixed physics- and knowledge-based potentials while requires much less computational resources. PremPLI can be used for guiding the design of ligand-binding proteins, identifying and understanding disease driver mutations, and finding potential resistance mutations for different drugs. PremPLI is freely available at <https://lilab.jysw.suda.edu.cn/research/PremPLI/> and allows to do large-scale mutational scanning.

INTRODUCTION

Protein-ligand interactions are of fundamental importance in myriad processes occurring in living organisms, triggering a multitude of signal transduction processes.¹⁻³ Many genetic diseases are caused by missense mutations through altering binding between small-molecule ligands and proteins.^{4,5} For example, the association between disease-related mutations in the type I collagen and the ligand binding sites was observed⁶ and the ligand-binding-domain mutations in human androgen receptor gene led to disrupted interaction between the N- and C-terminal domains.⁷ Moreover, drug resistance is the main cause of the failure of therapeutic drugs in clinical practice, particularly for the treatment of infectious diseases and cancer,⁸⁻¹⁰ and mutation in drug target is a critical mechanism that results in drug resistance.¹¹⁻¹³ For instance, in hormone-resistant breast cancer, mutations in estrogen receptor 1 gene (ESR1) are associated with acquired endocrine resistance,¹⁴ and two ESR1 ligand-binding site mutations produced partial resistance to the currently available endocrine therapies.¹⁵ During the last decade, next-generation sequencing techniques have detected vast amounts of genetic mutations in humans, leaving clinicians and researchers without knowledge of whether these mutations are associated with genetic diseases or the emergence of drug resistance. Experimental methods can accurately measure the effects of missense mutations on proteins and their complexes, but they are time-consuming and expensive and do not have the capability to tackle large amounts of data.¹⁶ Therefore, the development of reliable computational methods to reveal molecular effects of missense mutations would pave the way for the identification of pathogenic or drug resistance mutations and contribute to many biomedicine related fields and drug discovery.

Some attempts to model the effects of mutations on protein-ligand interactions have been made but with limited success and applicability. In summary, the methods can be classified into three main categories: (i) statistical/machine learning approaches using experimentally measured free energy data to parameterize regression models,¹⁷⁻¹⁹ such as mCSM-lig,¹⁷ which usually require little computational cost; (ii) simulation-based methods that rely on mixed physics- and knowledge-based potentials to sample side-chain rotamers with restrained backbone motion, such as Rosetta,²⁰ which have proven successful in the early stages of the discovery process;^{21,22} (iii) alchemical free energy calculations that introduce a series of intermediate *alchemical* states in a thermodynamic cycle,²³⁻²⁷ which have become increasingly popular in the lead optimization stage

of small molecule drug discovery and started being used prospectively by the pharmaceutical industry.^{28,29} Several studies have been conducted to examine the performance of these methods on different test cases.^{19,25,27} For instance, Aldeghi et al. evaluated the potential of all above three kinds of computational approaches to predict changes in binding affinity of eight different inhibitors with cancer target Abl kinase upon 144 mutations.¹⁹ Overall, the prediction performance of Rosetta and alchemical free energy calculations are better than the current available machine learning approaches but they are still not satisfied and need to be further improved in both accuracy and speed. Alchemical calculations considering the full conformational flexibility of protein-ligand complex and the discrete nature of solvent are much more computationally expensive compared to statistical and (semi)-empirical approaches, moreover the workflow of which is complex, tedious, and error-prone. Therefore, developing prediction methods with tradeoffs between computational cost and accuracy is urgently required.

With a large amount of experimental data available, many data-driven machine learning methods have been developed to assess the impact of mutations on protein stability³⁰⁻³⁶ and protein-protein interactions,³⁷⁻⁴⁷ including PremPS,³⁰ MutaBind,³⁹ and MutaBind2⁴⁰ introduced by us. In addition, we also proposed PremPDI⁴⁸ and PremPRI⁴⁹ to evaluate the effects of mutations on binding between protein and nucleic acid. These approaches have achieved relatively high predictive accuracy with low computational cost. However, very few machine learning methods were proposed to assess the impact of mutations on protein-ligand binding and their prediction accuracy is very limited.¹⁷⁻¹⁹ One reason hindering the development is due to the complexity of small molecule chemistry and binding interaction. Another major limitation is the availability of high-quality experimental data that can be used for training. Recently, a manually curated database Platinum was created,⁵⁰ which associates experimentally measured effects of mutations on protein-small molecule binding with three-dimension structures of the corresponding complexes, allowing us to propose a structure-guided computational method to estimate the affinity changes upon mutations quantitatively.

Therefore, we developed a new machine learning computational method, PremPLI, to estimate the effects of single point mutations on protein-ligand interactions by calculating the binding affinity changes quantitatively. PremPLI uses a random forest regression scoring function and consists of 11 sequence- and structure-based features. It performs well across different types of cross-validation and independent tests, with similar predictive accuracy to Rosetta and

alchemical free energy calculations but significantly lower computational costs. PremPLI can be used to aid in the development of new drugs to combat rising drug resistance, finding disease-causing or cancer driver missense mutations, and the design of proteins with novel ligand-binding functionalities and specificities.

METHODS

Experimental datasets used for training PremPLI

Platinum,⁵⁰ a manually curated database, includes experimentally measured effects of mutations (binding affinity changes) on structurally defined protein-ligand complexes. Binding affinity is calculated as $\Delta G = RT\ln(K_D) = RT\ln(K_i) = RT\ln(K_m)$, and the binding affinity change upon mutation is defined as $\Delta\Delta G = \Delta G_{mutant} - \Delta G_{wild\ type}$. The positive and negative values of $\Delta\Delta G$ correspond to the mutations decreasing and increasing binding affinity respectively. A total of 797 single mutations with experimentally determined complex structures and binding affinity changes were included in the Platinum database, among which 151 mutations had both wild-type and mutant structures. Following the anti-symmetric property, that is, Gibbs free energy change introduced by a forward mutation (*wild type* \rightarrow *mutant*) plus the change induced by its reverse mutation (*mutant* \rightarrow *wild type*) should be equal to zero, so we took the mutant complex as being the “wild type” and inverted the sign of the affinity change, resulting in 948 mutations.

Next, we removed following entries from the Platinum database: (a) ligands or mutations cannot be found in the given complex structures; (b) complexes with modified residues at protein-ligand binding interface. If any heavy atom of a residue was located within 5 Å from any heavy atom of ligands, we defined this residue as an interface residue; (c) mutations with missing coordinates; (d) mutations occurring at metal coordination sites; (e) structures with only C α atomic coordinates; (f) structures with a resolution lower than 3 Å. Besides, we only retained the interactions between one protein and one type of ligand. For mutations with multiple $\Delta\Delta G_{exp}$, we first preserved all values, resulting in 859 single mutations (it will be referred to as S859). Second, only one $\Delta\Delta G_{exp}$ value is chosen according to the following criteria in the order specified below: (1) pH is closest to neutral and/or the temperature (T) is closest to room temperature; (2) the value of $\Delta\Delta G_{exp}$ calculated by K_D has the highest priority, followed by K_i ; (3) the experimental method of fluorimetry is preferred over other measurements. As a result, 796 unique single mutations from

360 complex structures were retained (it will be referred to as S796). The majority of complexes contain only one mutation (Fig. S1a). More information about the training set of S796 is shown in Fig. S1 and Table S1.

The last step is to process PDB coordinate files of complexes, which were obtained from the Protein Data Bank (PDB).⁵¹ First, the biological assembly was used for each complex. Second, only coordinates of proteins and ligands studied were retained. Third, the ligand was further removed from the PDB file when the ratio of the number of its contact heavy atoms to total heavy atoms is less than 0.5. We defined an atom in ligand as a contact atom when it is located within 6 Å from any heavy atom in protein. That is, only ligands that bind relatively strongly to protein were retained. As a result, four types of protein-ligand complex structures were included in the training set: monomer with one ligand; monomer with two or more ligands; homomer with one ligand; homomer with two or more ligands (Fig. S2). Please note, we only study the interactions between one protein and one type of ligand, but they may have multiple chains or multiple ligands.

Experimental datasets used for testing

Recently, Hauser et al. compiled a benchmark dataset consisting of reliable inhibitor ΔpIC_{50} data for 144 clinically identified mutants of human kinase Abl and examined the potential for alchemical free-energy calculations to predict resistance of these mutations to eight FDA-approved tyrosine kinase inhibitors (TKIs).²⁵ None of the mutations or protein-ligand complexes in this benchmark dataset were included in our training dataset. The binding free energy change ($\Delta\Delta G_{exp}$) can be derived from the following equation:

$$\Delta\Delta G_{exp} = RT \ln \frac{K_{i,mut}}{K_{i,WT}} \approx RT \ln \frac{IC_{50,mut}}{IC_{50,WT}}$$

Names, chemical structures, and the distribution of experimental $\Delta\Delta G_{exp}$ for each inhibitor are provided in Fig. S3. In addition, Hauser et al. also provided eight processed complex structures (such as adding missing residues and loops) including two docking models, which were used to evaluate our method and perform comparisons with other methods (this dataset will be referred to as S144 including 144 mutations and eight processed complex structures). Since most users using our method will not or will not be able to produce the structures as Hauser et al. did for performing alchemical free-energy calculations in which the workflow is relatively complex and tedious, we also tested the performance of PremPLI on structures obtained from the Protein Data Bank directly

(129 mutations from six Abl-TKI complexes, it will be referred to as S129). This can also check the sensitivity of our method to the initial input structures. The criteria for processing the training dataset have been applied in checking the test sets.

