

FireProt 2.0: Web-based Platform for the Fully-Automated Design of Thermostable Proteins

Milos Musil^{1,2,3,*}, Andrej Jezik², Jana Horackova¹, Simeon Borko^{1,2,3}, Petr Kabourek^{1,3}, Jiri Damborsky^{1,3}, David Bednar^{1,3,*}

¹Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Masaryk University, Brno, Czech Republic; ²Department of Information Systems, Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic; ³International Clinical Research Centre, St. Anne's University Hospital Brno, Brno, Czech Republic

*corresponding authors: MM: imusilm@fit.vutbr.cz, AJ: andrejjezik@gmail.com, JH: jana.horackova@recetox.muni.cz, SB: simeon.borko@recetox.muni.cz, PK: peta.kabourek@seznam.cz, JD: 1441@mail.muni.cz, DB: 222755@mail.muni.cz, Website address: <http://loschmidt.chemi.muni.cz/fireprotweb>, (type of web server: computational workflow)

Keywords: ancestral, back-to-consensus, B-factor, epistasis, evolution, force-field, multiple-point mutant, protein engineering, saturation mutagenesis, thermostability

Abstract

Thermostable proteins find their use in numerous biomedical and biotechnological applications. However, the computational design of stable proteins often results in single-point mutations with a limited effect on protein stability. However, the construction of stable multiple-point mutants can prove difficult due to the possibility of antagonistic effects between individual mutations. FireProt protocol enables the automated computational design of highly stable multiple-point mutants. FireProt 2.0 builds on top of the previously published FireProt web, retaining the original functionality and expanding it with several new stabilization strategies. FireProt 2.0 integrates the AlphaFold database and the homology modelling for structure prediction, enabling calculations starting from a sequence. Multiple-point designs are constructed using the Bron-Kerbosch algorithm minimizing the antagonistic effect between the individual mutations. Users can newly limit the FireProt calculation to a set of user-defined mutations, run a saturation mutagenesis of the whole protein, or select rigidifying mutations based on B-factors. Evolution-based back-to-consensus strategy is complemented by ancestral sequence reconstruction. FireProt 2.0 is significantly faster and a reworked graphical user interface broadens the tool's availability even to users with older hardware. FireProt 2.0 is freely available at <http://loschmidt.chemi.muni.cz/fireprotweb>.

Introduction

The thermal stability of naturally occurring proteins often limits their applicability in basic research, biotechnology, and pharmaceutical industries¹. Protein engineering has emerged as a promising approach to overcome these limitations by custom tailoring proteins to enhance their usability, even in harsh industrial environments^{2,3}. Random mutagenesis yields reliable results, however, its natural success rate is considerably low for protein stability experiments⁴. Additionally, saturation mutagenesis of a significant part of the protein is often not feasible due to the costly and laborious protein characterization^{5,6}. Therefore, there is a growing demand⁵ for precise and effective predictors of protein stability.

Several *in silico* tools have been developed for such a task relying on either computationally demanding force-field calculations⁷⁻¹², or most notably one or multiple machine learning techniques¹³⁻

¹⁹. Physical effective energy functions are closely related to classical molecular mechanic force-fields allowing for a fundamental analysis of the molecular interactions²⁰. Physical force-fields are relatively accurate and capable of predicting the behaviour of the proteins under non-standard conditions such as elevated temperature, non-standard salinity, or non-physiological pH²¹. However, their versatility is burdened by high computational demands that can be lowered by partially or fully replacing terms in the energy-field equation by statistical potentials derived from the curated datasets of experimental protein structures projected into stability descriptors^{22,23}. While the predictive power of the force-field methods is unmatched by other approaches, their further development is limited by our current level of understanding of the physical and biochemical forces in the protein and the available computational resources²⁴. Moreover, the scoring functions neglect or approximate some terms that lower the accuracy of the prediction including protein solvation, dynamics, and evaluation of unfolded states necessary for good estimation of thermodynamic cycle. Additionally, force-field methods rely on the availability of high-quality 3D structures that are yet to be well-compensated by structure models significantly improved by novel methods like AlphaFold²⁵.

Machine learning represents a faster alternative to robust force-field calculations that can unearth new features and dependencies unbound by the current knowledge of biochemical forces. The applicability of the machine learning techniques is strongly dependent on the quality and diversity of training and testing datasets. Several databases^{26,27} and mutation datasets²⁸ were constructed over the years. However, they contain a relatively low number of mutations with significant overrepresentation of the destabilizing mutations and an insufficient number of some mutation types, e.g. the charge-changing mutations. Furthermore, these datasets usually do not consider the effect of experimental conditions on protein stability and other characteristics. The pH of the solution has a significant effect on the contribution of the charge-charge interactions that decrease with higher salt concentration²⁹. The differences in the experimental techniques can significantly exceed the average experimental error of 0.48 kcal/mol³⁰. A large portion of available protein stability data is also measured as a change in melting temperature, while the vast majority of predictive tools estimate the change of Gibbs free energy upon mutation. The free energy of the proteins varies non-linearly with temperature as the stability of proteins generally decreases at lower temperatures³¹. Therefore, thermostability data cannot be well utilized due to the correlation between melting temperature and free energy to be about 0.71³². Next to a limited amount of high-quality data, the recent study by Caldararu and coworkers has shown that only a few physico-chemical features significantly affect the predictions of the effect of mutations on protein stability³³.

In recent years, some of the tools have tried to overcome these issues of machine learning predictors by i) moving from standard methods such as random forests or support vector machines to convolutional neural networks^{34–38} and other deep learning techniques³⁹, ii) by utilizing more complex structural features^{40,41} and protein dynamics⁴², or iii) by including temperature dependant features to be able to predict changes in melting temperature³². Meta predictors also increase the accuracy of the machine learning predictions by combining multiple tools and models^{43,44}, or by completing machine learning models with statistical potentials^{45,46}. Furthermore, an apparent bias towards the destabilizing mutations caused by the unbalanced dataset^{47–49} is being addressed by including reverse mutations in the training datasets^{50,51}. However, most of the tools deal with only single-point mutations, and while several multiple-point predictors^{45,46,50,52,53} have surfaced recently, their limited reliability and potentially antagonistic effect of mutations make it still challenging to utilize these predictions to construct stable multiple-point mutants.

FireProt method⁵⁴, initially published in 2015, stands out as one of the few methods^{55,56} that take an opposite approach to the design of stable proteins. Rather than evaluating a set of user-defined

mutations, FireProt suggests potentially stabilizing mutations for the target protein. The web version of FireProt⁵⁷ made the strategy of automated design of multiple-point mutants accessible to a broader user community, analogously to the PROSS web server, developed by Sarel Fleischman and coworkers⁵⁸. The new version, FireProt 2.0, retains all the original functionality while introducing several new strategies and user experience improvements. The updated server provides users with two subsets of multiple-point designs. The low-risk designs are subjected to strict filtration of stabilizing mutations based on protein structure, multiple-sequence alignment, and physico-chemical properties. The high-risk designs are much more lenient and enforce only position-specific conditions. Both low- and high-risk multiple-point designs are constructed using a novel approach based on the Bron-Kerbosch algorithm⁵⁹ minimizing the antagonistic effect among individual mutations.

FireProt 2.0 allows users to limit calculations to a set of user-defined mutations, to run saturation mutagenesis of the entire protein, or to select modifications based on B-factors analysis⁶⁰. The fully automated prediction of stabilizing mutations based on ancestral sequence reconstruction^{61,62} is now available as a second evolution-based strategy. FireProt 2.0 also utilizes models of tertiary structures based on deep neural networks downloaded from the AlphaFold database²⁵ and homology models constructed using ProMod3⁶³. Compared to its predecessor, the updated version is significantly faster, reducing calculation time from over a week to several days. Reworked comprehensive graphical user interface with integrated Mol* visualizer⁶⁴ enhances the tool's usability for users with older hardware and offers a high degree of variability, enabling users to tailor their calculations to their intended stabilization goals. A vast range of optimization strategies broadens the tool's usability within the scientific, medical, and industrial communities.

Methods

FireProt 2.0 uses three main approaches for the identification of potentially stabilizing mutations: (i) energy-based approach, (ii) evolution-based (back-to-consensus), and (iii) ancestral reconstruction-based. The energy-based approach predicts the free energy change upon mutation by force-field calculations, while the evolution-based approach benefits from well-established back-to-consensus analysis. The ancestral design branch utilizes a fully automated ancestral sequence reconstruction workflow. Prior to the computationally demanding calculations in each approach, several predictive tools and database queries are employed to annotate the protein of interest with various sequence and structural information. This information is then utilized to narrow down the initial set of potential mutations. The entire workflow of the FireProt method is shown in Figure 1.

