

RESEARCH ARTICLE

# FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants

David Bednar<sup>1,2</sup>✉, Koen Beerens<sup>1</sup>✉, Eva Sebestova<sup>1</sup>, Jaroslav Bendl<sup>1,2,3</sup>, Sagar Khare<sup>4</sup>, Radka Chaloupkova<sup>1</sup>, Zbynek Prokop<sup>1,2,5</sup>, Jan Brezovsky<sup>1,2</sup>, David Baker<sup>6</sup>, Jiri Damborsky<sup>1,2,5\*</sup>

**1** Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Masaryk University, Brno, Czech Republic, **2** International Clinical Research Center, St. Anne's University Hospital Brno, Brno, Czech Republic, **3** Department of Information Systems, Faculty of Information Technology, Brno University of Technology, Brno, Czech Republic, **4** Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, New Jersey, United States of America, **5** Enantis, Ltd., Brno, Czech Republic, **6** Department of Biochemistry, University of Washington, Seattle, Washington, United States of America

✉ These authors contributed equally to this work.

\* [jiri@chemi.muni.cz](mailto:jiri@chemi.muni.cz)



**OPEN ACCESS**

**Citation:** Bednar D, Beerens K, Sebestova E, Bendl J, Khare S, Chaloupkova R, et al. (2015) FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLoS Comput Biol* 11(11): e1004556. doi:10.1371/journal.pcbi.1004556

**Editor:** Nikolay V. Dokholyan, University of North Carolina at Chapel Hill, UNITED STATES

**Received:** March 10, 2015

**Accepted:** September 14, 2015

**Published:** November 3, 2015

**Copyright:** © 2015 Bednar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The work was supported by the Grant Agency of the Czech Republic (P503/12/0572; <http://www.gacr.cz/>) and the Czech Ministry of Education of the Czech Republic (LO1214 and LH14027; <http://www.msmt.cz/>). DBe received a Brno Ph.D. Talent Scholarship funded by the Brno City Municipality and his stay at Rutgers University was supported by European Regional Development Fund – project FNUSA-ICRC (CZ.1.05/1.1.00/02.0123; <http://www.msmt.cz/>), European Social Fund and the state

## Abstract

There is great interest in increasing proteins' stability to enhance their utility as biocatalysts, therapeutics, diagnostics and nanomaterials. Directed evolution is a powerful, but experimentally strenuous approach. Computational methods offer attractive alternatives. However, due to the limited reliability of predictions and potentially antagonistic effects of substitutions, only single-point mutations are usually predicted *in silico*, experimentally verified and then recombined in multiple-point mutants. Thus, substantial screening is still required. Here we present FireProt, a robust computational strategy for predicting highly stable multiple-point mutants that combines energy- and evolution-based approaches with smart filtering to identify additive stabilizing mutations. FireProt's reliability and applicability was demonstrated by validating its predictions against 656 mutations from the ProTherm database. We demonstrate that thermostability of the model enzymes haloalkane dehalogenase DhaA and  $\gamma$ -hexachlorocyclohexane dehydrochlorinase LinA can be substantially increased ( $\Delta T_m = 24^\circ\text{C}$  and  $21^\circ\text{C}$ ) by constructing and characterizing only a handful of multiple-point mutants. FireProt can be applied to any protein for which a tertiary structure and homologous sequences are available, and will facilitate the rapid development of robust proteins for biomedical and biotechnological applications.

## Author Summary

Proteins are increasingly used in numerous biotechnological applications. A key property determining proteins' applicability is their stability under operating conditions. Natural proteins can be stabilized by modification of their structure. Methods of molecular biology

budget of the Czech Republic and project ICRC Human Bridge – "Support of Study Stays of Czech Researchers Abroad III: Young Talent Incubator" (CZ.1.07/2.3.00/20.0239; <http://www.msmt.cz/>). KB and JBr were supported by the "Employment of Best Young Scientists for International Cooperation Empowerment" (CZ.1.07/2.3.00/30.0037) project co-financed by the European Social Fund (<http://ec.europa.eu/esf/home.jsp>) and the state budget of the Czech Republic (<http://www.msmt.cz/>). The work of JBe was supported by the Research and Application of Advanced Methods in ICT project (FIT-S-14-2299; <http://www.fit.vutbr.cz/>). MetaCentrum and CERIT-SC are acknowledged for providing access to computing facilities (LM2010005 and CZ.1.05/3.2.00/08.0144; <http://www.msmt.cz/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** JD and ZP are founders of the university spin-off company Enantis Ltd. All other authors have no competing interests.

allow introduction of modifications—mutations—to the protein structure at will, but it is not straightforward where to mutate and which amino acid to introduce for better stability. Computational methods can be used for prediction of stabilizing mutations using computers. Current computational methods predict libraries of single-point mutations, which need to be constructed individually, tested and recombined, resulting in non-trivial experimental effort. Here we present a robust computational strategy for predicting multiple-point mutants, providing extremely stabilized proteins with a minimal experimental effort.

## Introduction

Proteins are increasingly used in biotechnological applications as therapeutics [1], diagnostics [2], nanomaterials [1] and biocatalysts [3]. Despite numerous advantages, the utility of proteins is frequently restricted by their limited stability under practical conditions, such as high temperatures, extreme pH, or the presence of organic solvents or proteases. Their thermostability is usually positively correlated with stability and performance in the presence of denaturing agents [4], expression yield [5], serum survival time [6] and shelf-life [7]. Thus, it is a key determinant of proteins' applicability in biotechnological processes. High temperatures may also be required to prevent bacterial contamination during enzymatic food processing [8]. Moreover, thermostable proteins can tolerate much larger numbers of mutations than mesophilic variants and show enhanced evolvability in protein engineering projects [9].

Protein engineering is frequently applied to obtain more stable proteins. If successful, such efforts typically enhance the melting temperature ( $T_m$ ) of engineered proteins by 2 to 15°C [7, 10]. Extremely stabilized proteins with even greater increases in melting temperature ( $\Delta T_m$ ) have been engineered by incorporating multiple mutations, and several outstanding increases of up to 35°C have been achieved using directed evolution methods [8]. However, these methods generally require extensive experiments, including screening up to  $10^8$  colonies of organisms expressing mutant variants to identify stable constructs, and appropriate high-throughput screening assays must be available [11]. A currently popular strategy is saturated mutagenesis of hotspots identified by (semi-)rational approaches [7, 8, 12], such as the most flexible residues [10]; tunnel-forming residues [13]; or residues at multimeric interfaces [14]. The selected hotspots are then subjected to site-saturation mutagenesis (while leaving the rest of the protein unchanged) to create smaller smart libraries, markedly reducing the required screening to thousands of colonies.

A long-sought alternative to screening-based approaches is reliable *in silico* design of stability-enhancing mutations. Numerous stable proteins have been computationally engineered via diverse approaches (singly or in combination), e.g., identification of back-to-consensus or ancestral mutations, calculation of changes in folding free energies upon mutation, introduction of disulfide bridges and elimination of highly flexible regions [7, 8, 12]. However, mutants generated using computational methods have rarely surpassed the 15°C  $\Delta T_m$  threshold of outstanding stabilization as a result of neutral, destabilizing or function-corrupting mutations that were predicted as stabilizing due to moderate accuracy of these methods [15, 16]. To overcome this obstacle and provide substantial stabilization, predicted mutations are usually introduced by site-directed mutagenesis and tested individually. The most viable mutations are then recombined in multiple-point mutants assuming they have additive effects, but this is often invalid due to antagonistic epistatic effects of individual mutations [17]. For all those reasons,

no computational method capable of directly designing highly stable multiple-point mutants has been previously published.

Here we introduce a strategy, FireProt, for computationally designing multiple-point mutants, enabling significant improvements of protein stability with minimal experimental effort. We demonstrate its power by stabilizing the model proteins haloalkane dehalogenase (HLD) DhaA and  $\gamma$ -hexachlorocyclohexane dehydrochlorinase LinA. The method's general applicability was further verified by validation against information from the ProTherm database [18], demonstrating that it can be used to identify stabilizing mutations in diverse proteins with known tertiary structures and homologous sequences allowing phylogenetic analysis, and thus should have broad utility in protein stabilization projects.

## Results

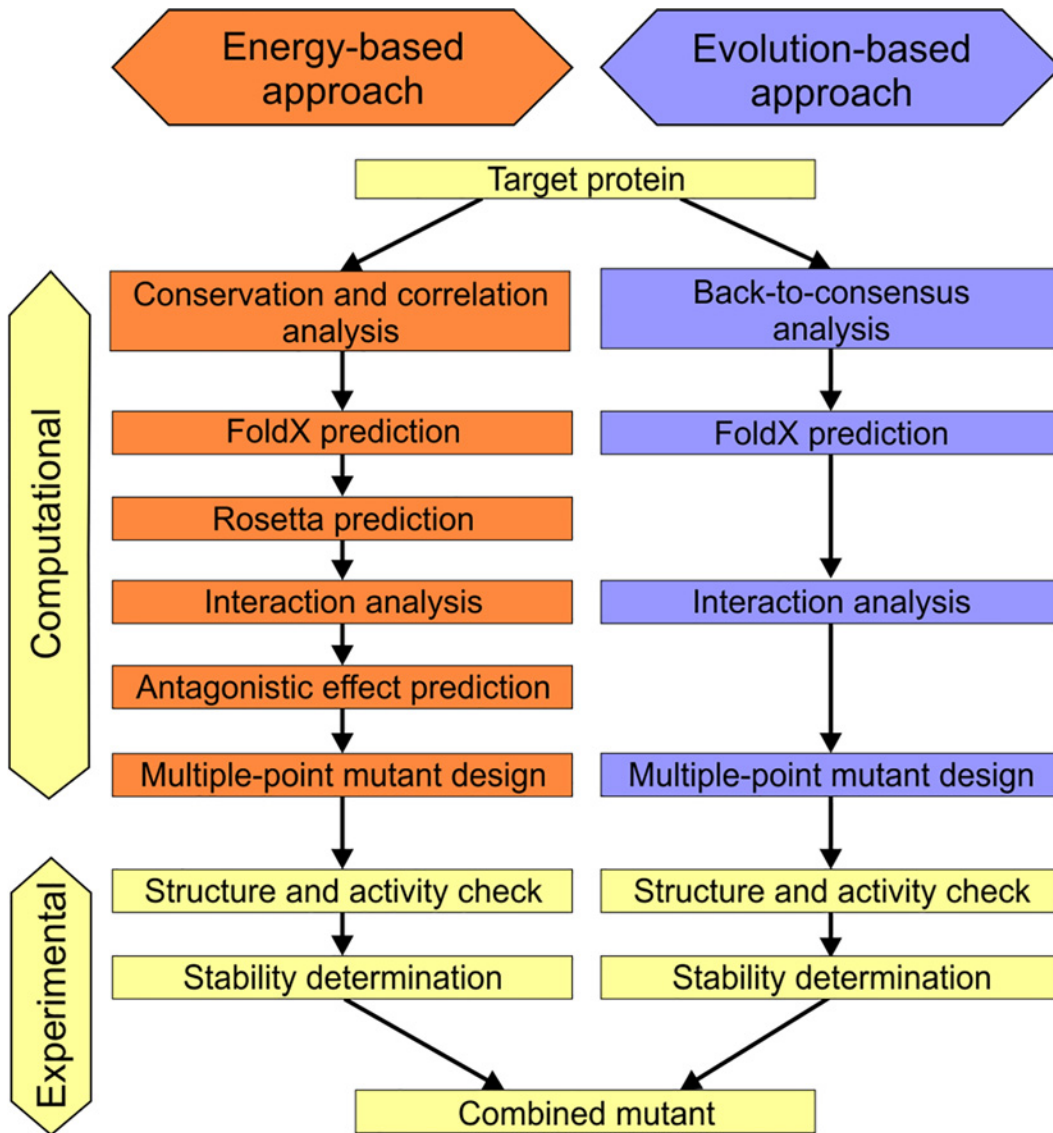
### Development of FireProt strategy for design of stabilizing multiple-point mutants