Another dataset proposed by Aldeghi et al.²⁷ includes 134 single and multiple mutations across 17 proteins and 27 ligands from the Platinum database. Only affinities determined by isothermal titration calorimetry and surface plasmon resonance were selected and proline mutations were excluded. Moreover, it is a diverse and challenging benchmark set, which includes large and flexible ligands, proteins with different folds, and many small-to-large/large-to-small and charge-changing mutations. Aldeghi et al. performed free energy calculations on these 134 mutations using first-principles statistical mechanics and Rosetta protocols, respectively. In our study, 25 multiple mutations were first removed from this benchmark. Then, ten single mutations for which ligand cannot be found in the given complex structures and protein interacts with multiple types of ligands were further removed. As a result, 99 single mutations from 42 complexes were retained (it will be referred to as S99), which is a subset of S796.

The 3D structures of protein-ligand complexes were either taken from Hauser et al. study²⁵ (S144 dataset) or from the Protein Data Bank (S796, S129 and S99 datasets). Mutant structures were produced using the BuildModel module of FoldX.³⁴ Missing heavy side-chain and hydrogen atoms in proteins were added via VMD program⁵² using the topology parameters of CHARMM36 force field.⁵³ Hydrogen atoms of ligands were added via Chimera.⁵⁴

The model of PremPLI

PremPLI employs Random Forest (RF) regression scoring function and is composed of 11 distinct features that contribute significantly to the quality of the model (Table S2).

- *PSSM*, a Position-Specific Scoring Matrix (PSSM) score for wild-type residue type at mutated site. PSSMs was calculated from PSI-BLAST⁵⁵ searches of protein sequences in NCBI non-redundant database and the resulting profile was constructed using the default parameters. ΔCS represents the change in conservation induced by a single mutation, which was calculated by PROVEAN program.⁵⁶
- $Prox^{wt}$ and $Prox^{mut}$ are the number of contacts between mutated site and ligands in wild-type and mutant structure, respectively. If the distance between two atoms is more than the Van der Waals interaction distance and within 5 angstroms, they were defined in contact.

These terms were calculated by Arpeggio.⁵⁷

- MWt presents the molecular weight of ligand which was calculated using XLOGP3.⁵⁸ N_{NO} is the number of nitrogen and oxygen atoms of ligand.
- H term accounts for the hydrophobicity of mutated site. The hydrophobicity for each type of amino acid residue was obtained from³⁸.
- P_{RKDE} is the fraction of charged residues (R, K, D and E) in the wild-type structure. $P_{RKDE} = \frac{N_R + N_K + N_D + N_E}{N_{All}}$, for instance, N_R is the number of all arginine residues, and N_{All} is the total number of amino acids. $P_Q = \frac{N_Q}{N_{All}}$, N_Q is the number of glutamine residues buried in the protein core. A residue is defined as buried if the ratio of solvent accessible surface area of this residue in the protein and in the extended tripeptide is less than 0.2.⁵⁹
- M_{AA1} is the matrix of single residue interchanges derived from spatially conserved motifs⁶⁰ and M_{AA2} is the matrix of amino acid substitutions produced by superposition of homologous protein structures.⁶¹ They represent the more/less favorable trends of residue exchange in protein structures, obtained from Amino Acid Index Database with identifiers of AZAE970102 and RISJ880101 (AAindex, <https://www.genome.jp/aaindex/>).

Cross-validation procedures

We first performed 10 times 5-fold and 10-fold cross-validations on the mutations from training dataset (named as CV1 and CV2), and then evaluated the performance of our approach on two low redundancy sets; low redundant at (i) complex (named as leave-one-complex-out validation) and (ii) ligand (named as leave-one-type-ligand-out validation). There are 360 complexes and 168 types of ligands in the training dataset (Fig. S1a and S1b). Namely, we leave all mutations from one complex or from the complexes having the same type of ligand out as a test set and use the remaining complexes/mutations to train the model, repeating this procedure for each complex or each type of ligand.

Statistical analysis and evaluation of performance

Pearson correlation coefficient (PCC) and root-mean-square error (RMSE) were used to measure the agreement between experimentally determined and predicted values of changes in binding affinity. T-test was used to assess whether the correlation coefficient is statistically

significant from zero. Hittner2003⁶² and Fisher1925⁶³ tests implemented in the R package *cocor*⁶⁴ were used to check whether the difference in correlation coefficients between PremPLI and other methods is significant. Hittner2003 and Fisher1925 tests are based on dependent groups with overlapping variable and independent groups respectively. RMSE (kcal mol⁻¹) is the standard deviation of the prediction errors, calculated by taking the square root of the average squared difference between predicted and experimental estimates of $\Delta\Delta G$.

To quantify the performance of PremPLI and other approaches in distinguishing resistance from other mutations, we performed Receiver Operating Characteristics (ROC) and Precision-Recall (PR) analyses. True positive rate/recall is defined as $TPR=TP/(TP+FN)$, false positive rate is defined as $FPR=FP/(FP+TN)$, and precision is defined as $PPV=TP/(TP+FP)$ (TP: true positive; TN: true negative; FP: false positive; FN: false negative). In addition, we also calculated Matthews correlation coefficient (MCC) that accounts for imbalances in the labeled dataset.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

RESULTS and DISCUSSION

Currently, the majority of studies for estimating the effects of mutations on protein-ligand binding are based on alchemical free energy calculations and Rosetta protocols. The prediction accuracy, especially the speed, are still not satisfied. Hence, we develop the machine-learning approach PremPLI, which significantly improves the predictive speed without compromising the accuracy.

PremPLI predictive model is composed of 11 features and built using Breiman's random forest regression algorithm implemented in the R randomForest package. Hyperparameter tuning in a balanced 5-fold cross validation determined that the optimal settings for the number of decision trees and the number of features considered by each tree when splitting a node are 300 and 3 respectively. As described in the Methods section, we compiled two datasets of S859 and S796, and the difference between them is using all values or only one selected binding affinity change for these mutations with multiple $\Delta\Delta G_{exp}$. The performance trained and tested on S859 and S796 are shown in Table S3 and no significant differences were observed. However, when tested on the blind set of S144, the Pearson correlation coefficients using the models trained by S796 and

S859 are 0.46 and 0.36, respectively, the difference is statistically significant (p -value < 0.01 , Hittner2003 test). Therefore, the unique single mutation dataset of S796 was used to parameterize PremPLI model. The Pearson correlation coefficient between experimental and calculated changes in binding affinity is 0.70 and the corresponding RMSE is $1.08 \text{ kcal mol}^{-1}$ when trained and tested on S796 (Fig. 1a). Besides, we also tried two other popular learning algorithms of Support Vector Machine (SVM) and eXtreme Gradient Boosting (XGBoost), and the results shown in Table S4 verified that the random forest regression algorithm shows the best performance.

Performance on four types of cross-validation

In supervised machine learning, overfitting is a common problem in which the model performs well on the training data but poorly on the new and unseen data. To prevent overfitting, cross-validation has been used to tune the PremPLI model, and in addition, only 11 features with significant contributions were selected. Here, we further performed four different types of cross-validation to identify how well our model performs on the unseen data (see Methods section for more details). Figure 1b shows the distribution of Pearson correlation coefficients for 10 times 5-fold (CV1) and 10-fold (CV2) cross-validations. The PCC of each cross-validation round exceeds 0.56 with an average value of ~ 0.70 for both CV1 and CV2 (Table S3). Upon the “leave-one-complex-out” procedure, the Pearson correlation coefficient between experimental and computed binding affinity changes reach the value of 0.55 and the corresponding RMSE is $1.26 \text{ kcal mol}^{-1}$ (Fig. 1c). Last, we performed a validation by leaving all complexes with the same type of ligand out of the training set (“leave-one-type-ligand-out” validation), and the correlation coefficient is 0.53, remaining statistically significant (Fig. 1d).

In addition, we analyzed 5% outliers to better understand strengths and limitations of PremPLI. Studentized residuals were used in detecting outliers. The performance after removing outliers is presented in Fig. S4a, showing significantly improved PCC after removing 5% outliers. Through the analysis (Fig. S4b and S4c), we found that the outliers are more likely to occur in complexes with lower-affinity ligand binding and at sites with higher number of proximal atoms and hydrogen bonds. Consistent with the observation from the study of¹⁷, outliers correspond to mutations with extreme experimental values (highly decreasing or increasing affinity). PremPLI could relatively correctly classify the highly decreasing mutations as decreasing but lose the ability to classify highly increasing mutations (Fig. S4d).