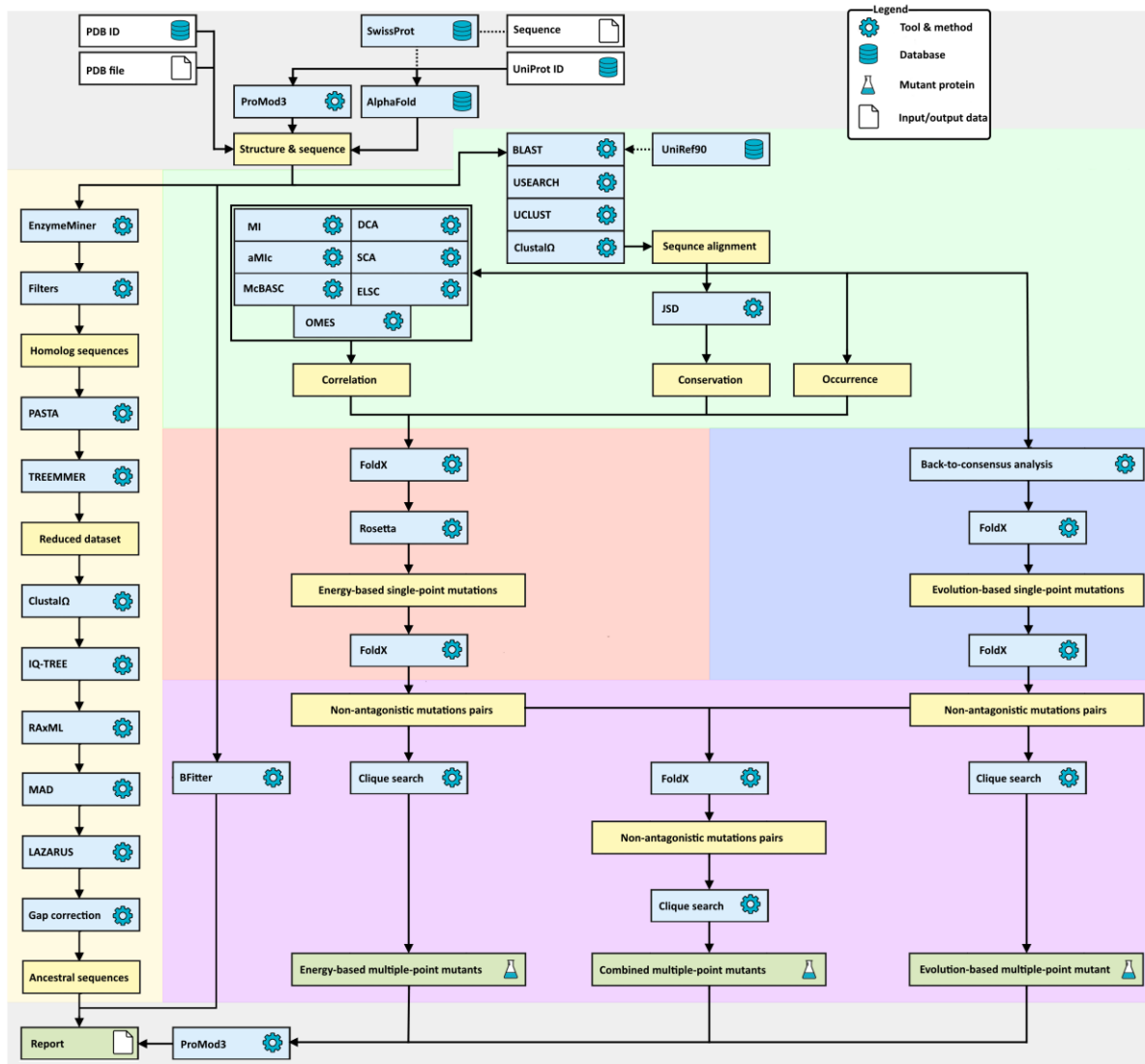


Figure 1: FireProt 2.0 can be started from either protein structure or protein sequence utilized for the construction of the homolog model (grey). Sequence conservation and evolutionary correlation scores are estimated using JSD and the consensus of seven predictive tools, respectively (green). Those annotations serve as filters for the following calculations separated into three main branches identifying potentially stabilizing single-point mutations: i) ancestral reconstruction employing a full FireProt-ASR workflow (yellow), ii) force-field predictions using FoldX and Rosetta (red), and iii) back-to-consensus analysis (blue). In the last step, individual sets of stabilizing single-point mutations are processed in pairs using FoldX, and implementation of the Bron-Kerbosch algorithm is utilized to identify cliques with the best energy evaluations to construct stable multiple-point mutants (purple). Finally, mutant structures are modelled by ProMod3 and provided to the user (grey).

Input and annotation of the target protein

The FireProt calculation can be initiated from either the structure or sequence of the target protein. The user can upload a PDB file or provide its PDB ID to provide structural information. The MakeMultimer tool (<http://watcut.uwaterloo.ca/tools/makemultimer/>) automatically generates the biological assembly, and the user can select the preferred one if multiple biological assemblies are available. Additionally, the calculation can be limited to user-defined chains. The user can provide the raw/FASTA amino acid sequence or UniProt⁶⁵ ID of the target protein as the sequence input. The AlphaFold database²⁵ is queried to get a model structure if the UniProt ID is provided. Otherwise, a

BLAST⁶⁶ search against the PDB database⁶⁷ is performed to obtain a structure template for the ProMod3⁶³ calculation.

To annotate the protein, the SwissProt database⁶⁸ is searched using the protein sequence obtained from the user input or PDB file to remove essential residues from further analysis. Next, the UniRef90 database⁶⁹ is searched by the BLAST algorithm to obtain a set of sequence homologs. Sequences with an identity outside of user-defined thresholds (30% and 90% by default) are excluded, and the remaining sequences are clustered using UCLUST⁷⁰ with a 90% identity threshold by default. The cluster representatives are sorted based on the BLAST query coverage, and up to 200 sequences are selected to construct a multiple-sequence alignment using Clustal Omega⁷¹. The conservation coefficient of each residue in the multiple-sequence alignment is estimated by calculating the Jensen-Shannon entropy⁷² and amino acid frequencies at the individual positions are stored for filtering in the next steps. Correlated positions are identified using a consensual decision of the OMES⁷³, MI⁷⁴, aMIc⁷⁵, DCA⁷⁶, SCA⁷⁷, ELSC⁷⁸, and McBASC⁷⁹ methods. Finally, the average B-factor values are computed for all protein residues⁸⁰.

Design of single-point mutations

The FireProt protocol has been expanded to include three branches for identifying potentially stabilizing single-point mutations: (i) energy-based, (ii) evolution-based, and (iii) ancestral reconstruction-based. The protein structure is amended and minimized using FoldX⁷ and Rosetta⁹ prior to force-field calculations. The filters are applied to exclude potentially deleterious mutations and accelerate the process. Suggested mutations in essential residues, conserved (conservation score 4 or lower), and correlated (consensus score 3.5 or higher by default) positions are considered unsafe and are omitted⁸¹⁻⁸⁷. Additional filters may also be applied, such as disallowing changes to amino acid charges on the protein surface⁸⁸ or omitting mutations not present in the multiple-sequence alignment. Saturation mutagenesis is then performed using the FoldX suite (PsmStability module, threshold set for -1 kcal/mol by default), and potentially stabilizing mutations are forwarded to the more comprehensive Rosetta calculations (ddg_monomer module, threshold set for -1.5 kcal/mol by default). The mutations passing both FoldX and Rosetta thresholds are considered for designing multiple-point mutants.

The evolution-based approach identifies potentially stabilizing mutations through the back-to-consensus analysis. Mutations are selected from either the majority or frequency ratio approach. The majority approach discovers mutations at positions where the consensus residue is present in at least 50% of all analysed sequences. In comparison, the ratio approach lowers the frequency threshold to 40% with the condition that the consensus residue is at least five times more represented than the wild-type amino acid. FoldX evaluates consensus mutations, and the stabilizing ones (threshold set for 0.5 kcal/mol by default) are listed as candidate mutations for designing multiple-point mutants.

