

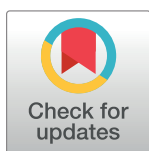
RESEARCH ARTICLE

# Predicting mutations deleterious to function in beta-lactamase TEM1 using MM-GBSA

Christopher Negron<sup>1</sup>, David A. Pearlman<sup>1</sup>, Guillermo del Angel<sup>2\*</sup>

**1** Schrödinger, New York, New York, United States of America, **2** Alexion Pharmaceuticals Inc., Boston, Massachusetts, United States of America

\* [Guillermo.delAngel@alexion.com](mailto:Guillermo.delAngel@alexion.com)



## Abstract

Missense mutations can have disastrous effects on the function of a protein. And as a result, they have been implicated in numerous diseases. However, the majority of missense variants only have a nominal impact on protein function. Thus, the ability to distinguish these two classes of missense mutations would greatly aid drug discovery efforts in target identification and validation as well as medical diagnosis. Monitoring the co-occurrence of a given missense mutation and a disease phenotype provides a pathway for classifying functionally disrupting missense mutations. But, the occurrence of a specific missense variant is often extremely rare making statistical links challenging to infer. In this study, we benchmark a physics-based approach for predicting changes in stability, MM-GBSA, and apply it to classifying mutations as functionally disrupting. A large and diverse dataset of 990 residue mutations in beta-lactamase TEM1 is used to assess performance as it is rich in both functionally disrupting mutations and functionally neutral/beneficial mutations. On this dataset, we compare the performance of MM-GBSA to alternative strategies for predicting functionally disrupting mutations. We observe that the MM-GBSA method obtains an area under the curve (AUC) of 0.75 on the entire dataset, outperforming all other predictors tested. More importantly, MM-GBSA's performance is robust to various divisions of the dataset, speaking to the generality of the approach. Though there is one notable exception: Mutations on the surface of the protein are the mutations that are the most difficult to classify as functionally disrupting for all methods tested. This is likely due to the many mechanisms available to surface mutations to disrupt function, and thus provides a direction of focus for future studies.

## OPEN ACCESS

**Citation:** Negron C, Pearlman DA, del Angel G (2019) Predicting mutations deleterious to function in beta-lactamase TEM1 using MM-GBSA. PLoS ONE 14(3): e0214015. <https://doi.org/10.1371/journal.pone.0214015>

**Editor:** Freddie Salsbury, Jr, Wake Forest University, UNITED STATES

**Received:** November 28, 2018

**Accepted:** March 5, 2019

**Published:** March 19, 2019

**Copyright:** © 2019 Negron et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** Schrödinger, Inc. and Alexion Pharmaceuticals funded the study. Schrödinger, Inc. provided support in the form of salary for authors CN and DP; Alexion Pharmaceuticals provided support in the form of salary for author GdA, but did not have any additional role in the study design, data collection, and analysis, decision to publish, or preparation of the

## Introduction

DNA sequencing has seen tremendous advances since the human genome was first sequenced in 2001 [1,2]. DNA sequencing cost is a great indicator of this progress as it has decreased by six orders of magnitude since the first sequenced human genome [3]. This in turn has facilitated the generation of large datasets of genetic variation [4–6]. Of particular note are the ExAC and gnomAD databases, which contain the full human exome (protein-coding region) for over 120,000 individuals [4]. However, despite these large collections of genetic variants,



manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing interests:** The authors have the following interests. Schrödinger, Inc. funded this study, provided software and employed the authors, Christopher Negron, and David A. Pearlman. Additionally, Alexion Pharmaceuticals co-funded this study and is the employer of Guillermo del Angel. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

predicting the functional effect of a genetic variant remains a significant challenge for missense mutations [7]. Missense mutations refer to a single-amino acid change to a gene.

Missense variants play a critical role in medical diagnosis and discovery since certain missense variants have been strongly correlated with disease susceptibility in several human cancers [8,9]. The most noteworthy examples are missense variants in tumor suppressor genes BRCA1 and BRCA2 [8,10]. However, the challenge of utilizing missense variants for diagnosis is that many missense variants are not deleterious to function, appearing in the human exome at 7.9% of sites [4]. Additionally, the rarity of any one specific missense variant makes drawing a statistical link between function and missense variant challenging [11,12]. Furthermore, rare diseases contain another barrier to obtaining significant statistical links due to the rare occurrence of the disease itself. Lastly, further complicating the matter is the fact that experimental assays to assess how a specific missense variant impacts protein function only covers a limited space [8]—thereby further complicating experimental validation.

Several algorithms have been developed for predicting the functional outcome of a missense variant [7,13–15]. Many of these algorithms rely on generating a multiple sequence alignment (MSA) using the protein of interest, along with closely related sequences, typically from other species. The fundamental premise of this analysis is that conserved residues in the MSA are less tolerant to mutation. Some of the more robust tools in this space also incorporate structural information such as accessible surface area, B-factor, and hydrophobic propensity. Despite some success, these tools have shown significant sensitivity to the input sequence alignment, and thus performance varies across different systems [16].

In this work, we seek to benchmark a purely physics-based approach that is independent of sequence alignment known as MM-GBSA [17,18]. MM-GBSA predicts the relative change in folding free energy of an amino acid mutation using an implicit solvent model combined with a molecular mechanics energy function. MM-GBSA indirectly captures mutations that disrupt function by detecting mutations that significantly disrupt the stability of a protein. The link between protein stability and function has long been known and explored [19]. A central postulate arising from this field of analysis has been that if a protein is significantly destabilized, this will result in the protein misfolding, and consequently the mutation will disrupt protein function.

