# Evaluation of Computational Tools for Thermodynamics and Structural Analysis of Protein Stability upon Point Mutation Prediction

**Alex D. Camargo [1],\*, Adriano V. Werhli [1] and Karina S. Machado [1]**

[1]   Universidade Federal o Rio Grande, Av. Itália – KM 8;      E-mails: alexcamargo@furg.br, werhli@furg.br, karina.machado@furg.br;

**\***   Author to whom correspondence should be addressed; E-mail: alexcamargo@furg.br; Tel.: +55 53-3233-6623; Fax: +55 53-3233-6652.

---

**Abstract:** In Bioinformatics, review of the state of the art about computational tools, including the interpretation of generated outputs and the restrictions of each software, contributes for choosing the best application to a specific problem. This way, an important research topic is the study of the impact of mutations in the treatment of complex diseases. Mutations have fundamental roles in evolution by introducing diversity into genomes, however, they can affect protein stability. Actually, researchers need accurate computational tools for prediction of how single point amino acid mutations affect the stability of a protein structure. Recent works show significant advances in predicting stability upon point mutation. This paper presents an evaluation of computational tools for thermodynamics and structural analysis of protein stability upon point mutation prediction. We choose to evaluate for thermodynamic analysis the software CUPSAT (Cologne University Protein Stability Analysis Tool) and mCSM (mutation Cutoff Scanning Matrix), and for structural analysis the software FoldX and Modeller. These software were chosen these software due to their popularity in this type of analyzes. In our proposed evaluation we verified the software outputs and evaluated the proximity to experimental results. As a case study we selected a set of 25 proteins extracted from: (i) MutaProt, which analyses pairs of PDB files whose members differ in one, or two, amino acids; (ii) ProTherm, database that contain experimentally determined thermodynamic parameters of protein stability. Each mutation in the datasets has attributes, as: PDB code, mutation, solvent accessibility, pH value, temperature and energy change ($\Delta\Delta G$). A stability prediction model was successfully created, and the majority of the point mutations were predicted successfully having a high correlation and low standard error.

## 1. Introduction

Mutations have fundamental roles in evolution of organisms by introducing diversity into genomes [7]. Methods for protein structure prediction have advanced rapidly in recent years. There are a wide range of strategies for estimating protein energy, most of the methods are based on the statistical analysis of known protein structures [2]. The core functionality of these computational methods is an energy function that calculates the free energy of the system [8].

The understanding about the mutations that affect protein stability often resulting in diease is an important subject. Accurate prediction of point mutations effect on protein stability that can appear upon mutagenesis is fundamental when it is necessary to understand the structure-function relationship of a protein or in the cases where a new protein needs to be designed [4].

In addition to the natural variations in single mutations on proteins among organisms, bioinformaticians frequently introduce single amino acid residue replacements by site-directed mutagenesis in the laboratory to explore structural and functional features of proteins [4].

Several recent papers focused on testing sophisticated potential functions for conformational search and development of new scoring functions for side-chain modeling, reporting improved accuracy compared to earlier approaches [8, 10].

This paper aims at evaluating computational tools for thermodynamics and structural analysis of protein stability upon point mutation prediction. We choose to evaluate for thermodynamic analysis the softwares CUPSAT (Cologne University Protein Stability Analysis Tool) [6] and mCSM (mutation Cutoff Scanning Matrix) [7]. For structural analysis the software FoldX [9] and Modeller [3] were chosen. Our choice for these tools is because they are commonly used in these kind of analyzes.

As a case study we selected a set of 50 proteins extracted from: (i) MutaProt, which analyses pairs of PDB files whose members differ in one, or two, amino acids [2]; (ii) ProTherm, a database that contain experimentally determined thermodynamic parameters of protein stability [5].

This paper is organized as follows: the obtained results and discussion are described in Section 2. Section 3 introduces the protein stability predictors. Finally, Section 4 presents the conclusion.

## 2. Results and Discussion

A good computational biology method to predict stability changes upon mutation will help in designing new or altered proteins with specific levels of stability, enzymatic activity and binding to other molecules [8]. However, the number of false positives and false negatives returned by the programs, is generally substantial [4].

In this study, the prediction performances were evaluated based on accuracy measure. Accuracy is defined as a percentage of correctly identified mutations on the total number of mutations.

For evaluation of computational tools for thermodynamics analysis we use the difference in the calculated free energies ($\Delta\Delta G$) between the mutant and the wild-type, well with it was observed structural changes by RMSD (Root-Mean-Square Deviation) also mutant and the wild-type. RMSD values are considered as reliable indicators of variability when utilized to

very similar proteins, like alternative conformations of the same protein [1].

For the set of selected proteins, none of the methods was able to accurately predict ΔΔGs for all mutations, as there is a significant deviation between experimental and calculated values. As seen in Table 1, often we are more interested to know whether a mutation is stabilizing or destabilizing, than to obtain the exact ΔΔG value.

In thermal experiments, it was observed that for the 25 mutations 88% were correctly predicted by mCSM to be either stabilizing or destabilizing. However, with a lower score, CUPSAT hit 72% of the predictions.

