



Article

# SAAMBE: Webserver to Predict the Change of Binding Free Energy Caused by Amino Acids Mutations

Marharyta Petukh <sup>1,†</sup>, Luogeng Dai <sup>1,2,†</sup> and Emil Alexov <sup>1,\*</sup>

<sup>1</sup> Computational Biophysics and Bioinformatics, Physics Department, Clemson University, Clemson, SC 29634, USA; mpetukh@clemson.edu (M.P.); luogend@g.clemson.edu (L.D.)

<sup>2</sup> Department of Computer Sciences, Clemson University, Clemson, SC 29634, USA

\* Correspondence: ealexov@g.clemson.edu; Tel.: +1-864-908-4796; Fax: +1-864-656-0805

† These authors contributed equally to this work.

Academic Editor: Tatyana Karabancheva-Christova

Received: 12 February 2016; Accepted: 7 April 2016; Published: 12 April 2016

**Abstract:** Predicting the effect of amino acid substitutions on protein–protein affinity (typically evaluated via the change of protein binding free energy) is important for both understanding the disease-causing mechanism of missense mutations and guiding protein engineering. In addition, researchers are also interested in understanding which energy components are mostly affected by the mutation and how the mutation affects the overall structure of the corresponding protein. Here we report a webserver, the Single Amino Acid Mutation based change in Binding free Energy (SAAMBE) webserver, which addresses the demand for tools for predicting the change of protein binding free energy. SAAMBE is an easy to use webserver, which only requires that a coordinate file be inputted and the user is provided with various, but easy to navigate, options. The user specifies the mutation position, wild type residue and type of mutation to be made. The server predicts the binding free energy change, the changes of the corresponding energy components and provides the energy minimized 3D structure of the wild type and mutant proteins for download. The SAAMBE protocol performance was tested by benchmarking the predictions against over 1300 experimentally determined changes of binding free energy and a Pearson correlation coefficient of 0.62 was obtained. How the predictions can be used for discriminating disease-causing from harmless mutations is discussed. The webserver can be accessed via [http://compbio.clemson.edu/saambe\\_webserver/](http://compbio.clemson.edu/saambe_webserver/).

**Keywords:** missense mutations; energy calculation; binding free energy; MM/PBSA method

## 1. Introduction

Practically every protein is involved in various binding processes [1], frequently with another proteins [2,3]. Altering such interactions via amino acid substitutions, naturally occurring or engineered, is expected to have significant impact on the wild type characteristics of the cell [4,5]. While such changes can, in principle, be experimentally measured, the cost and the required time are prohibitory for large-scale investigations. Because of that, the development of computational approaches is needed for large-scale modeling of the effects of amino acid substitutions on protein–protein binding [6–13] (for comparison of popular approaches see Reference [14]). The existing computational methods are typically clustered into two main groups: approaches utilizing sequence information and approaches considering structural features [7,8,15,16]. The main advantage of the approaches using sequence data is that they are fast, but the performance of such predictors strongly hinge on the training set of data. The structure-based approaches represent the other part of the spectrum. Among them the most rigorous, but computationally demanding, are the thermodynamic integration (IT) and the free energy perturbation (FEP) methods [17]. At the same time, the Molecular Mechanical Poisson-Boltzmann (Generalized Born)/Surface Accessible (MM/PB(GB)SA) approach

is a method that provides the details of the modeling while requiring reasonable computational time [18,19]. In the MM/PBSA method, the binding free energy change ( $\Delta\Delta G$ ) is modeled as a linear combination of several potential energies, including molecular mechanics energy, and polar and non-polar components of solvation energy.

Many disease-causing mutations affect protein–protein and protein–DNA interactions [4,9,20–25]. Thus, predicting binding free energy changes caused by amino acid mutations has implications for discriminating disease-causing mutations from harmless mutations [8,16,26]. Recently, it was demonstrated that the changes of the binding free energy are correlated with the propensity of a given mutation to be disease-causing [27], especially if the predicted energy changes are normalized to the wild type binding free energy [28]. Therefore, accurate predictions of the changes of the binding free energy are needed for detecting disease-causing mutations occurring in the human genome.

In this work we report a webserver, which utilizes a previously developed methodology termed Single Amino Acid Mutation based change in Binding free Energy (SAAMBE). SSAMBE combines two approaches: sequence- and structure-based methods [18]. It utilizes the MM/PBSA method along with an additional set of statistically delivered terms from numerical investigation of the physico-chemical properties of protein complexes (the corresponding entries are provided in Tables S1 and S2). The performance was tested against more than 1300 mutations with experimentally available changes of the binding free energy taken from 43 proteins [29] and resulted in a good agreement with experimental data (Pearson correlation coefficient 0.62). It should be pointed out that SAAMBE is fast enough to allow for large-scale calculations, since the average time for modeling a mutation is less than a minute.