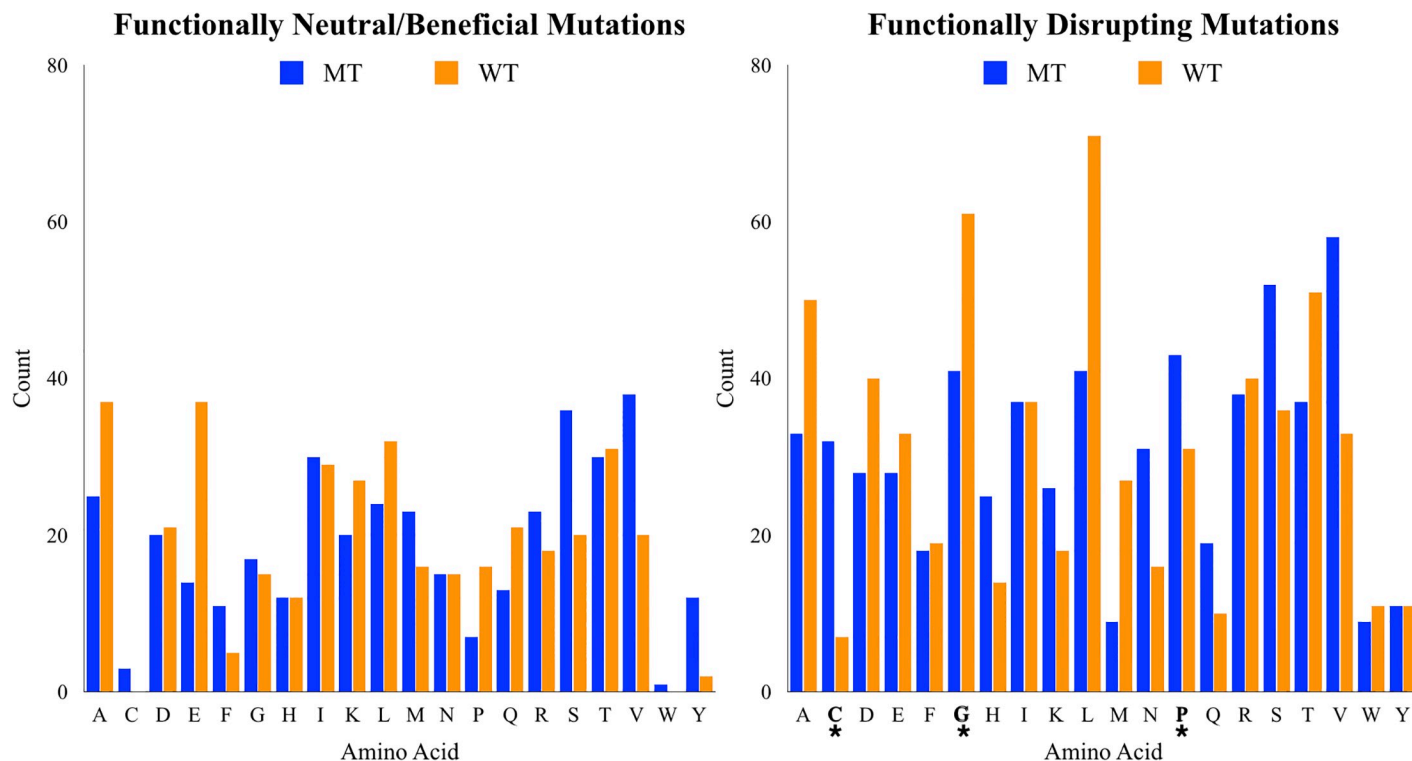
## Results

### Beta lactamase dataset

Datasets play an important role in any benchmark study. They need to be large enough to get meaningful statistics, and diverse enough to avoid bias in the data. Thus, we turned to a large scale mutational study of beta lactamase TEM-1 [20]. Beta lactamase TEM-1 is a bacterial enzyme responsible for hydrolyzing beta-lactam bonds that occur in beta-lactam antibiotics such as penicillin or amoxicillin. TEM-1 is well characterized both structurally and biochemically. And most importantly, the aforementioned study provides hundreds of examples of mutations that are both deleterious and neutral/beneficial to function. This in turn helps to mitigate bias in our study, and is the primary reason why we selected this system over human benchmarks, which lack substantial information on functionally neutral/beneficial mutations.

The beta lactamase TEM-1 dataset used in this study consists of 990 single-point mutations. Details of the functional assay performed on beta lactamase TEM-1 can be found in Jacquier *et al.*. Briefly, for each of the 990 mutations, cell growth was monitored at several concentrations of amoxicillin and the minimum concentration of amoxicillin required to inhibit cell growth was recorded, herein referred to as the minimum inhibitory concentration (MIC). The MIC measurement was converted to a MIC score via the following equation  $\log_2(\text{MIC}/500)$ .





**Fig 1. Frequency of appearance of amino acids in beta-lactamase dataset.** The amino acid distribution is plotted for the functionally neutral/beneficial mutations (left) and the functionally disrupting mutations (right). The blue bars in the plot show the number of mutations from that wildtype amino acid. And the orange bars show the number of mutations to that amino acid. Functionally disrupting mutations that occur a statistical significant number of times ( $p < 0.01$ ) are show in bold, with an asterisk under the amino acid.

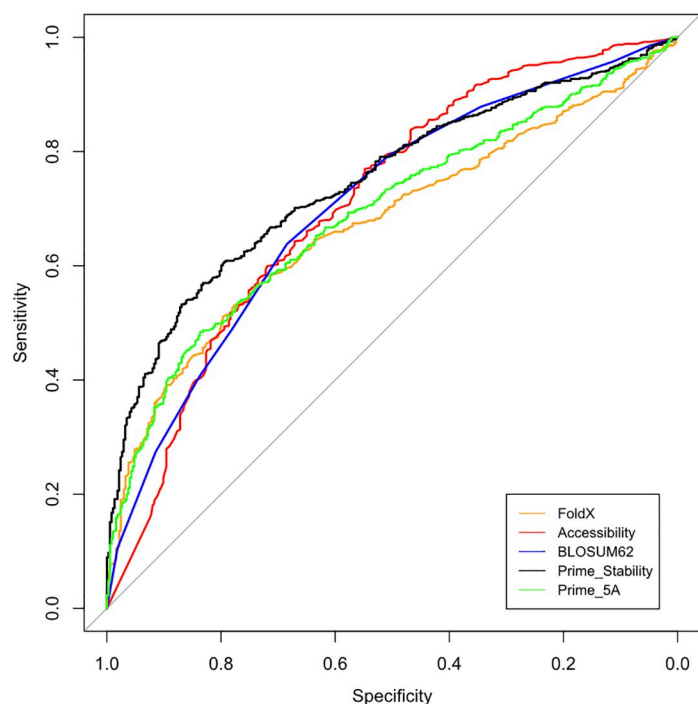
<https://doi.org/10.1371/journal.pone.0214015.g001>

500 mg/L refers to the MIC measurement of the wildtype sequence. As a result, a MIC score of 0 would represent activity on par with the wildtype enzyme. MIC scores less than 0 are classified as deleterious in this paper. And MIC scores of 0 or greater are classified as neutral/beneficial.

The amino acid distribution of the functionally deleterious mutations and functionally neutral/beneficial mutations are shown in Fig 1. One benefit of this binary classification scheme is that it allows us to compare the observed frequency of each amino acid in the deleterious set to the mutation probabilities that would be expected by random selection from a binomial distribution. This makes it easy to identify mutation types that are significantly enriched in the functionally disrupting set.