Table 2 shows the accuracy of the methods used for the three-dimensional structure prediction upon mutation point. The testing set consists of 25 pairs of known protein structures differing by a single mutation. The RMSD average between experimental and predicted structures values were 0,4Å for both methods, FoldX and Modeller.

**Table 1.** Accuracy of predicted change in ΔΔG upon point mutation.

| Method | Accuracy |
|--------|----------|
| CUPSAT | 72% |
| mCSM | 88% |

**Table 2.** Accuracy of predicted change in RMSD upon point mutation.

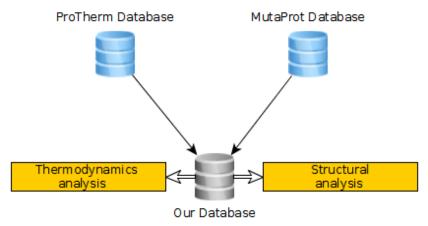| Method | Average | Standard deviation |
|--------|---------|--------------------|
| FoldX | 0,402Å | 0.322 |
| Modeller | 0,404Å | 0.323 |



**Figure 1.** After collecting information from the used two datasets, we integrated the results into a non redundant dataset of protein stability change effects data.

## 3. Materials and Methods

In the current study, we chose two different methods that were previously reported as being able to predict the effects in protein stability (ΔΔG) upon mutation: CUPSAT [6] and mCSM [7]. Also, we tested two other methods for the modeling of point mutation in protein structures: FoldX [9] and Modeller [3].

CUPSAT (Cologne University Protein Stability Analysis Tool) is a web tool to analyse and predict protein stability changes upon single amino acid point mutations [6].

The approach, called mutation Cutoff Scanning Matrix (henceforth called mCSM), encodes distance patterns between atoms to represent protein residue environments [7].

FoldX is an empirical force field that was developed for evaluation of the effect of mutations on the stability, folding and dynamics of proteins and nucleic acids [9].

Modeller is a method to model point mutations in protein structures with two cycles of the conjugate gradients: molecular dynamics with simulated annealing, and conjugate gradients phases [3].

The effects of mutations on protein stabilities were predicted using the default parameters of the analyzed tools. As shown in Figure 1, for the purpose of this job a data set of mutations was compiled from two sources. The first, called MutaProt, was a list of single mutations that was published previously by Eyal et al. [2]. The second set of single mutations was obtained from the ProTherm database [5]. Some of the mutations in the datasets were listed several times. Therefore, in such situations, it was filtered to exclude any mutation that is listed more than once, or has incomplete information.

## 4. Conclusions

This paper evaluated the accuracy of common methods used to predict stability changes in proteins upon mutation. In our proposed evaluation we have verified the software outputs and analyzed the proximity to experimental results. In general there was good agreement between the methods in predicting the direction of change when compared with the experimental data.

The validation tests with mCSM showed that 90% of the mutations were correctly predicted for thermal stability. To evaluate the structural rearrangements upon mutations we calculated the RMSD of backbone movements upon single mutation. The results from both methods were identical and, in addition, Modeller is relatively faster than many of the FoldX method.

All the tested computational methods showed a correct trend in their predictions, but failed in providing the precise values. In summary, the current computational methods are clearly good enough for most of the tasks they are used for.

## Acknowledgments

## Author Contributions
Guarantors of integrity of entire study, all authors.

**Conflicts of Interest**

The authors declare no conflict of interest.

**References**

1. Carugo, O., & Pongor, S. (2001). A normalized root- mean- spuare distance for comparing protein three- dimensional structures. Protein science, 10(7), 1470-1473.
2. Eyal, E., Najmanovich, R., Sobolev, V., & Edelman, M. (2001). MutaProt: a web interface for structural analysis of point mutations. Bioinformatics, 17(4), 381-382.
3. Feyfant, E., Sali, A., & Fiser, A. (2007). Modeling mutations in protein structures. Protein Science, 16(9), 2030-2041.
4. Khan, S., & Vihinen, M. (2010). Performance of protein stability predictors. Human mutation, 31(6), 675-684.
5. Kumar, M. S., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H., & Sarai, A. (2006). ProTherm and ProNIT: thermodynamic databases for proteins and protein–nucleic acid interactions. Nucleic Acids Research, 34(suppl 1), D204-D206.
6. Parthiban, V., Gromiha, M. M., & Schomburg, D. (2006). CUPSAT: prediction of protein stability upon point mutations. Nucleic acids research, 34(suppl 2), W239-W242.
7. Pires, D. E., Ascher, D. B., & Blundell, T. L. (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics, 30(3), 335-342.
8. Potapov, V., Cohen, M., & Schreiber, G. (2009). Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. Protein Engineering Design and Selection, 22(9), 553-560.
9. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: an online force field. Nucleic acids research, 33(suppl 2), W382-W388.
10. Tian, J., Wu, N., Chu, X., & Fan, Y. (2010). Predicting changes in protein thermostability brought about by single-or multi-site mutations. Bmc Bioinformatics, 11(1), 370.