Validation on independent test sets and comparison with other methods

First, the benchmark dataset of S144 was used to assess the predictive performance of PremPLI and compare with other computational approaches. Hauser et al. calculated relative changes in free energy for these 144 mutations using MD simulations combined with the solution of the Generalized Born equation calculated by Prime and alchemical free-energy perturbation calculations using FEP+, respectively.²⁵ More recently, Aldeghi et al. also used this dataset and tested performance of different methodologies including MD simulations with a nonequilibrium free energy calculation protocol, Rosetta, and machine learning (named as ML1 and ML2).¹⁹ For different prediction results of MD calculations under different force fields or Rosetta using different scoring functions, only the one showing the best performance was presented in our study. Table 1 provides the performance of six approaches tested on S144, which are PremPLI, FEP+ (MD calculation under the OPLS3 force field), R15 (Rosetta using the standard REF15 scoring function), Prime, and two machine learning methods of ML1 and mCSM-lig. The ML1 model trained on 484 single mutations from the Platinum database. mCSM-lig used a combination of different statistical potentials to predict $\Delta\Delta G$ values and was parameterized on 763 single mutations from Platinum database.¹⁷ The PCC and RMSE values shown in Table 1 and Fig. S5a indicate that R15 has the best performance, followed by PremPLI and FEP+ methods, and ML1 performs the worst. In addition, Aldeghi et al. also built a model of ML2 trained on the S144 and the PCC and RMSE are 0.57 and 0.68 kcal mol⁻¹, respectively, upon 8-fold nested cross-validation (similar to the leave-one-complex-out validation used in our study). To compare with ML2, we also used S144 as the training set and built a new model. The PCC reaches 0.70 and RMSE is 0.60 kcal mol⁻¹ upon leave-one-complex-out validation, outperforming ML2.

Second, since most users using machine learning methods, such as PremPLI or mCSM-lig, will not or will not be able to process complex structures as Hauser et al. did for performing alchemical free-energy calculations, the workflow of which is relatively complex and tedious, we also calculated binding affinity changes of PremPLI and mCSM-lig using structures obtained from the Protein Data Bank directly (S129, 129 mutations from six Abl-TKI complexes). The results presented in Table 1 demonstrate that PremPLI has significantly better performance than mCSM-lig and is not very sensitive to the initial input structures. Moreover, to further compare with mCSM-lig, we also retrained our model on S763 dataset (a training set of mCSM-lig, Table S1).

Even though our features selected were not based on this dataset, we still obtained a significantly higher correlation coefficient (PCC = 0.73 and RMSE = 1.07 kcal mol⁻¹) than mCSM-lig (PCC = 0.63 and RMSE = 2.06 kcal mol⁻¹),¹⁷ both of which were upon 10-fold cross-validation.

Third, we assessed the predictive performance of PremPLI on S99 dataset. Since S99 is a subset of S796, we retrained the model after removing all mutations in the overlapped complexes with S99 from the training dataset, and then tested it on S99 (named as PremPLI^C, trained on 671 mutations from 318 complexes). Binding affinity changes for S99 calculated by 11 Rosetta protocols, first-principles statistical mechanics in which MD simulations used six different force fields, and their combinations were available from the study of²⁷. The results of the best performing Rosetta protocol (R14) and MD calculation (A14), and their combination (RMD) were provided in Table 1 and Fig. S5b. The correlation coefficient of PremPLI^C is 0.69, significantly higher than other methods (p-value < 0.01, Hitter2003 test). On the S99, a diverse and challenging benchmark set, Rosetta did not perform as well as on the S144. In addition, we found that the prediction values of Rosetta for half of mutations from the S144 and S99 are around zero (Fig. S5).

Last, it is important to put the performances of approaches in the context of their computational cost. The running time of PremPLI for a single mutation per protein with ~ 400 residues is about ten minutes on a single CPU core, and it requires ~ 20 seconds for each additional mutation introduced in the same complex. For instance, calculating $\Delta\Delta G_{PremPLI}$ for a mutation and 26 mutations in Abl-axitinib complex takes ~10 and 20 minutes, respectively. However, each $\Delta\Delta G$ estimate takes up to 32 hours on a single CPU core and 72 hours on a single GPU core for Rosetta (R15) and FEP+ calculations, respectively (Table S5).

Performance on predicting resistance mutations

Generally, resistance can be defined according to the resistance fold (RF): RF ≤ 1, no resistance; 1 < RF ≤ 10, low resistance; RF > 10, resistance.^{19,25} Following the equation of $\Delta\Delta G_{exp} = RT \ln RF$, no resistance, low resistance and resistance mutations correspond to $\Delta\Delta G_{exp} \leq 0$, $0 < \Delta\Delta G_{exp} \leq 1.36$ and $\Delta\Delta G_{exp} > 1.36$ kcal mol⁻¹, respectively. Here we examined the potential for PremPLI and other methods to predict resistance mutations. ROC and PR curves using different approaches to distinguish resistance from no and low resistance mutations are shown in Fig. 2. The values of area under the receiver operating characteristics curve (AUC-ROC) and precision-recall curve (AUC-PR) and maximal MCC values are provided in Table 1, which are the most

relevant performance measures to examine when the objective is to identify resistance mutations rather than assessing their impact on ligand binding quantitatively. The maximal MCC value was calculated for each method through calculating the MCC across a range of thresholds. Given the fraction of resistance mutations in the datasets of S144 and S99, a random classifier would return an AUC-PR of 0.13 and 0.19, respectively. In terms of the results provided in Table 1 and Fig. 2, the best classifiers are PremPLI, Rosetta and MD calculations.

As are shown in Fig. S7, mutations located on binding interface have on average larger effects on protein-ligand interactions than non-interface mutations. The interface residues were defined if any heavy atoms of them are within 5 Å distance from any heavy atoms in ligands. Rosetta and MD calculations perform well on interface mutations but lose the ability to predict non-interface mutations, while PremPLI yields statistically significant positive correlations in predicting both interface and non-interface mutations (Fig. 3 and Fig. S7).

Performance on predicting different types of ligands

Previous studies have demonstrated that the impacts of mutations on different types of ligands cannot be equally well predicted.^{19,25} Here we further analyzed the predictive performance of PremPLI on each type of ligand. The S144 dataset is composed of eight TKIs, six of which have co-crystal structures and each has more than 20 mutations, and inhibitors of erlotinib and gefitinib adapted docking complex structures and each has less than 10 mutations (Fig. S3). Performance for each inhibitor calculated by six methods is shown in Fig. 4a and Fig. S8. PremPLI and FEP+ perform well on four TKIs with statistically significant PCC values, and Rosetta produces statistically significant PCCs for six TKIs including two using docking models. All approaches lose the ability to predict the effects of mutations on ponatinib binding. By observing experimental and predicted values (Fig. S8), we found that three mutations increasing Abl and axitinib binding are predicted by PremPLI as decreasing mutations. After excluding these three mutations, the correlation coefficients increase up to 0.43 for PremPLI (p-value < 0.05, t-test) but decrease for all other methods (Fig. 4b). Overall, Rosetta, FEP+, and PremPLI alternately present the best performance on different types of ligands, so it is necessary to develop new computational methods to complement each other.

In summary, Rosetta performs well on S144 dataset—with strong correlation, low absolute errors, and good classification performance, while it performs poorly on S99 dataset. MD

calculations with a nonequilibrium free energy calculation protocol perform robustly on different data with moderate prediction accuracy. However, for Rosetta or MD calculations, different scoring functions or force fields will present different prediction accuracies, and here we only chose and provided the results of the best performing protocols, so it would be an issue for applying them to unseen data. In addition, both approaches lose the ability to predict non-interface mutations and require heavy computational resources, especially for MD calculations. However, PremPLI presents similar predictive accuracy with the best performing Rosetta and MD calculation protocols while requiring much less computational resources. Therefore, a potentially integrated protocol might be PremPLI calculations for an initial large-scale mutational scan, followed by refinement of the most promising results via the combination of free energy calculations by Rosetta or MD simulations, which would allow for higher predictive power and provide dynamic insight into the impact of mutations.

Online Webserver

The 3D structure of a protein-ligand complex is required by the webserver, and the user could input Protein Data Bank (PDB) code or upload the coordinate file in the standard PDB format. Biological assemblies or asymmetric unit can be chosen and retrieved from the Protein Data Bank when the user inputs the PDB code (Fig. 5). Next, the server will display a 3D view colored by protein chains and types of ligands and provide the corresponding protein and ligand names (Fig. 5). Given that PremPLI only calculates the impact of mutations on the interaction between a protein and one type of ligand, one or multiple chains that must belong to one protein and one or multiple ligands that must belong to one type can be assigned to the following calculation. The last step is to select mutations and three options are provided: “Specify One or More Mutations”, in this option, users can not only perform calculations for the specified single mutations but also be allowed to view the mutated residue in the complex 3D structure; “Upload Mutation List”, allows users to upload a file including a list of mutations for performing large-scale mutational scans; “Alanine Scanning”, allows users to perform alanine scanning for the selected protein chain (Fig. 5).

For each mutation, the PremPLI server provides $\Delta\Delta G$ (kcal mol⁻¹), predicted binding affinity change induced by this mutation (positive and negative sign corresponds to mutations decreasing and increasing affinity, respectively), and location of the mutation (interface/non-interface).

Code availability. Source code of PremPLI can be found at the following URL:
<https://github.com/minghuilab/PremPLI>.

Data availability. All relevant data are publicly available: compiled experimental datasets and computational results that support our findings can be found at GitHub by following the URL:
<https://github.com/minghuilab/PremPLI>.

