FireProt: web server for automated design of thermostable proteins

Milos Musil[1,2,3,†], Jan Stourac[1,3,†], Jaroslav Bendl[1,2,3], Jan Brezovsky[1,3], Zbynek Prokop[1,3], Jaroslav Zendulka[2,4], Tomas Martinek[1,2,4], David Bednar[1,3,*] and Jiri Damborsky[1,3,*]

[1]Loschmidt Laboratories, Department of Experimental Biology, Masaryk University, Brno, Czech Republic, [2]Department of Information Systems, Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic, [3]International Centre for Clinical Research, St. Anne's University Hospital Brno, Brno, Czech Republic and [4]Centre of Excellence IT4Innovations, Technical University Ostrava, Ostrava

ABSTRACT

There is a continuous interest in increasing proteins stability to enhance their usability in numerous biomedical and biotechnological applications. A number of *in silico* tools for the prediction of the effect of mutations on protein stability have been developed recently. However, only single-point mutations with a small effect on protein stability are typically predicted with the existing tools and have to be followed by laborious protein expression, purification, and characterization. Here, we present FireProt, a web server for the automated design of multiple-point thermostable mutant proteins that combines structural and evolutionary information in its calculation core. FireProt utilizes sixteen tools and three protein engineering strategies for making reliable protein designs. The server is complemented with interactive, easy-to-use interface that allows users to directly analyze and optionally modify designed thermostable mutants. FireProt is freely available at http://loschmidt.chemi.muni.cz/fireprot.

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 (1). Protein engineering has revolutionized the utilization of naturally available proteins for different industrial applications by improving various protein features such as stability, activity or enantioselectivity to surpass their natural limitations. Protein stability is generally strongly correlated with its expression yield (2), half-life (3),

serum survival time (4) and performance in the presence of denaturing agents (5). Thus, stability is one of the key determinants of proteins applicability in biotechnological processes.

In the ideal case, the saturation mutagenesis would be applied to evaluate every possible mutation on every position of the engineered protein (6). However, such a search space would be enormous and the experimental evaluation can delay the design of truly thermostable protein for months or even years. Therefore, there are demands for effective and precise predictive computation of protein stability. To satisfy this goal a number of *in silico* tools have been developed recently. Some of these tools such as EASE-MM (7), I-Mutant (8) or mCSM (9) are based on machine learning techniques. Others are using so-called energetic functions. These programs can be further categorized into two groups. The first group utilizes a physical effective energy function for simulating the fundamental forces between atoms and is represented by the programs like Rosetta (10) and Eris (11). The second group is based on statistical potentials for which the energies are derived from frequencies of residues or atom contacts reported in the datasets of experimentally characterized protein mutants, e.g. PopMuSiC (12) and FoldX (13). 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 by units of degree (3,14). A much higher degree of stabilization can be achieved by constructing multiple-point mutants (15). We have recently developed the FireProt (16), combining energy- and evolution-based approaches for reliable design of stable multiple-point mutants. The protocol includes several preceding filters that accelerate the calculation by omitting potentially deleterious mutations. FireProt is currently

available only in a stand-alone format and requires extensive experience in bioinformatics to carry out all necessary steps of the work flow. Currently, we are aware of only one server for design of stable multiple-point mutants - PROSS (17), utilizing Rosetta modeling and phylogenetic sequence information in its computation core.

Here, we present a web version of FireProt for the automated design of thermostable proteins. FireProt integrates sixteen computational tools and utilizes both sequence and structural information. FireProt web server provides users with thermostable proteins, constructed by three distinct strategies: (i) evolution-based approach, utilizing back-to-consensus analysis; (ii) energy-based approach, evaluating change in free energy upon mutation and (iii) combination of both evolution-based and energy-based approaches. In our view, it is very important to have this integrated approach, since phylogenetic analysis enables identification of the mutations stabilized by entropy, which cannot be predicted by force field calculations (Beerens *et al.*, under review). The server allows users to include preferred mutations into the thermostable protein, to generate corresponding structures and sequences for gene syntheses. Compared to the previously published FireProt protocol (16), minimum effort and no bioinformatics knowledge is required from users to calculate and analyze the results. Furthermore, all input parameters and computational protocols were optimized to minimize otherwise highly time demanding procedure. The server was complemented with a graphical interface allowing users to directly analyze the protein of interest and design multiple-point mutants.