## 2. Results

The SAAMBE webserver is based on the SAAMBE algorithm [18], which predicts the changes of binding free energy caused by amino acid substitutions. The SAAMBE algorithm performance was previously reported [18] and it is shown that the algorithm achieves a Pearson correlation coefficient of 0.62 in a benchmark against more than 1300 experimentally-determined changes of binding free energy. The results of benchmarking are summarized in Table 1, where the correlation coefficient for specific cases is provided: (a) case 1, when a bulky residue is changed to a small one; (b) case 2, when the mutant (MT) residue is Ala, which is typically used for predicting protein “hot-spot”; and (c) case 3, the accuracy of predictions for mutations being in a particular structural region.

**Table 1.** Pearson correlation coefficient ( $R$ ), the  $y$ -intercept and the slope of linear fit between experimental and predicted binding free energy changes. The number of cases is provided in parentheses. The results are shown for different types of mutations and summarized at the bottom for all mutations. The results of 5-fold cross validation test are shown as well.

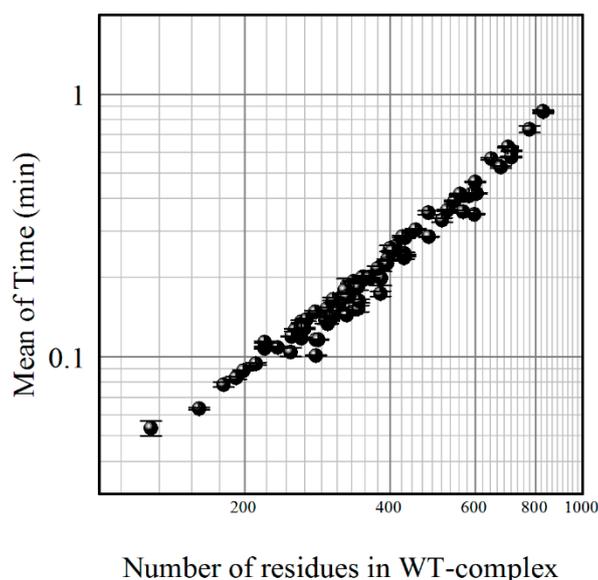
Type of Mutaton and Mutation Site Location	Pearson Correlation Coefficient ( $R$ )	$y$ -Intercept	Slope
Large-to-Small (173)	0.489	0.328	0.692
ALA-scanning (577)	0.488	0.268	0.695
COR, SUP (807)	0.461	0.351	0.813
RIM, SUR, INT (518)	0.478	−0.024	1.023
ALL (1326)	0.624 (0.716 $\pm$ 2SD, 0.603 CV)	1	−2.31 $\times$ 10 <sup>−5</sup>

$\pm$ 2SD: within two standard deviations; CV: 5-fold cross validation test. ALA: Alanine residue; COR: core; SUP: support; RIM: rim; SUR: surface; and INT: interior type of mutation site.

It can be seen that SAAMBE does not perform very well on particular types of substitutions, but overall performance is very good, achieving  $R = 0.716$  if cases within two standard deviations are considered. It was demonstrated in the original work [18] that such performance is much better than the performance of existing solutions, including BeAtMuSiC [30] and FoldX [31]. However, it should be mentioned that a recent paper [14] showed that these two approaches are not among the

best predictors. Instead, newly developed methods, such as mCSM-PPI [12] and those outlined in Reference [13], perform better.

One of the most important characteristics of any server performance is how fast the user receives the results. The SAAMBE method produces results in 0.22 min on average when employing 16 nodes on Clemson University Palmetto Supercomputer. The time of calculations depends on the number of residues in the protein complex; Figure 1 illustrates this, and shows that, even for very large complexes of more than 800 residues, the execution time is less than a minute.