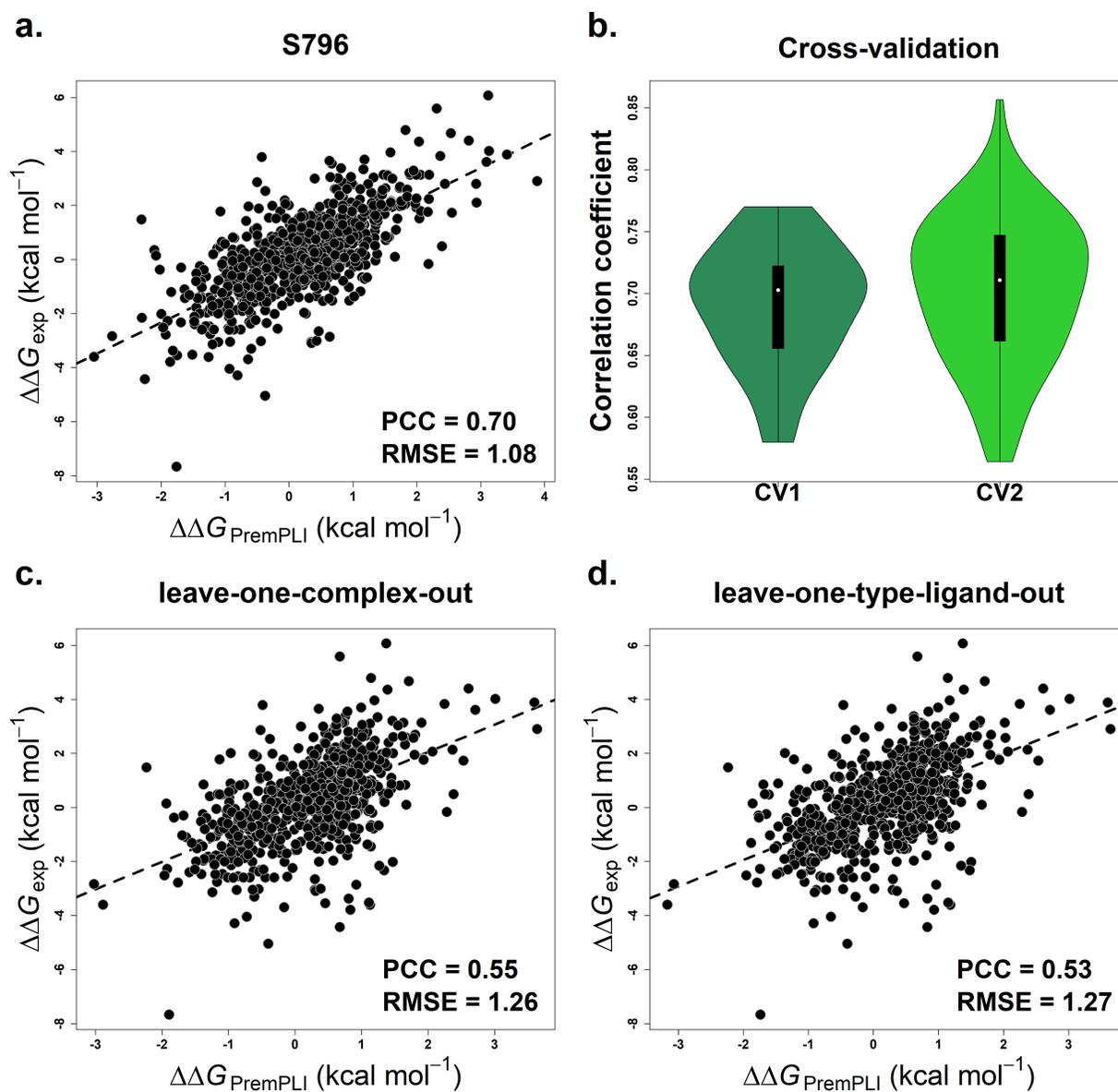
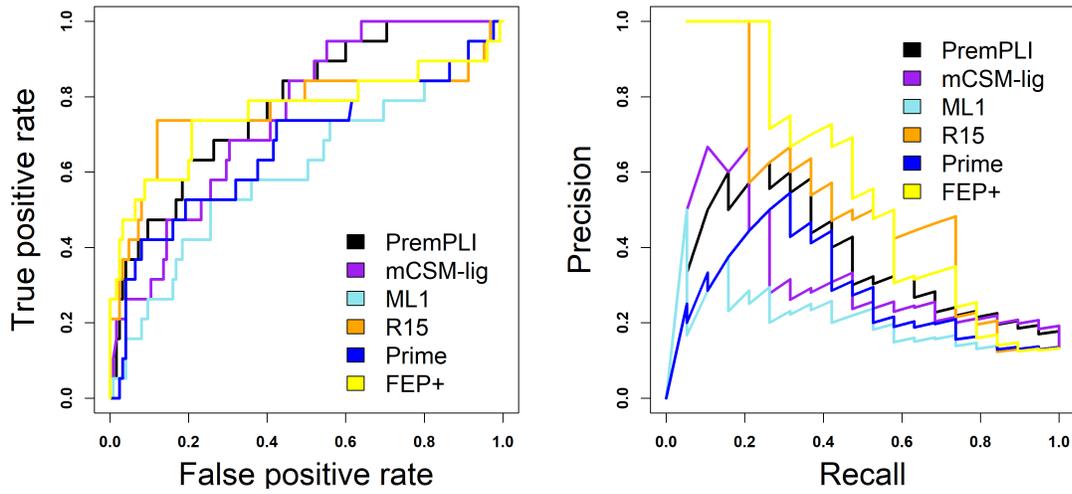


Figure 1. Pearson correlation coefficients between experimental and calculated changes in binding affinity for PremPLI trained and tested on S796 dataset (a), performing 10 times 5-fold and 10-fold cross-validations (CV1 and CV2) (b), performing leave-one-complex-out validation (CV3) (c) and leave-one-type-ligand-out validation (CV4) (d), respectively. PCC: Pearson correlation coefficient; RMSE (kcal mol⁻¹): root-mean-square error.

a. S144



b. S99

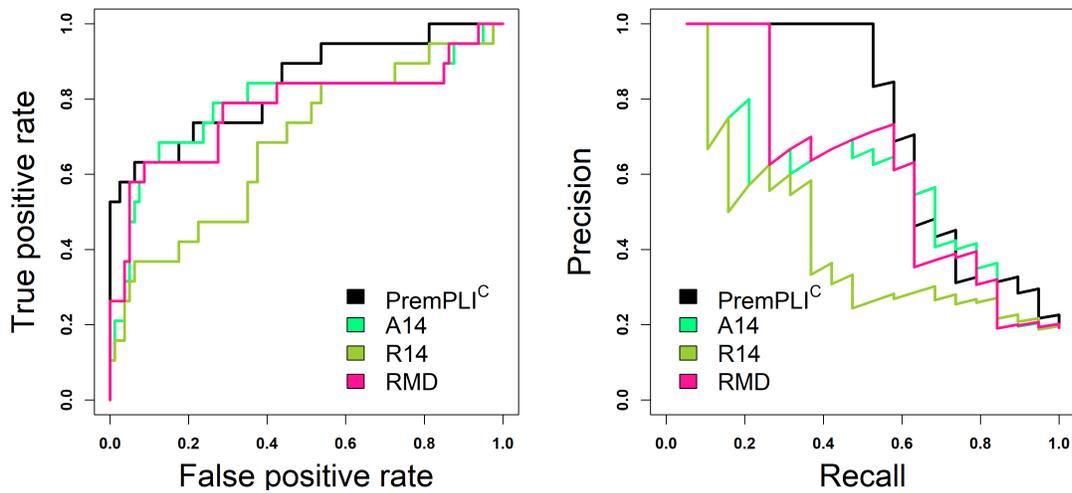


Figure 2. Receiver Operating Characteristics (ROC) and Precision recall (PR) curves for different methods to distinguish resistance from other mutations. The number of resistance mutations is 19 for both S144 and S99 datasets.

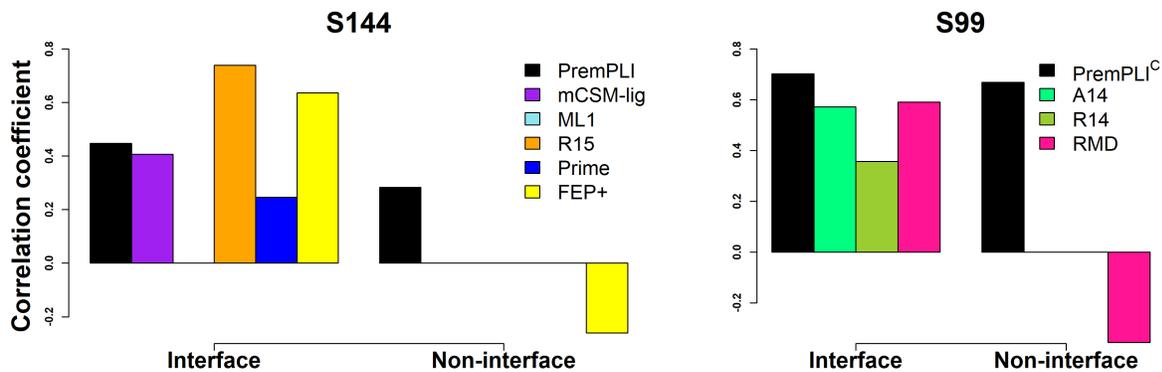
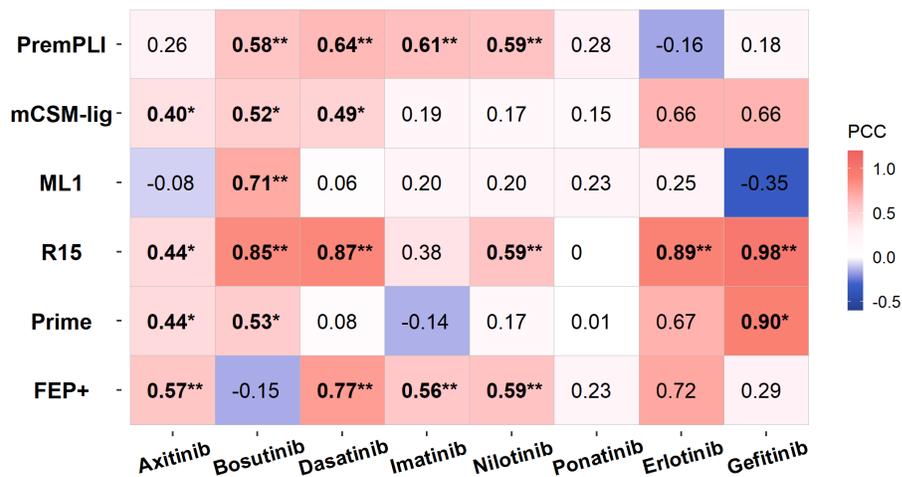