Interestingly, three types of mutations appeared significantly ( $p < 0.01$ ) more in the deleterious set than expected by random chance (Fig 1). These are mutations to cysteine, glycine, and proline, with cysteine being the most prominent enriched. Cysteine appears in the deleterious set 32 times while only appearing in the neutral/beneficial set 3 times. That is greater than a 10-fold enrichment in the deleterious set. Enrichment of cysteines in the deleterious set is likely tied to cysteine's ability to participate in disulfide bonds that prevent proper protein folding. Prolines appear ~6.1 times more frequently in the deleterious set than in the neutral/beneficial set. The enrichment of proline in the deleterious set may be tied to the fact that proline occupies a unique dihedral space among natural occurring amino acids which can result in a strained geometry when the proline mutation is inserted into a non-proline position. Lastly, mutations to glycine are observed 2.4 times more frequently in the deleterious set versus





Method	AUC	95% CI
<b>FoldX</b>	0.67	0.64-0.71
<b>Accessibility</b>	0.71	0.68-0.74
<b>BLOSUM62</b>	0.71	0.68-0.74
<b>Prime Stability</b>	0.75	0.72-0.78
<b>Prime Stability (5Å)</b>	0.69	0.66-0.72

**Fig 2. ROC curves for classifying functionally disrupting mutations on the beta lactamase dataset.** This dataset is composed of 990 single-point mutations with measured changes in their minimum inhibitory concentration (MIC). The five models include energy scores from FoldX, the solvent accessibility of a side chain, the score from the BLOSUM62 substitution matrix, the Prime Stability score, and the Prime stability score with a 5Å minimization cutoff. The area under the curve (AUC) for each ROC curve is shown on the right along with its 95% confidence interval (CI).

<https://doi.org/10.1371/journal.pone.0214015.g002>

the neutral/beneficial set. The small volume of the glycine side chain (a single hydrogen atom) likely creates large cavities in the core of a protein that in turn significantly destabilize the overall protein fold.

## Evaluating models

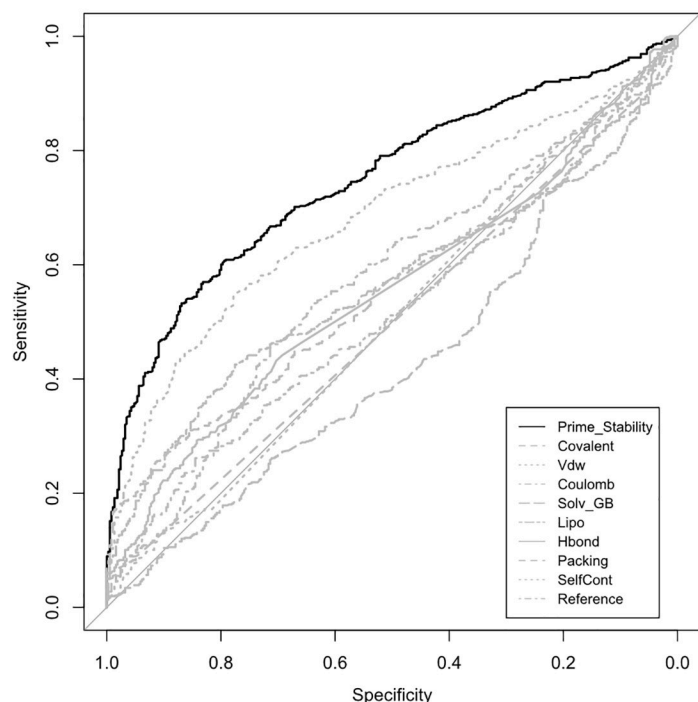
Looking beyond the statistical trends in the data, we sought to assess the performance of different models at classifying functionally deleterious amino acid mutations. To evaluate the models, we calculated a receiver operating characteristic (ROC) curve for each method. The area under the curve (AUC) is then measured to determine how well model separate deleterious mutations from the neutral/beneficial mutations. An AUC of 0.5 would represent a model that randomly classifies mutations as deleterious. An AUC of 1.0 would represent a model that perfectly agrees with the classification assigned by experiment.

Using AUC, we have benchmarked 5 different approaches for predicting mutations that are deleterious to function (Fig 2). The first approach is FoldX. FoldX is an empirical force field developed to predict changes in protein stability, and has been adapted for use in protein design [21,22]. FoldX obtains an AUC of 0.67 on this dataset, which is the lowest AUC among all the methods tested. Interestingly, a simple method such as the BLOSUM62 substitution matrix, a tool designed for generating sequence alignments, outperforms FoldX, obtaining an AUC of 0.71. This method works by predicting that the substitution of amino acids rarely observed in related protein families will result in a deleterious mutation. Another simple method, solvent accessibility, heretofore referred to as accessibility, also performs well on this dataset, with an AUC of 0.71. This method works by predicting that mutations at buried



positions are more prone to be deleterious than mutations at solvent exposed positions. Lastly, we have benchmarked Schrodinger's MM-GB/SA tool for predicting relative changes in protein stability, Prime [18]. Similar to FoldX, Prime predicts that deleterious mutations occur when significant increases in stability occur. Prime obtains the largest AUC, obtaining an AUC of 0.75. Lastly, we have benchmarked a variant of Prime that allowed greater side chain flexibility of neighboring residues by refining the side-chain of residues within 5 Å from the mutated side chain. Unexpectedly, this degrades Prime's performance, obtaining an AUC of only 0.69.

Additionally, we have examined how well each of the terms in the Prime Stability energy function perform individually. There are nine terms that are summed to calculate the Prime Stability score. The first term captures the energetics of covalent interactions (Covalent), such as bond angles and bond stretching. The second term models van der Waals interactions (VDW), which captures induced dipole interactions between atoms. The third term is a coulombic term (Coulomb), measuring electrostatic interactions. The fourth term is the generalized born term (Solv GB) which models the solvation and desolvation effects of amino acids. The fifth term is a measure of hydrophobic interactions with water, which is referred to as the Lipo term. The sixth term is a hydrogen bond term (Hbond). The seventh term is referred to as Packing and measures the quality of  $\pi$ - $\pi$  interactions. The eighth term is a measure of self-interaction (Self Cont) of a side chain. This term captures side-chain hydrogen bonds with backbone atoms. Residues such as asparagine, glutamine, and serine are examples of residues that engage in this type of interaction. Lastly, the ninth term represents the unfolding energy (Reference) for each amino acid. The total Prime Stability score outperforms all the individual components of the Prime energy function (Fig 3). Showing that the sum of all terms is a superior predictor of functionally disrupting mutations than any individual term.



Prime Term	AUC	95% CI
<b>Prime Stability</b>	0.75	0.72-0.78
<b>Covalent</b>	0.56	0.53-0.60
<b>Vdw</b>	0.68	0.65-0.71
<b>Coulomb</b>	0.57	0.54-0.61
<b>Solv GB</b>	0.44	0.40-0.47
<b>Lipo</b>	0.56	0.53-0.60
<b>Hbond</b>	0.55	0.52-0.59
<b>Packing</b>	0.50	0.48-0.52
<b>Self Cont</b>	0.51	0.47-0.54
<b>Reference</b>	0.52	0.48-0.55

**Fig 3. ROC curves for individual Prime terms on the beta lactamase dataset.** The ROC curve for the Prime Stability energy function (black) and individual terms of the Prime Stability energy function (gray). The table on the right list the corresponding area under the curve (AUC) values and the 95% confidence intervals for the AUC values.