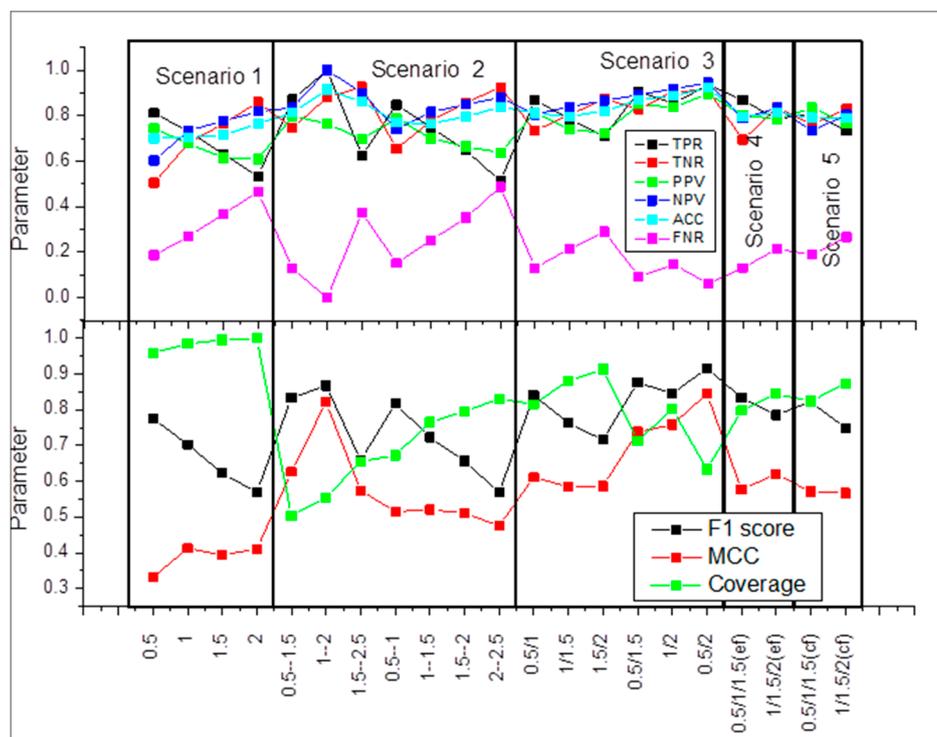


**Figure 1.** The computational time (mean value) of the Single Amino Acid Mutation based change in Binding free Energy (SAAMBE) algorithm as a function of the number of residues in the protein complex. Both axes are in log scale. WT: wild type complexes.

As mentioned above, the predicted binding free energy changes can be used to infer if mutations are disease-causing or harmless. Typically, this is done by setting up a particular threshold of the energy change and it is assumed that any mutation-causing effect larger than the threshold is disease-causing. Since, in this work, we do not introduce metrics for disease-association, the goal is to benchmark the performance of the SAAMBE method by setting up various thresholds for the calculated and experimental binding free energy change and to see how SAAMBE performs on matching experimentally observed large (large—above the specific cut-off) binding free energy changes. The performance is investigated using the procedure detailed in the [Materials and Methods](#) section. Several scenarios will be investigated. Below, we refer to experimentally-measured changes of the binding free energy and those calculated with SAAMBE. Further, we will apply one of the scenarios to test SAAMBE's ability to discriminate disease-causing and polymorphic mutations.

### 2.1. Scenario 1

The cut-off for experimental and calculated  $\Delta\Delta G$  is identical. In this case, the cut-off above which  $\Delta\Delta G$  is considered to be large is the same for the experimental and calculated  $\Delta\Delta G$ . We systematically varied the cut-off from 0.5 kcal/mol up to 2 kcal/mol in steps of 0.5 kcal/mol. The results are presented in Figure 2, scenario 1. It can be seen that the coverage is very good, reaching almost 100%, but the Matthew Correlation Coefficient (MCC) and F1 score are not impressive. At the same time, true positive ratio (TPR) and true negative ratio (TNR) are very good.



**Figure 2.** Benchmarking of statistical parameters as a function of various definitions (see [Materials and Method](#) section for details).

### 2.2. Scenario 2

The cut-off for experimental and calculated  $\Delta\Delta\Delta G$  is identical, but there is a gap between the threshold for large and small  $\Delta\Delta\Delta G$ . In terms of disease-causing and harmless mutations, this will correspond to the case for which there will be a “gray” zone of energy changes that cannot be associated, neither with disease, nor to be called harmless (similar approach was described in Reference [32] for evaluating protein stability changes). It can be seen that coverage is significantly lower as compared to scenario 1, but the MCC and F1 score are better (Figure 2, scenario 2).

### 2.3. Scenario 3

The cut-offs for experimental and calculated  $\Delta\Delta\Delta G$  are different. This scenario is applicable in the case when the method over- or under-predicts the experimentally observed  $\Delta\Delta\Delta G$ . The results are shown in Figure 2, scenario 3. It can be seen that this scenario achieves the best performance. The coverage reaches 0.8, without compromising MCC (MCC = 0.75) and F1 score (F1 score = 0.85) at cut-offs of 1 and 2 kcal/mol.

### 2.4. Scenario 4

The cut-offs for experimental and calculated  $\Delta\Delta\Delta G$  are different, but there is a difference in defining the bottom limit (see [Materials and Methods](#) section for details). In this testing, Figure 2, scenario 4, the coverage increases, but the MCC is low.

### 2.5. Scenario 5

The cut-offs for experimental and calculated  $\Delta\Delta\Delta G$  are different, but there is a difference in defining the bottom limit (see [Materials and Methods](#) section for details). As above, the coverage increases, but the MCC is low (Figure 2, scenario 5).

The above investigation indicates that, with proper definition of the cut-offs, the SAAMBE method can distinguish amino acid substitutions causing a large change of the binding free energy and to discriminate them from the substitutions causing minimal change. This paves the way for detecting disease-causing mutations, and which disease-causing effects are associated with protein–protein interactions, which will be demonstrated in the following paragraph.