MATERIALS AND METHODS

The basic workflow of FireProt strategy is outlined in Figure 1. In order to design a highly reliable thermostable multiple-point mutant, a protein defined by the user is annotated using several prediction tools and databases (Phase 1). With this knowledge in hand, energy- and evolution-based approach is applied to assemble a list of potentially stabilizing single-point mutations (Phase 2). Finally, three multiple-point mutants are generated in an additive manner, while removing potentially antagonistic effects of mutations (Phase 3).

Phase 1: Annotation of the protein

Initially, the user is requested to specify the protein structure, either by providing its PDB ID or by uploading a user-defined PDB file. The biological assembly of the target protein is then automatically generated by the MakeMultimer tool (http://watcut.uwaterloo.ca/tools/makemultimer/). Sequence homologs are obtained by performing a BLAST search (18) against the UniRef90 database (19), using the target protein sequence as an input query. Identified homologs are then aligned with the query protein using USEARCH (20), while sequences whose identity with the query is below or above the user defined thresholds (default: 30 and 90%) are excluded from the list of homologs. The remaining sequences are clustered using UCLUST (20), with a 90% identity threshold to remove close homologs. The cluster representatives are sorted based on the BLAST

query coverage and by default, the first 200 of them are used to create a multiple sequence alignment with Clustal Omega tool (21). The multiple sequence alignment is used to: (i) estimate the conservation coefficient of each residue position in the protein based on the Jensen–Shannon entropy (22); (ii) identify correlated positions employing a consensual decision of the OMES (23), MI (24), aMIc (25), DCA (26), SCA (27), ELSC (28), McBASC (29) and (iii) analyze amino acid frequencies at individual positions within the protein.

Phase 2: Prediction of single-point mutations

In accordance with the original FireProt protocol, 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 first, energy-based approach is employing FoldX and Rosetta tools that performed best on our testing dataset. Preceding filters accelerate the calculation by omitting potentially deleterious mutations. Prior to the identification of the single-point mutations itself, the target protein structure is amended and minimized. FoldX protocol is utilized to fill in the missing atoms in the residues and patched structure is consequently minimized with Rosetta minimization module. Conserved and correlated positions are immediately excluded from further analysis. It was observed that functional and structural constraints in proteins generally lead to the conservation of amino acid residues (30–33). Similarly, correlated residues ordinarily help to maintain protein function, folding or stability (34–36). Mutations conducted on these positions are therefore considered unsafe by current FireProt strategy, even though there is certainly a space for more sophisticated treatment of correlated positions, which will be further developed in future versions of FireProt server.

The remaining positions are subjected to saturation mutagenesis by using FoldX tool. Mutations with predicted ddG over given threshold (default: –1 kcal/mol) are steered away and rest is forwarded to Rosetta calculations. Finally, the mutations predicted by Rosetta as strongly stabilizing (default cut-off: –1 kcal/mol) are tagged as potential candidates for the design of the multiple-point mutants.

A high time demands of Rosetta analysis were one of the most excruciating issues with the original FireProt protocol. Even with the application of filters over 100 mutations was usually left for precise, but slow, Rosetta calculations. For this reason, we have evaluated several force fields and Rosetta protocols with the newly assembled dataset containing 1573 mutations from ProTherm database (37) and HotMuSiC dataset (38). Based on the results of the evaluations, the best trade-off between the time requirements and precision was selected. With Rosetta protocol 3, we have achieved more than tenfold increase in calculation speed while preserving high prediction accuracy. Details on dataset construction and protocols evaluation can be found in the Supplement 1 (Supplementary Tables S1–S5).