Figure 3. Pearson correlation coefficients between experimental and calculated $\Delta\Delta G$ values for interface and non-interface mutations. Only correlation coefficients statistically significantly different from zero are shown (p-value < 0.05, t-test)

a. S144



b. Abl-axitinib complex

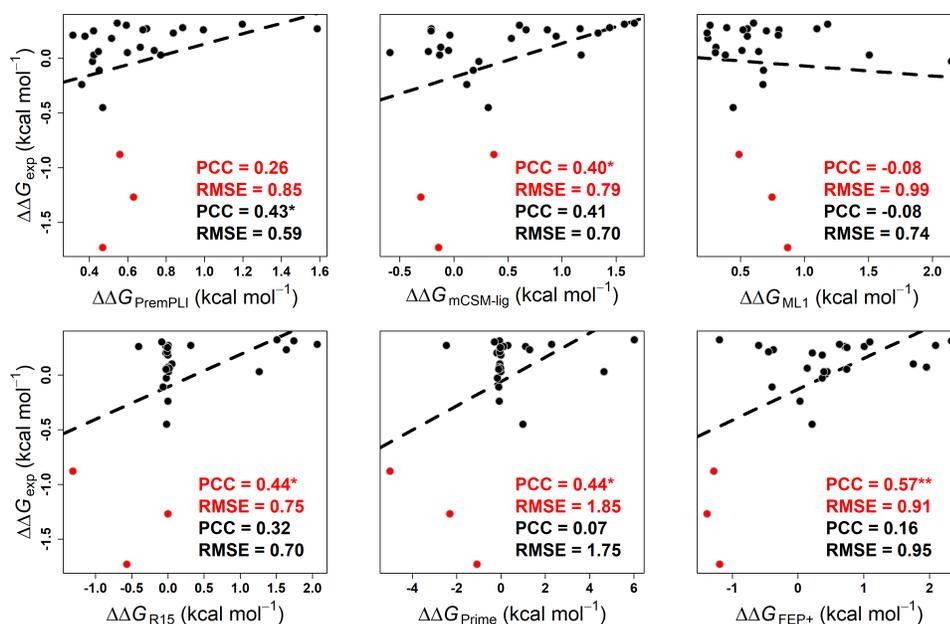


Figure 4. (a) Pearson correlation coefficient for each tyrosine kinase inhibitor. * and ** indicate statistically significant difference from zero in terms of PCC with p-value < 0.05 and p-value < 0.01 , respectively. (b) Performance for different methods tested on Abl-axitinib complex. PCC and RMSE in red: all 26 mutations; PCC and RMSE in black: mutations with three red dots removed.

PremPLI Method Help Results Download Contact

PremPLI - Predicting the Effects of Mutations on Protein-Ligand Interactions

$\Delta\Delta G_{bind} = \Delta G_{bind}^{mut} - \Delta G_{bind}^{wt}$
Arg → Ala

PremPLI predicts the effects of single mutations on protein-ligand interactions by calculating the binding affinity changes quantitatively. It can be used for guiding the design of ligand-binding proteins, identifying and understanding disease driver mutations, and finding potential resistance mutations for different drugs. The 3D structure of a protein-ligand complex is required for performing the prediction.

Step 1 - Select Protein-Ligand Complex

Input PDB code:
Example: 1OSS

Bioassembly 1st
Asymmetric Unit

Upload PDB file: No file selected
Format description for uploaded file

School of Biology & Basic Medical Sciences, Soochow University
199 Ren-Ai Road, Suzhou, Jiangsu, 215123 PR, China

Step 2 - Select Partners of Interaction

PDB id: 1OSS

Protein Chains
TRYPSIN A

Ligand Positions
BEN BENZAMIDINE A 246

Protein: TRYPSIN
Chains: **A**

Ligand: BEN (BENZAMIDINE)
Positions: **A 246**

Step 3 - Select Mutations

PDB id: 1OSS

Protein
Chain A: Trypsin

Ligand
BEN (BENZAMIDINE) A 246

Specify One or More Mutations: Example: Chain A V16A

Chain to Mutate	Residue	Mutant Residue	View in Structure
Chain A	Q 24 (GLN)	E (GLU)	<input type="button" value="View"/>
Chain A	K 82 (LYS)	H (HIS)	<input type="button" value="View"/>
Chain A	P 190 (PRO)	S (SER)	<input type="button" value="View"/>
Chain A	P 190 (PRO)	V (VAL)	<input type="button" value="View"/>

Add or Remove Mutations

School of Biology & Basic Medical Sciences, Soochow University
199 Ren-Ai Road, Suzhou, Jiangsu, 215123 PR, China

Job id: 2021031002563493343426069

Summary

PDB ID	Protein	Ligand	Start time (EST)	Processing time	Results
1OSS	A	BEN: A 246	2021-03-10 02:56:38	8 min	Download

Results

#	Mutated Chain	Mutation	$\Delta\Delta G$	Interface?
1	A	Q24E	1.27	No
2	A	K82H	0.59	No
3	A	P190S	-0.75	Yes
4	A	P190V	1.01	Yes

Figure 5. PremPLI server. Three steps and results pages are provided. “Processing time” refers to the running time of a job without counting the waiting time in the queue.

Table 1. Comparison of methods' performances on the datasets of S144, S129 and S99.

Dataset	Method	PCC	RMSE	AUC-ROC	AUC-PR	MCC
S144	PremPLI	0.46	0.77	0.78	0.36	0.40
	mCSM-lig ¹⁷	0.41	0.91	0.75	0.31	0.33
	ML1 ¹⁹	0.12 ^{a**}	0.87	0.61*	0.20	0.20
	R15 ¹⁹	0.67**	0.72	0.77	0.50	0.52
	Prime ²⁵	0.29	1.81	0.67	0.27	0.35
	FEP ²⁵	0.49	1.07	0.76	0.53	0.52
S129	PremPLI	0.48	0.78	0.81	0.35	0.39
	mCSM-lig ¹⁷	0.20 ^{b*}	1.06	0.55*	0.28	0.40
S99	PremPLI ^C	0.69	1.09	0.84	0.71	0.69
	A14 ²⁷	0.44**	1.35	0.78	0.55	0.54
	R14 ²⁷	0.33**	1.35	0.68	0.40	0.37
	RMD ²⁷	0.48*	1.23	0.77	0.59	0.58

All correlation coefficients are statistically significantly different from zero (p-value < 0.01, t-test) except ^ap-value = 0.14 and ^bp-value = 0.02. * and ** indicate statistically significant difference between PremPLI and other methods in terms of PCC (Hitter2003 test) and AUC-ROC (DeLong test) with p-value < 0.05 and p-value < 0.01, respectively. PremPLI^C: PremPLI was retrained after removing all mutations in the overlapped complexes with S99 from the training dataset. R15: Rosetta using the *flex_ddg* protocol and REF2015 scoring function. R14: Rosetta using the *flex_ddg* protocol and talaris2014 scoring function. A14: Amber14sb and GAFF(v2.1)/AM1-BCC force fields were used for proteins and ligands, respectively. RMD: the combination of R14 and A14. The results of more combinations are shown in Figure S6.