## 2.6. Case Studies

To illustrate the applicability of the SAAMBE webserver to detect disease-causing mutations, and to distinguish them from polymorphic mutations, we will investigate two proteins with available clinical data. These two cases were taken from the ClinVar database [33,34] to represent proteins with available 3D structures of hetero- and homo-complexes. It should be reiterated that disease-causing mechanisms may involve altering various biophysical characteristics of the corresponding macromolecule, including stability, interactions, and dynamics [8,16,26–28]. Thus, if one attempts to predict disease-causing mutation based on the effect of mutation on the binding free energy, while mutation affects mostly protein stability, then the prediction will not be correct. However, here, we emphasize the ability of predicting disease-causing mutations altering macromolecular interactions.

We begin the analysis with human ribonuclease inhibitor-angiogenin complex (EC 3.1.27.-), PDB ID 1a4y [35]. Missense mutations in angiogenin are associated with amyotrophic lateral sclerosis [36–38]. Currently, the ClinVar database provides sixteen amino acid changes, out of which fourteen can be mapped on the available experimental structure (1a4y) of ribonuclease-angiogenin. These fourteen mutations were subjected to the SAAMBE webserver to predict their effects on the binding free energy. Chains A and B in the original Protein Data Bank (PDB) file were used. The results are shown in Table 2.

**Table 2.** List of mutations and their positions, the effect provided by ClinVar, calculated  $\Delta\Delta G$  and the location of mutation site within protein interface (for interfacial regions definition, see Methods). WT: wild type; MT: mutant type complexes.

WT Residue	Position	MT Residue	Effect	$\Delta\Delta G$	Location
Q	36	L	Disease	0.00	SUP
Y	38	H	Disease	−0.42	INT
K	41	E	Disease	1.04	SUR
K	41	I	Disease	0.29	SUR
D	46	G	Disease	−1.25	SUR
S	52	N	Disease	0.00	RIM-COR
R	55	K	Disease	1.29	COR
C	63	W	Disease	0.07	INT-SUP
K	64	I	Disease	4.50	COR
I	70	V	Unclassified	0.42	INT
K	84	E	Polymorphism	0.81	SUR
P	136	L	Disease	−0.06	INT
V	137	I	Disease	0.79	SUP
H	138	R	Disease	−1.32	COR

This case has a polymorphic, an unclassified, and twelve disease-causing mutations. If one considers scenario 1, described above, and selects a cut-off of 1 kcal/mol (absolute value), then SAAMBE correctly predicts five disease-causing mutations and discriminates them from the polymorphic and unclassified mutations. The disease-causing effect of the remaining ten disease-causing mutations can be associated with altering other (different from protein affinity) characteristics of native proteins.

The second case study was done to assess SAAMBE's performance for mutations that are not located at a protein interface. For this purpose, we selected fructose 1,6-bisphosphate aldolase from human liver, PDB ID 1qo5 [39]. Mutations in this protein are associated with hereditary fructose intolerance [40–42]. The list contains eighteen mutations, out of which three are polymorphic and one

is unclassified. The molecule is a homo-dimer and a mutation should be introduced on both chains. However, currently, SAAMBE is only designed to predict single mutations. This made us introduce mutations separately on chain A and B and then to sum up the effects. Independently, using the in-house SAAMBE version, we tested this approach by introducing both mutations at the same time and compared the results obtained via simple summation of the effects. Indeed, a vast majority of the cases studied here resulted in a simple cumulative effect, justifying the simple summation approach (the root mean standard deviation (RMSD) between calculated binding free energies via simple summation and introducing both mutations was 0.3 kcal/mol). The results are shown in Table 3. Thus, adopting scenario 1 and selecting a cut-off of 1 kcal/mol, allows SAAMBE to correctly predict harmless and unclassified mutations. In addition, SAAMBE correctly predicts five of the disease-causing mutations, despite the fact that they are not located at the protein interface. It is speculated that the remaining ten disease-causing mutations affect protein stability or another important biophysical characteristics of the monomeric protein, rather than altering protein interactions.

**Table 3.** List of mutations and their positions, the effect provided by ClinVar, calculated  $\Delta\Delta\Delta G$ .  $\Delta\Delta\Delta G(A)$  and  $\Delta\Delta\Delta G(B)$  indicate calculations made with mutation introduced in chain A and B, respectively.  $\Delta\Delta\Delta G(A + B)$  is the sum of both predictions.

WT Residue	Position	MT Residue	Effect	$\Delta\Delta\Delta G(A)$	$\Delta\Delta\Delta G(B)$	$\Delta\Delta\Delta G(A + B)$
I	74	T	Disease	0.50	0.29	0.79
R	134	S	Polymorphism	0.32	0.64	0.96
C	135	R	Disease	0.48	0.01	0.48
W	148	R	Unclassified	0.13	0.74	0.88
A	150	P	Disease	0.52	0.74	1.27
A	175	D	Disease	0.56	0.52	1.08
C	178	R	Disease	0.46	0.00	0.46
P	185	R	Disease	0.03	-0.01	0.03
E	207	Q	Polymorphism	0.35	0.15	0.49
V	222	F	Disease	0.32	0.61	0.93
L	229	P	Disease	0.86	0.85	1.72
L	257	P	Disease	0.80	0.69	1.49
I	268	N	Polymorphism	0.42	0.54	0.96
L	284	P	Disease	0.85	0.85	1.69
R	304	Q	Disease	0.39	0.22	0.61
R	304	W	Disease	0.27	0.24	0.52
N	335	K	Disease	-0.12	0.00	-0.12
A	338	V	Disease	0.20	0.07	0.27