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
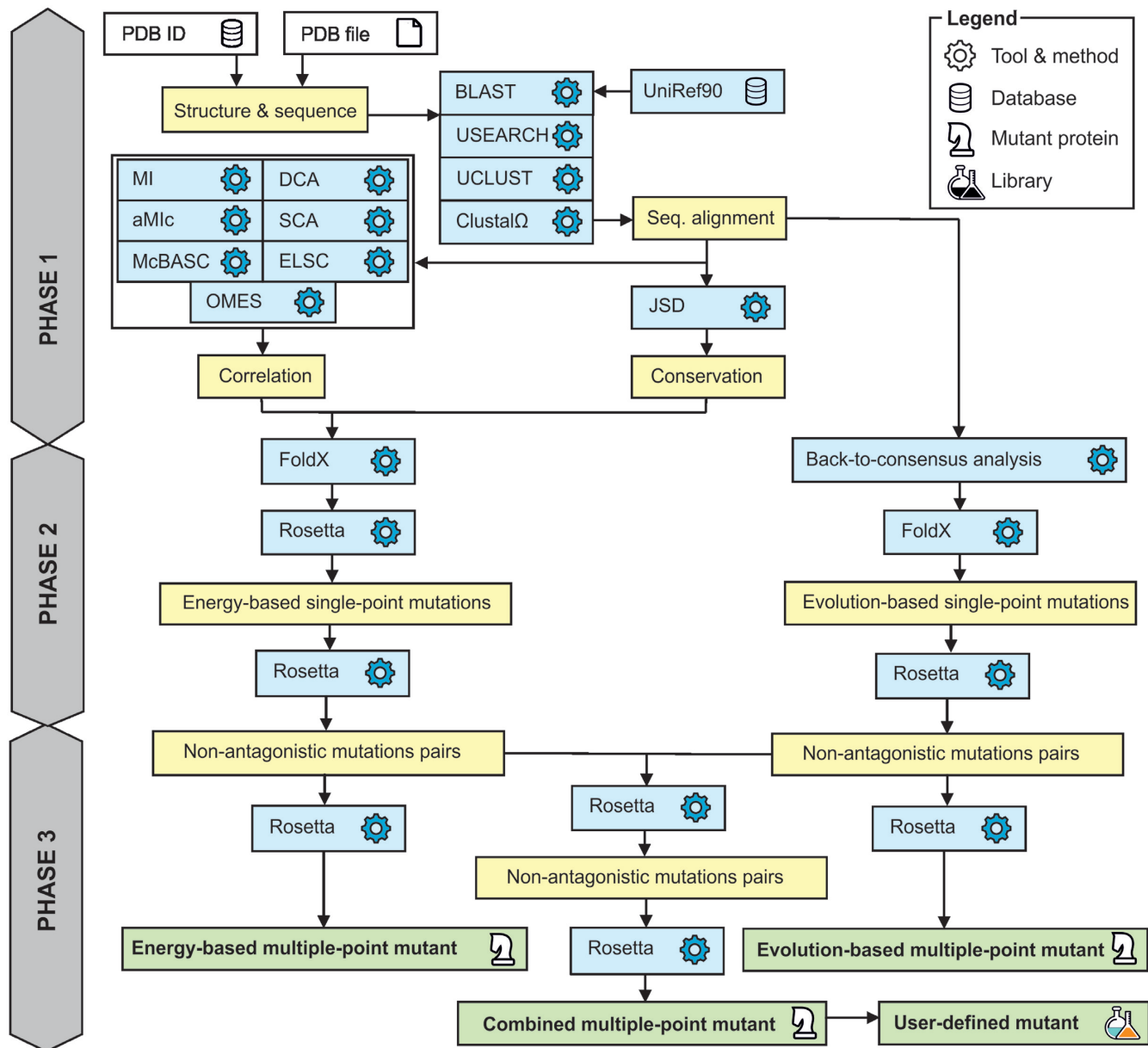
Figure 1. Workflow of FireProt strategy.