In the ancestral reconstruction-based approach, the phylogenetic tree and ancestral sequences are calculated using the FireProt-ASR method⁶². Initially, a reduced set of biologically relevant sequences is aligned, and the phylogenetic tree is constructed. The phylogenetic tree is then rooted using the minimal ancestral deviation algorithm⁸⁹, and the ancestral posterior probabilities for individual amino acids are calculated with the Lazarus package⁹⁰. A custom gap correction algorithm is then applied to fill gaps in the reconstructed ancestral sequences⁶². Finally, the ancestral sequence of each internal node between the query sequence and the root of the phylogenetic tree is pair-wise aligned to the query sequence of the target protein, and the substitutions are selected for further analysis as potentially stabilizing mutations.

Design of multiple-point mutations

FireProt provides users with up to six designs based on the selected analysis. These designs include energy-based, evolution-based, and ancestral reconstruction-based approaches and combined mutants constructed from multiple pools of potentially stabilizing mutations. The energy-based and combined mutants are offered in two variants, reflecting the confidence of predictions: (i) high-risk and (ii) low-risk designs. The high-risk mutants are designed for high-risk/high-reward scenarios, allowing for mutations not present in the sequence alignment and charge-changing mutations positioned on the protein surface. However, these mutations can negatively impact protein folding, and their energetic contributions can be difficult to estimate since force-fields can overestimate their effects on stability due to the bias caused by high concentrations of salts in protein crystallization. In contrast, the low-risk designs are constructed with all filters enabled to introduce only the safe mutations. These differentiated scenarios offer flexibility to the users, depending on preferred engineering strategy, allocated resources, access to gene synthesis technologies, availability of infrastructure for experimental validation, and experimental conditions. All FireProt calculations are processed for the physiological pH of 7, and therefore, high-risk mutations can be easily considered for the experiments conducted under non-physiological pH or in solutions with high salt concentration²⁹.

As multiple-point designs cannot be blindly constructed by a combination of individual stabilizing mutations due to the potential antagonistic effects, FireProt minimizes those risks by utilizing FoldX⁷ and Bron-Kerbosch algorithm⁵⁹. In the first step, FoldX separately evaluates all pairs of single-point mutations within a 10 Å range for the energy-based and evolution-based approaches using the BuildModel module. Modifications outside of the 10 Å range are considered additive without further analysis. A graph representation of the available mutations is then constructed using the free-energy predictions of individual mutations as nodes and the pair predictions as the edges of the full graph. Antagonistic edges (pairs of mutations with the worse energy compared to the higher stabilizing of the two) are removed from this graph. Bron-Kerbosch algorithm is then used to identify cliques in the graph, and the clique with the highest evaluation of individual mutations is selected for the multiple-point design. Once the energy- and evolution-based designs are completed, the procedure is repeated, considering only the pairs between mutations chosen in previous designs, thus constructing combined multiple-point mutants.

Description of the web server

Input

The FireProt web server requires the tertiary structure of the target protein, which can be provided as a PDB ID or a user-defined PDB file (Figure 2A). Alternatively, users can input a custom protein sequence or UniProt identifier to download the structure from the AlphaFold database or build the model by homology modelling. Users can also choose a predefined biological unit generated by MakeMultimer or limit the calculation to the set of manually selected chains.

FireProt offers two mutually exclusive modes – Automated design analysing the whole protein structure and User-defined analysis of particular positions. The *Automated mode* (Figure 2D) enables the identification of potentially stabilizing single-point mutations via the *Automated single-point* panel which enables saturation mutagenesis of all mutable positions, flexibility analysis using B-factors, and back-to-consensus analysis. Users can select low- and high-risk scenarios in the *Automated multiple-point* panel for the multiple-point designs to be constructed out of the pools of potentially stabilizing single-point mutations. It is also possible to start an ancestral sequence reconstruction analysis from

this panel. Alternatively, the *User-defined mode* (Figure 2F) is intended for users looking to evaluate a specific set of mutations in a shorter period. Multiple mutations can be marked for thorough analysis by all available tools for each selected position.

The user can also select essential residues from the SwissProt database or provide them manually (Figure 2B). The *Advanced settings* checkbox expands the list of adjustable parameters for (i) construction of the sequence dataset used for multiple-sequence alignment, (ii) identification of consensus residues, (iii) stability thresholds used by FoldX and Rosetta, and (iv) decision threshold employed in the analysis of correlated positions. *Advanced mode* allows expert users to fine-tune the parameters for their system of interest. However, default values were optimized to provide reliable results in most systems, and it is recommended not to change them in general scenarios.

Output

Upon submission, the user is redirected to the output page which can also be accessed later using a unique identifier assigned at the beginning of the calculation. The *Report panel* on the output page informs the user about the status of each step of the FireProt analysis and provides a log of any potential issues encountered during the calculation. Once all calculations are completed, the output page is updated with several new panels, as described below.

Protein visualisation

The web server provides an interactive visualization of the wild-type structure of the target protein using the Mol* library⁶⁴ (Figure 2C). This module allows users to switch between different protein visualization styles, analyse protein sequence, and highlight selected amino acids in the protein structure. In addition, the *Visualization module* is connected to the *Results panel*, which enables users to visualize selected mutations directly in the protein structure.

Results panel

The results of the FireProt analysis are presented in several tabs based on the set of analyses selected during the input phase (Figure 2E). The *All Residues* tab provides the available information for each residue. Additional tabs are dedicated to individual multiple-point designs: (i) the *Consensus tab* presents information from the back-to-consensus analysis, (ii) the *Energy low-risk* and the *Energy high-risk* tabs report multiple-point mutants resulting from two respective scenarios of the energy-based approach, (iii) the *Combined low- and high-risk* tabs presenting a combination of consensus mutations with the low- and high-risk energy-based mutants, and (iv) the *Ancestral design* tab showing single-point mutations obtained from the ancestral sequence reconstruction using the FireProt-ASR method.

Custom design

The *Custom Design* panel offers users the possibility to create their multiple-point designs by selecting mutations from the lists available in the individual tabs of the *Results panel*. After selecting the desired mutations by the plus button, users can generate a custom FASTA sequence by clicking on the *Generate sequence* button. In addition, users can download all available data related to their FireProt and user-defined designs by clicking on the *Download data* button. The downloaded .zip archive contains the PDB structure, multiple-sequence alignment, PDF report, raw data in .xml and .csv formats, as well as the FASTA sequences of all FireProt and user-defined designs.