### 3. Discussion

The SAAMBE webserver is a simple to use tool that utilizes new algorithms for the prediction of the change of the binding free energy caused by amino acid mutations. SAAMBE predicts, not only the binding free energy changes, but reports the changes of the corresponding energy components and provides energy-minimized structures of both the wild type (WT) and the mutant type (MT). This allows the users to carry out further structural analysis of the effects of the mutations.

To assess SAAMBE's ability to predict disease-causing mutations and to discriminate them from polymorphic mutations, we carried out analyses of two cases involving protein complexes. It was shown that SAAMBE can distinguish disease-causing and polymorphic mutations that affect protein interactions. However, since mutations may affect various biophysical characteristics of the corresponding protein, one should complement SAAMBE's predictions with investigations of protein stability, dynamics, hydrogen bonds, and other biologically important protein features.

#### 4. Materials and Methods

Definitions of mutation site locations: Here, we assign the location of mutated residues in the protein–protein complex using five distinctive categories (core (COR), support (SUP), rim (RIM), interior (INT) and surface (SUR)), as previously described [18]. This is done by calculating the relative solvent accessible surface area (SASA), which is the ratio between SASA of a residue in a protein and in water (rSASA). For example, rSASA = 1 corresponds to totally exposed residue in the protein. Thus, we calculate the SASA of the residue in the monomeric protein and term it rSASAm, and also in the complex and term it rSASAc. Finally, the term  $\Delta$ rSASA refers to their mutual difference. Based on this classification, a residue is considered to be at the interface if it is assigned to COR, SUP and RIM regions; and a residue is considered to be away from the interface if it is in SUR and INT regions. The last two locations, the RIM and SUR locations, refer to residue that is exposed to the water phase in the complex. The parameters of each of the above-mentioned definitions are provided in Table 4.

**Table 4.** Parameters of the residues location types in protein-protein complex.

Location	Interface	Solvent Exposure	rSASAm	rSASAc	$\Delta$ rSASA
COR	Yes	No	>25%	<25%	>0
SUP	Yes	No	<25%	<25%	>0
RIM	Yes	Yes	any	>25%	>0
INT	No	No	any	<25%	=0
SUR	No	Yes	any	>25%	=0

Typically, the accuracy of prediction of disease-causing mutations is evaluated via the ROC parameters, adopting particular cut-offs for true and false positives. Several quantities are evaluated with Equations (1)–(9), using the relationship between four quantities: true positive (tp), true negative (tn), false positive (fp), and false negative (fn).

$$\text{True Positive Rate (TPR, sensitivity)} = \frac{tp}{tp + fn} \quad (1)$$

$$\text{False Negative Rate (FNR)} = \frac{fp}{tfn + tp} \quad (2)$$

$$\text{True Negative Rate (TNR, specificity)} = \frac{tn}{tn + fp} \quad (3)$$

$$\text{Positive Predictive Value (PPV, precision)} = \frac{tp}{tp + fp} \quad (4)$$

$$\text{Negative Predictive Value (NPV)} = \frac{tn}{tn + fn} \quad (5)$$

$$\text{Accuracy (ACC)} = \frac{tp + tn}{tp + fp + tn + fn} \quad (6)$$

$$\text{F1 score} = \frac{2tp}{2tp + fp + fn} \quad (7)$$

$$\text{Matthews Correlation Coefficient (MCC)} = \frac{tp \cdot tn - fp \cdot fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}} \quad (8)$$

$$\text{Coverage} = \frac{tp + tn + fp + fn}{\text{Number of Cases in Database}} \quad (9)$$

Several scenarios are investigated and their definitions are provided in Table 5.

**Table 5.** Conditions used for calculating the number of tp, tn, fp, and fn cases for five scenarios. Here, A corresponds to the value of calculated change of the binding free energy ( $\Delta\Delta G_{\text{calc}}$ ), while B for experimentally determined change of the binding free energy ( $\Delta\Delta G_{\text{exp}}$ ).