<https://doi.org/10.1371/journal.pone.0214015.g003>



## Dividing the dataset

To better understand the limitations and strengths of the five methods presented in Fig 1, we attempted to divide the dataset using three approaches. The first approach divides the data set on the basis of whether a mutation conserves or changes the net charge on the protein. The second approach divides the data on the basis of whether a mutation substantially changes the volume occupied by the residue (big->small or small->big). We classify a volume change “significant” if it changes the volume of the sidechain by  $> =$  the volume of an alanine->valine mutation. And lastly, we separate mutations that occur on the surface of the protein from mutations that are buried in the protein, using the relative accessible surface area of the wild-type side chain to determine if it is solvent exposed [23].

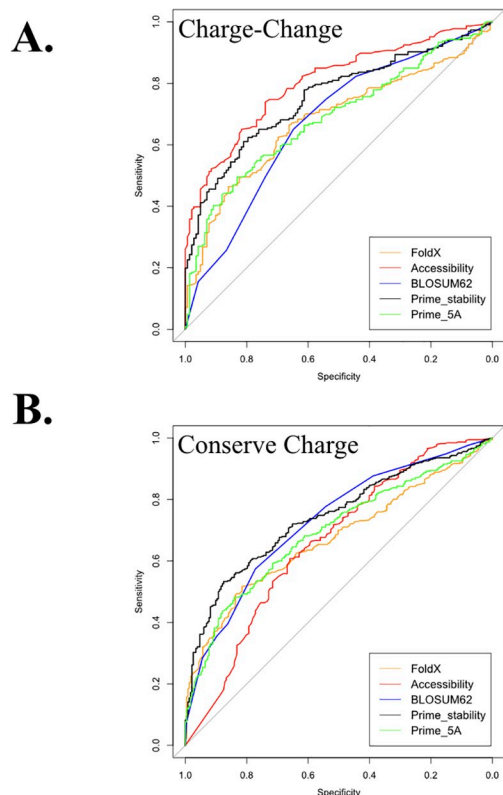
Mutations involving residues that are typically charged at neutral pH (D, E, R, K,) can be challenging for several reasons. For instance, depending on the protein environment, charged residues can occupy different protonation states, which can significantly alter their physiochemical properties. Additionally, positively charged residues, such as lysine and arginine, contain a larger number of rotatable bonds relative to other amino acids, making prediction of the side chain conformation for these positively charged residues challenging. Due to these same types of challenges, other physics-based algorithms have also been documented to struggle with predicting changes in stability for mutations that change the net charge of a protein [24]. Thus, we evaluated whether the performance of Prime would improve upon removal of mutations that change the net charge of the protein.

As seen in Fig 4A, Prime obtains an AUC of 0.75 on mutations changing the net charge of the protein. Surprisingly, this performance is on par with Prime’s performance on the entire dataset (AUC = 0.75) and very similar to the performance on the dataset involving mutations conserving the net charge of the protein (AUC = 0.74, Fig 4B). Only the accessibility metric showed substantial sensitivity to dividing the dataset in this way, obtaining an AUC of 0.66 on the entire conserve-charge set vs an AUC of 0.80 on mutations thought to change the net charge of the protein.

We divided the dataset to separate big to small mutations from small to big mutations to determine if the added flexibility of the side chain relaxation protocol in Prime would benefit small to large mutations. Small to big mutations are defined as mutations whose side chain volume increases by more than the volume of an alanine to valine mutation (This was inversely used for big to small mutations). Interestingly, the 5 Å refinement protocol in Prime negatively impacted performance for the small to big mutations, reducing the AUC of Prime from 0.78 to 0.69 (Fig 5). Also, and perhaps not unexpectedly, the added flexibility did not improve Prime’s performance for the big to small mutations. This suggests that not perturbing the crystal structure is beneficial relative to allowing minor alterations of the side chains of the structure. This has previously been observed by Kellogg et al. when modifying the sampling protocol of the Rosetta energy function [25]. It is also worth noting that the BLOSUM62 substitution matrix struggled with the small to big mutations (AUC = 0.63) relative to the entire dataset (AUC = 0.71).

It has been documented that positions at the surface of a protein are typically more tolerant to mutation than mutations at buried positions in a protein [26]. As a result, we were interested in whether the performance of the models tested in this study would be sensitive to whether a mutation appears at the surface of the protein or at a buried position (Fig 6). A position is defined as buried if the percent of solvent exposure of the wildtype amino acid is  $\leq 5\%$ . A position is defined as solvent exposed if the percent of solvent exposure of the wildtype residue is  $> 20\%$ . Predictably, the accessibility metric does not perform well at classifying functionally disrupting mutations at buried positions, getting an AUC value close to random





**Fig 4. ROC curves separated by mutations that change in charge from mutations that conserve charge.** (A) The charge-change dataset consisted of 369 mutations (B) The conserve-charge dataset consists of 623 mutations that conserve the net charge of the protein. The tables on the right list the area under the curve (AUC) values along with 95% confidence interval (CI) for the AUC values.

<https://doi.org/10.1371/journal.pone.0214015.g004>

Method	AUC	95% CI
<b>FoldX</b>	0.68	0.64-0.72
<b>Accessibility</b>	0.80	0.76-0.85
<b>BLOSUM62</b>	0.68	0.62-0.73
<b>Prime Stability</b>	0.75	0.71-0.80
<b>Prime Stability (5Å)</b>	0.69	0.63-0.74

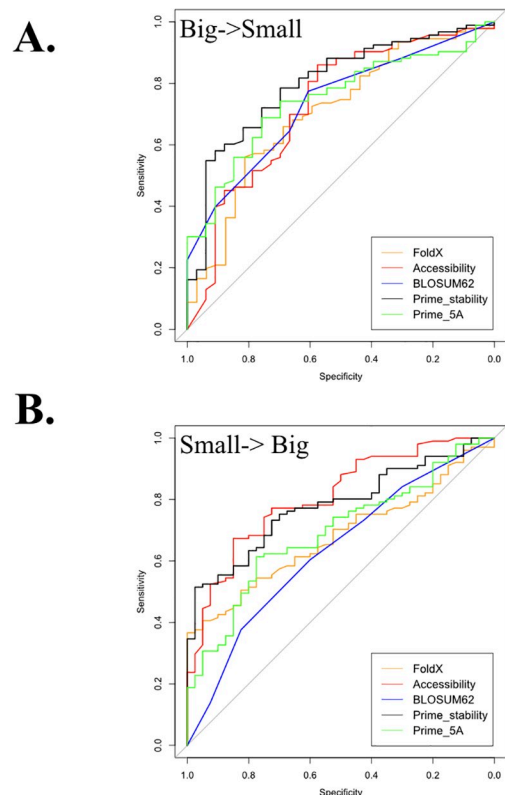
Method	AUC	95% CI
<b>FoldX</b>	0.67	0.64-0.70
<b>Accessibility</b>	0.66	0.61-0.70
<b>BLOSUM62</b>	0.73	0.69-0.77
<b>Prime Stability</b>	0.74	0.70-0.78
<b>Prime Stability (5Å)</b>	0.69	0.65-0.73