The FireProt strategy for protein stabilization is based on combining the best multiple-point mutants obtained from predictions of  $\Delta\Delta G$  following mutation from a set of crystal structures and evolutionary information derived from multiple sequence alignment (Fig 1). Additional pre- and post-processing filters are applied in both approaches to improve prediction reliability and reduce the required computational effort.

**Dataset construction.** FireProt's capacity to identify the most stabilizing single point mutations either by energy- or evolution-based approaches was evaluated using the dataset derived from the ProTherm database employing the records with absolute  $\Delta\Delta G$  values  $\geq 0.5$  kcal.mol<sup>-1</sup>. This value was selected since the estimated experimental error for measurements of  $\Delta\Delta G$  is about 0.48 kcal.mol<sup>-1</sup> [19]. When multiple  $\Delta\Delta G$  values were available for a single mutation, the value determined closest to the standard experimental conditions was retained. Ten most mutated proteins from the ProTherm database, representing all four protein structural classes (S1 Table), were selected to increase the chance that predicted mutations will be experimentally characterized. The total number of possible mutations in these 10 proteins was 30,058, but good quality experimental data were available for only about 2% of the mutations, yielding the dataset of 656 mutations (119 stabilizing and 537 destabilizing) at 337 different positions (S1 Table).

**Optimization of energy-based approach.** We evaluated the performance of four prediction tools: FoldX [20], Rosetta [21], ERIS [22] and CUPSAT [23], using the ProTherm dataset (S2 Table). The purpose was to select suitable combination of tools and the thresholds of predicted energy change upon mutation that would achieve very high precision (the ratio between the truly stabilizing mutations and mutations predicted as stabilizing) and simultaneously very low false positive rate (the fraction of destabilizing mutations incorrectly predicted as stabilizing from all truly destabilizing mutations). The stability change thresholds were tested in the range between -2.5 and +2.5 kcal.mol<sup>-1</sup> with the step size 0.5 kcal.mol<sup>-1</sup>. According to this evaluation, Rosetta and FoldX achieved the highest precision, 76% and 67%, respectively, when using the thresholds of -2 kcal.mol<sup>-1</sup> and -1 kcal.mol<sup>-1</sup> (S2 Table). The false positive rates for these thresholds for Rosetta and FoldX were 1% and 2%, respectively (S2 Table).

**Evaluation of energy-based approach.** Combination of the prediction by the two best performing tools using their optimal thresholds with conservation analysis by Rate4Site tool [24] in the energy-based FireProt approach performed notably better than either Rosetta or FoldX alone: FireProt achieved 100% precision and 0% false positive rate. Such high reliability in prediction was achieved at the expense of the number of recognized stabilizing mutations. Rosetta and FoldX correctly identified 16 and 20 truly stabilizing mutations out of 21 and 30



**Fig 1. Workflow of the FireProt method.** Individual steps involved in the energy- and evolution-based approaches

doi:10.1371/journal.pcbi.1004556.g001

mutations predicted by these tools as stabilizing (S2 Table). However, only 8 mutations were agreed upon as stabilizing by both methods following the FireProt strategy. Conservation analysis discarded 2 false positive mutations since they targeted immutable positions with high CONSURF conservation grade ( $\geq 8$ ) [25, 26]. The remaining 6 mutations selected by FireProt as stabilizing were all true positives. In addition to 6 mutations included in the evaluation, the FireProt predicted other 101 stabilizing mutations for 10 most mutated proteins for which, however, experimental data were not available (S3 Table).

**Evaluation of evolution-based approach.** In addition to the energy-based approach, the evolution-based approach of FireProt strategy was also evaluated using the ProTherm dataset. The back-to-consensus method [27] identified six potentially stabilizing mutations in the 10 most mutated proteins for which experimental data were available. Four of these mutations were true positives and two were false positives. The subsequent application of the FoldX filter (Fig 1) correctly discarded one of the false positive mutations, giving a precision of 80%.

**Evaluation of prediction for multiple-point mutants.** The last step in the development of the FireProt workflow before its application towards a design of thermostable proteins was to evaluate Rosetta's ability to predict the stability effects for multiple-point mutations. We collected a consistent dataset of previously measured stability changes in 46 mutants of DhaA enzyme for this purpose. Using this dataset a correlation of 0.81 between the predicted  $\Delta\Delta G$  and experimentally determined  $T_m$  values was observed, suggesting that Rosetta could be employed for this purpose (S1 Fig).

## Design of thermostable haloalkane dehalogenase DhaA

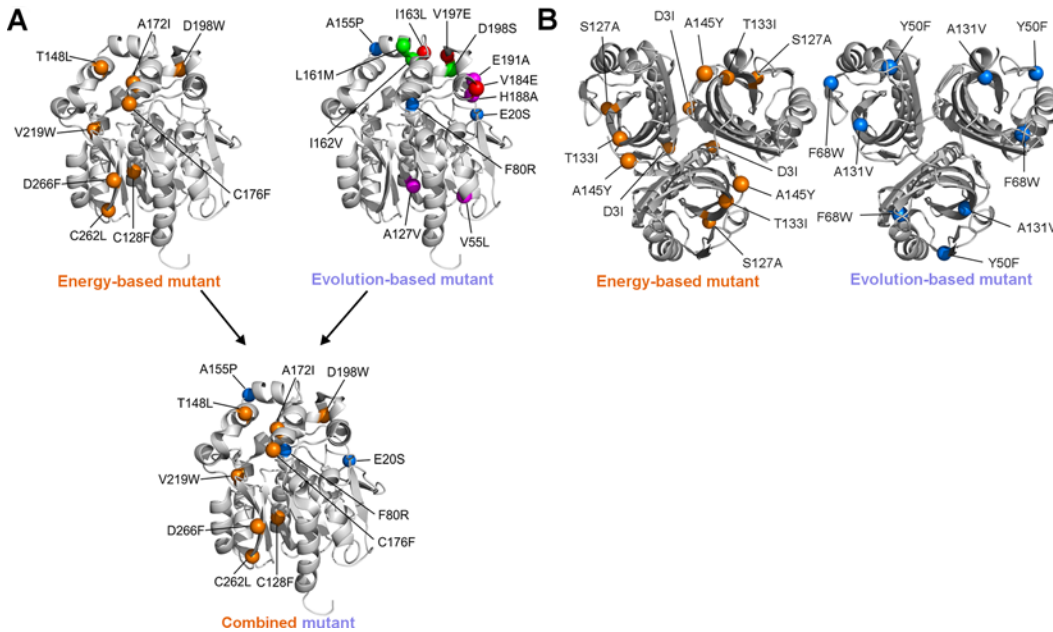
DhaA enzyme was selected as the first model protein due to the wealth of knowledge available on mutants engineered towards higher thermostability, prolonged half-life and stability in organic co-solvents that enables quantitative comparison of their performance with the mutants designed by FireProt [13, 28].

**Energy-based approach.** Out of 5,529 possible single-point mutations, 1,919 were at positions with high evolutionary conservation (CONSURF grade  $\geq 8$ ) or exhibiting evolutionary correlation with other residues (Correlated Mutation Analysis score  $\geq 0.8$  calculated by Comulculator tool [29]), indicating that these positions are functionally important. All these mutations were discarded to avoid major changes in the enzymes' activity or substrate specificity. FoldX analysis of  $\Delta\Delta G$  for the remaining 3,610 mutations identified 151 potentially stabilizing single-point mutations. Rosetta calculation was then applied to further decrease the number of potentially false positives among these mutations, resulting in 22 promising mutations (S4 Table). A mutation disrupting a salt bridge and five with antagonistic effects ( $\Delta\Delta G$  of double-point mutants  $> -3.0$  kcal.mol<sup>-1</sup>) were also discarded, leaving 16 potentially stabilizing mutations. Only the most favorable mutations at each position were further analyzed for their mutual interactions, leaving a final set of eight potentially non-antagonistic stabilizing mutations: C128F, T148L, A172I, C176F, D198W, V219W, C262L and D266F (Fig 2). The Rosetta predicted high stability of the recombined multiple-point mutant DhaA112 carrying all these mutations (Table 1).