ROC Parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
tp	$A \geq x, B \geq x,$ $\text{sign}(A) = \text{sign}(B)$	$A \geq y, B \geq y,$ $\text{sign}(A) = \text{sign}(B)$	$A \geq x, B \geq y,$ $\text{sign}(A) = \text{sign}(B)$	$A \geq x, B \geq y, \text{sign}(A)$ $= \text{sign}(B)$	$A \geq y, B \geq x, \text{sign}(A)$ $= \text{sign}(B)$
tn	$A < x, B < x$	$A < x, B < x$	$A < y, B < x$	$A < z, B < y$	$A < y, B < z$
fp	$A \geq x, B < x$	$A \geq y, B < x$	$A \geq y, B < x$	$A \geq x, B < y$	$A \geq y, B < z$
fn	$A < x, B \geq x$	$A < x, B \geq y$	$A < x, B \geq y$	$A < z, B \geq y$	$A < y, B \geq x$
example	0.5	0.5-1	0.5/1.5	0.5/1/1.5 (ef)	0.5/1/1.5 (cf)
parameter values	$x = 0.5$	$x = 0.5, y = 1$	$x = 0.5, y = 1$	$x = 0.5, y = 1, z = 1.5$	$x = 0.5, y = 1, z = 1.5$

## 5. Webserver Architecture

### 5.1. Overview of SAAMBE Webserver

The design of the SAAMBE server can be described by three basic components, the client or user interface, the server and the job backend. The client interface is implemented using HTML and JavaScript). It provides the user a form to fill out various parameters for a job to be submitted to the server, as well as a button to upload the PDB file. Once the job submitted, the user will be redirected to the result page. The result page will refresh itself every 30s to get the latest results of the backend running job. The server part analyzes the parameters and writes the job and parameter files to job backend, assembles the Protein Data Bank (PDB) [43] files and starts the job remotely. The server is implemented using PHP and hosted by the Apache webserver. The job backend executes the job within a Palmetto cluster. The job handling is implemented using Python Below we describe each component of the processes in detail.

### 5.2. Client Interface

This is a web page where the user inputs all necessary information to complete a server job. The user must supply a PDB file by uploading a PDB file from their local file system. The job parameters include a partner selection part: partner 1 and 2; a mutation part: position, chain, original amino acid and mutated amino acid. Partner 1, partner 2, position, and chain are provided by the user, along with the original and mutated amino acid. A help file is provided as well, along with a particular example.

### 5.3. Server

The server acts as a middle man, it gets the PDB file and parameters from the user, puts it to the job backend, and starts the job execution in the job backend. It checks and gets the job results from the job backend upon the user's query for the results.

### 5.4. Job Backend

Once the user has submitted a request, a job is submitted with a PBS command and executed by the Palmetto cluster. The Palmetto cluster is a supercomputer and it can give results to a user in a short time.

### 5.5. Results

If all user inputs are correct, then the computed results of their request are returned to them. The basic output for a successful job includes three files, the output txt file, the energy minimized mutant and wild type structures: MT\_min PDB file and WT\_min PDB file. If the job failed, then the output includes an error txt file and the user is expected to find the reason for the failure from this file. If the job is still running, the result page will also inform the user that the job is running.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/17/4/547/s1>.

**Acknowledgments:** This work has been supported by a grant from the National Institutes of Health, grant number R01 GM093937. We thank Jesse Kennedy, Taylor Kimmitt and Min Zhang for the help in the development of the SAAMBE webserver.

**Author Contributions:** Marharyta Petukh conducted the research and wrote the paper; Luogeng Dai designed the webserver; Emil Alexov designed the research and wrote the paper.

