FireProt: web server for automated design of thermostable proteins

Miloš Musil[123], Jan Štourač[12], Jaroslav Bendl[123], Jan Brezovský[12], Zbyněk Prokop[12], Jaroslav Zendulka[3], Tomáš Martínek[3], David Bednář[12], and Jiří Damborský[12]

[1] Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Faculty of Science, Masaryk University
Kamenice 5, Bld. A13, 625 00 Brno, Czech Republic
imusilm@fit.vutbr.cz
[2] International Centre for Clinical Research, St. Anne's University Hospital Brno
Pekařská 53, 656 91 Brno, Czech republic
[3] Department of Information Systems, Faculty of Information Technology, Brno University of Technology
Božetěchova 2, 612,66 Brno

Abstract. There is a continuous interest in increasing proteins stability to enhance their usability in numerous biomedical and biotechnological applications. Only a single-point mutations with a small impact on protein stability are typically predicted with the existing tools, while more complex multiple-point mutants are necessary for useful stabilization. Here, we present FireProt, a web server for automated design of multiple-point thermostable mutants that combines structural and evolutionary information in its calculation core. The server is complemented with interactive, easy-to-use interface that allows users to directly analyze and optionally modify designed thermostable mutants.

Keywords: protein stability · thermostable proteins · hybrid method · protein engineering

1 Introduction

Proteins are widely used in numerous biomedical and biotechnological applications. However, naturally occurring proteins cannot usually withstand the harsh industrial environment, since they are mostly evolved to function at mild conditions. Therefore, there is a continuous interest in increasing proteins stability to enhance their industrial usability. A number of in silico tools for the prediction of the effect of mutations on proteins have been recently developed to improve their stability, thus surpassing their natural limitations. However, due to the potentially antagonistic effect of mutations, only single-point mutations are usually predicted in silico and have to be followed by laborious and costly protein expression, purification and characterization. Single-point mutations typically

enhance the melting temperature of target proteins only by units of degree. A much higher degree of stabilization can be achieved by constructing multiple-point mutants. Unfortunately, the automated design of multiple-point mutants is uneasy, due to the existence of the potential antagonistic effects of individual mutations. To address these issues, FireProt web was developed [1].

2 Dataset

One of the most crucial properties of every predictive tool is the construction of the reliable training dataset. Our dataset was mainly obtained from the ProTherm database [2]. ProTherm is the most extensive collection of thermodynamic parameters such as Gibbs free energy, heat capacity or enthalpy, measured as a difference between wild-type and mutant protein.

ProTherm was lastly updated in February 2013 and up to date it contains nearly 26,000 entries of both single- and multiple-point mutants designed over 740 unique proteins. However, even though ProTherm database is the most common source of data used for the training of the existing tools, in its current state it suffers from several serious issues such as missing values, opposite signs or various inconsistencies in measures. To deal with this issues, ProTherm had to be manually currated and all the redundancies were discarded during the process of the final dataset construction. Furthermore, we have also omitted mutations with the measured change of Gibbs free energy from the interval of (-0.5,0.5) since the average experimental error of measurement is about 0.48 kcal/mol [3].

Our training dataset was later expanded with experimental data obtained from the HotMuSiC dataset [4] and the final set consists of 1,573 mutations (349 stabilizing, 1,224 destabilizing) designed over 103 unique proteins. Included proteins were completed with corresponding 3D structure in the highest quality available (based on resolution and R-factor).

3 FireProt method

The web server integrates sixteen computational tools and utilizes both sequence and structural information in the process. At first, potentially stabilizing single-point mutations are identified via two separate branches: one relying on the estimation of the change of free energy upon mutation and second utilizing back-to-consensus approach. The energy-based approach is employing FoldX [5] and Rosetta [6] tools to estimate free energy via the statistical and physical force fields. Furthermore, preceding filters accelerate the calculation by omitting the high risk residues – ones that are located in the highly conserved or correlated regions of the protein. It was observed that functional and structural constraints in proteins generally lead to the conservation of amino acid residues and similarly, correlated residues ordinarily help to maintain protein function, folding and stability. Mutations conducted on these positions are therefore considered unsafe and excluded from the design of stable proteins. The second approach is based on the information obtained from multiple sequence alignment. The

most common amino acid in each position of protein sequence often provides a non-negligible effect on protein stability. Thus, FireProt implements majority and frequency ratio approach to identify mutations at positions where wild-type amino acid differs from the most prevalent one.

