RESEARCH ARTICLE

Application of Rigidity Theory to the Thermostabilization of Lipase A from *Bacillus subtilis*

Prakash Chandra Rathi[1], Alexander Fulton[2], Karl-Erich Jaeger[2,3]*, Holger Gohlke[1]*

1 Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Düsseldorf, Germany, 2 Institute of Molecular Enzyme Technology, Heinrich-Heine-University, Düsseldorf, Germany, 3 Institute of Bio- and Geosciences IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich, Germany

* karl-erich.jaeger@fz-juelich.de (KEJ); gohlke@uni-duesseldorf.de (HG);

Abstract

Protein thermostability is a crucial factor for biotechnological enzyme applications. Protein engineering studies aimed at improving thermostability have successfully applied both directed evolution and rational design. However, for rational approaches, the major challenge remains the prediction of mutation sites and optimal amino acid substitutions. Recently, we showed that such mutation sites can be identified as structural weak spots by rigidity theory-based thermal unfolding simulations of proteins. Here, we describe and validate a unique, ensemble-based, yet highly efficient strategy to predict optimal amino acid substitutions at structural weak spots for improving a protein's thermostability. For this, we exploit the fact that in the majority of cases an increased structural rigidity of the folded state has been found as the cause for thermostability. When applied prospectively to lipase A from *Bacillus subtilis*, we achieved both a high success rate (25% over all experimentally tested mutations, which raises to 60% if small-to-large residue mutations and mutations in the active site are excluded) in predicting significantly thermostabilized lipase variants and a remarkably large increase in those variants' thermostability (up to 6.6°C) based on single amino acid mutations. When considering negative controls in addition and evaluating the performance of our approach as a binary classifier, the accuracy is 63% and increases to 83% if small-to-large residue mutations and mutations in the active site are excluded. The gain in precision (predictive value for increased thermostability) over random classification is 1.6-fold (2.4-fold). Furthermore, an increase in thermostability predicted by our approach significantly points to increased experimental thermostability ($p < 0.05$). These results suggest that our strategy is a valuable complement to existing methods for rational protein design aimed at improving thermostability.

Author Summary

Protein thermostability is a crucial factor for biotechnological enzyme applications. However, performance studies of computational approaches for predicting effects of mutations on protein (thermo)stability have suggested that there is still room for improvement. We describe and validate a novel and unique strategy to predict optimal amino acid substitutions at structural weak spots. At variance with other rational approaches, we exploit the fact that in the majority of cases an increased structural rigidity of the folded state is the underlying cause for thermostability. When applied prospectively on lipase LipA from *Bacillus subtilis*, a high success rate in predicting thermostabilized lipase variants and a remarkably large increase in their thermostability is achieved. This demonstrates the value of the novel strategy, which extends the existing portfolio of methods for rational protein design.

Introduction

Thermostability is a crucial factor for a wealth of biotechnological enzyme applications [1,2]. Protein engineering aimed at improving thermostability is thus an important field of research in biotechnology [3,4]. There, methods of directed evolution are usually applied, which mimic natural evolution [5–8]. However, directed evolution is limited in that out of the extraordinarily large number of possible variant proteins, only a small subset can be experimentally tested [9]. Alternatively, rational approaches have been successfully pursued [10–13] but the major challenge here remains the prediction of mutation sites and the optimal amino acid substitution at such sites [14,15].

As to the prediction of mutation sites, we developed the rigidity theory-based Constraint Network Analysis (CNA) approach [16–21] (available as a web service at http://cpclab.uni-duesseldorf.de/cna/ [16–21]), which identifies residues in a protein that are structural "weak spots". For this, a protein is modeled as a network of sites (atoms) and constraints (covalent and noncovalent interactions) [22]. Rigid atom clusters and flexible regions in between are then rigorously determined by rigidity analysis [23–25]. By successively removing non-covalent constraints from the network, the thermal unfolding of the protein is simulated (Fig 1a and 1b) [16,18,19,26]. From the unfolding trajectory, a phase transition temperature T_p is identified, which relates to the (thermodynamic) thermostability, as are the weak spots (Fig 1c). Mutating such weak spots should likely improve a protein's thermostability [16,18,19].