provides a non-negligible effect on protein stability ([39–42]). Therefore, FireProt implements majority and frequency ratio approach to identify mutations at positions where the wild-type amino acid differs from the most prevalent one. By default, the single out mutations are located in the positions where the consensus residue is present in at least 50% of all analyzed sequences (majority method) or where consensus residue frequency is 40% and is at least five times more frequent than the wild-type amino acid (frequency ratio method). These thresholds were chosen in accordance to the previously published HotSpot Wizard method ([43]). Selected mutations are evaluated by FoldX and the stabilizing ones are listed as candidate mutations for the engineering of multiple-point mutant.

Phase 3: Design of thermostable protein

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 effects between individual mutations, we cannot combine individual mutations blindly.

To avoid possible clashes, FireProt strategy is trying to minimize antagonistic effects by utilizing Rosetta. In the first step, all pairs of single-point mutations within the range of 10 Å are evaluated separately for energy- and evolution-based approach. Once change in free energy is obtained for all residue pairs, FireProt starts to introduce them into the multiple-point mutant in the order based on their predicted

stability, excluding the mutations that are colliding with already included mutations. Algorithm stops once there are no mutations left or the stabilizing effect of analyzed pair drops below defined threshold.

Upon the completion of previous step, procedure is repeated this time considering only the pairs between the mutations chosen for the construction of energy- and evolution-based mutants. Finally, structures of all three mutants are modeled using the Rosetta protocol 16.

DESCRIPTION OF THE WEB SERVER

Input

The only required input to the web server is a tertiary structure of the protein of interest, provided either as a PDB ID or a user-defined PDB file. The user can then choose a predefined biological unit generated by the MakeMultimer tool or manually select chains for which the calculation should be performed. The calculations can be configured in either basic or advanced mode.

In the basic mode, user is allowed to change the setting of BLAST search and alignment construction. The advanced mode expands the list of modifiable parameters by the ones connected with: (i) the identification of consensus residues by majority and frequency ratio approach, (ii) the thresholds used by FoldX and Rosetta prediction tools and (iii) the decision threshold employed in the consensual analysis of correlated positions. Advanced mode allows expert users to fine-tune the parameters of calculation according to studied systems. However, the presented default values are optimized to provide reliable results for most of the systems and we therefore do not advice their change in the general scenarios.

Output

Upon submission, a unique identifier is assigned to each job to track the calculation and the 'Results browser' informs the user about the status of the individual steps in the FireProt workflow (Figure 2B). Once the job is finished, users can either directly download the results in the .zip archive or navigate themselves into the 'Results page' for further analysis. The 'Results page' is intuitively organized into several panels as described below.

Protein visualization. The wild-type and the mutant structure is interactively visualized in the web browser (Figure 2D) utilizing the Jsmol applet (http://wiki.jmol.org/index.php/JSmol). Users can switch between different protein visualization styles and also highlight selected amino acids in the protein structure. Residues that were included into energy-based mutant are colored in orange, evolution-based mutations are in blue and all other residues are in gray. User selected residues that were not part of any mutant are underlined in red.

Mutant overview. The 'Mutant overview' panel is organized into four tabs (Figure 2A). The first three tabs provide information about mutations included into combined, energy-based and evolution-based mutant. The checkbox,

allowing users to visualize the chosen residues in Jsmol applet, can be found in each row together with all data relevant for a given computational approach. The last tab contains the list of all residues in the wild-type structure. While 'wild-type' tab is active, the wild-type structure is visualized in Jsmol applet instead of the mutated one and the user is allowed to introduce user-defined mutations into multiple-point mutant via the 'plus' icon in the last column.