In total, three protein designs are provided by FireProt strategy. The first design includes only the mutations from energy-based approach, the second contains the mutations suggested by the evolution-based approach and the third is the combination of both. Naturally, because of potentially antagonistic effect between individual mutations, these mutations cannot be combined blindly. To avoid possible clashes, FireProt strategy is trying to minimize antagonistic effects by utilizing Rosetta. The physical force field is used to evaluate all pairs of single-point mutations within the range of 10 Å and subsequently the mutations are introduced into the stable protein via the edge- based graph algorithm. In this way, the colliding mutations are excluded from protein and algorithm stops once there are no mutations left or the stabilizing effect of analyzed pair of mutations drops below defined threshold. Detail worklow of the FireProt web can be found in the Figure

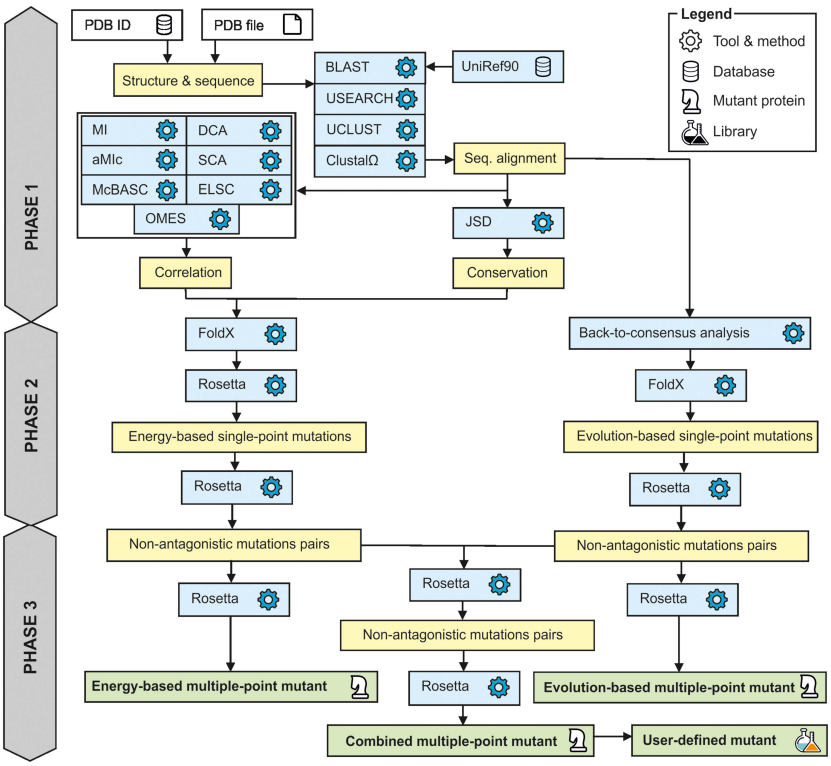

Fig. 1. Complete workflow of the FireProt web.

A high time demands of Rosetta analysis were one of the most excruciating issues with the FireProt protocol. Even with the application of preceding filters over 100 mutations is usually left for precise, but slow, Rosetta calculations. For this reason, several force fields and protocols were evaluated on the dataset of 1,573 mutations. Based on the results of these evaluations, the best trade-off between the time requirements and precision was selected. Moreover, the FireProt time demands were drastically reduced by massive parallelization via the use of MetaCentrum. These measures have lowered the calculation time from several weeks to a single day.

4 Conclusions

FireProt is a web server that provides users with a one-stop-shop solution for the design of thermostable multiple-point mutants. It combines energy- and evolution-based information in its calculation core and all the protocols were highly optimized to increase the calculation speed, while maintaining the prediction accuracy. The server is complemented by an easy-to-use graphical interface that allows users to interactively analyze individual mutations selected as a part of energy- of evolution-based approach together with the ability to design their own multiple-point mutants on top of our robust strategy. The web server is freely available at `https://loschmidt.chemi.muni.cz/fireprot/`.

References

1. Musil, M., Stourac, J.,Bendl, J., Brezovsky, J.,Prokop, Z., Zendulka, J.,Martinek, T., Bednar, D., Damborsky, J.: FireProt: web server for automated design of thermostable proteins. Nucleic Acids Research **45**(W1), W393–399 (2017)
2. Bava, KA., Gromiha, MM., Uedaira, H., Kitajima, K., Sarai, A.: ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Research **32**, D120–121 (2004)
3. Khatum, J., Khare, SD., Dokholyan, NV.: Can contact potentials reliably predict stability of proteins? Journal of Molecular Biology **366**(5), 1223–1238 (2004)
4. Pucci, F., Bourgeas, R., Rooman, M.: Predicting protein thermal stability changes upon point mutations using statistical potentials: introducing HoTMuSiC. Scientific Reports **6**, 23257 (2016)
5. Guerois, R., Nielsen, JE., Serrano, L.: Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. Journal of Molecular Biology **320**, 369–387 (2002)
6. Kellogg, EH., Leaver-Fay, A., Baker, D.: Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins **79**, 830–838 (2011)