Here, we describe and validate a novel and unique strategy based on the CNA approach to predict optimal amino acid substitutions at these weak spots. At variance with other rational approaches that rely upon calculating free energies for predicting effects of mutations on a protein's thermostability [27–33], we exploit the fact that in the majority of cases an increased structural rigidity of the folded state has been identified as the underlying cause for thermostability [34]. To this end, we add a highly efficient, ensemble-based second step by generating structural models of single-point site-saturation mutations at identified weak spots, filtering the models with respect to their structural quality, and screening for variants with increased structural rigidity (Fig 1d–1f, see below for detailed descriptions). Using the recently developed ENT^{FNC} approach [35] that performs rigidity analyses on an *ensemble* of *network topologies* generated from a single input structure using *fuzzy network constraints*, rather than a structural ensemble, this second step only takes about 1 h on a single core per variant and can be performed in parallel for multiple variants. We applied this strategy prospectively on lipase LipA from *Bacillus subtilis* (*Bs*LipA); *Bs*LipA has considerable biotechnological importance [36,37] and has been extensively studied with respect to thermostability [6,15,38–43], which
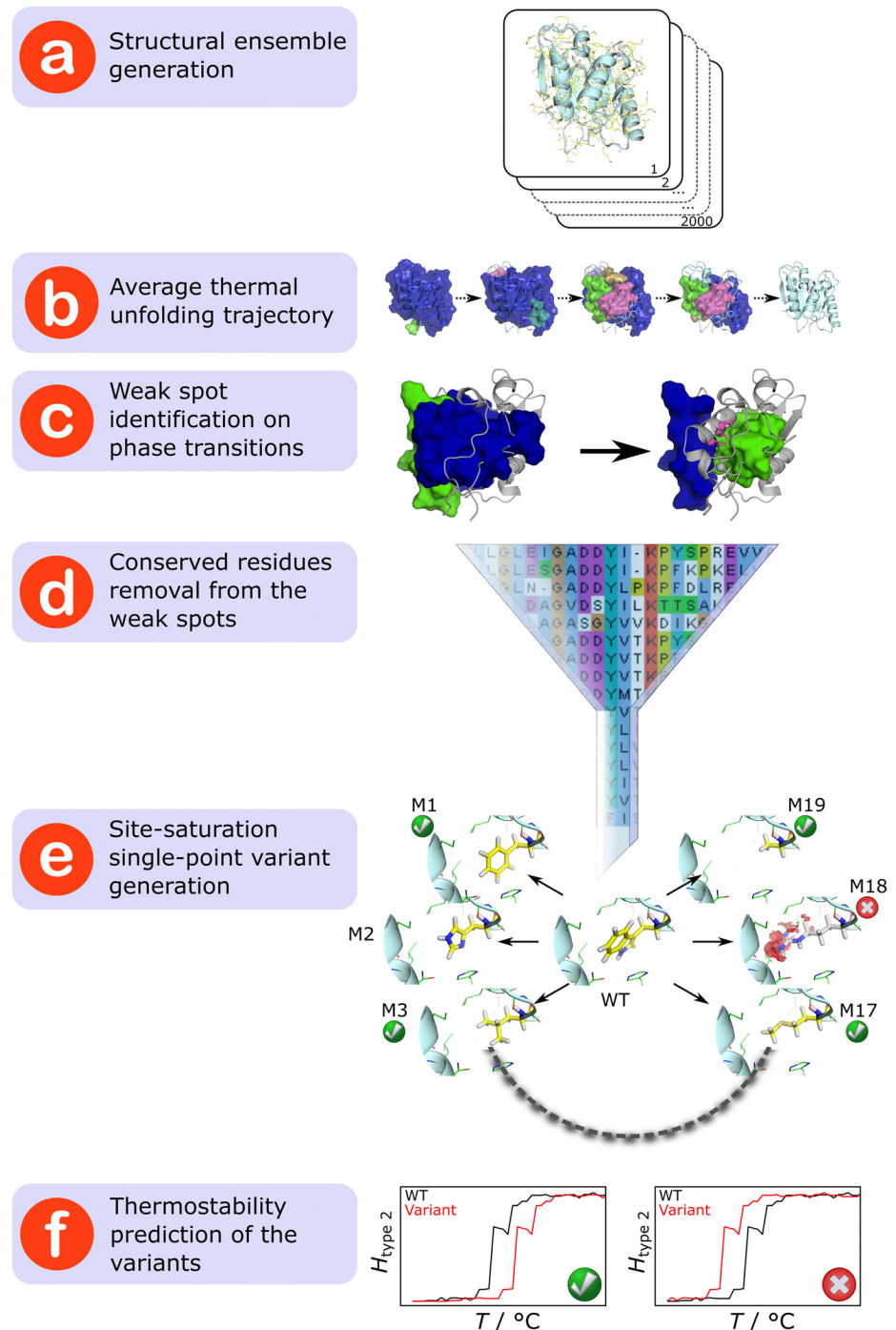
Fig 1. Strategy to rationally predict mutations that increase structural rigidity and thermostability.
(a) A structural ensemble of the respective protein is generated by MD simulations. (b) The average thermal unfolding trajectory depicting a decomposition into rigid clusters (in the order of decreasing size colored in blue, green, magenta, cyan, orange, and violet) for each step of the unfolding simulation is created by subjecting the structural ensemble to CNA. (c) For every major transition during the thermal unfolding, weak spot residues (depicted as a sphere for the C_α atom and sticks for the side-chain) are identified. (d) Weak spot residues that are highly conserved across the protein family ($\geq$ 80% identity) are removed from the weak spot list. (e) For the rest, structures of site-saturation single-point variants (termed M1–M19) are generated. Mutations that lead to energetically unfavorable structures (indicated by red discs around the mutated residue in the case of M18) are discarded. (f) For each variant, the phase transition temperature T_p is computed using CNA; a higher T_p value than that of the WT protein indicates a thermostabilizing mutation.