(AUC = 0.47). In contrast, the BLOSUM62 matrix does exceptionally well at classifying mutations as functionally disrupting at buried positions, resulting in an AUC of 0.84. Prime's performance on the buried set is very close to its performance on the entire dataset, resulting in an AUC of 0.78 vs. an AUC of 0.75 on the entire dataset. Added flexibility in Prime did not help Prime with buried mutations, resulting in an AUC of 0.73. Interestingly, all algorithms significantly struggle with classifying disrupting mutations on the surface of the protein. The AUCs range from 0.61 to 0.67.

## Conclusion

The MM-GBSA method in Prime is able to outperform all other methods tested in this study for predicting functionally disrupting mutations. And, perhaps more importantly, Prime works more robustly across various divisions of the dataset than the other methods tested in this study. The broad consistent performance of Prime likely reflects the fact that Prime is a physics-based approach. Excluding surface exposed mutations, the AUC values of Prime range from 0.74 to 0.80, which is a range of only 0.06. This is equal to the range of the 95% confidence interval on the entire dataset (Fig 2). However, the AUC values for the BLOSUM62 matrix range from 0.63 to 0.84. This is a variation of 0.21, a value that is more than three times larger than Prime's variation. The range of AUC values obtained for the accessibility metric is even larger in magnitude, ranging from 0.47 (a nearly random model) to 0.81. The scoring function in FoldX, similar to Prime, uses physics-based terms. And thus, similar to Prime,





Method	AUC	95% CI
<b>FoldX</b>	0.71	0.63-0.78
<b>Accessibility</b>	0.73	0.62-0.84
<b>BLOSUM62</b>	0.73	0.64-0.82
<b>Prime Stability</b>	0.80	0.71-0.89
<b>Prime Stability (5Å)</b>	0.75	0.66-0.84

Method	AUC	95% CI
<b>FoldX</b>	0.68	0.62-0.74
<b>Accessibility</b>	0.81	0.73-0.88
<b>BLOSUM62</b>	0.63	0.53-0.73
<b>Prime Stability</b>	0.78	0.70-0.86
<b>Prime Stability (5Å)</b>	0.69	0.60-0.78

**Fig 5. ROC curves separated by mutations that decrease in volume from mutations that increase in volume.** (A) The big->small dataset consists of 126 mutations that decrease the net volume of the side chain (see [Methods](#)). (B) The small->big dataset consists of 141 mutations that increase the volume of the mutated side chain (see [Methods](#)). The tables on the right list the area under the curve (AUC) values along with corresponding 95% confidence interval (CI) for the AUCs.

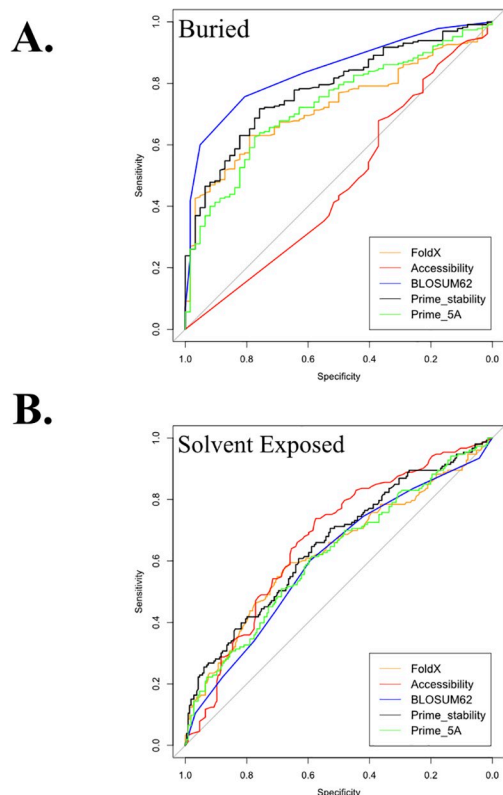
<https://doi.org/10.1371/journal.pone.0214015.g005>

FoldX also achieves a consistent performance, with AUC values ranging from 0.67 to 0.72. However, Prime is always able to outperform FoldX on all divisions of the dataset.

Beyond benchmarking Prime, we have attempted to improve Prime's performance by introducing flexibility in the protein side chains within 5 Å of the residue being mutated. Intriguingly, this degrades performance across all divisions of the dataset, even for mutations that significantly increase the volume of the side chain being mutated. Others have made similar observations when validating scoring functions for protein stability. As observed in Kellogg et al., to properly introduce protein flexibility into the Rosetta energy function it is required to allow backbone atoms of the protein to move in conjunction with side chain motion [25]. Therefore, future studies using molecular-dynamics based methods such as free energy perturbation [27,28], which allow for complete protein flexibility, should also be benchmarked to determine how well these approaches can predict functionally disrupting mutations based on changes in protein stability.

When discussing the generality of Prime at predicting mutations that will disrupt function it is also paramount to discuss its shortcomings. Prime struggles on surface exposed positions, obtaining an AUC only of 0.65 (Fig 6). However, it is worth pointing out that when Prime predicts that a surface mutation does destabilize the protein (and thus disrupts function), the prediction is likely correct, as shown by the high precision and specificity scores in Fig 7 using two separate cutoff values for Prime. Prime's limitation is that when it predicts that a mutation does not disrupt stability (and function) this may or may not be correct, as reflected by the low sensitivity values for two different Prime cutoffs.





Method	AUC	95% CI
<b>FoldX</b>	0.72	0.68-0.77
<b>Accessibility</b>	0.47	0.39-0.56
<b>BLOSUM62</b>	0.84	0.79-0.89
<b>Prime Stability</b>	0.78	0.72-0.84
<b>Prime Stability (5Å)</b>	0.73	0.66-0.79

Method	AUC	95% CI
<b>FoldX</b>	0.63	0.59-0.67
<b>Accessibility</b>	0.67	0.62-0.73
<b>BLOSUM62</b>	0.61	0.55-0.67
<b>Prime Stability</b>	0.65	0.60-0.71
<b>Prime Stability (5Å)</b>	0.62	0.56-0.68

**Fig 6. ROC curves separated by mutations that are buried in the protein from solvent exposed mutations.** (A) The buried dataset consists of 292 mutations that are buried in the core of beta lactamase ( $\leq 5\%$  solvent exposed). (B) The solvent exposed dataset consists of 367 mutations that are solvent exposed ( $>20\%$  solvent exposed). The tables on the right list the area under the curve (AUC) values along with corresponding 95% confidence interval (CI) for the AUCs.