**Evolution-based approach.** The back-to-consensus approach employing simple consensus and frequency ratio predictions identified 42 potentially stabilizing substitutions (S5–S8 Tables). Of these, 22 were excluded by FoldX predictions as potentially destabilizing ( $\Delta\Delta G > 0.5$  kcal.mol<sup>-1</sup>), and seven to preserve residues with side-chains involved in important interactions. In total, 13 back-to-consensus mutations passed all of the applied filters and were combined into four multiple-point mutants (Fig 2). These mutations were combined according to their different origin from two applied consensus techniques using either representative MSA of the HLD-II subfamily or MSA of whole HLD family (see Methods for more details). DhaA100 featured the I136L, V184E and V197E mutations, which were predicted by both simple consensus and frequency ratio analyses of the representative MSA of the HLD-II subfamily. DhaA101 contained the mutations E20S, F80R and A155P, which were predicted solely by the frequency ratio analysis, while DhaA102 included the mutations L161M, I162V and D198S, which were predicted solely by the simple consensus analysis. Finally, DhaA103 contained the V55L, A127V, H188A and E191A mutations, which were predicted by simple consensus analysis of the representative MSA of the whole HLD family. Interestingly, none of these mutants was predicted to be more stable than the wild-type by Rosetta (Table 1).

**Characterization of mutants designed by FireProt.** Expression and purification of all constructed mutants resulted in good yields and protein purity. Far-UV CD spectra of wild-type and mutants show that none of the mutations caused significant changes in secondary structure (S2A Fig) and all tested variants were active with 1-iodohexane at 37°C (Table 1).





**Fig 2. Location of stabilizing mutations in designed enzymes.** A) Locations of substitutions in energy-based, evolution-based and combined mutants of DhaA enzyme. Substitutions in the multiple-point mutant designed by the energy-based approach (DhaA112) are represented as orange spheres, while substitutions in multiple-point mutants designed by the evolution-based approach are represented as red (DhaA100), blue (DhaA101), green (DhaA102) and magenta (DhaA103) spheres. Mutations in the combined mutant (DhaA115) are colored in orange and blue in correspondence with their original mutants (DhaA112 and DhaA101). B) Locations of substitutions in energy-based, and evolution-based mutants of LinA enzyme. Substitutions in the multiple-point mutant designed by the energy-based approach (LinA01) are represented as orange spheres, while substitutions in multiple-point mutant designed by the evolution-based approach (LinA02) are represented as blue spheres.

doi:10.1371/journal.pcbi.1004556.g002

The constructed variants' thermostability was determined by thermally-induced denaturation using DSC (Table 1) and CD spectroscopy (S2A Fig). A substantial increase in melting temperatures ( $\Delta T_m$  16.2°C) was observed for variant DhaA112 designed by the energy-based approach, indicating strong stabilization effects of the introduced mutations (Table 1). The

**Table 1. Characteristics of predicted multiple-point mutants of DhaA.**

Method	Protein	Mutations	Rosetta $\Delta\Delta G$ (kcal.mol <sup>-1</sup> )	DSC		Activity at 37°C <sup>b</sup> (nmol s <sup>-1</sup> mg <sup>-1</sup> )
				$T_m$ (°C)	$\Delta T_m$ (°C)	
	DhaAwt	- <sup>a</sup>	- <sup>a</sup>	49.0 ± 0.7	- <sup>a</sup>	18.0
Energy-based	DhaA112	C128F + T148L + A172I + C176F + D198W + V219W + C262L + D266F	-32.0 ± 1.4	65.2 ± 0.1	+16.2	5.5
Evolution-based	DhaA100	I136L + V184E + V197E	1.1 ± 0.3	48.5 ± 0.4	-0.6	9.4
	DhaA101	E20S + F80R + A155P	0.8 ± 0.1	58.6 ± 0.3	+9.6	49.3
	DhaA102	L161M + I162V + D198S	2.3 ± 0.7	48.1 ± 0.1	-0.9	9.4
	DhaA103	V55L + A127V + H188A + E191A	3.0 ± 1.0	51.4 ± 0.1	+2.3	6.5
Combined	DhaA115	E20S + F80R + C128F + T148L + A155P + A172I + C176F + D198W + V219W + C262L + D266F	-32.4 ± 1.0	73.6 ± 0.1	+24.6	5.6

<sup>a</sup> not applicable

<sup>b</sup> activity determined with 1-iodohexane at 37°C and pH 8.6;  $\Delta\Delta G$ —predicted change in Free Gibbs Energy; DSC—Differential Scanning Calorimetry; DhaA115 combines mutations of DhaA101 and DhaA112

doi:10.1371/journal.pcbi.1004556.t001

effect of the consensus substitutions was moderate—only two tested variants, DhaA101 and DhaA103, showed positive thermostabilization effects ( $\Delta T_m$  9.6 and 2.3°C), while the mutations in DhaA100 and DhaA102 had neutral or destabilizing effects (Table 1). Combining the best energy-based mutant (DhaA112) with the best mutant identified by the evolution-based approach (DhaA101) produced a final mutant, DhaA115. Effects of the evolution-based substitutions were complementary and additive with those predicted by the energy-based approach, giving an outstanding increase in thermostability:  $\Delta T_m$  24.6°C (Table 1).

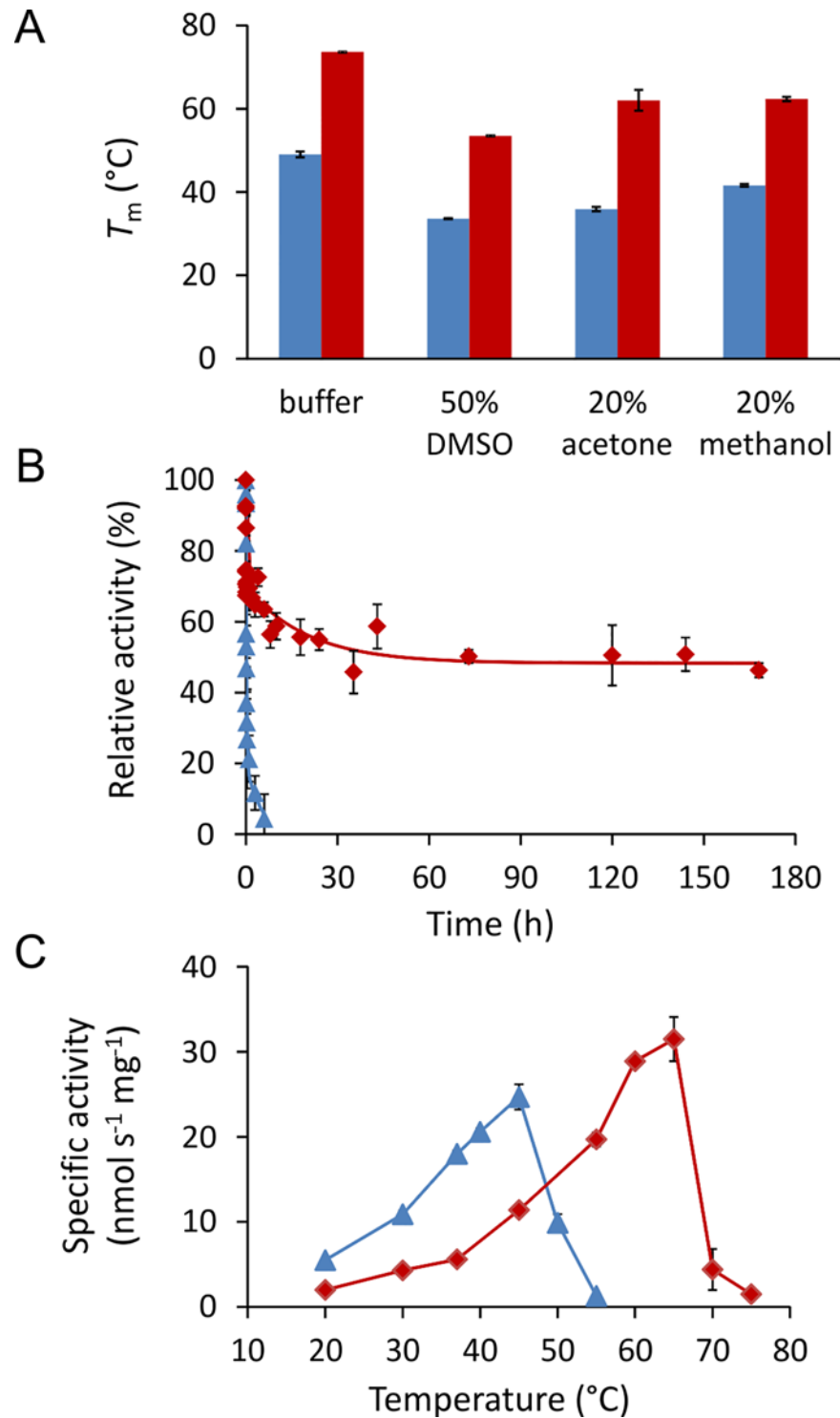
**Characterization of the final combined mutant.** The combined mutant DhaA115 was characterized in detail, in terms of its thermostability in the presence of organic co-solvents, half-life at elevated temperature and temperature profile. The  $T_m$  determinations show that the mutations had stabilizing effects in the presence of three organic co-solvents comparable to those in the pure buffer ( $\Delta T_m$ : 20 to 26°C; Fig 3A). The enhanced thermostability was also reflected in a strongly improved half-life at 60°C (Fig 3B). After seven days incubation at 60°C, the mutant DhaA115 still retained about 50% of its initial activity, while the wild-type became inactivated within six hours. Two inactivation phases were observed for all DhaA variants: rapid initial inactivation followed by a slower decay of activity (Fig 3B). The wild-type lost around 80% of its activity in the first fast phase, while the mutations in DhaA115 reduced the loss during the first inactivation phase to only 30%. The rate of inactivation during the first phase was comparable for both wild-type and DhaA115, while the rate during the second phase was dramatically (100-fold) slower for the mutant. Similar effects were observed with a previously reported stable variant of DhaA63 (denoted Dha8) [28] constructed using Gene Site-Saturation Mutagenesis (S3 Fig). The apparent optimal temperature shifted from 45°C for the wild-type enzyme to 65°C for the mutant DhaA115, and its specific activity with 1-iodohexane at optimum temperature was 28% higher, but the shape of the temperature profile remained largely unchanged (Fig 3C). Steady-state kinetic constants of the two enzymes determined at their respective suboptimal temperatures with 1-iodohexane revealed comparable catalytic properties (Table 2).