General information. The 'FireProt protocol design' panel provides users with general information about the target protein and the designs constructed by FireProt strategy, such as a number of mutations and estimated change in free energy (Figure 2C).

Mutant designer. The 'Mutant designer' panel allows the user to design own multiple-point mutant by managing mutations divided into energy- and evolution-based subset. If all mutations in the subset have their predicted energy values assigned, a total change in Gibbs free energy is immediately estimated assuming simple additivity. Users can also generate an amino acid sequence from the designed multiple-point mutant that combines mutations included into energy- and evolution-based subsets. All prepared designs can be downloaded in one .zip archive (Figure 2E).

EXPERIMENTAL VALIDATION

The original FireProt strategy was experimentally verified with three proteins (haloalkane dehalogenase DhaA, PDB ID 4E46; γ-hexachlorocyclohexane dehydrochlorinase LinA, PDB ID 3A76; and fibroblast growth factor 2, PDB ID 4OEE) and provided respective stabilization of proteins $\Delta T_m = 25$, 21 and 15°C (Table 1). The original protocol was modified to enable fully automated calculation at the reasonable time, while maintaining high prediction accuracy (Supplementary Table S6). Prediction of eight multiple-point mutants using this modified protocol was validated using the data of FRESCO (44) and identified mutations were compared with another online protein stabilization tool PROSS (17). FireProt and PROSS showed similar predictive power, correctly identifying 29 and 20 potentially stabilizing positions, respectively (Supplementary Table S7).