doi:10.1371/journal.pcbi.1004754.g001

makes *Bs*LipA a prominent model system. Out of 589 *Bs*LipA variants screened *in silico*, twelve were suggested for experimental testing. Of these, three showed a significant increase of up to 6.6°C in thermostability with respect to the wild-type enzyme (WT). We thus achieved both a high success rate in predicting thermostabilized lipase variants and a remarkably large increase in the thermostability of such variants. This demonstrates the value of the novel strategy, which extends the existing portfolio of methods for rational protein design aimed at improving thermostability.

Results

Predicting thermostabilizing mutations

*Bs*LipA has a minimal α/β hydrolase fold in which a central parallel β-sheet of six β-strands is surrounded by six α-helices [44]. For identifying weak spots on *Bs*LipA, a thermal unfolding simulation was carried out by CNA on an ensemble of 2000 WT *Bs*LipA structures extracted from a molecular dynamics (MD) trajectory of 100 ns length (Fig 1a). The ensemble-based CNA was pursued to increase the robustness of the rigidity analyses [19,35,45]. The unfolding trajectory (Figs 1b and 2) reveals the early segregation of loops from the largest rigid cluster, followed by the segregation of α-helices and, finally, the segregation and disintegration of the β-sheet region. This order of segregation is in agreement with experimental findings on the unfolding of other α/β hydrolase proteins [46,47]. The realistic description of WT *Bs*LipA thermal unfolding encouraged us to identify weak spots at major phase transitions along the unfolding trajectory (Fig 1c). By visual inspection of the unfolding trajectory, we identified five major transitions (T1–T5) at which helices αA, αF, αD and αE, αB, αC as well as the central beta sheet segregate from the largest rigid cluster at temperatures 316, 318, 334, 336, and 338 K, respectively (Table 1 and Fig 2).