## Design of thermostable $\gamma$ -hexachlorocyclohexane dehydrochlorinase LinA

$\gamma$ -hexachlorocyclohexane dehydrochlorinase LinA enzyme was selected as the second model protein to illustrate broader applicability of FireProt strategy to other proteins of very different characteristics: (i) LinA is natively homotrimer (DhaA is monomer), (ii) LinA monomers form  $\alpha$ + $\beta$  barrel fold (DhaA possesses  $\alpha$ / $\beta$ -hydrolase fold), (iii) LinA is mainly composed of  $\beta$ -sheets ( $\alpha$ -helices and  $\beta$ -sheets and equally represented in DhaA) and (iv) LinA is with 156 amino acids two-times shorter (DhaA has 294 residues).

**Energy-based approach.** Out of 2,689 possible single-point mutations, 1,533 passed the evolutionary conservation or correlation filter. FoldX analysis of  $\Delta\Delta G$  for the remaining mutations identified 68 potentially stabilizing single-point mutations. Subsequent Rosetta calculation further decreased the number of promising mutations to 8 (S9 Table). A mutation D19M disrupting a salt bridge and T133L with antagonistic effect with residue D3 were also discarded, leaving 6 potentially stabilizing mutations at four positions. Only the most favorable mutations at each position were further analyzed: D3I, S127Y, T133I and A145H (Fig 3B). The Rosetta predicted high stability of the recombined multiple-point mutant LinA01 carrying all these four mutations (Table 3).

**Evolution-based approach.** The back-to-consensus identified 15 potentially stabilizing substitutions (S10 Table). Of these, 10 were excluded as destabilizing due to FoldX predictions. Mutation K20Y touches the halide-stabilizing residue and F113Y has a negative effect on



**Fig 3. Biochemical properties of DhaA wild-type and the final mutant DhaA115.** A) Melting temperatures of DhaA wild-type (blue) and DhaA115 (red) in the presence of indicated solvents. B) Half-life of DhaA wild-type (blue) and DhaA115 (red) determined at 60°C and pH 8.6 with the substrate 1-iodohexane. C) Temperature profiles of DhaA wild-type (blue) and DhaA115 (red) determined at pH 8.6 with the substrate 1-iodohexane.

doi:10.1371/journal.pcbi.1004556.g003



**Table 2. Steady-state kinetic constants of DhaA wild-type and the final mutant Dha115 determined with 1-iodohexane at 37°C and 57°C, respectively, and pH 8.6.**

Enzyme	$k_{cat}$ (s <sup>-1</sup> )	$K_{0.5}$ (μM)	$n$	$K_{si}$ (mM)
DhaAwt	2.47 ± 0.01	12.00 ± 1.14	1.80 ± 0.04	0.53 ± 0.01
DhaA115	2.85 ± 0.03	5.00 ± 2.31	1.89 ± 0.01	-

$K_{0.5}$ —concentration of substrate at half maximal velocity,  $k_{cat}$ —catalytic constant,  $n$ —Hill coefficient  $K_{si}$ —substrate inhibition constant

doi:10.1371/journal.pcbi.1004556.t002

enzyme activity according to Uniprot database. Remaining 3 back-to-consensus mutations passed all of the applied filters and were combined into three-point mutant: Y50F, F68W and A131V (Fig 3B).

**Characterization of mutants designed by FireProt.** Expression and purification of all constructed mutants resulted in good protein yields and purity. Comparison of the far-UV CD spectra of LinA wild-type and its mutants show that none of the mutations caused significant changes in secondary structure (S2B Fig). Both LinA variants retained similar level of specific activity as LinAwt (Table 3) showing that the introduced mutations did not alter activity negatively. The thermostability of the constructed LinA variants was determined by thermally-induced denaturation using both DSC (Table 3) and CD spectroscopy (S2B Fig). Similar to the energy-based DhaA variant, the energy-based LinA variant (LinA01) showed a substantial increase in melting temperatures ( $\Delta T_m$  20.9°C) again showing the strong stabilization effects of the introduced mutations (Table 3). The evolution-based mutant (LinA02) showed small decrease in thermostability ( $\Delta T_m$  -3.7°C) indicating that mutations identified by consensus methods are not conserved to preserve the stability of the enzyme (Table 3). No combined mutant was constructed due to the absence of stable evolution-based mutations.

### Structural basis of mutation effect

Visual inspection of mutant structures coupled with detailed analysis of their individual energy terms calculated by Rosetta provided indications of the possible structural basis of protein stabilization by mutations of DhaA115 and LinA01 (S11 Table). These mutations were introduced to various locations in the protein structure with different types of secondary structures.

**Stabilizing mutations in DhaA.** Out of 11 mutations, 3 residues are lining a main transport tunnel, 3 mutated residues were buried in the protein core and 5 are exposed to solvent on the protein surface (Fig 2A and S11 Table). 8 mutations identified by the energy-based approach introduce bulkier, more hydrophobic residues (S11 Table) that probably enhance

**Table 3. Characteristics of predicted multiple-point mutants of LinA.**

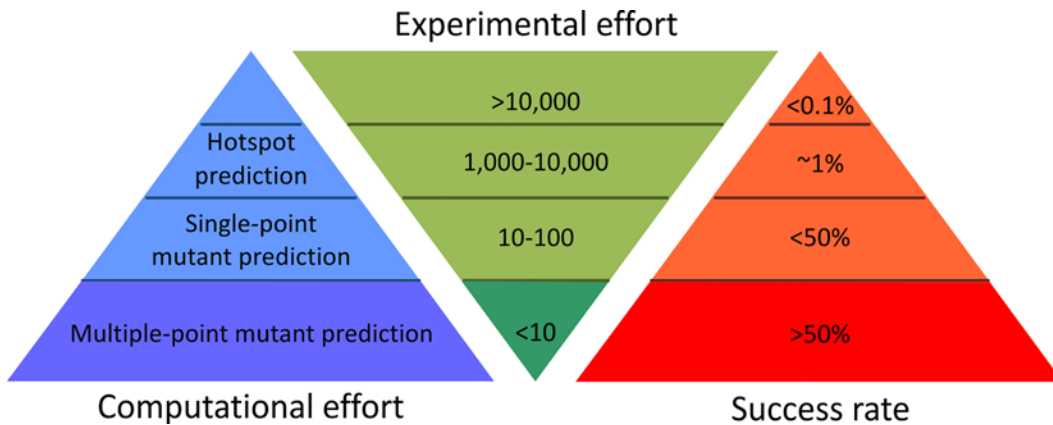
Method	Protein	Mutations	Rosetta $\Delta\Delta G$ (kcal.mol <sup>-1</sup> )	DSC		Activity at 30°C <sup>b</sup>	
				$T_m$ (°C)	$\Delta T_m$ (°C)	(μmol s <sup>-1</sup> mg <sup>-1</sup> )	
	LinAwt	- <sup>a</sup>	- <sup>a</sup>	41.4 ± 0.1	- <sup>a</sup>	0.21 (0.12 mM) <sup>c</sup>	1.91 (0.38 mM) <sup>c</sup>
Energy-based	LinA01	D3I + S127Y + T133I + A145H	-31.4	62.3 ± 0.4	+20.9	0.32 (0.12 mM) <sup>c</sup>	1.29 (0.34 mM) <sup>c</sup>
Evolution-based	LinA02	Y50F + F68W + A131V	0.1	37.7 ± 0.2	-3.7	0.17 (0.11 mM) <sup>c</sup>	ND

<sup>a</sup> not applicable

<sup>b</sup> activity determined with γ-hexachlorocyclohexane at 30°C and pH 8.6

<sup>c</sup> initial γ-HCH concentration is given since it affects determined specific activity;  $\Delta\Delta G$ —predicted change in Free Gibbs Energy; DSC—Differential Scanning Calorimetry; ND, not determined

doi:10.1371/journal.pcbi.1004556.t003



**Fig 4. Schematic comparison of protein stabilization methods.** Examples of representative methods with their characteristics and success rates are presented in [S12 Table](#).

doi:10.1371/journal.pcbi.1004556.g004

stability by improving packing of atoms in the protein interior and/or strengthening hydrophobic interactions. The contributions to stability of the mutations proposed by the evolution-based approach are more difficult to explain. The A155P mutation (at the fourth most flexible position in the protein structure) could increase rigidity by introducing proline to the affected loop and account for most of the observed stability improvement, while effects of the E20S and F80R mutations are probably neutral or restructuring the charged network on the surface of DhaA ([S11 Table](#)).

**Stabilizing mutations in LinA.** Out of 4 stabilizing mutations, 2 are buried in the protein core and 2 are exposed on the protein surface ([Fig 2B](#) and [S11 Table](#)). In correspondence with observation for stabilizing mutations introduced into DhaA enzyme, all 4 mutations identified by the energy-based approach for LinA introduced bulkier and more hydrophobic residues ([S11 Table](#)).

## Discussion

The last decade has seen significant advances towards more rational approaches to reduce the experimental effort required to engineer highly stable proteins ([Fig 4](#) and [S12 Table](#)). As a contribution to these efforts we have developed a hybrid strategy integrating energy-based and evolution-based approaches, with smart filtering of mutations that are destabilizing or may impair enzymes' functions, enabling the identification of additively stabilizing substitutions in multiple-point mutants. It is essential to correctly configure all of the tools used in both the energy- and evolution-based approaches of the FireProt workflow in order to achieve robust and reliable predictions. Therefore, individual steps of the workflow were verified using a dataset featuring diverse proteins from the ProTherm database. The predictions carried out for 656 mutations confirmed the FireProt's precision: the energy- and evolution-based approaches identified stabilizing mutations with success rates of 100% and 80%, respectively. Strikingly, only one stabilizing mutation that exceeded our thresholds was identified by both approaches, suggesting that they are highly complementary. The potential downside of the stringent conditions imposed to avoid false positives was that 92% of the available stabilizing mutations were discarded. However, the remaining correctly identified stabilizing mutations should be more than sufficient to construct highly stable catalysts ([S3 Table](#)).