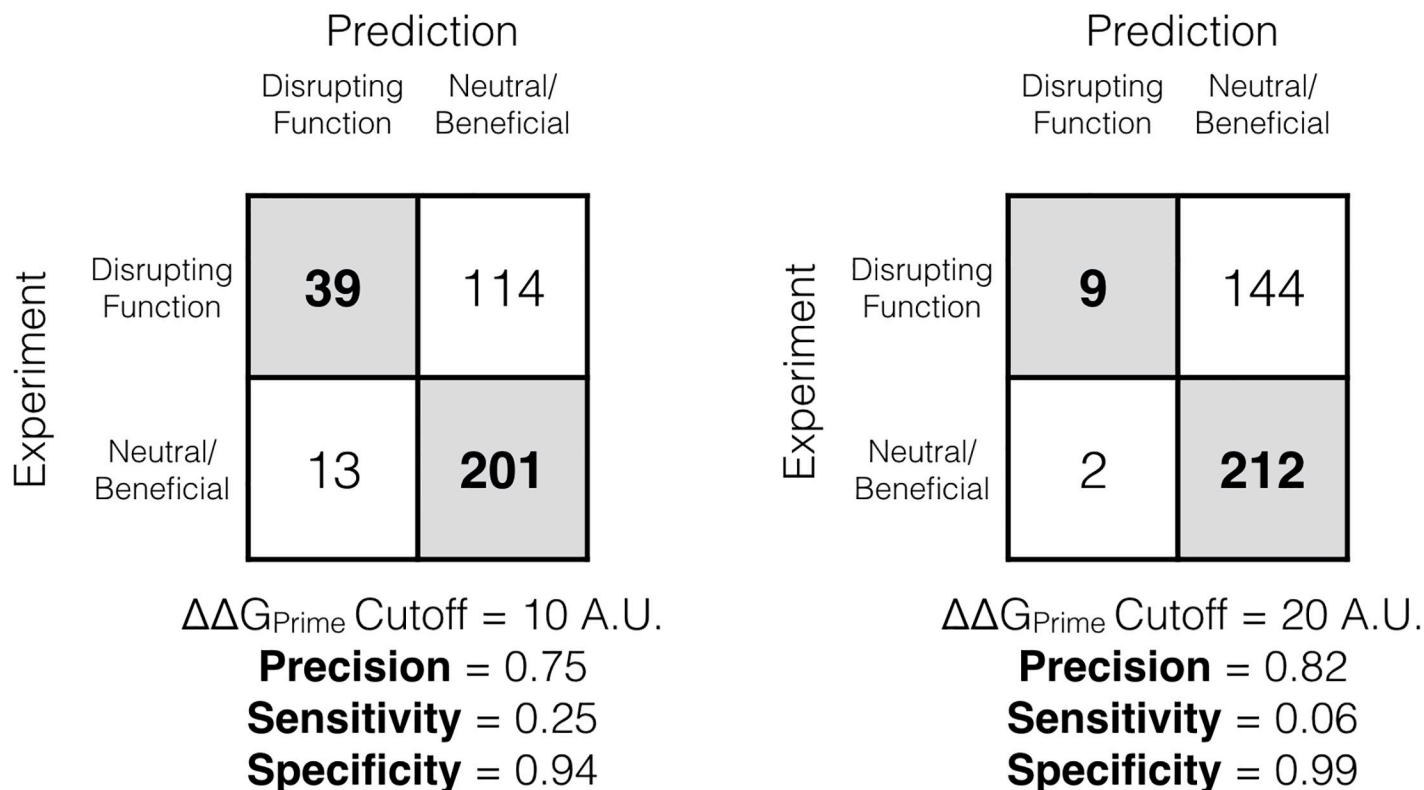
<https://doi.org/10.1371/journal.pone.0214015.g006>

The main challenge for mutations on the surface is that these mutations have several mechanisms for disrupting function. For instance, cysteine mutations on the surface may engage in disulfide bridges that prevent the protein from performing its function. And this may be why mutations to cysteine are 10 times more likely to result in functional disruption than to be neutral/beneficial. Additionally, mutations at the surface of a protein can interfere with crucial protein-protein interactions or protein-substrate interactions, which cannot be predicted from a crystal structure of just a protein monomer. However, in the case of beta-lactamase, we can use the crystal structure of beta lactamase in complex with an acylation transition state analog to approximate changes in binding affinity to the natural substrate. To do this while still accounting for changes in protein stability we took the maximum Prime score between the predicted change in stability and the predicted change in affinity for each mutation. The paradigm of this approach is that if a mutation is predicted to have a minor change in stability but is predicted to significantly destabilize substrate binding affinity the mutation would still be classified as functionally disrupting and vice versa. Unfortunately, this did not significantly improve the results, changing the AUC value from 0.74 to 0.75 (Fig 8). As a result, it appears that for this dataset, loss of affinity to substrate plays a nominal role.

Beyond the challenge of surface mutations, there are additional factors impeding the prediction of functionally disrupting mutations via protein stability predictions. For example, molecular chaperones have been shown to guide the folding of a protein whose unfolded state is actually lower in energy than its folded state [29]. This is done through kinetically trapping the protein in the folded state. And fascinatingly, these chaperones have been known to act



# Surface Mutations



**Fig 7. Two confusion matrices classifying functionally disrupting mutations on the surface using the Prime stability score.** The left matrix uses a Prime energy cutoff of 10 prime energy units to classify functionally disrupting mutations from neutral/beneficial mutations. The matrix on the right uses a Prime energy cutoff of 20 prime energy units. The precision, sensitivity, and specificity are shown below the matrices.