Weak spot residues were then identified as those residues that are in the neighborhood of the largest rigid cluster from which they segregate at the respective major transition. These residues are particularly promising for increasing *Bs*LipA's thermostability considering that their mutation can improve the interaction strength with the largest rigid cluster and, hence, delay the disintegration of that cluster with increasing temperature. In total, 36 weak spots were found, which are located on α-helices and loops joining α-helices and β-strands (Fig 2). The weak spot residues are very diverse in size (ranging from Gly to Trp) and physicochemical properties (charged, uncharged polar, and hydrophobic) (Table 1). Of these, weak spot residues at highly conserved sequence positions were discarded (Figs 1d and S1; Table 1) because conserved residues are usually important for function and/or stability of a protein and, hence, should not be mutated [48,49].

For each of the remaining 31 weak spots (~17% of all *Bs*LipA residues), computational site saturation mutagenesis was performed by generating structures of all possible single-point amino acid substitutions using the SCWRL program (Fig 1e) [50]. SCWRL constructs variant models by predicting backbone-dependent side-chain conformations with the help of a rotamer library. This resulted in 589 single point variants. 67 variant structures were discarded based on the evaluation of residue-wise non-local interaction energies by the ANOLEA server (S1 Fig) [51,52]. In such structures, the mutation apparently does not fit into the environment of the other residues.

The remaining 522 variants were subjected to thermal unfolding simulations on ensembles of network topologies using the ENT^{FNC} approach [35] implemented in CNA. Differences in the phase transition temperatures $\Delta T_p = T_p$ (variant) $- T_p$ (WT) were averaged over 1000 simulations started from different network topologies generated for each variant (see "Materials and Methods section"; Fig 1f). A map of ΔT_p values of all variants is shown in S1 Fig. In total,
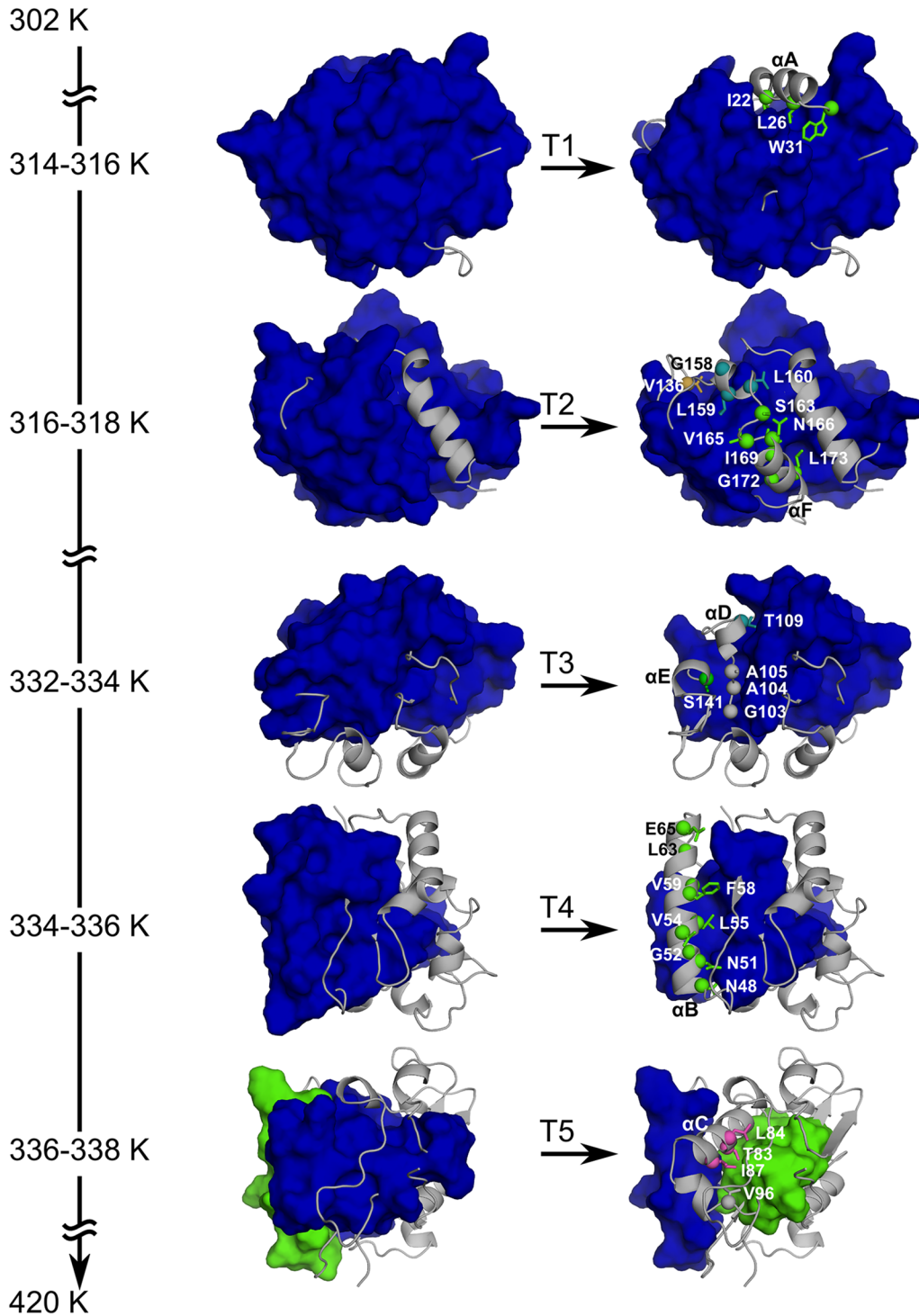
Fig 2. Thermal unfolding trajectory of WT *Bs*LipA showing transitions for which weak spot residues were identified. Uniformly colored bodies represent rigid clusters; for clarity, only the largest rigid cluster (blue) is shown for the first four transitions (T1–T4), and the two largest rigid clusters (blue and green) are shown for the last transition (T5). C_α atoms of the identified weak spot residues are shown as spheres, and side-chain atoms are shown in stick representation. Weak spot residues are colored according to the rigid cluster they are part of (rigid clusters are assigned blue, green, magenta, cyan, and orange colors in the descending order of their size in terms of the number of residues); a weak spot residue colored in gray indicates that it is part of a rigid cluster composed of less than three residues. Weak spot residues that are highly conserved in the multiple sequence alignment of the lipase family are not shown. Important helices that segregate from the largest rigid cluster at the respective transition are labeled.