When the energy-based approach was applied to DhaA and LinA enzymes, the removal of conserved and correlated positions from analysis helped to avoid modification of structurally

and functionally important residues, thereby greatly reducing the number of possible mutations requiring evaluation by computationally intensive free energy calculation. Since FoldX computation is about an order of magnitude faster than Rosetta, it was applied as a pre-filter, further reducing numbers of mutations to be analyzed by Rosetta. Regarding the prediction of multiple-point mutants, simple recombination of the most promising mutants could weaken stabilization, since strong antagonistic effects were detected even at the level of double-point mutants. The thermostability enhancement for the eight- and four-point mutants predicted by this approach, DhaA112 ( $\Delta T_m$  16°C) and LinA01 ( $\Delta T_m$  21°C), both exceeded the threshold for outstanding stabilization, although none of the introduced mutations optimized either hydrogen bonds or charge-charge interactions. This may be due to sampling limited rotamer libraries during the calculations and the requirement for both FoldX and Rosetta to unambiguously evaluate selected mutations as stabilizing. FoldX and Rosetta employ simplified scoring functions and despite using three protein structures for analysis, only limited protein flexibility is allowed, implying that it should be possible to supplement mutations proposed by free energy calculations with beneficial substitutions identified using different principles.

To this end, additional mutations were selected by the evolution-based approach. The mutations predicted by the back-to-consensus method were filtered by FoldX to discard mutations proposed due to function-related evolutionary constraints rather than structural stabilization. This filtering step proved to be very important as over half of the mutations were discarded as potentially destabilizing. Interestingly, all five multiple-point mutants DhaA100-DhaA103 and LinA02 were predicted as destabilizing by Rosetta and had to be tested experimentally. While this prediction was accurate for three of them (DhaA100, DhaA102 and LinA02), the other two mutants (DhaA101 and DhaA103) were clearly more stable than the wild-type. This result suggests that some underlying principles important for stability detected by the back-to-consensus method are not captured by the applied Rosetta protocol. We speculate that these may include larger backbone rearrangements, interactions with ions present in the solvent, or other entropic contributions that are not well accounted for in the current protocols. Experimental characterization of these mutants by microcalorimetry, temperature-jump stopped-flow and protein crystallography is currently on-going in our laboratory. Despite its lower reliability, the evolution-based approach should still be considered as useful supplement to the energy-based approach, potentially enabling further improvement in the stability of designed proteins. The final 11-point mutant DhaA115 arising from this hybrid prediction strategy is one of the most stable HLD protein known to date ( $\Delta T_m > 24^\circ\text{C}$ ) [13, 30].

We have compared our strategy against several methods providing exceptional protein stabilization (S12 Table). The experimentally intensive protocols of directed evolution and hot-spot predictions can provide engineered enzymes with comparable enhancement. However, since their success rate is generally below 1%, stable proteins can only be obtained after extensive screening. Notably, two of these studies also focused on improving stability of the enzyme DhaA. In one, an eight-point mutant DhaA was obtained with a  $\Delta T_m$  of 18°C after screening all 121,000 possible variants [28]. We have obtained a clearly superior enzyme after experimental evaluation of as few as six mutants, highlighting the importance of removing mutations with antagonistic and uncertain stabilizing effects. In the other study performed with DhaA, four hotspots in an access tunnel were experimentally randomized, requiring experimental screening of 5,000 mutations [13], and the  $\Delta T_m$  for the best four-point mutant was 19°C.

Highly stable proteins have been obtained by *in silico* prediction of stabilizing effects of single-point mutations in four recently published studies [15, 31, 32, 33]. In one, 67 variants of epoxide hydrolase with mutations identified as potentially stabilizing by the FRESKO method were experimentally tested, 24 were reportedly more stable than the parent protein, and the variant with the best permutation of mutations had remarkably enhanced thermostability

( $\Delta T_m$  36°C) [15]. Much of this enhancement arose from disulfide bridges at the dimer interface, making this approach particularly suitable for multimeric proteins. In another of the studies, four out of six engineered methionine aminopeptidases designed by the Rosetta<sub>VIP</sub> method were found to be stabilizing and a combined five-point mutant reportedly had a  $\Delta T_m$  of 18°C [31]. The authors noted that their final construct is still less stable than the most thermostable native aminopeptidases and that the method is particularly effective for mutagenesis of buried residues around internal cavities. In the other study, a 12-point mutant of Tobacco 5-epi-aristolochene synthase was generated using the SCADS method with an impressive  $\Delta T_m$  (45°C), but at the expense of 98% of catalytic activity at the optimal temperature [32]. In comparison to the methods applied in these and other relevant studies (S12 Table), FireProt affords a reduction of experimental screening effort due to robust identification of stabilizing mutations and ensuring their additivity. In addition, it has promising applicability to diverse proteins, potentially all proteins with known tertiary structure and homologous sequences, due to the diverse locations of introduced mutations and universal applicability of underlying principles.

In summary, the presented hybrid strategy FireProt affords rapid design of stable proteins. Consideration of the additivity of identified potentially beneficial mutations enables prediction of multiple-point mutants with significantly enhanced stability. Despite a dramatic reduction in experimental effort, the workflow provided two proteins with outstanding stability. One of them a HLD with greater thermostability than all known HLD enzymes, either obtained from thermophilic organisms or engineered using extensive combinatorial screening. Furthermore, owing to the smart filtering, this strategy is affordable by users with limited access to powerful computer facilities. In addition, implementation of the FireProt strategy in the web-based protein engineering tool Hotspot Wizard [34] is currently on-going in our laboratory to ensure user convenience.

## Methods

### Bioinformatics analysis

**Construction of multiple sequence alignments.** Sequences of six experimentally characterized HLDs—DhaA, LinB, DrbA, DmbC, DhIA and DmbB—or the sequence of LinA were used as queries for PSI-BLAST [35] searches against the nr NCBI database (version July-2009 and May-2015, respectively) [36], with threshold *E*-values of  $10^{-10}$  and  $10^{-15}$  for the initial BLAST search and inclusion of a sequence in the position-specific matrix, respectively. Sequences collected after three PSI-BLAST iterations were clustered by CD-HIT [37] at 90% identity threshold. The resulting dataset (including 8,226 sequences for DhaA and 946 for LinA) was clustered with CLANS [38] using default parameters and varying threshold *P*-values. Sequences clustering with query at the *P*-value of  $10^{-29}$  were extracted and aligned with MUSCLE [39]. All artificial, incomplete or divergent sequences were removed. Final multiple sequence alignment (MSA) of LinA contained 13 sequences. The prepared MSA of the HLD protein family comprised 168 sequences. All sequences (47) belonging to the HLD-II subfamily were then extracted to create a HLD-II subfamily dataset and aligned with MUSCLE. To reduce possible bias toward highly similar sequences, UniqueProt [40] was used (with a HSSP cut-off value of 40) to select representative sequences from both datasets. The created representative HLD family and HLD-II subfamily datasets comprised 87 and 27 sequences, respectively. Each representative dataset was then aligned with MUSCLE [39].

**Analysis of evolutionary conservation and correlation.** The MSA of the whole HLD-II subfamily or the MSA of LinA was used to estimate the level of conservation at individual positions. Normalized evolutionary rates for each amino acid site of the MSA were calculated by the Bayesian method implemented in Rate4Site v2.01 [24] with the WAG evolutionary model

[41] and maximum likelihood optimization of branch lengths using a gamma model with four discrete categories. Calculated normalized evolutionary rates were converted to CONSURF conservation grades [25, 26]. Positions with a grade  $\geq 8$  were considered immutable. The MSA of the whole HLD family or MSA of LinA was used to identify positions with correlated mutations (and a threshold Correlated Mutation Analysis score  $\geq 0.8$ ), by applying the 3DM database's Comulator online tool [29].

**Back-to-consensus analysis.** Back-to-consensus mutations were selected by analyzing both representative MSAs of HLDs and the MSA of LinA by simple consensus and frequency ratio approaches [27]. Residues from poorly aligned regions (DhaA residues 1–18, 131–179 and 279–293 from the HLD family MSA, and residues 1–14 from the HLD-II subfamily MSA) were excluded from the analysis. The simple consensus analysis was performed using the consensus cut-off of 0.5, meaning that a given residue must be present at a given position in at least 50% of all analyzed sequences to be assigned as the consensus residue. Two cut-offs were simultaneously applied in the frequency ratio analyses: a frequency cut-off of 0.2 for the maximum allowed ratio between target and conserved residue frequencies, and a minimal frequency of 0.4 for the most conserved residue.

**Compilation of a validation dataset.** The validation dataset was compiled from ProTherm records that include experimentally determined differences between the Gibbs free energies of folding for mutant proteins and the corresponding wild type ( $\Delta\Delta G$ ). To increase the reliability of the analysis, only records with absolute  $\Delta\Delta G$  values of  $\geq 0.5$  kcal.mol<sup>-1</sup> were included; the experimental error for measurements of  $\Delta\Delta G$  is estimated to be about 0.48 kcal.mol<sup>-1</sup> [19]. When multiple  $\Delta\Delta G$  values were available for a single mutation, only the value determined under the experimental conditions closest to the physiological pH of 7 was retained. The dataset was also limited to mutations in the ten most mutated proteins in ProTherm. Multiple sequence alignments for each protein in the dataset were constructed using a protocol similar to that applied in the analysis of the model enzyme DhaA. However, an automatic multi-step procedure was developed to circumvent the need to manually select PSI-BLAST queries. First, the name of the relevant protein family was found in the SCOP database [42]. Then, all members of the same protein family were clustered by CD-HIT using an identity threshold of 90%. Finally, up to five representatives of each resulting cluster were selected at random to constitute the set of PSI-BLAST queries.

## Molecular modeling

**Preparation of protein structures.** Crystal structures of wild-type DhaA (PDB ID: 1BN6, 1BN7 and 1CQW) and wild-type LinA (PDB ID: 3A76) were downloaded from the RCSB PDB database [43]. PyMOL v1.4 [44] was used to model substitutions A172V, I209L and G292A in the crystal structures of DhaA to ensure their correspondence with DhaA from *Rhodococcus rhodochrous* (GI number 3114657). The crystal structures were then prepared for predictions by removing ligands and water molecules. Missing side-chain atoms were added by the <RepairPDB> module of FoldX 3.0 [20]. Repaired structures were minimized by the minimize\_with\_cst module of Rosetta [21], with: both backbone and side-chains optimization enabled (*—sc\_min\_only false*), distance for full atom pair potential set to 9 Å (*—fa\_max\_dis 9.0*), standard weights for the score function and a constraint weight of 1 (*—constraint\_weight 1.0*). Output from the minimization was used to create a constraint file by script convert\_to\_cst\_file.sh.