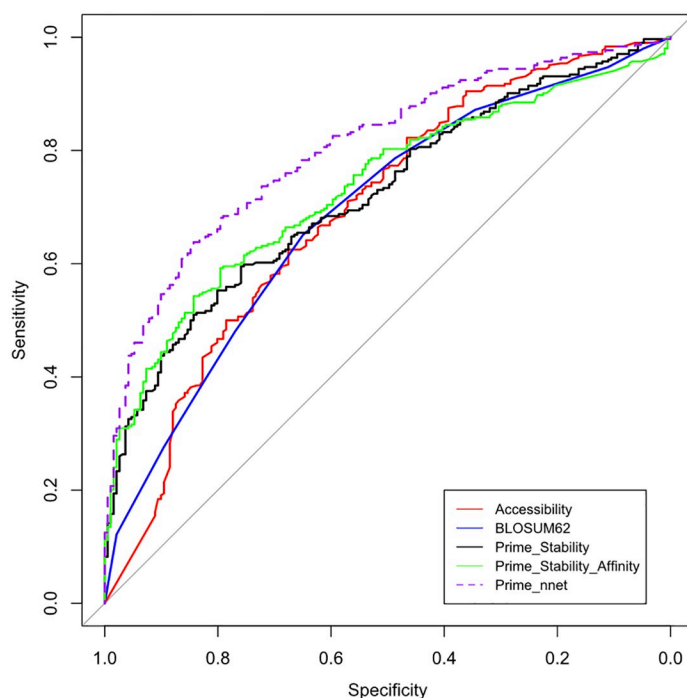
<https://doi.org/10.1371/journal.pone.0214015.g007>

heterogeneously on genetic variants of a protein causing the connection between protein stability and function to be further blurred [30].

Another significant limitation to applying Prime to predict functionally disrupting mutations is the ability to obtain an accurate structure of the protein target of interest. In this study, the 1BTL crystal structure provided a promising starting point to build a structure-based model of the beta-lactamase protein, only requiring hydrogen atoms to be added and optimized. However, despite the large size of the protein databank, the majority of proteins have not been crystallized [31]. Homology modeling provides a path to extend the protein databank for proteins closely related in sequence to those that have already been crystallized, and future studies should look into how well Prime will predict functionally disrupting mutations using homology models.

Finally, using a single-layer neural network, we integrated the best performing predictors on the entire beta lactamase dataset, which are accessibility, BLOSUM62, and Prime Max, into a single machine learning (ML) model. We trained the ML model by randomly sampling half the 990 single-point mutations, and testing on the remaining half. Interestingly, the model obtains an AUC of 0.84 (Fig 8). It is not clear if the ML model will generalize beyond the beta-lactamase dataset. However, it demonstrates the ability to unify the models under a single score. And such an approach may assist with integrating models that capture other





Method	AUC	95% CI
Accessibility	0.72	0.68-0.77
BLOSUM62	0.71	0.66-0.75
Prime Stability	0.74	0.70-0.78
Prime Max (Stability+Affinity)	0.75	0.71-0.79
ML Model	0.84	0.80-0.87

**Fig 8. ROC curves using 495 mutations randomly sampled from the beta lactamase dataset.** Prime Max refers to the maximum prime energy between the predicted change in stability and predicted change in affinity. The machine learning (ML) model refers to a single layer neural network trained on the other 495 mutations not included in this dataset. The area under the curve (AUC) values along with the corresponding 95% confidence intervals are shown in the table on the right.

<https://doi.org/10.1371/journal.pone.0214015.g008>

mechanisms for disrupting function outside of protein stability—such as those described above. Lastly, despite the limitations described, accurate models of protein stability do capture a significant amount of the functionally disrupting mutations in the beta-lactamase dataset.

## Materials and methods

### Preparing 1BTL and 1AXB crystal structures

Schrodinger's Protein Preparation Wizard (PrepWizard) was used to prepare all PDB structures for Prime [32]. The PrepWizard assigns bond orders, predicts protonation states, samples Asn/Gln/His flip states, removes select crystallographic waters, optimizes the H-bond network, and minimizes the structure. Default values were used for all parameters except the -propka\_pH flag, which was set to pH 7.2. The following command line was used:

```
$SCHRODINGER/utilities/prepwizard -fillsidechains -propka_pH
<pH> -NOJOBID <Input.pdb> <Output.mae>
```

### Running Residue Scanning

The Residue Scanning functionality in BioLuminate (18) version 18–3 was used to By default, For this work, the default sampling protocol was used via the following command line:

```
$SCHRODINGER/run $SCHRODINGER/mmshare- v32017/python/
scripts/residue_scanning_backend.py jobname $jobname -fast-
res_file <Resfile_name> -receptor_asl <Chain_name> -refine_mut
prime_residue -dist 0.00 <Input_file.mae> -NJOBS 1 -NOJOBID
```



Due to limitations of the Residue Scanning application, mutations breaking covalent bonds in the structure were given a Prime score of 1000.0.

### Splitting the beta-lactamase dataset

As described, the beta-lactamase dataset has been divided using three approaches. The first approach separates the mutations that change the net charge of the protein from those that conserve the net charge of the protein. For generating the dataset of net-charge-changing mutations, mutations involving lysine, arginine, aspartic acid, and glutamic acid are used. If residues are mutated to or from these amino acids they are considered to change the net charge of the protein, unless the mutation remains positive (i.e. a lysine to an arginine or an arginine to a lysine) or remains negative (i.e. an aspartic acid to a glutamic acid or a glutamic acid to an aspartic acid). Mutations are considered to conserve the net charge of a protein if they do not involve lysine, arginine, aspartic acid, or a glutamic acid. With the exception being for mutations between lysine and arginine or between glutamic acid and aspartic acid.

To split the dataset by mutations that significantly increase (small->big) or decrease (big->small) in volume, we use van der Waals volumes reported by Darby and Creighton (1993) [33]. Changes in volume larger than an alanine to valine mutation ( $37 \text{ \AA}^3$ ) are considered to be a significant change in volume, and vice versa for mutations decreasing in volume. More stringent filters were considered, however, these resulted in unreliably small datasets.

Lastly, to classify mutations as buried or solvent exposed, the relative solvent accessible surface area is calculated for each position in the wildtype structure. This was done by dividing the solvent accessibility (Accessibility) metric by the maximum solvent accessible surface area of that amino acid in a tri-peptide according to Miller et al. (1987) [23]. Relative solvent accessible surface areas  $\leq 0.05$  were classified as buried. And relative solvent accessible surface areas  $> 0.20$  were considered solvent exposed. Glycine residues were excluded from this analysis.

### MIC score, FoldX, Accessibility, and BLOSUM62

The values for the MIC score, FoldX, Accessibility, and BLOSUM62 were obtained from Jacquier et al. [20].

### Calculating the area under the curve (AUC) and the machine learning (ML) model

R version 3.3.0 was used to calculate the receiver operating characteristic curves (ROC). R was used to calculate the area under the curve, the 95% confidence intervals, train a neural network [34]. Specifically the pROC package and the nnet package in R were used [35,36].

### Supporting information

**S1 Table. Model scores for each mutation.**  
(CSV)

### Acknowledgments

The authors acknowledge the Schrödinger support and development team for their helpful discussions.