doi:10.1371/journal.pcbi.1004754.g002

Table 1. Phase transition points at which weak spot residues are identified during the thermal unfolding simulation of *Bs*LipA.

Phase transition	Temperature of the phase transition[a]	Major secondary structures segregating from the giant rigid cluster	Weak spot residues[b]
T1	314–316	αA	I22, L26, W31
T2	316–318	αF	**D133**, V136, G158, L159, L160, S163, V165, N166, I169, G172, L173
T3	332–334	αD and αE	G103, G104, A105, **N106**, T109, S141
T4	334–336	αB	N48, N51, G52, V54, L55, F58, V59, **V62**, L63, E65, **T66**, **V71**
T5	336–338	αC and central β sheet	T83, L84, I87, V96

[a] In K.

[b] Residues in bold are highly conserved in the multiple sequence alignment; see the main text for details.

doi:10.1371/journal.pcbi.1004754.t001

this procedure yielded a predicted thermostabilization with respect to WT *Bs*LipA for 75 out of the 522 mutations (~14%) investigated. In order to further reduce the number of mutations for experimental validation only the mutation with the highest ΔT_p was chosen from all mutations with $\Delta T_p > 1$ K at a weak spot. The sole exception is G104 located in the active site, for which two mutations were chosen. This resulted in twelve lipase variants of which the most are associated with weak spot residues on helix αB identified during the late transition T4 (Table 2; S1 Fig).

As a negative control, we also predicted 10 variants with negative ΔT_p, i.e., where a mutation according to the thermal unfolding simulations leads to a decrease in thermostability with respect to WT (S1 Table). Six of these mutations were chosen from the above analyses of 522 variants such that they have the most negative ΔT_p; four were chosen with the most negative ΔT_p from analyses of variants with a mutation not at a weak spot.