**Prediction of stability effects by FoldX.** Stability effects of all possible single-point mutations were estimated using the <BuildModel> module of FoldX. Calculations were performed five times for each mutation following the recommended protocol (pH 7, temperature 298 K, ion strength 0.050 M, VdWDesign 2). Mutations with predicted  $\Delta\Delta G$  averaged over all three



analyzed structures smaller than  $-1.0 \text{ kcal.mol}^{-1}$  were considered as stabilizing, while a tighter criterion was applied for detecting potentially destabilizing mutations ( $\Delta\Delta G > 0.5 \text{ kcal.mol}^{-1}$ ).

**Prediction of stability effects by Rosetta.** Protocol 16 incorporating backbone flexibility within the `ddg_monomer` module of Rosetta was applied according to Kellogg et al. [21]. The soft-repulsive design energy function (`—soft_rep_design_weights`) was used for repacking side-chains (`—sc_min_only false`). Optimization was performed on each whole protein without distance restriction (`—local_opt_only false`). The previously created constraint file was used during backbone minimization (`—min_cst true`). Three rounds of optimization with increasing weight on the repulsive term (`—ramp_repulsive true`) were applied. The minimum energies from 20 and 50 iterations were used as the final parameters describing the stability effects of single- and multiple-point mutations, respectively. Mutations with  $\Delta\Delta G$  averaged over all three analyzed DhaA structures or three chains of LinA  $< -2.0 \text{ kcal.mol}^{-1}$  were considered as stabilizing. The additivity of stabilizing mutations was evaluated by predicting the stability of variants with all pairs of potentially stabilizing single-point mutations. Mutation pairs for which the respective double-point mutants showed predicted  $\Delta\Delta G > -3.0 \text{ kcal.mol}^{-1}$  were considered as antagonistic. The cumulative mutants were then prepared by combining mutually additive single-point mutations starting with the most stabilizing mutation. If there were more than one stabilizing mutations at the same position, the most favorable mutation was used.

**Analyses of interactions.** Selected mutations were visually analyzed in PyMOL. All three crystal structures of DhaA and three chains of LinA were analyzed for the presence of side-chains involved in intra-molecular salt bridges by the ESBRI server [45] and additional intra-molecular interactions using the Protein Interactions Calculator server [46]. An interaction had to be present in at least one of the analyzed structures or chains to be considered as important.

## Construction of mutants and biochemical characterization

**Subcloning.** All reagents and primers were purchased from Sigma-Aldrich unless otherwise specified. Restriction enzymes, T4 DNA ligase and accompanying buffers were purchased from New England Biolabs. Genes encoding tested *Rhodococcus rhodochrous* DhaA (Uniprot: P0A3G2) mutants were cloned in the pET21b (EMD Biosciences) (DhaAwt, 101, 110–112, 115–116) or pAQN vector [47] (DhaA63, 100–103) using the restriction enzyme pairs *NdeI/HindIII* or *BamHI/HindIII*, respectively, followed by ligation with T4 DNA ligase according to the supplier's protocol. The plasmid pET28-LinAwt encoding the wild-type LinA from *Sphingobium japonicum* UT26 (Uniprot: P51697) was a gift from Dr. Yuji Nagata [48]. The genes encoding for the mutants were cloned in the pET28b (EMD Biosciences) using the restriction enzymes *NdeI* and *EcoRI*, followed by ligation with T4 DNA ligase (Promega) according to the supplier's protocol. Correct integration of the genes was verified by sequencing (GATC Biotech) and analyzed using Clone Manager Professional (Sci-Ed Software) and BioEdit (Ibis Biosciences) software. DhaA and LinA variants were expressed with a C-terminal and N-terminal His<sub>6</sub>-tag, respectively, to facilitate purification.

**Enzyme expression and purification.** The His<sub>6</sub>-tagged DhaA and LinA mutants were overexpressed in *Escherichia coli* BL21(DE3) cells as previously described [47]. Proteins were purified using Ni-NTA Superflow Cartridges (Qiagen) and a previously described method [49]. Protein concentration was determined by assays with the Bradford reagent (Sigma-Aldrich). The purity of purified proteins was checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie Brilliant Blue R-250 staining.

**Enzyme activity assays with DhaA.** Reaction mixtures containing 12  $\mu\text{L}$  of substrate in 12 mL 100 mM glycine buffer (pH 8.6) were preheated at 37°C for 30 min, 240  $\mu\text{L}$  of purified

enzyme (0.4–1.0 mg.mL<sup>-1</sup>) was added to initiate the reaction, and it was monitored by withdrawing samples at periodic intervals (0–30 min). The samples were immediately mixed with 35% nitric acid to terminate the reaction. The released halide ions were measured spectrophotometrically at 460 nm after reaction with mercuric thiocyanate and ferric ammonium sulfate [50]. Dehalogenating activity was quantified as the rate of halide product formation per unit time. Temperature profiles of DhaAwt and DhaA115 were evaluated by measuring their activity, as described above, at temperatures ranging from 20°C–75°C in three independent replicates. The operational stability was evaluated by measuring residual activity after incubating 1 mL enzyme samples (1.0 mg.mL<sup>-1</sup>) at 60°C in a heat block (Biosan Pst-100 HL).

Residual activity was determined using a MicroLab StarLet Manuload Liquid Handling Robot (Hamilton). For the residual activity measurements, 1 mL of 100 mM Glycine buffer (pH 8.6) with 1 µL 1,2-dibromoethane was incubated at 37°C for 30 min then 50 µL of heat-treated enzyme solution was added to start the reaction (DhaAwt 0.1 mg.mL<sup>-1</sup>, DhaA115 and DhaA63 1.0 mg.mL<sup>-1</sup>). Samples (100 µL) were taken before enzyme addition (0 min) and after 5, 10 and 15 min reaction time. Samples were then transferred to wells of a MTP microplate containing 10 µL 35% H<sub>3</sub>NO<sub>4</sub> for inactivation. After all samples had been collected halide product was detected as described earlier. OD<sub>460nm</sub> was then measured using a Sunrise microplate reader (Tecan). Dehalogenation activities were quantified by the slope of the regression between the product concentration and time.

**Enzyme activity assay with LinA.** The activity of the LinA variants was tested with γ-hexachlorocyclohexane (γ-HCH) at 30°C and analyzed using GC-MS. Saturating γ-HCH substrate mixtures in 1 mL 100 mM glycine buffer (pH 8.6) were preheated at 30°C for 30 min. 10 µL of purified enzyme (8.9–28.8 µg.mL<sup>-1</sup>) was added to initiate the reaction that was monitored by withdrawing 1 µL samples at 15 min periodic intervals (0–75 min). The samples were immediately analyzed by GC-MS. Gas chromatograph equipped with the PAL robotic tool change system enabled fully automatized sample preparations, organic extractions and analysis. The consumption of particular substrates was quantified using gas chromatography (Trace 1300, Thermo Scientific, USA) equipped with capillary column TG-SQC (30m x 0.25mm x 0.25µm, Thermo Scientific, USA) and connected with mass spectrometer (ISQ™ LT Single Quadrupole, Thermo Scientific, USA). The 1 µL samples were injected into split-splitless inlet at 250°C, with split ratio 1:50. The temperature program was isothermal at 40°C for 1 min, followed by increase to 250°C at 20°C.min<sup>-1</sup> and hold for 4 min. The flow of carrier gas (He) was 1 mL.min<sup>-1</sup>. The MS was operated at SCAN mode (30 to 320 amu). The temperature of ion source and GC-MS transfer line was 200°C and 250°C, respectively. Dehydrochlorination activities were quantified following the decrease in concentration of γ-HCH over time.

**Steady-state kinetic measurements.** Substrate to product conversion by the action of DhaAwt and DhaA115 was monitored by using the isothermal titration microcalorimeter VP-ITC (MicroCal, Piscataway, USA) at 37°C and 57°C, respectively. The substrate 1-iodohexane was dissolved in 100 mM glycine buffer (pH 8.6) and the solution was allowed to reach thermal equilibrium in the reaction cell (1.4 mL). The reaction was initiated by injecting 10 µL of enzyme solution containing either 22 µM DhaAwt or 24 µM DhaA115 into the reaction cell. Enzymes were dialyzed overnight against the same glycine buffer. The measured rate of heat change was assumed to be directly proportional to the velocity of the enzymatic reaction according to the Eq (1)

$$\frac{dQ}{dt} = -\Delta H V \frac{d[S]}{dt} \quad (1)$$

where  $\Delta H$  is the enthalpy of the reaction,  $[S]$  is the substrate concentration, and  $V$  is the volume of the cell.  $\Delta H$  was determined by titrating the substrate into the reaction cell containing the

enzyme. Each reaction was allowed to proceed to completion. The integrated total heat of a reaction was divided by the amount of injected substrate. The evaluated rate of substrate depletion ( $-d[S]/dt$ ) and corresponding substrate concentrations were then fitted by nonlinear regression to kinetic models using Origin 6.1 (OriginLab, Massachusetts, USA).

**Circular dichroism (CD) spectroscopy.** CD spectra (190 to 260 nm) were obtained from samples of the purified enzymes ( $0.20\text{--}0.25\text{ mg}\cdot\text{mL}^{-1}$  in 50 mM phosphate buffer, pH 7.5, in a 0.1 cm quartz cuvette) to confirm their correct folding, using a Chirascan CD Spectrometer equipped with a Peltier thermostat (Applied Photophysics). Each presented spectrum is the baseline-corrected average of 5–10 scans. Mean residue ellipticity ( $\Theta_{\text{MRE}}$ ) was calculated using Eq (2)

$$\Theta_{\text{MRE}} = \frac{\Theta_{\text{obs}} \cdot M_w \cdot 100}{n \cdot c \cdot l} \quad (2)$$

where  $\Theta_{\text{obs}}$  is observed ellipticity in degrees,  $M_w$  is the protein molecular weight,  $n$  is number of residues,  $l$  is the cell path length (0.1 cm),  $c$  is the protein concentration and the factor 100 originates from conversion of the molecular weight to  $\text{mg}\cdot\text{dmol}^{-1}$  [51].