CONCLUSIONS AND OUTLOOK

FireProt is a web server that provides users with a one-stop-shop solution for the design of thermostable multiple-point mutant proteins. In comparison with the standalone FireProt strategy (16), all default parameters and computational protocols were optimized to increase the calculation speed, while maintaining the prediction accuracy. The designs produced by the FireProt workflow were experimentally verified and thus users can obtain highly reliable thermostable proteins with minimal experimental effort. 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- or evolution-based approach together with the ability to design their own multiple-point mutants on top of our robust strategy.
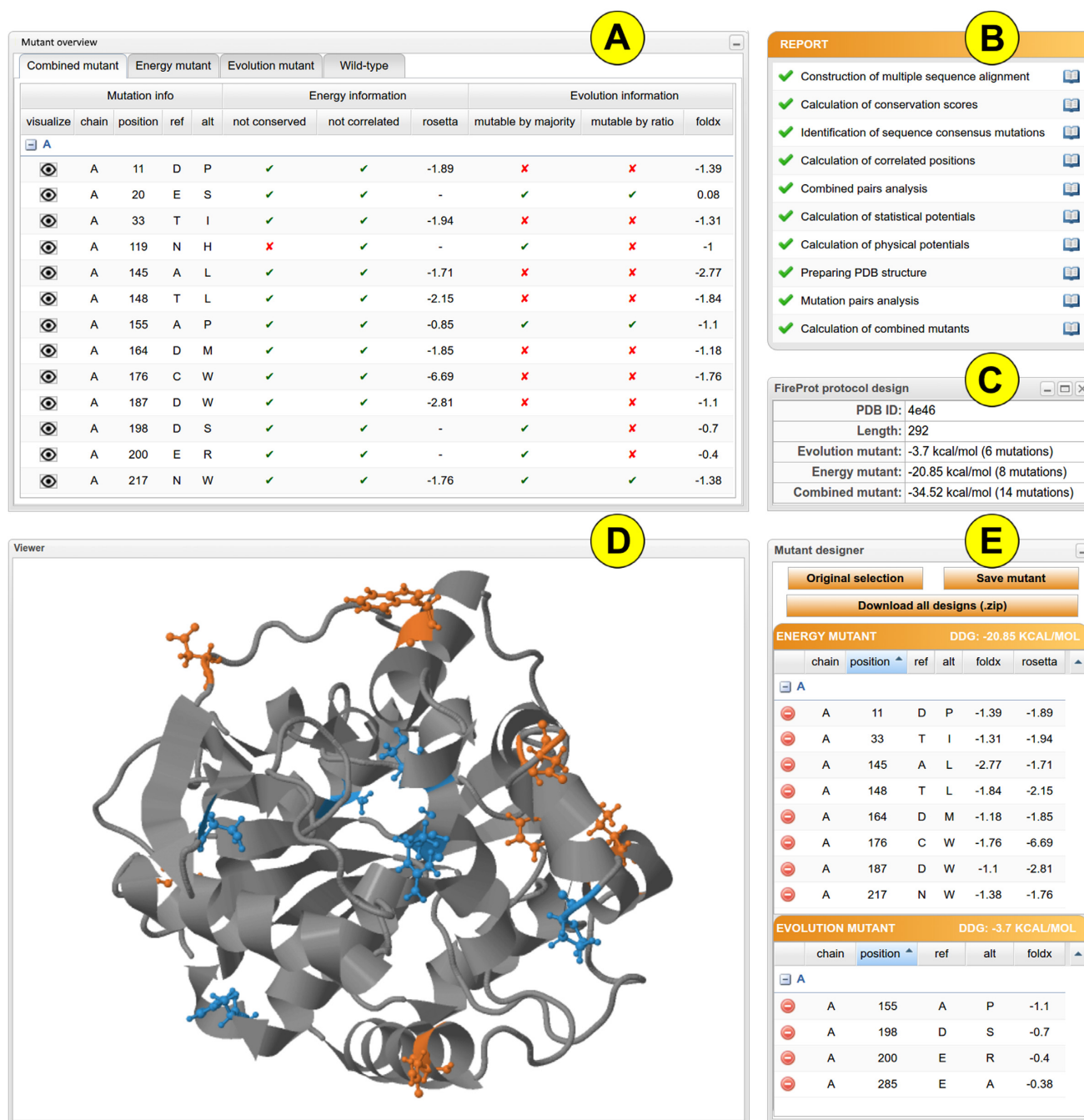
Figure 2. FireProt's graphical user interface showing the results obtained for the haloalkane dehalogenase DhaA (PDB ID: 4e46). (**A**) The 'Mutant overview' panel provides a list of mutations introduced into protein structure. (**B**) The 'Report' panel shows the status of calculation in the individual steps of the computational pipeline. (**C**) The 'Protocol design' panel provides general information about FireProt designs. (**D**) The JSmol ´Viewer´ allows interactive visualization of the protein. (**E**) The 'Mutant designer' panel enables manual adjustment of a new combined mutant.

Table 1. Experimental validation of FireProt strategy

Protein PDB ID	Energy-based mutations	Evolution-based mutations	ΔT_m [°C]
4E46	8	3	+25
3A76	4	3	+21
4OEE	4	2	+15

The automation of the whole procedure makes the process of the design of thermostable proteins accessible to users without any prior expertise in bioinformatics since it eliminates the need to select, install and evaluate tools, optimize their parameters, and interpret intermediate results. However, the energy-based approach of the FireProt strategy depends on the quality of provided protein structure and therefore the prediction accuracy might be compromised in the case of low-resolution structures or homology models.