References

- (1) Beato, M.; Chávez, S.; Truss, M. Transcriptional regulation by steroid hormones. *Steroids* **1996**, *61* (4), 240.
- (2) Ronnett, G. V.; Moon, C. G proteins and olfactory signal transduction. *Annual review of physiology* **2002**, *64*, 189.
- (3) Missale, C.; Nash, S. R.; Robinson, S. W.; Jaber, M.; Caron, M. G. Dopamine receptors: from structure to function. *Physiological reviews* **1998**, *78* (1), 189.
- (4) Nemethova, M.; Radvanszky, J.; Kadasi, L.; Ascher, D. B.; Pires, D. E. V.; Blundell, T. L.; Porfirio, B.; Mannoni, A.; Santucci, A.; Milucci, L. et al. Twelve novel HGD gene variants identified in 99 alkaptonuria patients: focus on 'black bone disease' in Italy. *Eur J Hum Genet* **2016**, *24* (1), 66.
- (5) Kim, P.; Zhao, J.; Lu, P.; Zhao, Z. mutLBSgeneDB: mutated ligand binding site gene DataBase. *Nucleic acids research* **2017**, *45* (D1), D256.
- (6) Di Lullo, G. A.; Sweeney, S. M.; Korkko, J.; Ala-Kokko, L.; San Antonio, J. D. Mapping the ligand-binding sites and disease-associated mutations on the most abundant protein in the human, type I collagen. *The Journal of biological chemistry* **2002**, *277* (6), 4223.
- (7) Jääskeläinen, J.; Deeb, A.; Schwabe, J. W.; Mongan, N. P.; Martin, H.; Hughes, I. A. Human androgen receptor gene ligand-binding-domain mutations leading to disrupted interaction between the N- and C-terminal domains. *Journal of Molecular Endocrinology* **2006**, *36* (2), 361.
- (8) Holohan, C.; Van Schaeybroeck, S.; Longley, D. B.; Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nature reviews. Cancer* **2013**, *13* (10), 714.
- (9) Crystal, A. S.; Shaw, A. T.; Sequist, L. V.; Fribolet, L.; Niederst, M. J.; Lockerman, E. L.; Frias, R. L.; Gainor, J. F.; Amzallag, A.; Greninger, P. et al. Patient-derived models of acquired resistance can identify effective drug combinations for cancer. *Science* **2014**, *346* (6216), 1480.
- (10) Housman, G.; Byler, S.; Heerboth, S.; Lapinska, K.; Longacre, M.; Snyder, N.; Sarkar, S. Drug resistance in cancer: an overview. *Cancers* **2014**, *6* (3), 1769.
- (11) Carter, T. A.; Wodicka, L. M.; Shah, N. P.; Velasco, A. M.; Fabian, M. A.; Treiber, D. K.; Milanov, Z. V.; Atteridge, C. E.; Biggs, W. H., 3rd; Edeen, P. T. et al. Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (31), 11011.
- (12) Juchum, M.; Günther, M.; Laufer, S. A. Fighting cancer drug resistance: Opportunities and challenges for mutation-specific EGFR inhibitors. *Drug resistance updates: reviews and commentaries in antimicrobial and anticancer chemotherapy* **2015**, *20*, 12.
- (13) Wensing, A. M.; Calvez, V.; Günthard, H. F.; Johnson, V. A.; Paredes, R.; Pillay, D.; Shafer, R. W.; Richman, D. D. 2017 Update of the Drug Resistance Mutations in HIV-1. *Topics in antiviral medicine* **2016**, *24* (4), 132.
- (14) Angus, L.; Beijs, N.; Jager, A.; Martens, J. W. M.; Sleijfer, S. ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients. *Cancer Treatment Reviews* **2017**, *52*, 33.
- (15) Fanning, S. W.; Mayne, C. G.; Dharmarajan, V.; Carlson, K. E.; Martin, T. A.; Novick, S. J.; Toy, W.; Green, B.; Pancharukhi, S.; Katzenellenbogen, B. S. et al. Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *eLife* **2016**, *5*.

- (16) Ghai, R.; Falconer, R. J.; Collins, B. M. Applications of isothermal titration calorimetry in pure and applied research--survey of the literature from 2010. *Journal of Molecular Recognition* **2012**, *25* (1), 32.
- (17) Pires, D. E.; Blundell, T. L.; Ascher, D. B. mCSM-lig: quantifying the effects of mutations on protein-small molecule affinity in genetic disease and emergence of drug resistance. *Scientific reports* **2016**, *6*, 29575.
- (18) Wang, D. D.; Ou-Yang, L.; Xie, H.; Zhu, M.; Yan, H. Predicting the impacts of mutations on protein-ligand binding affinity based on molecular dynamics simulations and machine learning methods. *Computational and structural biotechnology journal* **2020**, *18*, 439.
- (19) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Predicting Kinase Inhibitor Resistance: Physics-Based and Data-Driven Approaches. *ACS central science* **2019**, *5* (8), 1468.
- (20) Alford, R. F.; Leaver-Fay, A.; Jeliaskov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K. et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of chemical theory and computation* **2017**, *13* (6), 3031.
- (21) Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L. et al. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **2013**, *501* (7466), 212.
- (22) Röthlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O. et al. Kemp elimination catalysts by computational enzyme design. *Nature* **2008**, *453* (7192), 190.
- (23) Wright, D. W.; Hall, B. A.; Kenway, O. A.; Jha, S.; Coveney, P. V. Computing Clinically Relevant Binding Free Energies of HIV-1 Protease Inhibitors. *Journal of chemical theory and computation* **2014**, *10* (3), 1228.
- (24) Rapp, C.; Kalyanaraman, C.; Schiffmiller, A.; Schoenbrun, E. L.; Jacobson, M. P. A molecular mechanics approach to modeling protein-ligand interactions: relative binding affinities in congeneric series. *Journal of chemical information and modeling* **2011**, *51* (9), 2082.
- (25) Hauser, K.; Negron, C.; Albanese, S. K.; Ray, S.; Steinbrecher, T.; Abel, R.; Chodera, J. D.; Wang, L. Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Communications biology* **2018**, *1*, 70.
- (26) Ono, F.; Chiba, S.; Isaka, Y.; Matsumoto, S.; Ma, B.; Katayama, R.; Araki, M.; Okuno, Y. Improvement in predicting drug sensitivity changes associated with protein mutations using a molecular dynamics based alchemical mutation method. *Scientific reports* **2020**, *10* (1), 2161.
- (27) Aldeghi, M.; Gapsys, V.; de Groot, B. L. Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. *ACS central science* **2018**, *4* (12), 1708.
- (28) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J. et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *Journal of the American Chemical Society* **2015**, *137* (7), 2695.
- (29) Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *Journal of chemical information and modeling* **2015**, *55* (11), 2411.

- (30) Chen, Y.; Lu, H.; Zhang, N.; Zhu, Z.; Wang, S.; Li, M. PremPS: Predicting the impact of missense mutations on protein stability. *PLoS computational biology* **2020**, *16* (12), e1008543.
- (31) Getov, I.; Petukh, M.; Alexov, E. SAAFEC: Predicting the Effect of Single Point Mutations on Protein Folding Free Energy Using a Knowledge-Modified MM/PBSA Approach. *International journal of molecular sciences* **2016**, *17* (4), 512.
- (32) Dehouck, Y.; Grosfils, A.; Folch, B.; Gilis, D.; Bogaerts, P.; Rooman, M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics* **2009**, *25* (19), 2537.
- (33) Pires, D. E.; Ascher, D. B.; Blundell, T. L. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **2014**, *30* (3), 335.
- (34) Guerois, R.; Nielsen, J. E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology* **2002**, *320* (2), 369.
- (35) Savojardo, C.; Fariselli, P.; Martelli, P. L.; Casadio, R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. *Bioinformatics* **2016**, *32* (16), 2542.
- (36) Li, G.; Panday, S. K.; Alexov, E. SAAFEC-SEQ: A Sequence-Based Method for Predicting the Effect of Single Point Mutations on Protein Thermodynamic Stability. *International journal of molecular sciences* **2021**, *22* (2), 606.
- (37) Li, M.; Petukh, M.; Alexov, E.; Panchenko, A. R. Predicting the Impact of Missense Mutations on Protein-Protein Binding Affinity. *Journal of chemical theory and computation* **2014**, *10* (4), 1770.
- (38) Petukh, M.; Li, M.; Alexov, E. Predicting Binding Free Energy Change Caused by Point Mutations with Knowledge-Modified MM/PBSA Method. *PLoS computational biology* **2015**, *11* (7), e1004276.
- (39) Li, M.; Simonetti, F. L.; Goncarenco, A.; Panchenko, A. R. MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. *Nucleic acids research* **2016**, *44* (W1), W494.
- (40) Zhang, N.; Chen, Y.; Lu, H.; Zhao, F.; Alvarez, R. V.; Goncarenco, A.; Panchenko, A. R.; Li, M. MutaBind2: Predicting the Impacts of Single and Multiple Mutations on Protein-Protein Interactions. *iScience* **2020**, *23* (3), 100939.
- (41) Li, G.; Pahari, S.; Murthy, A. K.; Liang, S.; Fragoza, R.; Yu, H.; Alexov, E. SAAMBE-SEQ: a sequence-based method for predicting mutation effect on protein-protein binding affinity. *Bioinformatics* **2020**, DOI:10.1093/bioinformatics/btaa761.
- (42) Dehouck, Y.; Kwasigroch, J. M.; Rooman, M.; Gilis, D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. *Nucleic acids research* **2013**, *41* (Web Server issue), W333.
- (43) Petukh, M.; Dai, L.; Alexov, E. SAAMBE: Webserver to Predict the Change of Binding Free Energy Caused by Amino Acids Mutations. *International journal of molecular sciences* **2016**, *17* (4), 547.
- (44) Xiong, P.; Zhang, C.; Zheng, W.; Zhang, Y. BindProfX: Assessing Mutation-Induced Binding Affinity Change by Protein Interface Profiles with Pseudo-Counts. *Journal of molecular biology* **2017**, *429* (3), 426.