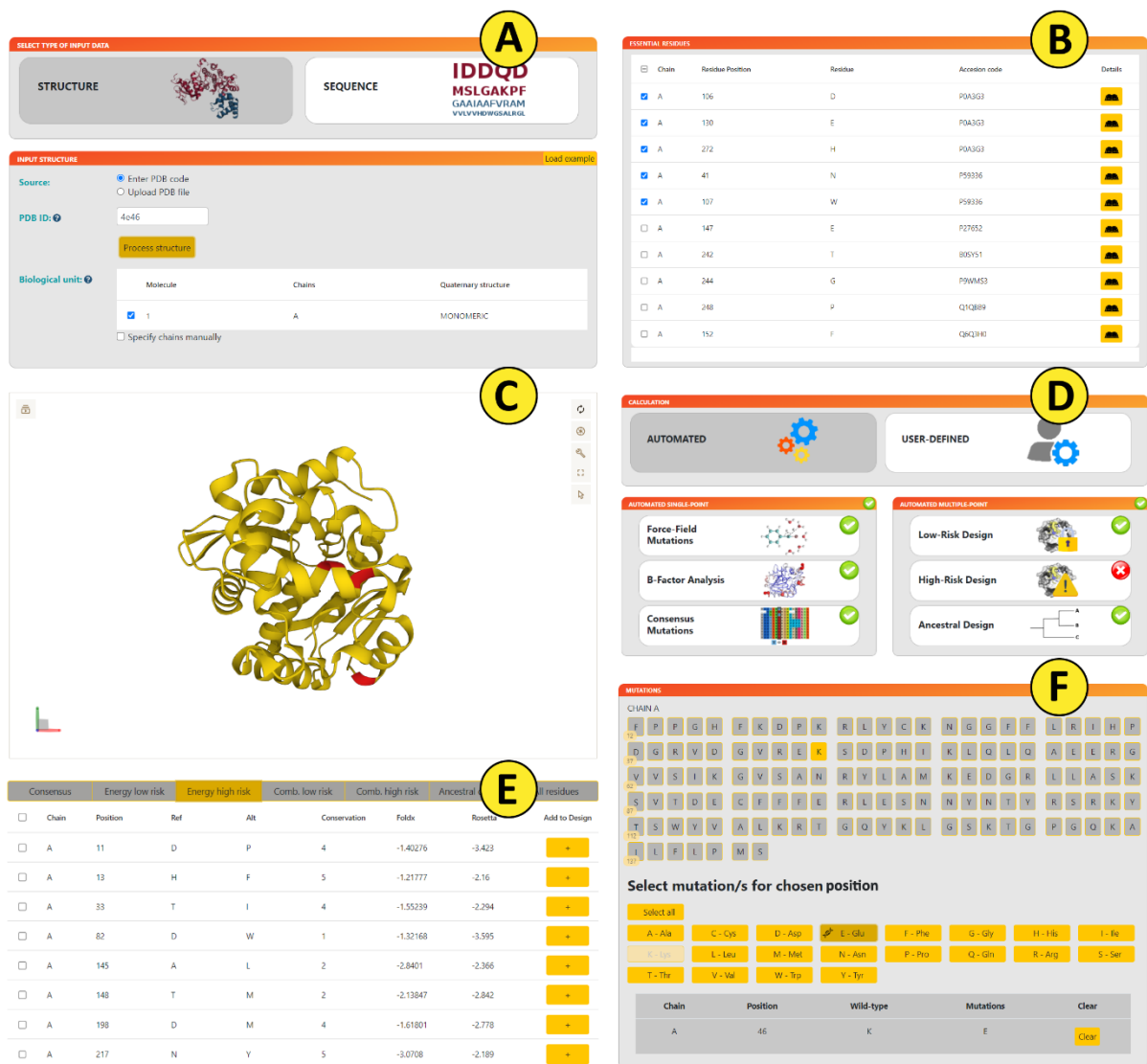


Figure 2: Graphical user interface of FireProt presenting the analysis of the haloalkane dehalogenase DhaA (PDB ID 4E46). (A) Input page allowing users to start from either protein sequence or structure. (B) The table reporting the essential residues identified in the SwissProt database. (C) The Mol* visualization of the wild-type structure of the protein of interest. (D) Analysis selection showing options of the fully automated calculation. (E) Energy high-risk table providing a list of potentially stabilizing mutations with their conservation scores and free energy predictions. (F) Selection of the positions and mutations in the user-defined mode.

Validation

To evaluate the robustness of our web server, we have collected an experimental dataset of nine highly mutated proteins from FireProt^{DB26} counting 861 mutations (Supplementary Table S1) separated from the original S1573 dataset utilized for FireProt parametrization⁵⁷. Using this dataset, we have calculated sensitivity, specificity, and false discovery rates of individual steps of the FireProt workflow and compared them to some of the novel tools for the prediction of the effect of mutations on protein stability (Table 1). While both FoldX and Rosetta show relatively high sensitivity and specificity, their false discovery rates stand at around 50 % meaning that about half of the predicted potentially stabilizing mutations are destabilizing. Using the consensual prediction of both FoldX⁷ and Rosetta⁹, we can observe a notable decrease in sensitivity with every third predicted stabilizing mutation to be a false positive. The false positive rate can be completely mitigated by filtering out

conserved and correlated residues and by moving the decision threshold from 0 kcal/mol to default values of -1.0 and 1.5 kcal/mol for FoldX and Rosetta, respectively. This comes with the price of significantly lower sensitivity as only about 10 % of mutations pass all of the criteria as compared to the FoldX or Rosetta alone. However, low sensitivity with a nullified false discovery rate is the intended goal of the FireProt protocol as the mutations suggested to the user are generally safe to design as multiple-point mutants. In Table 1, we have also compared FoldX and Rosetta with DDGun⁵⁰ and DynaMut2⁴² showing still preserving superiority of the force-field approaches over the recent machine learning tools.

Table 1: Evaluation of the individual steps of the energy approach of the FireProt method. The first two lines show the sensitivity, specificity, and false discovery rates of FoldX and Rosetta with a decision threshold set at 0 kcal/mol. With FoldX+Rosetta, only the mutations passing both FoldX and Rosetta thresholds are considered stabilizing. The next two lines show the results of the combination of FoldX and Rosetta with decision thresholds set to -1 and 1.5 kcal/mol, respectively, and with including conservation and correlation scores as additional filters. DDGun and DynaMut2 are included for comparison with novel machine learning tools.

TYPE	TP	FP	TN	FN	SENSITIVITY	SPECIFICITY	FALSE DISCOVERY RATE
FOLDX	80	94	641	46	0.635	0.872	0.540
ROSETTA	68	63	672	58	0.540	0.914	0.481
FOLDX+ROSETTA	54	27	708	72	0.429	0.963	0.333
+THRESHOLDS	11	0	735	115	0.087	1.000	0.000
+CONSERVATION	7	0	735	119	0.056	1.000	0.000
DDGUN	70	153	582	56	0.556	0.792	0.686
DYNAMUT2	59	86	649	68	0.468	0.883	0.593

In the next step, we have compared the results of FireProt with PROSS⁵⁶, the only direct competitor for automated design of multiple-point mutants (Supplementary Table S2). Across nine proteins in the testing dataset, PROSS identified 140 mutations in the total of 81 protein designs (9 designs per protein with different levels of confidence). Out of the 140 mutations, 35 mutations were also identified by FireProt and 22 more positions were suggested with different amino acids. Considering only the top half of the safer designs provided by PROSS, 23 out of 36 mutations were also identified by FireProt. PROSS predictions largely overlap with FireProt's evolution-based designs as 63 out of 140 mutations (14 out of 36 in the safe designs) are charge-affecting mutations. Charged mutations are generally prohibited in FireProt's energy-based design as charged residues on the protein surface are often targeted by force-field methods while playing an important role in protein folding⁹¹. However, charge-affecting mutations are still viable in the FireProt's high-risk designs and can be commonly found in the consensus design relying on evolution rather than force-field methods. Mutations obtained from the ancestral sequence reconstruction were considered only if the mutation was present in multiple ancestral designs due to the large amount of mutations observed close to the root of the phylogenetic tree.

Finally, we have collected over twenty studies from the literature, where the original FireProt method (equivalent to the *Combined high-risk* design) was utilized for the protein stabilization (Supplementary Table S3). Listed studies report an increase of melting temperature by up to 14°C⁹² and up to 160-fold improvement of the protein's half-life⁹³. The optimal temperature of wild-type proteins ranges from 35 to 80°C. Baseline FireProt protocol was utilized to obtain a highly stabilized haloalkane dehalogenase DhaA by introducing mutations strengthening the interactions around the protein's access tunnel while preserving its catalytic activity⁹⁴. The ancestral sequence reconstruction algorithm implemented in the new version of the FireProt web was also reported to improve the

stability of haloalkane dehalogenase by up to 24°C⁶². Furthermore, a recent study by Livada et. al.⁹⁵ evaluated the method on 56 ancestral sequences from which 94% have shown higher or similar stability compared to the wild-type proteins with 56% of the ancestral proteins improving the melting temperature by more than 5°C with an average improvement of 9°C. Concerning the consensus designs, the proportion of the destabilizing mutations has been estimated to be approximately 40% among all characterized variants^{96,97}.

Conclusions

FireProt web⁵⁷ is a fully automated tool that enables even inexperienced users without prior knowledge of bioinformatics tools and biological systems of interest to design a set of stable mutant proteins. It eliminates the need to install and set up the individual tools as the default parameters and computational protocols have been optimized to accelerate the calculation while maintaining prediction accuracy. FireProt 2.0 represents a significant advancement in the FireProt method⁵⁴, which has already been used to analyse over 6,500 proteins since its first release in 2017. This analysis led to a design of more stable mesophilic and thermophilic proteins with increased melting temperature by up to 14°C⁹² and up to 160-fold improvement of the protein's half-life⁹³. The new version of the FireProt web was also complemented with the ancestral sequence reconstruction algorithm that was previously utilized for the successful stabilization of haloalkane dehalogenase⁶², *Renilla* luciferase⁹⁸, and Ene reductase⁹⁵. Furthermore, users are offered with an extended range of available strategies, including low-and high-risk scenarios of the original designs. This feature allows users to fine-tune their risk/reward ratio by enabling filters that make the final designs safer yet more conservative to produce. The previous FireProt version was often used to design and experimentally characterize only single-point mutants. Therefore, FireProt provides all data for the selection of single-point mutations and also information obtained from B-factor analysis⁸⁰. FireProt analysis can now also be started from the protein sequence only, using the AlphaFold database²⁵ and ProMod3⁶³ modelling, which broadens the number of target proteins by several orders of magnitude. The updated graphical user interface, which employs current technologies, has expanded the tool's availability to users with older hardware, enhancing its usability for the wide range of scientific, medical, and industrial communities. Finally, compared to its predecessor, FireProt 2.0 computational demands were significantly decreased due to the reworked multiple-point mutant construction algorithm from up to a week to less than a day for an average size protein of 300 amino acids. However, the actual calculation time can be higher due to the queue occupancy and the current load on the computational resources. The robustness of the web server was thoroughly tested on a set of 42 diverse proteins.