**Conflicts of Interest:** The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. Counihan, J.L.; Ford, B.; Nomura, D.K. Mapping proteome-wide interactions of reactive chemicals using chemoproteomic platforms. *Curr. Opin. Chem. Biol.* **2016**, *30*, 68–76. [[CrossRef](#)] [[PubMed](#)]
2. Liu, Z.P.; Chen, L. Proteome-wide prediction of protein–protein interactions from high-throughput data. *Protein Cell* **2012**, *3*, 508–520. [[CrossRef](#)] [[PubMed](#)]
3. Tuncbag, N.; Gursoy, A.; Nussinov, R.; Keskin, O. Predicting protein–protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nat. Protoc.* **2011**, *6*, 1341–1354. [[CrossRef](#)] [[PubMed](#)]
4. Nishi, H.; Tyagi, M.; Teng, S.; Shoemaker, B.A.; Hashimoto, K.; Alexov, E.; Wuchty, S.; Panchenko, A.R. Cancer missense mutations alter binding properties of proteins and their interaction networks. *PLoS ONE* **2013**, *8*, e66273.
5. Panchenko, A.R.; Babu, M.M. Editorial overview: Linking protein sequence and structural changes to function in the era of next-generation sequencing. *Curr. Opin. Struct. Biol.* **2015**, *32*, viii–x. [[CrossRef](#)] [[PubMed](#)]
6. Nussinov, R.; Panchenko, A.R.; Przytycka, T. Physics approaches to protein interactions and gene regulation. *Phys. Biol.* **2011**, *8*, 030301. [[CrossRef](#)] [[PubMed](#)]
7. Alexov, E. Protein–protein interactions. *Curr. Pharm. Biotechnol.* **2008**, *9*, 55–56. [[CrossRef](#)] [[PubMed](#)]
8. Kucukkal, T.G.; Yang, Y.; Chapman, S.C.; Cao, W.; Alexov, E. Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. *Int. J. Mol. Sci.* **2014**, *15*, 9670–9717. [[CrossRef](#)] [[PubMed](#)]
9. Teng, S.; Madej, T.; Panchenko, A.; Alexov, E. Modeling effects of human single nucleotide polymorphisms on protein–protein interactions. *Biophys. J.* **2009**, *96*, 2178–2188. [[CrossRef](#)] [[PubMed](#)]
10. Teng, S.; Michonova-Alexova, E.; Alexov, E. Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. *Curr. Pharm. Biotechnol.* **2008**, *9*, 123–133. [[CrossRef](#)] [[PubMed](#)]
11. Zhang, Z.; Miteva, M.A.; Wang, L.; Alexov, E. Analyzing effects of naturally occurring missense mutations. *Comput. Math. Methods Med.* **2012**, *2012*, 805827. [[CrossRef](#)] [[PubMed](#)]
12. Pires, D.E.; Ascher, D.B.; Blundell, T.L. mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **2014**, *30*, 335–342. [[CrossRef](#)] [[PubMed](#)]
13. Moal, I.H.; Moretti, R.; Baker, D.; Fernandez-Recio, J. Scoring functions for protein–protein interactions. *Curr. Opin. Struct. Biol.* **2013**, *23*, 862–867. [[CrossRef](#)] [[PubMed](#)]
14. Ascher, D.B.; Jubb, H.C.; Pires, D.E.V.; Ochi, T.; Higuero, A.; Blundell, T.L. Protein–Protein Interactions: Structures and Druggability. In *Multifaceted Roles of Crystallography in Modern Drug Discovery*; Springer Netherlands: Amsterdam, The Netherlands, 2015.
15. Alexov, E.; Sternberg, M. Understanding molecular effects of naturally occurring genetic differences. *J. Mol. Biol.* **2013**, *425*, 3911–3913. [[CrossRef](#)] [[PubMed](#)]
16. Stefl, S.; Nishi, H.; Petukh, M.; Panchenko, A.R.; Alexov, E. Molecular mechanisms of disease-causing missense mutations. *J. Mol. Biol.* **2013**, *425*, 3919–3936. [[CrossRef](#)] [[PubMed](#)]
17. Bash, P.A.; Singh, U.C.; Brown, F.K.; Langridge, R.; Kollman, P.A. Calculation of the relative change in binding free energy of a protein–inhibitor complex. *Science* **1987**, *235*, 574–576. [[CrossRef](#)] [[PubMed](#)]
18. Petukh, M.; Li, M.; Alexov, E. Predicting binding free energy change caused by point mutations with knowledge-modified MM/PBSA method. *PLoS Comput. Biol.* **2015**, *11*, e1004276. [[CrossRef](#)] [[PubMed](#)]