Table 2. *Bs*LipA variants with positive ΔT_p characterized experimentally.

*Bs*LipA variant[a]	Location of the mutation on secondary structure element	Phase transition of weak spot identification	Predicted ΔT_p[b]	T'_{50}[c]
Wild-type	-	-	-	49.10
I22W	αA	T1	2.80	44.89
N51F	αB	T4	4.30	46.05
G52M	αB	T4	16.47	49.59
V54H	αB	T4	2.09	54.80
L55F	αB	T4	3.48	47.62
F58I	αB	T4	2.27	55.65
V59F	αB	T4	11.95	49.44
I87W	αC	T5	4.91	-[d]
V96S	β6	T5	2.36	52.65
G104I	Loop β6- αD	T3	1.98	-[d]
G104L	Loop β6- αD	T3	5.07	-[d]
L160H	αF	T2	2.25	43.30

[a] Variants highlighted in bold show a significant increase in T'_{50} compared to WT.

[b] Difference phase transition temperatures T_p (variant) – T_p (WT); in °C.

[c] The temperature at which the fraction of the activity to the initial activity (at 40°C) is 50% after incubating for 30 min; in °C.

[d] No activity after 30 min incubation at temperatures of 40–60°C.

doi:10.1371/journal.pcbi.1004754.t002

Experimental characterization of thermostability

Initially, specific activities of WT BsLipA and the twelve variants (Table 2) for hydrolysis of p-nitrophenyl-palmitate (pNPP) were measured at temperatures between 40 and 60°C after keeping them at the respective temperatures for 5 min. WT BsLipA showed the highest specific activity (246 U/mg) among all BsLipA variants at the temperature of maximum activity T_{max} (40°C) (S2 Fig). At temperatures above 55°C, the activity begins to drop, which is probably due to an unfolding already within 5 min of preincubation. However, two variants, F58I and V96S, showed higher activities than the WT at temperatures above 58°C (S2 Fig), which may originate from them being more stable at high temperatures.

Next, thermostability was assessed by measuring the activity of each BsLipA variant at temperatures between 40 and 60°C after incubating the respective variant at these temperatures for 30 min. Three variants, V54H, F58I, and V96S, were more thermostable than WT; they consistently showed higher activities than the WT at temperatures above 48°C (Figs 3a and S3). The largest differences between thermostabilities of WT and variants of BsLipA was observed at 53.5°C where the activities of V54H and V96S were twice as a high as that of the WT, and the activity of F58I was four times higher (Fig 3). The kinetic constants of these variants were derived from initial rate measurements for hydrolysis of p-nitrophenyl-decanoate (pNPD) at 40°C (see S1 Text). No significant impact on the Michaelis constant (K_M) was observed, and the turnover numbers (k_{cat}) were reduced by at most 25% (Table 3). Thus, the thermostability of the variants has been increased without significantly influencing k_{cat} / K_M at 40°C. Still, two of the three thermostable variants showed lower activities than WT at temperatures below ~45°C (Fig 3). This may have been caused by a rigidification of the lipase structure in the thermostable variants (see section "Analysis of thermostability changes at the structural level" below), which may also influence the flexibility of the active site. Similarly, in a series of five orthologs of 2-deoxy-D-ribose-5-phosphate aldolase (DERA) from psychrophilic, mesophilic, and hyperthermophilic organisms investigated by us recently in terms of biochemical, structural, and rigidity properties, an anticorrelation between specific activity at temperatures ≤ 40°C and experimental or computed melting temperature was observed [53]. In that study, both the analysis of local rigidity by CNA and B-factor analysis of X-ray structures provided independent clues that psychrophilic DERAs have a more flexible environment of the substrate binding pocket. Thus, it may depend on the actual operating temperature of an enzymatic process whether it is worth to apply thermostable variants with increased activities at high(er) temperatures only.