In the future, FireProt will incorporate new strategies, such as checking the back-bone integrity using ProteinMPNN⁹⁹ or removing potentially destabilizing mutations using the novel machine-learning model trained over the manually curated protein stability data from the recently established database FireProt DB²⁶. Design of disulfide bridges and reinforcing protein multimerization by methods like RF Diffusion¹⁰⁰ may further extend the functionalities of the FireProt web server and the charge interactions can be optimized using the methods like TKSA-MC¹⁰¹⁻¹⁰³. Furthermore, predictions of stabilizing mutations can be combined with designs aiming at improving protein solubility and suppressing protein aggregation, or by considering the changes in the accessible surface area¹⁰⁴. Those further improvements could potentially mitigate the issue of stability/solubility⁸⁸/activity^{105,106}/heat capacity¹⁰⁷ trade-off designing versatile and well-applicable proteins.

Data availability

FireProt 2.0 is a web server available at <https://loschmidt.chemi.muni.cz/fireprotweb/>.

Funding

The authors would like to express their thanks to the Czech Ministry of Education [TEAMING - CZ.02.1.01/0.0/0.0/17_043/0009632; ESFRI ELIXIR LM2023055; ESFRI RECETOX RI LM2023069; ESFRI eINFRA LM2018140]; the Technology Agency of the Czech Republic [TREND FW03010208; RETEMED TN02000122]; European Union [TEAMING 857560]; Brno University of Technology [FIT-S-23-8209]. The research was further supported by the project National Institute for Neurology Research [Programme EXCELES, ID Project No. LX22NPO5107) funded by the European Union — Next Generation EU]. Funding for open access charge: European Union's Horizon 2020 Research and Innovation Framework Programme [857560].

Conflict of interest statement.

None declared.

References

- (1) Zamost, B. L.; Nielsen, H. K.; Starnes, R. L. Thermostable Enzymes for Industrial Applications. *Journal of Industrial Microbiology* **1991**, *8* (2), 71–81. <https://doi.org/10.1007/BF01578757>.
- (2) Schmidt-Dannert, C.; Arnold, F. H. Directed Evolution of Industrial Enzymes. *Trends in Biotechnology* **1999**, *17* (4), 135–136. [https://doi.org/10.1016/S0167-7799\(98\)01283-9](https://doi.org/10.1016/S0167-7799(98)01283-9).
- (3) Kumar, A.; Singh, S. Directed Evolution: Tailoring Biocatalysts for Industrial Applications. *Critical Reviews in Biotechnology* **2013**, *33* (4), 365–378. <https://doi.org/10.3109/07388551.2012.716810>.
- (4) Buß, O.; Rudat, J.; Ochsenreither, K. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Computational and Structural Biotechnology Journal* **2018**, *16*, 25–33. <https://doi.org/10.1016/j.csbj.2018.01.002>.
- (5) Acevedo-Rocha, C. G.; Reetz, M. T.; Nov, Y. Economical Analysis of Saturation Mutagenesis Experiments. *Sci Rep* **2015**, *5* (1), 10654. <https://doi.org/10.1038/srep10654>.
- (6) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2013**, *2* (2), 83–92. <https://doi.org/10.1021/sb300037w>.
- (7) 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), W382–W388. <https://doi.org/10.1093/nar/gki387>.
- (8) Parthiban, V.; Gromiha, M. M.; Schomburg, D. CUPSAT: Prediction of Protein Stability upon Point Mutations. *Nucleic Acids Research* **2006**, *34* (suppl_2), W239–W242. <https://doi.org/10.1093/nar/gkl190>.
- (9) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79* (3), 830–838. <https://doi.org/10.1002/prot.22921>.
- (10) Yin, S.; Ding, F.; Dokholyan, N. V. Eris: An Automated Estimator of Protein Stability. *Nat Methods* **2007**, *4* (6), 466–467. <https://doi.org/10.1038/nmeth0607-466>.
- (11) Benedix, A.; Becker, C. M.; de Groot, B. L.; Cafilisch, A.; Böckmann, R. A. Predicting Free Energy Changes Using Structural Ensembles. *Nat Methods* **2009**, *6* (1), 3–4. <https://doi.org/10.1038/nmeth0109-3>.

- (12) Pandurangan, A. P.; Ochoa-Montaño, B.; Ascher, D. B.; Blundell, T. L. SDM: A Server for Predicting Effects of Mutations on Protein Stability. *Nucleic Acids Research* **2017**, *45* (W1), W229–W235. <https://doi.org/10.1093/nar/gkx439>.
- (13) Folkman, L.; Stantic, B.; Sattar, A.; Zhou, Y. EASE-MM: Sequence-Based Prediction of Mutation-Induced Stability Changes with Feature-Based Multiple Models. *Journal of Molecular Biology* **2016**, *428* (6), 1394–1405. <https://doi.org/10.1016/j.jmb.2016.01.012>.
- (14) Teng, S.; Srivastava, A. K.; Wang, L. Sequence Feature-Based Prediction of Protein Stability Changes upon Amino Acid Substitutions. *BMC Genomics* **2010**, *11* (Suppl 2), S5. <https://doi.org/10.1186/1471-2164-11-S2-S5>.
- (15) Wainreb, G.; Wolf, L.; Ashkenazy, H.; Dehouck, Y.; Ben-Tal, N. Protein Stability: A Single Recorded Mutation Aids in Predicting the Effects of Other Mutations in the Same Amino Acid Site. *Bioinformatics* **2011**, *27* (23), 3286–3292. <https://doi.org/10.1093/bioinformatics/btr576>.
- (16) Cheng, J.; Randall, A.; Baldi, P. Prediction of Protein Stability Changes for Single-Site Mutations Using Support Vector Machines. *Proteins: Structure, Function, and Bioinformatics* **2006**, *62* (4), 1125–1132. <https://doi.org/10.1002/prot.20810>.
- (17) Huang, L.-T.; Gromiha, M. M.; Ho, S.-Y. iPTREE-STAB: Interpretable Decision Tree Based Method for Predicting Protein Stability Changes upon Mutations. *Bioinformatics* **2007**, *23* (10), 1292–1293. <https://doi.org/10.1093/bioinformatics/btm100>.
- (18) Savojardo, C.; Fariselli, P.; Martelli, P. L.; Casadio, R. INPS-MD: A Web Server to Predict Stability of Protein Variants from Sequence and Structure. *Bioinformatics* **2016**, *32* (16), 2542–2544. <https://doi.org/10.1093/bioinformatics/btw192>.
- (19) Chen, C.-W.; Lin, M.-H.; Liao, C.-C.; Chang, H.-P.; Chu, Y.-W. iStable 2.0: Predicting Protein Thermal Stability Changes by Integrating Various Characteristic Modules. *Computational and Structural Biotechnology Journal* **2020**, *18*, 622–630. <https://doi.org/10.1016/j.csbj.2020.02.021>.
- (20) Lazaridis, T.; Karplus, M. Effective Energy Functions for Protein Structure Prediction. *Current Opinion in Structural Biology* **2000**, *10* (2), 139–145. [https://doi.org/10.1016/S0959-440X\(00\)00063-4](https://doi.org/10.1016/S0959-440X(00)00063-4).
- (21) *The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design | Journal of Chemical Theory and Computation*. https://pubs.acs.org/doi/abs/10.1021/acs.jctc.7b00125?casa_token=cPdu-xJyoJ4AAAAA:RVfDOSmHwdcdoMr6dU_LK47Za5OOCp23KOaH8T6E4ha7BL1swyly1fOT7tkb9hOK3nLkFyH1cpUNdXzl (accessed 2020-09-09).
- (22) Dehouck, Y.; Gilis, D.; Rooman, M. A New Generation of Statistical Potentials for Proteins. *Biophysical Journal* **2006**, *90* (11), 4010–4017. <https://doi.org/10.1529/biophysj.105.079434>.
- (23) Liu, H. On Statistical Energy Functions for Biomolecular Modeling and Design. *Quant Biol* **2015**, *3* (4), 157–167. <https://doi.org/10.1007/s40484-015-0054-x>.
- (24) Musil, M.; Konegger, H.; Hon, J.; Bednar, D.; Damborsky, J. Computational Design of Stable and Soluble Biocatalysts. *ACS Catal.* **2019**, *9* (2), 1033–1054. <https://doi.org/10.1021/acscatal.8b03613>.
- (25) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Židek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Research* **2022**, *50* (D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
- (26) Stourac, J.; Dubrava, J.; Musil, M.; Horackova, J.; Damborsky, J.; Mazurenko, S.; Bednar, D. FireProtDB: Database of Manually Curated Protein Stability Data. *Nucleic Acids Research* **2021**, *49* (D1), D319–D324. <https://doi.org/10.1093/nar/gkaa981>.
- (27) Nikam, R.; Kulandaisamy, A.; Harini, K.; Sharma, D.; Gromiha, M. M. ProThermDB: Thermodynamic Database for Proteins and Mutants Revisited after 15 Years. *Nucleic Acids Research* **2021**, *49* (D1), D420–D424. <https://doi.org/10.1093/nar/gkaa1035>.