## Author Contributions

**Conceptualization:** Christopher Negron.

**Data curation:** Christopher Negron.

**Investigation:** Christopher Negron.

**Software:** David A. Pearlman.

**Supervision:** David A. Pearlman, Guillermo del Angel.

**Writing – original draft:** Christopher Negron.

**Writing – review & editing:** Christopher Negron, David A. Pearlman, Guillermo del Angel.

## References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001; 409(6822):860–921. <https://doi.org/10.1038/35057062> PMID: 11237011
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001; 291(5507):1304–51. <https://doi.org/10.1126/science.1058040> PMID: 11181995
3. Reuter JA, Spacek D V., Snyder MP. High-Throughput Sequencing Technologies. *Mol Cell*. 2016; 58(4):586–97.
4. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016; 536(7616):285–91. <https://doi.org/10.1038/nature19057> PMID: 27535533
5. Land M, Hauser L, Jun S-R, Nookaew I, Leuze MR, Ahn T-H, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics*. 2015; 15(2):141–61. <https://doi.org/10.1007/s10142-015-0433-4> PMID: 25722247
6. Forster SC, Browne HP, Kumar N, Hunt M, Denise H, Mitchell A, et al. HPMCD: The database of human microbial communities from metagenomic datasets and microbial reference genomes. *Nucleic Acids Res*. 2016; 44(D1):D604–9. <https://doi.org/10.1093/nar/gkv1216> PMID: 26578596
7. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, et al. Comparison of predicted and actual consequences of missense mutations. *Proc Natl Acad Sci*. 2015; 112(37):E5189–98. <https://doi.org/10.1073/pnas.1511585112> PMID: 26269570
8. Zhou Xi, Iversen Edwin S. Jr. and P G. Classification of Missense Mutations of Disease Genes. *J Am Stat Assoc*. 2005; 100:51–60. <https://doi.org/10.1198/016214504000001817> PMID: 18418466
9. Foley SB, Rios JJ, Mgbemena VE, Robinson LS, Hampel HL, Toland AE, et al. Use of Whole Genome Sequencing for Diagnosis and Discovery in the Cancer Genetics Clinic. *EBioMedicine*. Elsevier B.V.; 2015; 2(1):74–81.
10. Guidugli L, Shimelis H, Masica DL, Pankratz VS, Lipton GB, Singh N, et al. Assessment of the Clinical Relevance of BRCA2 Missense Variants by Functional and Computational Approaches. *Am J Hum Genet*. 2018; 102(2):233–48. <https://doi.org/10.1016/j.ajhg.2017.12.013> PMID: 29394989
11. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational Landscape of Metastatic Cancer Revealed from Prospective Clinical Sequencing of 10,000 Patients. *Nat Med*. 2017; 23(6):703–13. <https://doi.org/10.1038/nm.4333> PMID: 28481359
12. Hauser K, Negron C, Albanese SK, Ray S, Steinbrecher T, Abel R, et al. Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. *Commun Biol*. Springer US; 2018; 1(1):70.
13. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002; 30(17):3894–900. PMID: 12202775
14. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Vol. 7, *Current Protocols in Human Genetics*. 2015. Unit7.20.
15. Liu M, Watson LT, Zhang L. HMMvar-func: A new method for predicting the functional outcome of genetic variants. *BMC Bioinformatics*. BMC Bioinformatics; 2015; 16(1):1–10.
16. Hicks S, Wheeler DA, Plon SE, Kimmel M. Prediction of Missense Mutation Functionality Depends on both the Algorithm and Sequence Alignment Employed. *Hum Mutat*. 2011; 32(6):661–8. <https://doi.org/10.1002/humu.21490> PMID: 21480434



17. Massova I, K PA. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J Am Chem Soc.* 1999; 121:8133–8143.
18. Beard H, Cholleti A, Pearlman D, Sherman W, Loving KA. Applying physics-based scoring to calculate free energies of binding for single amino acid mutations in protein-protein complexes. *PLoS One.* 2013; 8(12):1–11.
19. Pakula AA, Young VB, Sauer RT. Bacteriophage lambda cro mutations: effects on activity and intracellular degradation. *Proc Natl Acad Sci.* 1986; 83(23):8829–33. PMID: [2947238](#)
20. Jacquier H, Birgy A, Le Nagard H, Mechulam Y, Schmitt E, Glodt J, et al. Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci.* 2013; 110(32):13067–72. <https://doi.org/10.1073/pnas.1215206110> PMID: [23878237](#)
21. Guerois R, Nielsen JE, Serrano L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *Journal Mol Biol.* 2002; 320(2):369–87.
22. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: An online force field. *Nucleic Acids Res.* 2005; 33(SUPPL. 2):382–8.
23. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. *Journal Mol Biol.* 1987; 196(3):641–56.
24. Steinbrecher T, Zhu C, Wang L, Abel R, Negron C, Pearlman D, et al. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *J Mol Biol. Elsevier Ltd;* 2017; 429(7):948–63.
25. Kellogg EH, Leaver-Fay A, Baker D. Changes in Protein Structure and Stability. 2011; 79(3):830–8.
26. Ramsey DC, Scherrer MP, Zhou T, Wilke CO. The relationship between solvent accessibility and evolutionary rate in protein evolution. *Genetics.* 2011; 188(2):479–88. <https://doi.org/10.1534/genetics.111.128025> PMID: [21467571](#)
27. Zwanzig RW. High Temperature Equation of State By a Perturbation Method. *J Chem Phys.* 1954; 22(8):1420–1426.
28. Ford MC, Babaoglu K. Examining the Feasibility of Using Free Energy Perturbation (FEP+) in Predicting Protein Stability. 2017;
29. Pauwels K, Molle IV, Tommassen J G P. Chaperoning Anfinsen: the steric foldases. *Mol Microbiol.* 2007; 64(4):917–22. <https://doi.org/10.1111/j.1365-2958.2007.05718.x> PMID: [17501917](#)
30. Karras GI, Yi S, Sahni N, Fischer M, Xie J, Vidal M, et al. HSP90 Shapes the Consequences of Human Genetic Variation. *Cell.* 2017; 168(5):856–866.e12. <https://doi.org/10.1016/j.cell.2017.01.023> PMID: [28215707](#)
31. Khafizov K, Madrid-Aliste C, Almo SC, Fiser A. Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative. *Proc Natl Acad Sci.* 2014; 111(10):3733–8. <https://doi.org/10.1073/pnas.1321614111> PMID: [24567391](#)
32. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des.* 2013; 27(3):221–34. <https://doi.org/10.1007/s10822-013-9644-8> PMID: [23579614](#)
33. Darby NJ, Creighton TE. *Protein Structure.* New York: IRL Press at Oxford University Press; 1993.
34. Team (R Core). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Vienna, Austria; 2016.
35. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011; 12:77. <https://doi.org/10.1186/1471-2105-12-77> PMID: [21414208](#)
36. Venables WN, Ripley BD. *Modern Applied Statistics with S.* Fourth Edi. New York: Springer; 2002.