Finally, the thermostability of BsLipA variants was quantified by T'_{50} values; these values report on the temperature at which the fraction of the activity to the initial activity (at 40°C) is 50% after incubation for 30 min. This is different from the T_{50} values normally used for characterizing the thermostability of proteins [15,54,55] in that the activity here is measured at the temperature of incubation, not at room temperature after cooling. T'_{50} thus reports on the thermo-tolerance of an enzyme during operational bioprocesses carried out at elevated temperatures for a longer duration of time, e.g., as done in the lipid processing industry [56]. The three variants V54H, F58I, and V96S showed T'_{50} values higher by 5.7, 6.6, and 3.6°C, respectively, than WT BsLipA (Fig 3c; Table 2). The predicted ΔT_p values for these variants were similar to each other, in agreement with the similar T'_{50} values found, but at the lower end of all predicted ΔT_p (Table 2).

For the variants used as a negative control (S1 Table) [57], the thermostability was quantified by T'_{50} values; these values report on the temperature of incubation for 20 min after which the fraction of the activity at room temperature to the initial activity is 50%. With respect to the T'_{50} values used above, a significant and very good correlation was obtained for T'_{50} (see S1 Text)

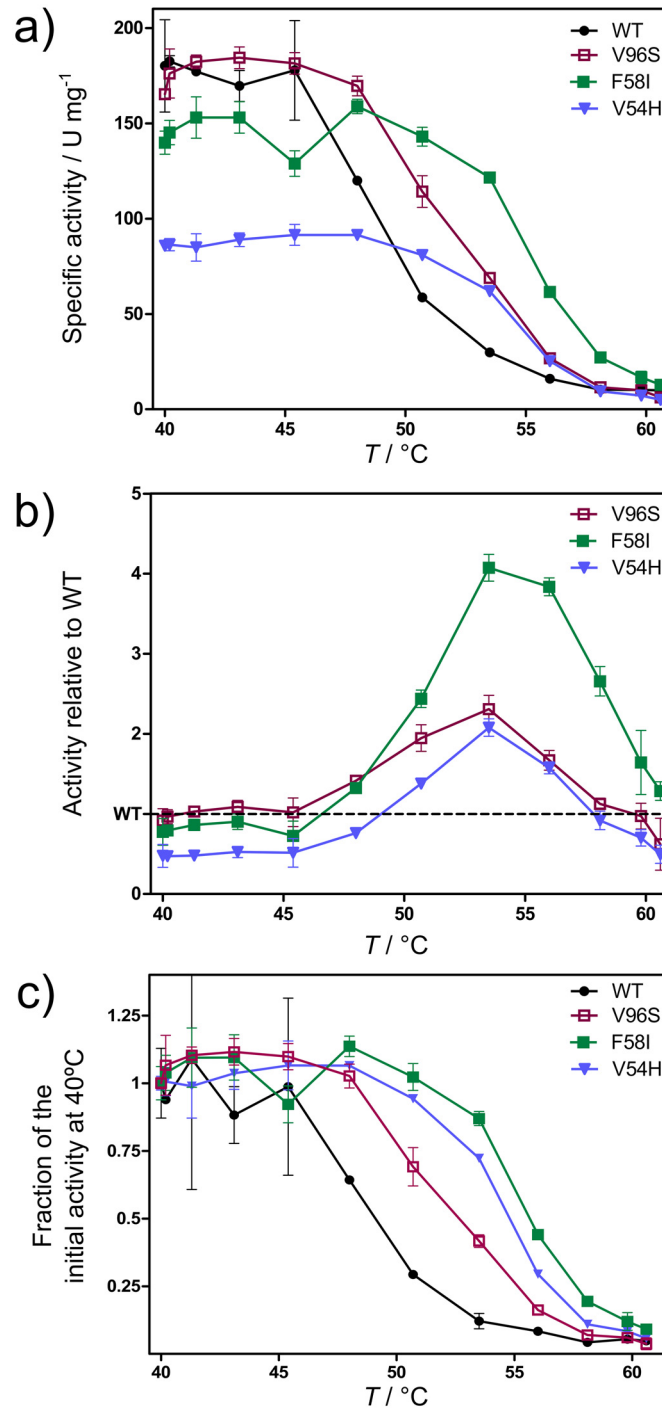**Fig 3. Thermostability of WT (black) and variants V54H (blue), F58I (green), and V96S (red) shown as activity *vs.* temperature curves.** The activity was measured at indicated temperatures after incubating for 30 min at these temperatures. Curves show absolute specific activity (a), activity normalized by the activity of WT (b), and the fraction of the activity to the initial activity at 40°C (c).