- (45) Rodrigues, C. H. M.; Myung, Y.; Pires, D. E. V.; Ascher, D. B. mCSM-PPI2: predicting the effects of mutations on protein-protein interactions. *Nucleic acids research* **2019**, *47* (W1), W338.
- (46) Geng, C.; Vangone, A.; Folkers, G. E.; Xue, L. C.; Bonvin, A. iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations. *Proteins* **2019**, *87* (2), 110.
- (47) Pahari, S.; Li, G.; Murthy, A. K.; Liang, S.; Fragoza, R.; Yu, H.; Alexov, E. SAAMBE-3D: Predicting Effect of Mutations on Protein-Protein Interactions. *International journal of molecular sciences* **2020**, *21* (7), 2563.
- (48) Zhang, N.; Chen, Y.; Zhao, F.; Yang, Q.; Simonetti, F. L.; Li, M. PremPDI estimates and interprets the effects of missense mutations on protein-DNA interactions. *PLoS computational biology* **2018**, *14* (12), e1006615.
- (49) Zhang, N.; Lu, H.; Chen, Y.; Zhu, Z.; Yang, Q.; Wang, S.; Li, M. PremPRI: Predicting the Effects of Missense Mutations on Protein-RNA Interactions. *International journal of molecular sciences* **2020**, *21* (15), 5560.
- (50) Pires, D. E.; Blundell, T. L.; Ascher, D. B. Platinum: a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes. *Nucleic acids research* **2015**, *43* (Database issue), D387.
- (51) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic acids research* **2000**, *28* (1), 235.
- (52) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *Journal of molecular graphics* **1996**, *14* (1), 33.
- (53) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry, B* **1998**, *102* (18), 3586.
- (54) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **2004**, *25* (13), 1605.
- (55) Bhagwat, M.; Aravind, L. PSI-BLAST tutorial. *Methods in molecular biology (Clifton, N.J.)* **2007**, *395*, 177.
- (56) Choi, Y.; Sims, G. E.; Murphy, S.; Miller, J. R.; Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PloS one* **2012**, *7* (10), e46688.
- (57) Jubb, H. C.; Higuero, A. P.; Ochoa-Montano, B.; Pitt, W. R.; Ascher, D. B.; Blundell, T. L. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *Journal of molecular biology* **2017**, *429* (3), 365.
- (58) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge. *Journal of chemical information and modeling* **2007**, *47* (6), 2140.
- (59) Hou, Q.; Kwasigroch, J. M.; Rooman, M.; Pucci, F. SOLart: a structure-based method to predict protein solubility and aggregation. *Bioinformatics* **2020**, *36* (5), 1445.
- (60) Azarya-Sprinzak, E.; Naor, D.; Wolfson, H. J.; Nussinov, R. Interchanges of spatially neighbouring residues in structurally conserved environments. *Protein engineering* **1997**, *10* (10), 1109.

- (61) Risler, J. L.; Delorme, M. O.; Delacroix, H.; Henaut, A. Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new and efficient scoring matrix. *Journal of molecular biology* **1988**, *204* (4), 1019.
- (62) Hittner, J. B.; May, K.; Silver, N. C. A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of general psychology* **2003**, *130* (2), 149.
- (63) Edwards, A. W. F. R.A. Fischer, statistical methods for research workers, first edition (1925). *Landmark Writings in Western Mathematics 1640-1940* **2005**, DOI:10.1016/B978-044450871-3/50148-0, 856.
- (64) Diedenhofen, B.; Musch, J. cocor: a comprehensive solution for the statistical comparison of correlations. *PloS one* **2015**, *10* (3), e0121945.

Figures

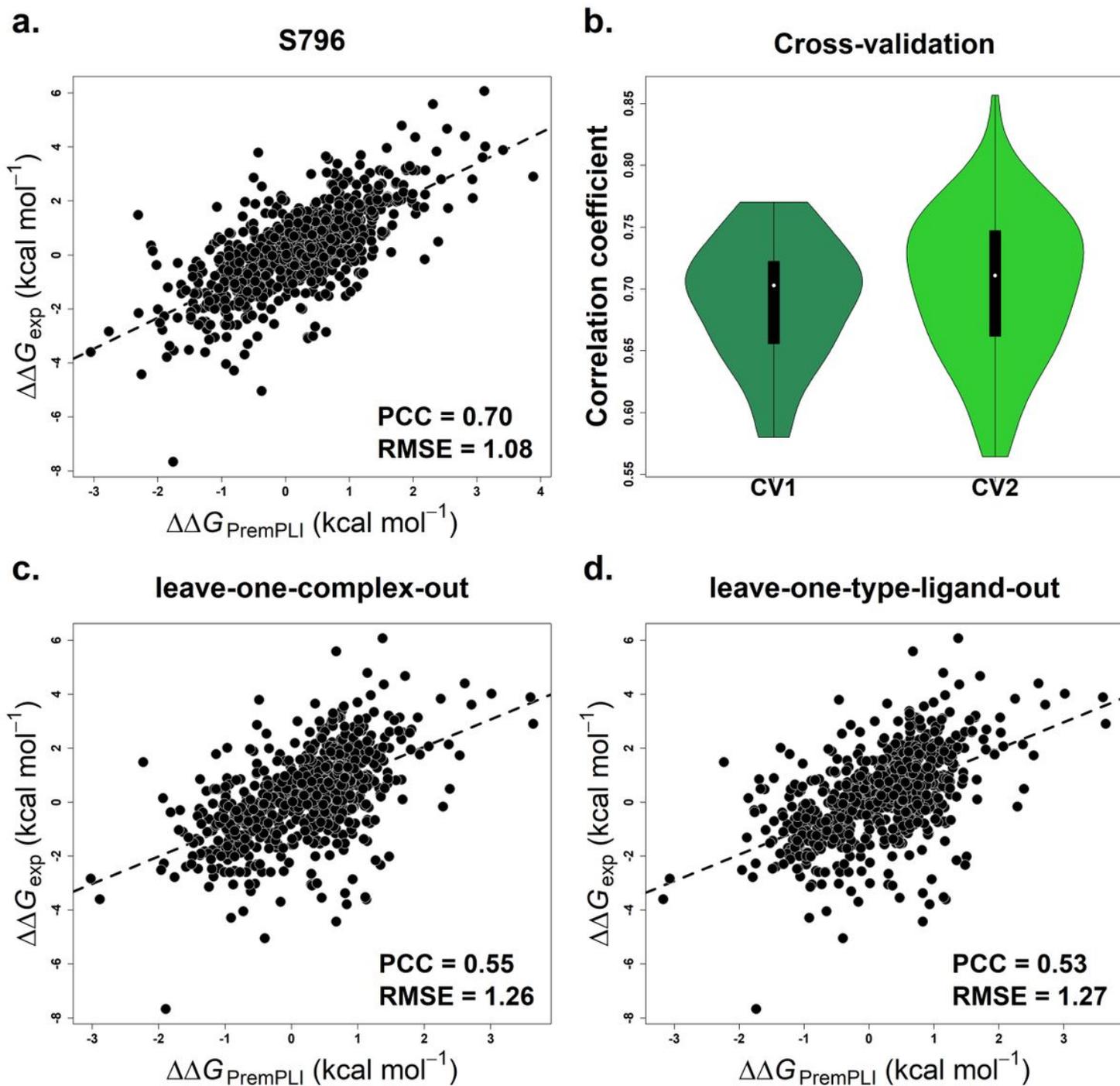
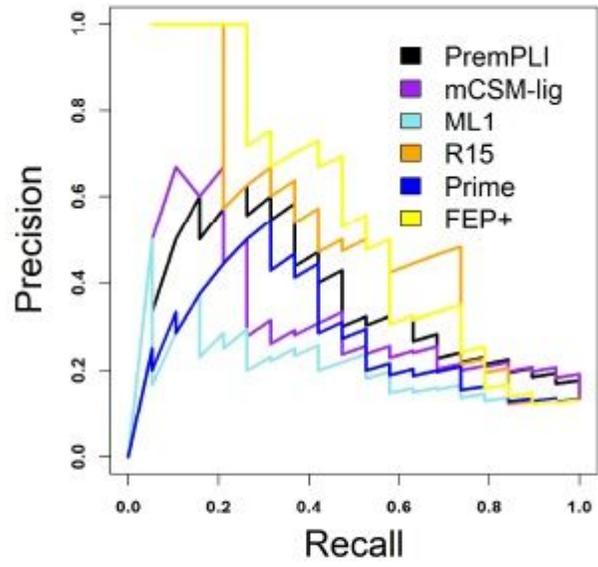
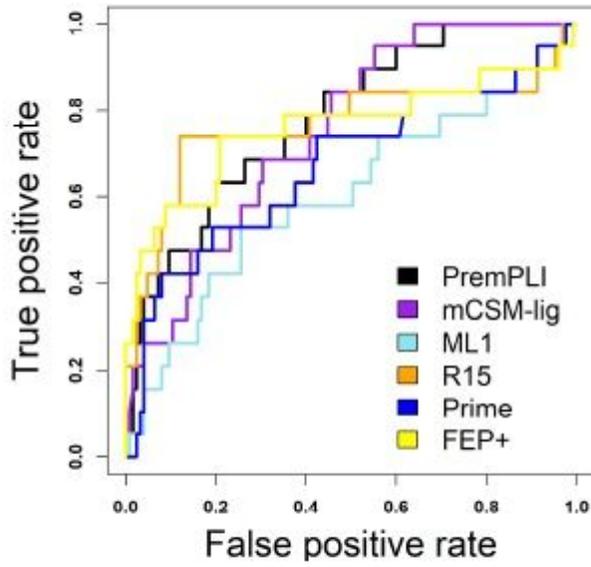


Figure 1

Pearson correlation coefficients between experimental and calculated changes in binding affinity for PremPLI trained and tested on S796 dataset (a), performing 10 times 5-fold and 10-fold cross-validations (CV1 and CV2) (b), performing leave-one-complex-out validation (CV3) (c) and leave-one-type-ligand-out validation (CV4) (d), respectively. PCC: Pearson correlation coefficient; RMSE (kcal mol⁻¹): root-mean-square error.