19. Li, M.; Petukh, M.; Alexov, E.; Panchenko, A.R. Predicting the impact of missense mutations on protein–protein binding affinity. *J. Chem. Theory Comput.* **2014**, *10*, 1770–1780. [[CrossRef](#)] [[PubMed](#)]
20. Peng, Y.; Suryadi, J.; Yang, Y.; Kucukkal, T.G.; Cao, W.; Alexov, E. Mutations in the KDM5C ARID Domain and Their Plausible Association with Syndromic Claes-Jensen-Type Disease. *Int. J. Mol. Sci.* **2015**, *16*, 27270–27287. [[CrossRef](#)] [[PubMed](#)]
21. Petukh, M.; Wu, B.; Stefl, S.; Smith, N.; Hyde-Volpe, D.; Wang, L.; Alexov, E. Chronic Beryllium Disease: Revealing the role of beryllium ion and small peptides binding to HLA-DP2. *PLoS ONE* **2014**, *9*, e111604. [[CrossRef](#)] [[PubMed](#)]
22. Takano, K.; Liu, D.; Tarpey, P.; Gallant, E.; Lam, A.; Witham, S.; Alexov, E.; Chaubey, A.; Stevenson, R.E.; Schwartz, C.E.; *et al.* An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity. *Hum. Mol. Genet.* **2012**, *21*, 4497–4507. [[CrossRef](#)] [[PubMed](#)]
23. Witham, S.; Takano, K.; Schwartz, C.; Alexov, E. A missense mutation in CLIC2 associated with intellectual disability is predicted by *in silico* modeling to affect protein stability and dynamics. *Proteins* **2011**, *79*, 2444–2454. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, Z.; Norris, J.; Schwartz, C.; Alexov, E. *In silico* and *in vitro* investigations of the mutability of disease-causing missense mutation sites in spermine synthase. *PLoS ONE* **2011**, *6*, e20373. [[CrossRef](#)] [[PubMed](#)]
25. Zhang, Z.; Teng, S.; Wang, L.; Schwartz, C.E.; Alexov, E. Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Hum. Mutat.* **2010**, *31*, 1043–1049. [[CrossRef](#)] [[PubMed](#)]
26. Kucukkal, T.G.; Petukh, M.; Li, L.; Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 18–24. [[CrossRef](#)] [[PubMed](#)]
27. Petukh, M.; Kucukkal, T.G.; Alexov, E. On human disease-causing amino acid variants: Statistical study of sequence and structural patterns. *Hum. Mutat.* **2015**, *36*, 524–534. [[CrossRef](#)] [[PubMed](#)]
28. Peng, Y.; Alexov, E. Investigating the linkage between disease-causing amino acid variants and their effect on protein stability and binding. *Proteins* **2016**, *84*, 232–239. [[CrossRef](#)] [[PubMed](#)]
29. Moal, I.H.; Fernandez-Recio, J. SKEMPI: A structural kinetic and energetic database of mutant protein interactions and its use in empirical models. *Bioinformatics* **2012**, *28*, 2600–2607. [[CrossRef](#)] [[PubMed](#)]
30. Dehouck, Y.; Kwasigroch, J.M.; Rooman, M.; Gilis, D. BeAtMuSiC: Prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res.* **2013**, *41*, W333–W339. [[CrossRef](#)] [[PubMed](#)]
31. 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*, W382–W388. [[CrossRef](#)] [[PubMed](#)]
32. Khan, S.; Vihinen, M. Performance of protein stability predictors. *Hum. Mutat.* **2010**, *31*, 675–684. [[CrossRef](#)] [[PubMed](#)]
33. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Hoover, J.; *et al.* ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **2016**, *44*, D862–D868. [[CrossRef](#)] [[PubMed](#)]
34. Landrum, M.J.; Lee, J.M.; Riley, G.R.; Jang, W.; Rubinstein, W.S.; Church, D.M.; Maglott, D.R. ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **2014**, *42*, D980–D985. [[CrossRef](#)] [[PubMed](#)]
35. Papageorgiou, A.C.; Shapiro, R.; Acharya, K.R. Molecular recognition of human angiogenin by placental ribonuclease inhibitor—An X-ray crystallographic study at 2.0 Å resolution. *EMBO J.* **1997**, *16*, 5162–5177. [[CrossRef](#)] [[PubMed](#)]
36. Pan, L.; Deng, X.; Ding, D.; Leng, H.; Zhu, X.; Wang, Z. Association between the Angiogenin (ANG) K171 variant and amyotrophic lateral sclerosis risk in Caucasian: A meta-analysis. *Neurol. Sci.* **2015**, *36*, 2163–2168. [[CrossRef](#)] [[PubMed](#)]
37. Higashida, K.; Tsukie, T.; Fukazawa, H.; Fujikura, M.; Ono, S. Immunohistochemical studies of angiogenin in the skin of patients with amyotrophic lateral sclerosis. *J. Neurol. Sci.* **2013**, *326*, 18–23. [[CrossRef](#)] [[PubMed](#)]
38. Zou, Z.Y.; Wang, X.N.; Liu, M.S.; Sun, Q.; Li, X.G.; Cui, L.Y.; Kong, J. Identification of a novel missense mutation in angiogenin in a Chinese amyotrophic lateral sclerosis cohort. *Amyotroph. Lateral Scler.* **2012**, *13*, 270–275. [[CrossRef](#)] [[PubMed](#)]
39. Dalby, A.R.; Tolan, D.R.; Littlechild, J.A. The structure of human liver fructose-1,6-bisphosphate aldolase. *Acta Crystallogr. D Biol. Crystallogr.* **2001**, *57*, 1526–1533. [[CrossRef](#)] [[PubMed](#)]

40. Coffee, E.M.; Tolan, D.R. Mutations in the promoter region of the aldolase B gene that cause hereditary fructose intolerance. *J. Inherit. Metab. Dis.* **2010**, *33*, 715–725. [[CrossRef](#)] [[PubMed](#)]
41. Davit-Spraul, A.; Costa, C.; Zater, M.; Habes, D.; Berthelot, J.; Broue, P.; Feillet, F.; Bernard, O.; Labrune, P.; Baussan, C. Hereditary fructose intolerance: Frequency and spectrum mutations of the aldolase B gene in a large patients cohort from France—Identification of eight new mutations. *Mol. Genet. Metab.* **2008**, *94*, 443–447. [[CrossRef](#)] [[PubMed](#)]
42. Gruchota, J.; Pronicka, E.; Korniszewski, L.; Stolarski, B.; Pollak, A.; Rogaszewska, M.; Ploski, R. Aldolase B mutations and prevalence of hereditary fructose intolerance in a Polish population. *Mol. Genet. Metab.* **2006**, *87*, 376–378. [[CrossRef](#)] [[PubMed](#)]
43. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. The worldwide Protein Data Bank (wwPDB): Ensuring a single, uniform archive of PDB data. *Nucleic Acids Res.* **2007**, *35*, D301–D303. [[CrossRef](#)] [[PubMed](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).