- (28) Sasidharan Nair, P.; Vihinen, M. VariBench: A Benchmark Database for Variations. *Hum Mutat* **2013**, *34* (1), 42–49. <https://doi.org/10.1002/humu.22204>.
- (29) Martins De Oliveira, V.; Godoi Contessoto, V. D.; Bruno Da Silva, F.; Zago Caetano, D. L.; Jurado De Carvalho, S.; Pereira Leite, V. B. Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models. *Biophysical Journal* **2018**, *114* (1), 65–75. <https://doi.org/10.1016/j.bpj.2017.11.012>.
- (30) Khatun, J.; Khare, S. D.; Dokholyan, N. V. Can Contact Potentials Reliably Predict Stability of Proteins? *Journal of Molecular Biology* **2004**, *336* (5), 1223–1238. <https://doi.org/10.1016/j.jmb.2004.01.002>.
- (31) Kazlauskas, R. Engineering More Stable Proteins. *Chemical Society Reviews* **2018**, *47* (24), 9026–9045. <https://doi.org/10.1039/C8CS00014J>.
- (32) Pucci, F.; Bourgeas, R.; Rومان, M. Predicting Protein Thermal Stability Changes upon Point Mutations Using Statistical Potentials: Introducing HoTMuSiC. *Sci Rep* **2016**, *6* (1), 23257. <https://doi.org/10.1038/srep23257>.
- (33) *Three Simple Properties Explain Protein Stability Change upon Mutation | Journal of Chemical Information and Modeling*. https://pubs.acs.org/doi/full/10.1021/acs.jcim.1c00201?casa_token=zSyrG2zAYgYAAAAA%3ADfWLjMPQGxVUoNZeQOiKH_EMXT48h3cldr8VoUUD37IXz8sJHaGVAv9H-g--J-ZkSOeyq3FlgPempssR (accessed 2023-06-29).
- (34) Fang, X.; Huang, J.; Zhang, R.; Wang, F.; Zhang, Q.; Li, G.; Yan, J.; Zhang, H.; Yan, Y.; Xu, L. Convolution Neural Network-Based Prediction of Protein Thermostability. *J Chem Inf Model* **2019**, *59* (11), 4833–4843. <https://doi.org/10.1021/acs.jcim.9b00220>.
- (35) Cang, Z.; Wei, G.-W. TopologyNet: Topology Based Deep Convolutional and Multi-Task Neural Networks for Biomolecular Property Predictions. *PLOS Computational Biology* **2017**, *13* (7), e1005690. <https://doi.org/10.1371/journal.pcbi.1005690>.
- (36) Blaabjerg, L. M.; Kassem, M. M.; Good, L. L.; Jonsson, N.; Cagiada, M.; Johansson, K. E.; Boomsma, W.; Stein, A.; Lindorff-Larsen, K. Rapid Protein Stability Prediction Using Deep Learning Representations. *eLife* **2023**, *12*, e82593. <https://doi.org/10.7554/eLife.82593>.
- (37) Zhou, Y.; Pan, Q.; Pires, D. E. V.; Rodrigues, C. H. M.; Ascher, D. B. DDMut: Predicting Effects of Mutations on Protein Stability Using Deep Learning. *Nucleic Acids Research* **2023**, gkad472. <https://doi.org/10.1093/nar/gkad472>.
- (38) Shroff, R.; Cole, A. W.; Morrow, B. R.; Diaz, D. J.; Donnell, I.; Gollihar, J.; Ellington, A. D.; Thyer, R. A Structure-Based Deep Learning Framework for Protein Engineering. *bioRxiv* November 8, 2019, p 833905. <https://doi.org/10.1101/833905>.
- (39) Jung, F.; Frey, K.; Zimmer, D.; Mühlhaus, T. DeepSTABp: A Deep Learning Approach for the Prediction of Thermal Protein Stability. *International Journal of Molecular Sciences* **2023**, *24* (8), 7444. <https://doi.org/10.3390/ijms24087444>.
- (40) Quan, L.; Lv, Q.; Zhang, Y. STRUM: Structure-Based Prediction of Protein Stability Changes upon Single-Point Mutation. *Bioinformatics* **2016**, *32* (19), 2936–2946. <https://doi.org/10.1093/bioinformatics/btw361>.
- (41) Cao, H.; Wang, J.; He, L.; Qi, Y.; Zhang, J. Z. DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks. *J. Chem. Inf. Model.* **2019**, *59* (4), 1508–1514. <https://doi.org/10.1021/acs.jcim.8b00697>.
- (42) Rodrigues, C. H. M.; Pires, D. E. V.; Ascher, D. B. DynaMut2: Assessing Changes in Stability and Flexibility upon Single and Multiple Point Missense Mutations. *Protein Science* **2021**, *30* (1), 60–69. <https://doi.org/10.1002/pro.3942>.
- (43) Pires, D. E. V.; Ascher, D. B.; Blundell, T. L. DUET: A Server for Predicting Effects of Mutations on Protein Stability Using an Integrated Computational Approach. *Nucleic Acids Research* **2014**, *42* (W1), W314–W319. <https://doi.org/10.1093/nar/gku411>.
- (44) Fariselli, P.; Martelli, P. L.; Savojardo, C.; Casadio, R. INPS: Predicting the Impact of Non-Synonymous Variations on Protein Stability from Sequence. *Bioinformatics* **2015**, *31* (17), 2816–2821. <https://doi.org/10.1093/bioinformatics/btv291>.