In the future, we plan to implement new strategies such as a design based on the analysis of correlated positions that would contribute to the construction of the final combined mutant, elimination of highly flexible regions and introduction of disulfide bridges. Also, we plan to equip FireProt with several new filters, e.g. exclusion of the amino acids located in the close neighborhoods of the active sites or the ones participating in oligomerization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Modarres,H.P., Mofrad,M.R. and Sanati-Nezhad,A. (2016) Protein thermostability engineering. *RSC Adv.*, **6**, 115252–115270.
2. Ferdjani,S., Ionita,M., Roy,B., Dion,M., Djeghaba,Z., Rabiller,C. and Tellier,C. (2011) Correlation between thermostability and stability of glycosidases in ionic liquid. *Biotechnol. Lett.*, **33**, 1215–1219.
3. Wijma,H.J., Floor,R.J. and Janssen,D.B. (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol.*, **23**, 588–594.
4. Gao,D., Narasimhan,D.L., Macdonald,J., Brim,R., Ko,M.C., Landry,D.W., Woods,J.H., Sunahara,R.K. and Zhan,C.G. (2009) Thermostable variants of cocaine esterase for long-time protection against cocaine toxicity. *Mol. Pharmacol.*, **75**, 318–323.
5. Polizzi,K.M., Bommarius,A.S., Broering,J.M. and Chaparro-Riggers,J.F. (2007) Stability of biocatalysts. *Curr. Opin. Chem. Biol.*, **11**, 220–225.
6. Gray,K.A., Richardson,T.H., Kretz,K., Short,J.M., Bartnek,F., Knowles,R., Kan,L., Swanson,P.E. and Robertson,D.E. (2001) Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.*, **343**, 607–617.
7. Folkman,L., Stantic,B., Sattar,A. and Zhou,Y. (2016) EASE-MM: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. *J. Mol. Biol.*, **428**, 1394–1405.
8. Capriotti,E., Fariselli,P. and Casadio,R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, W306–W310.
9. Pires,D.E., Ascher,D.B. and Blundell,T.L. (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics*, **30**, 335–342.
10. Kellogg,E.H., Leaver-Fay,A. and Baker,D. (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins*, **79**, 830–838.
11. Yin,S., Ding,F. and Dokholyan,N.V. (2007) Modeling backbone flexibility improves protein stability estimation. *Structure*, **15**, 1567–1576.
12. Dehouck,Y., Grosfils,A., Folch,B., Gilis,D., Bogaerts,P. and Rooman,M. (2009) Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics*, **25**, 2537–2543.
13. Guerois,R., Nielsen,J.E. and Serrano,L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.*, **320**, 369–387.
14. Gumulya,Y. and Reetz,M.T. (2011) Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways. *ChemBioChem*, **12**, 2502–2510.
15. Bommarius,A.S. and Paye,M.F. (2013) Stabilizing biocatalysts. *Chem. Soc. Rev.*, **42**, 6534–6565.
16. Bednar,D., Beerens,K., Sebestova,E., Bendl,J., Khare,S., Chaloupkova,R., Prokop,Z., Brezovsky,J., Baker,D. and Damborsky,J. (2015) FireProt: energy- and evolution-based computational design of thermostable multiple-point mutants. *PLoS Comput. Biol.*, **11**, e1004556.
17. Goldenzweig,A., Goldsmith,M., Hill,S.E., Gertman,O., Laurino,P., Ashani,Y., Dym,O., Unger,T., Albeck,S., Prilusky,J. *et al.* (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell*, **63**, 337–346.
18. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
19. Suzek,B.E., Wang,Y., Huang,H., McGarvey,P.B. and Wu,C.H. (2015) UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, **31**, 926–932.
20. Edgar,R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
21. Sievers,F., Wilm,A., Dineen,D., Gibson,T.J., Karplus,K., Li,W., Lopez,R., McWilliam,H., Remmert,M., Soding,J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol. Syst. Biol.*, **7**, 539.
22. Capra,J.A. and Singh,M. (2007) Predicting functionally important residues from sequence conservation. *Bioinformatics*, **23**, 1875–1882.
23. Kass,I. and Horovitz,A. (2002) Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins*, **48**, 611–617.
24. Korber,B.T.M., Farber,R.M., Wolpert,D.H. and Lapedes,A.S. (1993) Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 7176–7180.
25. Lee,B.C. and Kim,D. (2009) A new method for revealing correlated mutations under the structural and functional constraints in proteins. *Bioinformatics*, **25**, 2506–2513.
26. Weigt,M., White,R.A., Szurmant,H., Hoch,J.A. and Hwa,T. (2008) Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 67–72.
27. Lockless,S.W. and Ranganathan,R. (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, **286**, 295–299.
28. Dekker,J.P., Fodor,A., Aldrich,R.W. and Yellen,G. (2004) A perturbation-based method for calculating explicit likelihood of

evolutionary co-variance in multiple sequence alignments. *Bioinformatics*, **20**, 1565–1572.