a. S144



b. S99

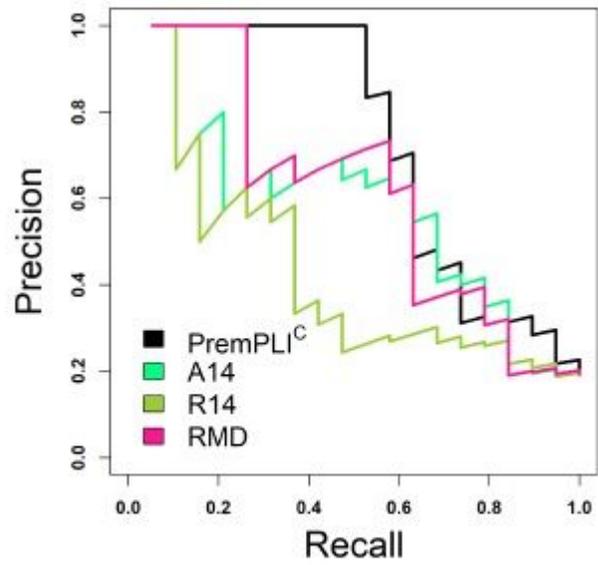
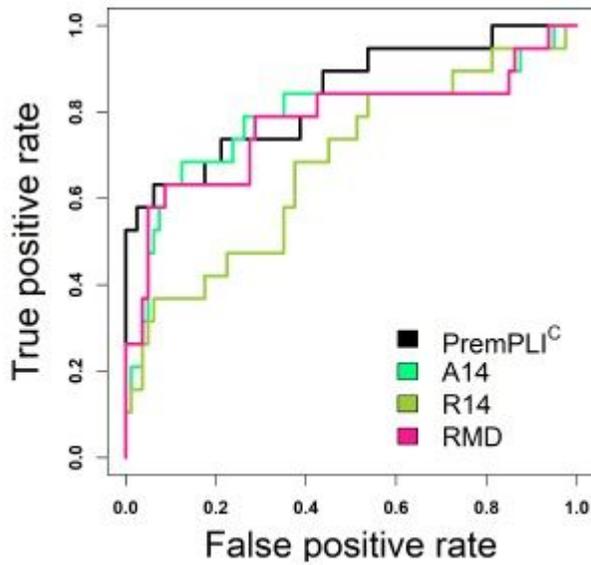


Figure 2

Receiver Operating Characteristics (ROC) and Precision recall (PR) curves for different methods to distinguish resistance from other mutations. The number of resistance mutations is 19 for both S144 and S99 datasets.

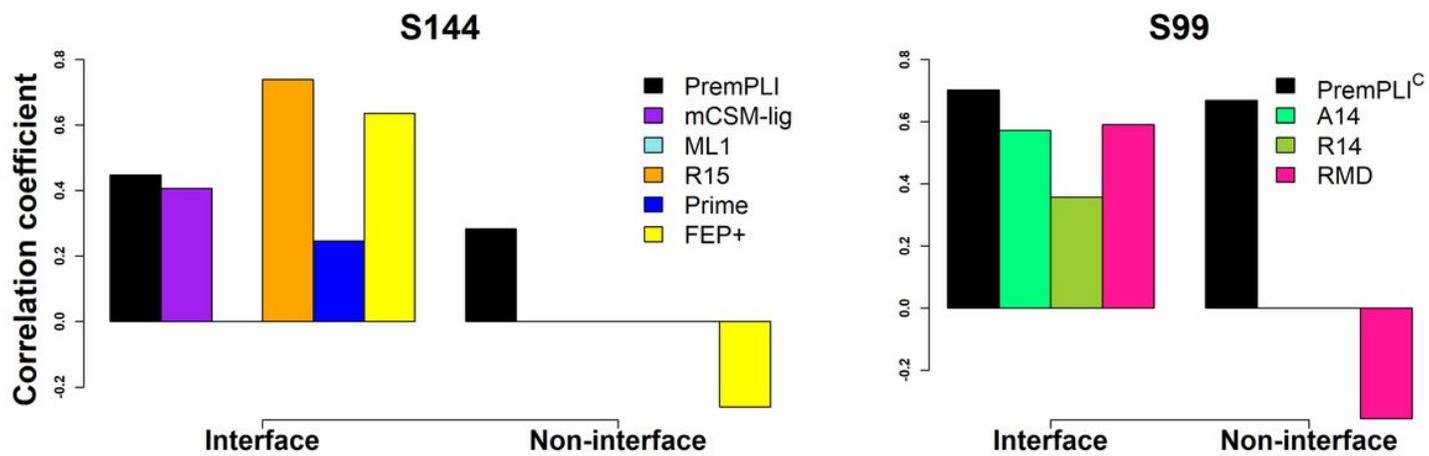
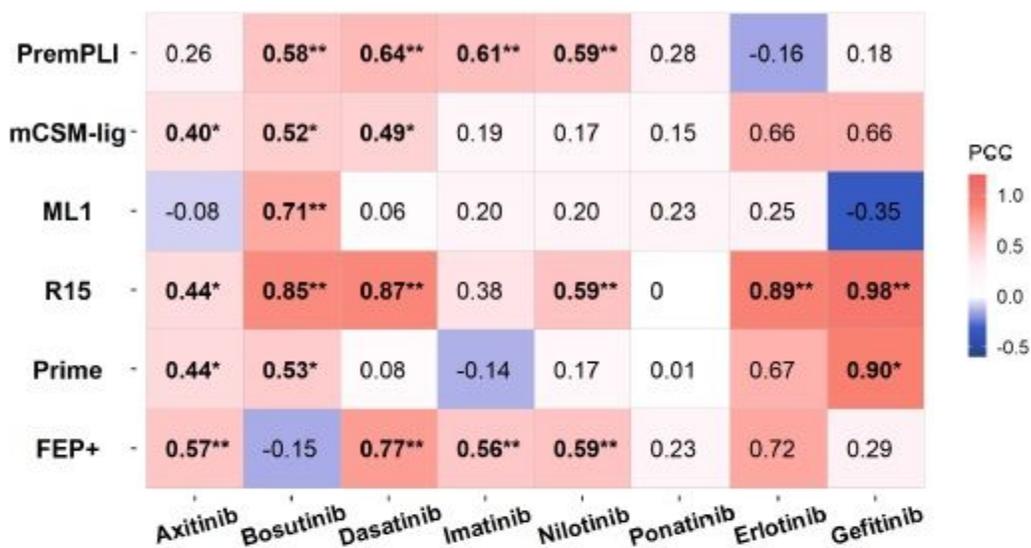


Figure 3

Pearson correlation coefficients between experimental and calculated $\Delta\Delta G$ values for interface and non-interface mutations. Only correlation coefficients statistically significantly different from zero are shown (p-value < 0.05, t-test)

a. S144



b. Abl-axitinib complex

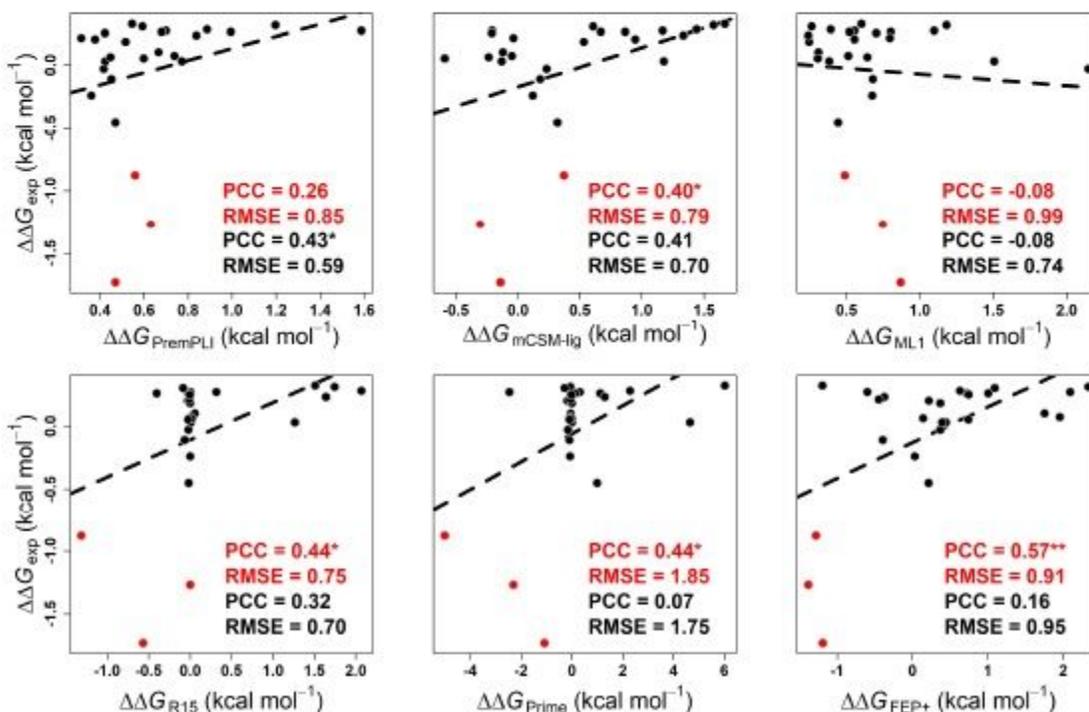


Figure 4

(a) Pearson correlation coefficient for each tyrosine kinase inhibitor. * and ** indicate statistically significant difference from zero in terms of PCC with p-value < 0.05 and p-value < 0.01, respectively. (b) Performance for different methods tested on Abl-axitinib complex. PCC and RMSE in red: all 26 mutations; PCC and RMSE in black: mutations with three red dots removed.