- (45) Laimer, J.; Hofer, H.; Fritz, M.; Wegenkittl, S.; Lackner, P. MAESTRO - Multi Agent Stability Prediction upon Point Mutations. *BMC Bioinformatics* **2015**, *16* (1), 116. <https://doi.org/10.1186/s12859-015-0548-6>.
- (46) Witvliet, D. K.; Strokach, A.; Giraldo-Forero, A. F.; Teyra, J.; Colak, R.; Kim, P. M. ELASPIC Web-Server: Proteome-Wide Structure-Based Prediction of Mutation Effects on Protein Stability and Binding Affinity. *Bioinformatics* **2016**, *32* (10), 1589–1591. <https://doi.org/10.1093/bioinformatics/btw031>.
- (47) Montanucci, L.; Savojardo, C.; Martelli, P. L.; Casadio, R.; Fariselli, P. On the Biases in Predictions of Protein Stability Changes upon Variations: The INPS Test Case. *Bioinformatics* **2019**, *35* (14), 2525–2527. <https://doi.org/10.1093/bioinformatics/bty979>.
- (48) Thiltgen, G.; Goldstein, R. A. Assessing Predictors of Changes in Protein Stability upon Mutation Using Self-Consistency. *PLOS ONE* **2012**, *7* (10), e46084. <https://doi.org/10.1371/journal.pone.0046084>.
- (49) Usmanova, D. R.; Bogatyreva, N. S.; Ariño Bernad, J.; Eremina, A. A.; Gorshkova, A. A.; Kanevskiy, G. M.; Lonishin, L. R.; Meister, A. V.; Yakupova, A. G.; Kondrashov, F. A.; Ivankov, D. N. Self-Consistency Test Reveals Systematic Bias in Programs for Prediction Change of Stability upon Mutation. *Bioinformatics* **2018**, *34* (21), 3653–3658. <https://doi.org/10.1093/bioinformatics/bty340>.
- (50) Montanucci, L.; Capriotti, E.; Frank, Y.; Ben-Tal, N.; Fariselli, P. DDGun: An Untrained Method for the Prediction of Protein Stability Changes upon Single and Multiple Point Variations. *BMC Bioinformatics* **2019**, *20* (14), 335. <https://doi.org/10.1186/s12859-019-2923-1>.
- (51) Pucci, F.; Bernaerts, K.; Teheux, F.; Gilis, D.; Rooman, M. Symmetry Principles in Optimization Problems: An Application to Protein Stability Prediction★. *IFAC-PapersOnLine* **2015**, *48* (1), 458–463. <https://doi.org/10.1016/j.ifacol.2015.05.068>.
- (52) Li, Y.; Fang, J. PROTS-RF: A Robust Model for Predicting Mutation-Induced Protein Stability Changes. *PLOS ONE* **2012**, *7* (10), e47247. <https://doi.org/10.1371/journal.pone.0047247>.
- (53) Tian, J.; Wu, N.; Chu, X.; Fan, Y. Predicting Changes in Protein Thermostability Brought about by Single- or Multi-Site Mutations. *BMC Bioinformatics* **2010**, *11* (1), 370. <https://doi.org/10.1186/1471-2105-11-370>.
- (54) *FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants* | *PLOS Computational Biology*. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004556> (accessed 2023-04-14).
- (55) Floor, R. J.; Wijma, H. J.; Colpa, D. I.; Ramos-Silva, A.; Jekel, P. A.; Szymański, W.; Feringa, B. L.; Marrink, S. J.; Janssen, D. B. Computational Library Design for Increasing Haloalkane Dehalogenase Stability. *Chembiochem* **2014**, *15* (11), 1660–1672. <https://doi.org/10.1002/cbic.201402128>.
- (56) Weinstein, J. J.; Goldenzweig, A.; Hoch, S.; Fleishman, S. J. PROSS 2: A New Server for the Design of Stable and Highly Expressed Protein Variants. *Bioinformatics* **2021**, *37* (1), 123–125. <https://doi.org/10.1093/bioinformatics/btaa1071>.
- (57) 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 Res* **2017**, *45* (W1), W393–W399. <https://doi.org/10.1093/nar/gkx285>.
- (58) Goldenzweig, A.; Goldsmith, M.; Hill, S. E.; Gertman, O.; Laurino, P.; Ashani, Y.; Dym, O.; Unger, T.; Albeck, S.; Prilusky, J.; Lieberman, R. L.; Aharoni, A.; Silman, I.; Sussman, J. L.; Tawfik, D. S.; Fleishman, S. J. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* **2016**, *63* (2), 337–346. <https://doi.org/10.1016/j.molcel.2016.06.012>.
- (59) Bron, C.; Kerbosch, J. Algorithm 457: Finding All Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16* (9), 575–577. <https://doi.org/10.1145/362342.362367>.
- (60) *Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability* | *Chemical Reviews*.

https://pubs.acs.org/doi/full/10.1021/acs.chemrev.8b00290?casa_token=nwJpnVwTe6sAAAAA%3AVhQPGvEsSJeqlBJzaVavZJilaNKQnwGBGH2eTpFKaVW2c0FMoNSrlaN99ITsIfqRgf4_JXaas6WlLsZ1 (accessed 2023-06-29).

- (61) Thornton, J. W. Resurrecting Ancient Genes: Experimental Analysis of Extinct Molecules. *Nat Rev Genet* **2004**, *5* (5), 366–375. <https://doi.org/10.1038/nrg1324>.
- (62) Musil, M.; Khan, R. T.; Beier, A.; Stourac, J.; Konegger, H.; Damborsky, J.; Bednar, D. FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. *Brief Bioinform* **2020**. <https://doi.org/10.1093/bib/bbaa337>.
- (63) Studer, G.; Tauriello, G.; Bienert, S.; Biasini, M.; Johner, N.; Schwede, T. ProMod3—A Versatile Homology Modelling Toolbox. *PLOS Computational Biology* **2021**, *17* (1), e1008667. <https://doi.org/10.1371/journal.pcbi.1008667>.
- (64) Sehnal, D.; Bittrich, S.; Deshpande, M.; Svobodová, R.; Berka, K.; Bazgier, V.; Velankar, S.; Burley, S. K.; Koča, J.; Rose, A. S. Mol* Viewer: Modern Web App for 3D Visualization and Analysis of Large Biomolecular Structures. *Nucleic Acids Research* **2021**, *49* (W1), W431–W437. <https://doi.org/10.1093/nar/gkab314>.
- (65) UniProt Consortium, T. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Research* **2018**, *46* (5), 2699–2699. <https://doi.org/10.1093/nar/gky092>.
- (66) Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10* (1), 421. <https://doi.org/10.1186/1471-2105-10-421>.
- (67) Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. Protein Data Bank (PDB): Database of Three-Dimensional Structural Information of Biological Macromolecules. *Acta Cryst D* **1998**, *54* (6), 1078–1084. <https://doi.org/10.1107/S09074444998009378>.
- (68) Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.-C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. The SWISS-PROT Protein Knowledgebase and Its Supplement TrEMBL in 2003. *Nucleic Acids Research* **2003**, *31* (1), 365–370. <https://doi.org/10.1093/nar/gkg095>.
- (69) Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H.; the UniProt Consortium. UniRef Clusters: A Comprehensive and Scalable Alternative for Improving Sequence Similarity Searches. *Bioinformatics* **2015**, *31* (6), 926–932. <https://doi.org/10.1093/bioinformatics/btu739>.
- (70) Edgar, R. C. Search and Clustering Orders of Magnitude Faster than BLAST. *Bioinformatics* **2010**, *26* (19), 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- (71) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Molecular Systems Biology* **2011**, *7* (1), 539. <https://doi.org/10.1038/msb.2011.75>.
- (72) Capra, J. A.; Singh, M. Predicting Functionally Important Residues from Sequence Conservation. *Bioinformatics* **2007**, *23* (15), 1875–1882. <https://doi.org/10.1093/bioinformatics/btm270>.
- (73) Kass, I.; Horovitz, A. Mapping Pathways of Allosteric Communication in GroEL by Analysis of Correlated Mutations. *Proteins* **2002**, *48* (4), 611–617. <https://doi.org/10.1002/prot.10180>.
- (74) Korber, B. T.; Farber, R. M.; Wolpert, D. H.; Lapedes, A. S. 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* **1993**, *90* (15), 7176–7180. <https://doi.org/10.1073/pnas.90.15.7176>.
- (75) Lee, B.-C.; Kim, D. A New Method for Revealing Correlated Mutations under the Structural and Functional Constraints in Proteins. *Bioinformatics* **2009**, *25* (19), 2506–2513. <https://doi.org/10.1093/bioinformatics/btp455>.
- (76) Weigt, M.; White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. Identification of Direct Residue Contacts in Protein-Protein Interaction by Message Passing. *Proc Natl Acad Sci U S A* **2009**, *106* (1), 67–72. <https://doi.org/10.1073/pnas.0805923106>.