Thermal unfolding of the enzyme variants was followed by monitoring the ellipticity at 222 nm over the temperature range of 20°C to 90°C, with a resolution 0.2°C, at a heating rate at 1°C/min. Recorded denaturation curves of tested enzyme were fitted to sigmoid curves (Boltzmann) using OriginPro8 software (OriginLab, Massachusetts, USA). The melting temperatures ( $T_m$ ) were evaluated from the collected data as a midpoint ( $x_0$ ) of the normalized thermal transition.

**Differential scanning calorimetry (DSC).** Melting temperatures of the purified enzymes were determined by monitoring their heat capacity in solution ( $1.0\text{ mg}\cdot\text{mL}^{-1}$ ) in 50 mM aqueous phosphate buffer (pH 7.5) and in the presence of three cosolvents: 20% acetone, 20% methanol and 50% DMSO (v/v). The measurements were acquired, after degassing, at temperatures from 20 to 100°C using the VP-capillary differential scanning calorimetry (DSC) system (MicroCal) and a  $1^\circ\text{C}\cdot\text{min}^{-1}$  heating rate. The melting point of each protein was determined as the temperature at which the heat capacity curve peaked [52].

## Supporting Information

**S1 Fig. Correlation between  $\Delta T_m$  and  $\Delta\Delta G$  values of 46 DhaA mutants.** The experimentally characterized homogenous set of DhaA mutants [1, 2] employed during validation of the Rosetta approach is shown as blue squares. The red line represents the trend in the experimentally characterized mutants (correlation coefficient, 0.81).  
(TIF)

**S2 Fig. Far-UV CD spectra of the tested mutants and determined melting temperatures.** A) Variants of haloalkane dehalogenase DhaA. B) Variants of  $\gamma$ -hexachlorocyclohexane dehydrochlorinase LinA. The melting temperatures ( $T_m$ ) were evaluated as midpoints of the normalized thermal transitions.  
(TIF)

**S3 Fig. Half-life of DhaA63 determined at 60°C in 50 mM phosphate buffer pH 7.5.**  
(TIF)

**S1 Table. Composition of single-point mutation dataset derived from ProTherm database.**  
(PDF)

**S2 Table. Performance of four evaluated prediction tools at different decision thresholds.**  
(PDF)

**S3 Table. Stabilizing mutations selected for the 10 most mutated proteins from ProTherm.**  
(PDF)

**S4 Table. Results of the energy-based analysis of DhaA.**  
(PDF)

**S5 Table. Results of the frequency ratio analysis of the HLD-II subfamily.**  
(PDF)

**S6 Table. Results of the simple consensus analysis of the HLD-II subfamily.**  
(PDF)

**S7 Table. Results of the frequency ratio analysis of the HLD family.**  
(PDF)

**S8 Table. Results of the simple consensus analysis of the HLD family.**  
(PDF)

**S9 Table. Results of the energy-based analysis of LinA.**  
(PDF)

**S10 Table. Results of the consensus analysis of LinA.**  
(PDF)

**S11 Table. Predicted effects of DhaA115 and LinA01 mutations on its stability.**  
(PDF)

**S12 Table. Examples of methods providing enzymes with outstanding stabilization.**  
(PDF)

**S1 Text. Supporting references.**  
(PDF)

## Acknowledgments

We kindly thank Dr. Yuji Nagata (Graduate School of Life Sciences, Tohoku University, Japan) for providing the pET28-LinAwt plasmid.

## Author Contributions

Conceived and designed the experiments: DBe KB ES JBe RC ZP JBr SK DBa JD. Performed the experiments: DBe KB ES JBe RC JBr. Analyzed the data: DBe KB ES JBe SK RC ZP JBr DBa JD. Wrote the paper: DBe KB ZP JBr JD.

## References

1. Khoury GA, Smadbeck J, Kieslich CA, Floudas CA. (2014) Protein folding and *de novo* protein design for biotechnological applications. *Trends Biotechnol.* 32: 99–109. doi: [10.1016/j.tibtech.2013.10.008](https://doi.org/10.1016/j.tibtech.2013.10.008) PMID: [WOS:000331491100008](https://pubmed.ncbi.nlm.nih.gov/25000000/).
2. Hartmann M, Roeraade J, Stoll D, Templin MF, Joos TO. (2009) Protein microarrays for diagnostic assays. *Anal Bioanal Chem.* 393: 1407–1416. doi: [10.1007/s00216-008-2379-z](https://doi.org/10.1007/s00216-008-2379-z) PMID: [WOS:000263522900005](https://pubmed.ncbi.nlm.nih.gov/19000000/).
3. Schmid A, Hollmann F, Park JB, Buhler B. (2002) The use of enzymes in the chemical industry in Europe. *Curr Opin Biotechnol.* 13: 359–366. doi: [10.1016/s0958-1669\(02\)00336-1](https://doi.org/10.1016/s0958-1669(02)00336-1) PMID: [WOS:000177248000013](https://pubmed.ncbi.nlm.nih.gov/12000000/).

4. Polizzi KM, Bommaris AS, Broering JM, Chaparro-Riggers JF. (2007) Stability of biocatalysts. *Curr Opin Chem Biol.* 11: 220–225. doi: [10.1016/j.cbpa.2007.01.685](https://doi.org/10.1016/j.cbpa.2007.01.685) PMID: [WOS:000246089100017](https://pubmed.ncbi.nlm.nih.gov/1500246089100017/).
5. Ferdjani S, Ionita M, Roy B, Dion M, Djeghaba Z, Rabiller C, et al. (2011) Correlation between thermostability and stability of glycosidases in ionic liquid. *Biotechnol Lett.* 33: 1215–1219. doi: [10.1007/s10529-011-0560-5](https://doi.org/10.1007/s10529-011-0560-5) PMID: [WOS:000291655200020](https://pubmed.ncbi.nlm.nih.gov/1855200020/).
6. Gao D, Narasimhan DL, Macdonald J, Brim R, Ko M-C, Landry DW, et al. (2009) Thermostable variants of cocaine esterase for long-time protection against cocaine toxicity. *Mol Pharmacol.* 75: 318–323. doi: [10.1124/mol.108.049486](https://doi.org/10.1124/mol.108.049486) PMID: [WOS:000262623400007](https://pubmed.ncbi.nlm.nih.gov/18500262623400007/).
7. Wijma HJ, Floor RJ, Janssen DB. (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr Opin Struct Biol.* 23: 588–594. doi: [10.1016/j.sbi.2013.04.008](https://doi.org/10.1016/j.sbi.2013.04.008) PMID: [WOS:000323590500015](https://pubmed.ncbi.nlm.nih.gov/23590500015/).
8. Bommaris AS, Paye MF. (2013) Stabilizing biocatalysts. *Chem Soc Rev.* 42: 6534–6565. doi: [10.1039/c3cs60137d](https://doi.org/10.1039/c3cs60137d) PMID: [WOS:000321570200017](https://pubmed.ncbi.nlm.nih.gov/21570200017/).
9. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci U S A.* 103: 5869–5874. doi: [10.1073/pnas.0510098103](https://doi.org/10.1073/pnas.0510098103) PMID: [WOS:000236896200043](https://pubmed.ncbi.nlm.nih.gov/16236896200043/).
10. Gumulya Y, Reetz MT. (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. doi: [10.1002/cbic.201100412](https://doi.org/10.1002/cbic.201100412) PMID: [WOS:000297160300018](https://pubmed.ncbi.nlm.nih.gov/20297160300018/).
11. Seitz T, Thoma R, Schoch GA, Stihle M, Benz J, D'Arcy B, et al. (2010) Enhancing the stability and solubility of the glucocorticoid receptor ligand-binding domain by high-throughput library screening. *J Mol Biol.* 403: 562–577. doi: [10.1016/j.jmb.2010.08.048](https://doi.org/10.1016/j.jmb.2010.08.048) PMID: [WOS:000284440500006](https://pubmed.ncbi.nlm.nih.gov/20284440500006/).
12. Damborsky J, Brezovsky J. (2014) Computational methods for designing and engineering biocatalysts. *Curr Opin Chem Biol.* 19: 8–16. doi: [10.1016/j.cbpa.2013.12.003](https://doi.org/10.1016/j.cbpa.2013.12.003) PMID: [24780274](https://pubmed.ncbi.nlm.nih.gov/24780274/)
13. Koudelakova T, Chaloupkova R, Brezovsky J, Prokop Z, Sebestova E, Hesseler M, et al. (2013) Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angew Chem Int Ed.* 52: 1959–1963. doi: [10.1002/anie.201206708](https://doi.org/10.1002/anie.201206708) PMID: [WOS:000314654000011](https://pubmed.ncbi.nlm.nih.gov/231465400011/).
14. Bosshart A, Panke S, Bechtold M. (2013) Systematic optimization of interface interactions increases the thermostability of a multimeric enzyme. *Angew Chem Int Ed.* 52: 9673–9676. doi: [10.1002/anie.201304141](https://doi.org/10.1002/anie.201304141) PMID: [WOS:000323829600015](https://pubmed.ncbi.nlm.nih.gov/2323829600015/).
15. Wijma HJ, Floor RJ, Jekel PA, Baker D, Siewert MJ, Janssen DB. (2014) Computationally designed libraries for rapid enzyme stabilization. *Protein Eng Des Sel.* 27: 49–58. doi: [10.1093/protein/gzt061](https://doi.org/10.1093/protein/gzt061) PMID: [24402331](https://pubmed.ncbi.nlm.nih.gov/24402331/)
16. Komor RS, Romero PA, Xie CB, Arnold FH. (2012) Highly thermostable fungal cellobiohydrolase I (Cel7A) engineered using predictive methods. *Protein Eng Des Sel.* 25: 827–833. doi: [10.1093/protein/gzs058](https://doi.org/10.1093/protein/gzs058) PMID: [WOS:000311670800004](https://pubmed.ncbi.nlm.nih.gov/2311670800004/).
17. Reetz MT, Soni P, Acevedo JP, Sanchis J. (2009) Creation of an amino acid network of structurally coupled residues in the directed evolution of a thermostable enzyme. *Angew Chem Int Ed.* 48: 8268–8272. doi: [10.1002/anie.200904209](https://doi.org/10.1002/anie.200904209) PMID: [WOS:000271302900015](https://pubmed.ncbi.nlm.nih.gov/20271302900015/).
18. Kumar MDS, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, et al. (2006) ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res.* 34: D204–D206. doi: [10.1093/nar/gkj103](https://doi.org/10.1093/nar/gkj103) PMID: [WOS:000239307700043](https://pubmed.ncbi.nlm.nih.gov/16239307700043/).
19. Khatun J, Khare SD, Dokholyan NV. (2004) Can contact potentials reliably predict stability of proteins? *J Mol Biol.* 336: 1223–1238. doi: [10.1016/j.jmb.2004.01.002](https://doi.org/10.1016/j.jmb.2004.01.002) PMID: [WOS:000189289300019](https://pubmed.ncbi.nlm.nih.gov/189289300019/).
20. Guerois R, Nielsen JE, 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. doi: [10.1016/s0022-2836\(02\)00442-4](https://doi.org/10.1016/s0022-2836(02)00442-4) PMID: [WOS:000176849700019](https://pubmed.ncbi.nlm.nih.gov/176849700019/).
21. Kellogg EH, Leaver-Fay A, Baker D. (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Protein Struct Funct Bioinformatics.* 79: 830–838. doi: [10.1002/prot.22921](https://doi.org/10.1002/prot.22921) PMID: [WOS:000287784000011](https://pubmed.ncbi.nlm.nih.gov/20287784000011/).
22. Yin S, Ding F, Dokholyan NV. (2007) Eris: an automated estimator of protein stability. *Nature Methods.* 4: 466–467. doi: [10.1038/nmeth0607-466](https://doi.org/10.1038/nmeth0607-466) PMID: [WOS:000246922900004](https://pubmed.ncbi.nlm.nih.gov/18246922900004/).
23. Parthiban V, Gromiha MM, Schomburg D. (2006) CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* 34: W239–W242. doi: [10.1093/nar/gkl190](https://doi.org/10.1093/nar/gkl190) PMID: [WOS:000245650200050](https://pubmed.ncbi.nlm.nih.gov/16245650200050/).
24. Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. (2002) Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics.* 18: S71–77. PMID: [12169533](https://pubmed.ncbi.nlm.nih.gov/12169533/)



25. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, et al. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* 33: W299–W302. doi: [10.1093/nar/gki370](https://doi.org/10.1093/nar/gki370) PMID: [WOS:000230271400059](https://pubmed.ncbi.nlm.nih.gov/16051600/).
26. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 38: W529–W533. doi: [10.1093/nar/gkq399](https://doi.org/10.1093/nar/gkq399) PMID: [WOS:000284148900085](https://pubmed.ncbi.nlm.nih.gov/20513125/).
27. Lehmann M, Pasamontes L, Lassen SF, Wyss M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim Biophys Acta, Protein Structure M.* 1543: 408–415. doi: [10.1016/S0167-4838\(00\)00238-7](https://doi.org/10.1016/S0167-4838(00)00238-7) PMID: [WOS:000166500900011](https://pubmed.ncbi.nlm.nih.gov/10811111/).
28. Gray KA, Richardson TH, Kretz K, Short JM, Bartnek F, Knowles R, et al. (2001) Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv Synth Catal.* 343: 607–616.
29. Kuipers RKP, Joosten H-J, Verwiël E, Paans S, Akerboom J, van der Oost J, et al. (2009) Correlated mutation analyses on super-family alignments reveal functionally important residues. *Protein Struct Funct Bioinformatics.* 76: 608–616. doi: [10.1002/prot.22374](https://doi.org/10.1002/prot.22374) PMID: [WOS:000267963900008](https://pubmed.ncbi.nlm.nih.gov/19080008/).
30. Koudelakova T, Bidmanova S, Dvorak P, Pavelka A, Chaloupkova R, Prokop Z, et al. (2013) Haloalkane dehalogenases: Biotechnological applications. *Biotechnol J.* 8: 32–45. doi: [10.1002/biot.201100486](https://doi.org/10.1002/biot.201100486) PMID: [WOS:000312989500012](https://pubmed.ncbi.nlm.nih.gov/22811112/).
31. Borgo B, Havranek JJ. (2012) Automated selection of stabilizing mutations in designed and natural proteins. *Proc Natl Acad Sci U S A.* 109: 1494–1499. doi: [10.1073/pnas.1115172109](https://doi.org/10.1073/pnas.1115172109) PMID: [WOS:000299731400037](https://pubmed.ncbi.nlm.nih.gov/22811112/).
32. Diaz JE, Lin C-S, Kunishiro K, Feld BK, Avrantinis SK, Bronson J, et al. (2011) Computational design and selections for an engineered, thermostable terpene synthase. *Protein Sci.* 20: 1597–1606. doi: [10.1002/pro.691](https://doi.org/10.1002/pro.691) PMID: [WOS:000294383900013](https://pubmed.ncbi.nlm.nih.gov/21811113/).
33. Floor RJ, Wijma HJ, Colpa DI, Ramos-Silva A, Jekel PA, Szymanski W, et al. (2014) Computational library design for increasing haloalkane dehalogenase stability. *ChemBioChem.* 15: 1659–1671. doi: [10.1002/cbic.201402128](https://doi.org/10.1002/cbic.201402128) PMID: [WOS:000339502900017](https://pubmed.ncbi.nlm.nih.gov/24811117/).
34. Pavelka A, Chovancova E, Damborsky J. (2009) HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* 37: W376–W383. doi: [10.1093/nar/gkp410](https://doi.org/10.1093/nar/gkp410) PMID: [ISI:000267889100066](https://pubmed.ncbi.nlm.nih.gov/19080066/).
35. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
36. Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvermin V, et al. (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 37: D5–D15. doi: [10.1093/nar/gkn741](https://doi.org/10.1093/nar/gkn741) PMID: [WOS:000261906200002](https://pubmed.ncbi.nlm.nih.gov/19080002/).
37. Li WZ, Godzik A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics.* 22: 1658–1659. doi: [10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158) PMID: [WOS:000238905700017](https://pubmed.ncbi.nlm.nih.gov/16080017/).
38. Frickey T, Lupas A. (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics.* 20: 3702–3704. PMID: [15284097](https://pubmed.ncbi.nlm.nih.gov/15284097/).
39. Edgar RC. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 5: 1–19. doi: [10.1186/1471-2105-5-113](https://doi.org/10.1186/1471-2105-5-113) PMID: [WOS:000223920500001](https://pubmed.ncbi.nlm.nih.gov/150800001/).
40. Mika S, Rost B. (2003) UniqueProt: creating representative protein sequence sets. *Nucleic Acids Res.* 31: 3789–3791. doi: [10.1093/nar/gkg620](https://doi.org/10.1093/nar/gkg620) PMID: [WOS:000183832900111](https://pubmed.ncbi.nlm.nih.gov/12811111/).
41. Whelan S, Goldman N. (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 18: 691–699. PMID: [WOS:000168427400001](https://pubmed.ncbi.nlm.nih.gov/1168427400001/).
42. Murzin AG, Brenner SE, Hubbard T, Chothia C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol.* 247: 536–540. doi: [10.1016/S0022-2836\(05\)80134-2](https://doi.org/10.1016/S0022-2836(05)80134-2) PMID: [WOS:A1995QQ85600003](https://pubmed.ncbi.nlm.nih.gov/1199599500003/).
43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. (2000) The Protein Data Bank. *Nucleic Acids Res.* 28: 235–242. PMID: [10592235](https://pubmed.ncbi.nlm.nih.gov/10592235/)
44. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC.
45. Costantini S, Colonna G, Facchiano AM. (2008) ESBRI: A web server for evaluating salt bridges in proteins. *Bioinformation.* 3: 137–138. PMID: [19238252](https://pubmed.ncbi.nlm.nih.gov/19238252/)
46. Tina KG, Bhadra R, Srinivasan N. (2007) PIC: Protein Interactions Calculator. *Nucleic Acids Res.* 35: W473–W476. doi: [10.1093/nar/gkm423](https://doi.org/10.1093/nar/gkm423) PMID: [WOS:000255311500089](https://pubmed.ncbi.nlm.nih.gov/17500255311500089/).

47. Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, et al. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat Chem Biol.* 5: 727–733. doi: [10.1038/nchembio.205](https://doi.org/10.1038/nchembio.205) PMID: [ISI:000270039900010](https://pubmed.ncbi.nlm.nih.gov/18811111/)
48. Okai M, Kubota K, Fukuda M, Nagata Y, Nagata K, Tanokura M. (2009) Crystallization and preliminary X-ray analysis of gamma-hexachlorocyclohexane dehydrochlorinase LinA from *Sphingobium japonicum* UT26. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications.* 65: 822–824. doi: [10.1107/s1744309109026645](https://doi.org/10.1107/s1744309109026645) PMID: [ISI:000268517300020](https://pubmed.ncbi.nlm.nih.gov/19311111/).
49. Stepankova V, Damborsky J, Chaloupkova R. (2013) Organic co-solvents affect activity, stability and enantioselectivity of haloalkane dehalogenases. *Biotechnol J.* 8: 719–729. doi: [10.1002/biot.201200378](https://doi.org/10.1002/biot.201200378) PMID: [WOS:000320031900013](https://pubmed.ncbi.nlm.nih.gov/23111111/).
50. Iwasaki I, Utsumi S, Ozawa T. (1952) New colorimetric determination of chloride using mercuric thiocyanate and ferric ion. *Bull Chem Soc Jpn.* 25: 226.
51. Kelly SM, Jess TJ, Price NC. (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta.* 1751: 119–139. doi: [10.1016/j.bbapap.2005.06.005](https://doi.org/10.1016/j.bbapap.2005.06.005) PMID: [WOS:000231428500001](https://pubmed.ncbi.nlm.nih.gov/16111111/).
52. Ladbroke BD, Williams RM, Chapman D. (1968) Studies on lecithin-cholesterol-water interactions by differential scanning calorimetry and X-ray diffraction. *Biochim Biophys Acta.* 150: 333–340. doi: [10.1016/0005-2736\(68\)90132-6](https://doi.org/10.1016/0005-2736(68)90132-6) PMID: [WOS:A1968B096400001](https://pubmed.ncbi.nlm.nih.gov/16611111/).