29. Valencia,A. (2003) Multiple sequence alignments as tools for protein structure and function prediction. *Compar. Funct. Genomics*, **4**, 424–427.

30. Benner,S.A. and Gerloff,D. (1991) Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: a prediction of the structure of the catalytic domain of protein kinases. *Adv. Enzyme Regul.*, **31**, 121–181.

31. Brenner,S. (1988) The molecular evolution of genes and proteins: a tale of two serines. *Nature*, **334**, 528–530.

32. Cooperman,B.S., Baykov,A.A. and Lahti,R. (1992) Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Trends Biochem. Sci.*, **17**, 262–266.

33. Howell,N. (1989) Evolutionary conservation of protein regions in the protonmotive cytochrome b and their possible roles in redox catalysis. *J. Mol. Evol.*, **29**, 157–169.

34. Gobel,U., Sander,C., Schneider,R. and Valencia,A. (1994) Correlated mutations and residue contacts in proteins. *Proteins*, **18**, 309–317.

35. Neher,E. (1994) How frequent are correlated changes in families of protein sequences? *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 98–102.

36. Taylor,W.R. and Hatrick,K. (1994) Compensating changes in protein multiple sequence alignments. *Protein Eng.*, **7**, 341–348.

37. Bava,K.A., Gromiha,M.M., Uedaira,H., Kitajima,K. and Sarai,A. (2004) ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res.*, **32**, D120–D121.

38. Pucci,F., Bourgeas,R. and Rooman,M. (2016) Predicting protein thermal stability changes upon point mutations using statistical potentials: introducing HoTMuSiC. *Scientific Rep.*, **6**, 23257.

39. Amin,N., Liu,A.D., Ramer,S., Aehle,W., Meijer,D., Metin,M., Wong,S., Gualfetti,P. and Schellenberger,V. (2004) Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Select.*, **17**, 787–793.

40. Lehmann,M., Loch,C., Middendorf,A., Studer,D., Lassen,S.F., Pasamontes,L., van Loon,A.P. and Wyss,M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. *Protein Eng.*, **15**, 403–411.

41. Pey,A.L., Rodriguez-Larrea,D., Bomke,S., Dammers,S., Godoy-Ruiz,R., Garcia-Mira,M.M. and Sanchez-Ruiz,J.M. (2008) Engineering proteins with tunable thermodynamic and kinetic stabilities. *Proteins*, **71**, 165–174.

42. Sullivan,B.J., Nguyen,T., Durani,V., Mathur,D., Rojas,S., Thomas,M., Syu,T. and Magliery,T.J. (2012) Stabilizing proteins from sequence statistics: the interplay of conservation and correlation in triosephosphate isomerase stability. *J. Mol. Biol.*, **420**, 384–399.

43. Bendl,J., Stourac,J., Sebestova,E., Vavra,O., Musil,M., Brezovsky,J. and Damborsky,J. (2016) HotSpot wizard 2.0: automated design of site-specific mutations and smart libraries in protein engineering. *Nucleic Acids Res.*, **44**, W479–W487.

44. Floor,R.J.1, Wijma,H.J., Colpa,D.I., Ramos-Silva,A., Jekel,P.A., Szymański,W., Feringa,B.L., Marrink,S.J. and Janssen,D.B. (2014) Computational library design for increasing haloalkane dehalogenase stability. *ChemBioChem*, **15**, 1660–1672.