- (77) Lockless, S. W.; Ranganathan, R. Evolutionarily Conserved Pathways of Energetic Connectivity in Protein Families. *Science* **1999**, *286* (5438), 295–299. <https://doi.org/10.1126/science.286.5438.295>.
- (78) Dekker, J. P.; Fodor, A.; Aldrich, R. W.; Yellen, G. A Perturbation-Based Method for Calculating Explicit Likelihood of Evolutionary Co-Variance in Multiple Sequence Alignments. *Bioinformatics* **2004**, *20* (10), 1565–1572. <https://doi.org/10.1093/bioinformatics/bth128>.
- (79) Olmea, O.; Rost, B.; Valencia, A. Effective Use of Sequence Correlation and Conservation in Fold Recognition. *J Mol Biol* **1999**, *293* (5), 1221–1239. <https://doi.org/10.1006/jmbi.1999.3208>.
- (80) Reetz, M. T.; Carballeira, J. D.; Vogel, A. Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angew Chem Int Ed Engl* **2006**, *45* (46), 7745–7751. <https://doi.org/10.1002/anie.200602795>.
- (81) Benner, S. A.; Gerloff, D. 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* **1991**, *31*, 121–181. [https://doi.org/10.1016/0065-2571\(91\)90012-b](https://doi.org/10.1016/0065-2571(91)90012-b).
- (82) Brenner, S. The Molecular Evolution of Genes and Proteins: A Tale of Two Serines. *Nature* **1988**, *334* (6182), 528–530. <https://doi.org/10.1038/334528a0>.
- (83) Cooperman, B. S.; Baykov, A. A.; Lahti, R. Evolutionary Conservation of the Active Site of Soluble Inorganic Pyrophosphatase. *Trends Biochem Sci* **1992**, *17* (7), 262–266. [https://doi.org/10.1016/0968-0004\(92\)90406-y](https://doi.org/10.1016/0968-0004(92)90406-y).
- (84) Howell, N. Evolutionary Conservation of Protein Regions in the Protonmotive Cytochrome b and Their Possible Roles in Redox Catalysis. *J Mol Evol* **1989**, *29* (2), 157–169. <https://doi.org/10.1007/BF02100114>.
- (85) Göbel, U.; Sander, C.; Schneider, R.; Valencia, A. Correlated Mutations and Residue Contacts in Proteins. *Proteins* **1994**, *18* (4), 309–317. <https://doi.org/10.1002/prot.340180402>.
- (86) Neher, E. How Frequent Are Correlated Changes in Families of Protein Sequences? *Proc Natl Acad Sci U S A* **1994**, *91* (1), 98–102. <https://doi.org/10.1073/pnas.91.1.98>.
- (87) Taylor, W. R.; Hatrick, K. Compensating Changes in Protein Multiple Sequence Alignments. *Protein Eng* **1994**, *7* (3), 341–348. <https://doi.org/10.1093/protein/7.3.341>.
- (88) Broom, A.; Jacobi, Z.; Trainor, K.; Meiering, E. M. Computational Tools Help Improve Protein Stability but with a Solubility Tradeoff. *J. Biol. Chem.* **2017**, *292* (35), 14349–14361. <https://doi.org/10.1074/jbc.M117.784165>.
- (89) Tria, F. D. K.; Landan, G.; Dagan, T. Phylogenetic Rooting Using Minimal Ancestor Deviation. *Nature Ecology & Evolution* **2017**, *1* (1), 1–7. <https://doi.org/10.1038/s41559-017-0193>.
- (90) Hanson-Smith, V.; Kolaczowski, B.; Thornton, J. W. Robustness of Ancestral Sequence Reconstruction to Phylogenetic Uncertainty. *Mol Biol Evol* **2010**, *27* (9), 1988–1999. <https://doi.org/10.1093/molbev/msq081>.
- (91) Contessoto, V. G.; de Oliveira, V. M.; de Carvalho, S. J.; Oliveira, L. C.; Leite, V. B. P. NTL9 Folding at Constant pH: The Importance of Electrostatic Interaction and pH Dependence. *J. Chem. Theory Comput.* **2016**, *12* (7), 3270–3277. <https://doi.org/10.1021/acs.jctc.6b00399>.
- (92) Yu, Z.; Yu, H.; Xu, J.; Wang, Z.; Wang, Z.; Kang, T.; Chen, K.; Pu, Z.; Wu, J.; Yang, L.; Xu, G. Enhancing Thermostability of Lipase from *Pseudomonas Alcaligenes* for Producing l -Menthhol by the CREATE Strategy. *Catalysis Science & Technology* **2022**, *12* (8), 2531–2541. <https://doi.org/10.1039/D2CY00082B>.
- (93) Pongpamorn, P.; Watthaisong, P.; Pimviriyakul, P.; Jaruwat, A.; Lawan, N.; Chitnumsub, P.; Chaiyen, P. Identification of a Hotspot Residue for Improving the Thermostability of a Flavin-Dependent Monooxygenase. *ChemBioChem* **2019**, *20* (24), 3020–3031. <https://doi.org/10.1002/cbic.201900413>.
- (94) Markova, K.; Chmelova, K.; Marques, S.; Carpentier, P.; Bednar, D.; Damborsky, J.; Marek, M. Decoding the Intricate Network of Molecular Interactions of a Hyperstable Engineered Biocatalyst. *Chemical Science* **2020**, *11* (41), 11162–11178. <https://doi.org/10.1039/D0SC03367G>.

- (95) *Ancestral Sequence Reconstruction Enhances Gene Mining Efforts for Industrial Enzymes Reductases by Expanding Enzyme Panels with Thermostable Catalysts* | *ACS Catalysis*.
<https://pubs.acs.org/doi/full/10.1021/acscatal.2c03859> (accessed 2023-10-12).
- (96) Steipe, B.; Schiller, B.; Plückthun, A.; Steinbacher, S. Sequence Statistics Reliably Predict Stabilizing Mutations in a Protein Domain. *Journal of Molecular Biology* **1994**, *240* (3), 188–192. <https://doi.org/10.1006/jmbi.1994.1434>.
- (97) Porebski, B. T.; Buckle, A. M. Consensus Protein Design. *Protein Engineering, Design and Selection* **2016**, *29* (7), 245–251. <https://doi.org/10.1093/protein/gzw015>.
- (98) Chaloupkova, R.; Liskova, V.; Toul, M.; Markova, K.; Sebestova, E.; Hernychova, L.; Marek, M.; Pinto, G. P.; Pluskal, D.; Waterman, J.; Prokop, Z.; Damborsky, J. Light-Emitting Dehalogenases: Reconstruction of Multifunctional Biocatalysts. *ACS Catal.* **2019**, *9* (6), 4810–4823. <https://doi.org/10.1021/acscatal.9b01031>.
- (99) *Robust deep learning-based protein sequence design using ProteinMPNN* | *Science*.
<https://www.science.org/doi/full/10.1126/science.add2187> (accessed 2023-04-14).
- (100) Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; Bortoli, V. D.; Mathieu, E.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. Broadly Applicable and Accurate Protein Design by Integrating Structure Prediction Networks and Diffusion Generative Models. *bioRxiv* December 14, 2022, p 2022.12.09.519842. <https://doi.org/10.1101/2022.12.09.519842>.
- (101) Contessoto, V. G.; de Oliveira, V. M.; Fernandes, B. R.; Slade, G. G.; Leite, V. B. P. TKSA-MC: A Web Server for Rational Mutation through the Optimization of Protein Charge Interactions. *Proteins: Structure, Function, and Bioinformatics* **2018**, *86* (11), 1184–1188. <https://doi.org/10.1002/prot.25599>.
- (102) Ngo, K.; Bruno da Silva, F.; Leite, V. B. P.; Contessoto, V. G.; Onuchic, J. N. Improving the Thermostability of Xylanase A from *Bacillus Subtilis* by Combining Bioinformatics and Electrostatic Interactions Optimization. *J. Phys. Chem. B* **2021**, *125* (17), 4359–4367. <https://doi.org/10.1021/acs.jpcc.1c01253>.
- (103) De Godoi Contessoto, V.; Ramos, F. C.; De Melo, R. R.; De Oliveira, V. M.; Scarpassa, J. A.; De Sousa, A. S.; Zanphorlin, L. M.; Slade, G. G.; Leite, V. B. P.; Ruller, R. Electrostatic Interaction Optimization Improves Catalytic Rates and Thermostolerance on Xylanases. *Biophysical Journal* **2021**, *120* (11), 2172–2180. <https://doi.org/10.1016/j.bpj.2021.03.036>.
- (104) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J. Denaturant m Values and Heat Capacity Changes: Relation to Changes in Accessible Surface Areas of Protein Unfolding. *Protein Science* **1995**, *4* (10), 2138–2148. <https://doi.org/10.1002/pro.5560041020>.
- (105) Miller, S. R. An Appraisal of the Enzyme Stability-Activity Trade-Off. *Evolution* **2017**, *71* (7), 1876–1887. <https://doi.org/10.1111/evo.13275>.
- (106) Siddiqui, K. S. Defying the Activity–Stability Trade-off in Enzymes: Taking Advantage of Entropy to Enhance Activity and Thermostability. *Critical Reviews in Biotechnology* **2017**, *37* (3), 309–322. <https://doi.org/10.3109/07388551.2016.1144045>.
- (107) Ruller, R.; Deliberto, L.; Ferreira, T. L.; Ward, R. J. Thermostable Variants of the Recombinant Xylanase a from *Bacillus Subtilis* Produced by Directed Evolution Show Reduced Heat Capacity Changes. *Proteins: Structure, Function, and Bioinformatics* **2008**, *70* (4), 1280–1293. <https://doi.org/10.1002/prot.21617>.