Table 3. Kinetic parameters of *Bs*LipA variants and wild-type.

Variant	K_M[a] (μM)	k_{cat} (s^{-1})	k_{cat} / K_M ($μM^{-1} * s^{-1}$)
Wild-type	34.72 ± 6.49	926.40 ± 38.81	26.68 ± 6.10
V54H	40.02 ± 8.64	784.60 ± 38.97	19.51 ± 5.18
F58I	36.71 ± 7.83	690.50 ± 33.35	18.80 ± 4.91
V96S	32.30 ± 7.39	785.00 ± 39.73	24.30 ± 6.79

[a] Kinetic parameters were derived from experiments conducted at 40°C using *p*NPD as substrate.

doi:10.1371/journal.pcbi.1004754.t003

For nine out of ten variants, significantly lower thermostabilities were measured, with the largest decrease being 7.3°C for the N48R variant (S1 Table).

Analysis of thermostability changes at the structural level

The three thermostable variants involve mutations at weak spots identified at later phase transitions T4 and T5 during the thermal unfolding simulation. This finding supports our previous reasoning that it is the late phase transition(s) involving the final decay of the rigid core during thermal unfolding that mostly determine(s) the thermodynamic thermostability of a protein [16,18,19]. Accordingly, mutations that strengthen contacts of weak spot residues identified at late phase transitions should particularly improve thermostability. A sound discussion of this implication requires X-ray structural data of the variants, which is not yet available. Still, using the modeled variant structures, we observed that the three variants V54H, F58I, and V96S do have in general stronger "rigid contacts" between neighboring residues than the WT (a "rigid contact" denotes that two residues belong to one rigid cluster): On average, the mutations V54H, F58I, and V96S increased the strength of rigid contacts of neighboring residues by 2.0, 1.2, and 0.4 K, respectively, compared to WT (S4 Fig; see section "Constraint Network Analysis: Local rigidity indices" for an explanation how these values were calculated).

Considering the most thermostable variant F58I in more detail, the strengthening holds true for local contacts as well as contacts that arise from a long-range stabilization. As to local contacts, Ile at position 58 along with residues of the neighboring loop β4-αB (A38, V39, D40) are part of a rigid cluster, which persists to a temperature ~3 K higher than the rigid cluster formed by F58 of WT and the same loop residues (Figs 4a, 4b, S4b and S6a). The persistence at higher temperature results from a better side-chain packing (Fig 4c). In particular, in variant F58I, V39 forms four hydrophobic contacts with three different residues (V7, S16, F41), whereas in WT it only forms two such hydrophobic contacts (Fig 4c). However, not all F58I mutation-induced changes lead to stabilization (Fig 4d). As to contacts that arise from a long-range stabilization, residues of several pairs of secondary structure (αA/β strands 3,4,5; αB/αC; loop αB-β5/loop αC-β6; loop αC-β6/loop αD-β7) remain part of one rigid cluster for temperatures 2–5 K higher in the variant F58I than in WT (Figs 4d, S4b and S5b–S5e). This demonstrates the inherent long-range aspect to rigidity percolation [23,45,58–60], i.e., a local change on one end of a network can affect the stability all across the network.

Recently, we described the unfolding pathway of *Bs*LipA in detail as deduced from thermal unfolding simulations [61] (see also page 9, Fig 3 in that publication). We observed that α-helices αD and αE first segregate to form individual small rigid clusters, followed by αA and αF. The giant rigid cluster at this temperature is formed by the central β-sheet region and the two helices αB and αC. Next, the β-sheet region becomes sequentially flexible, beginning with β4 and β8, followed by the remaining β-strands in the order β3, β7, and β5–β6, finally leading to a completely flexible β-sheet region. As described above, several of these secondary structural
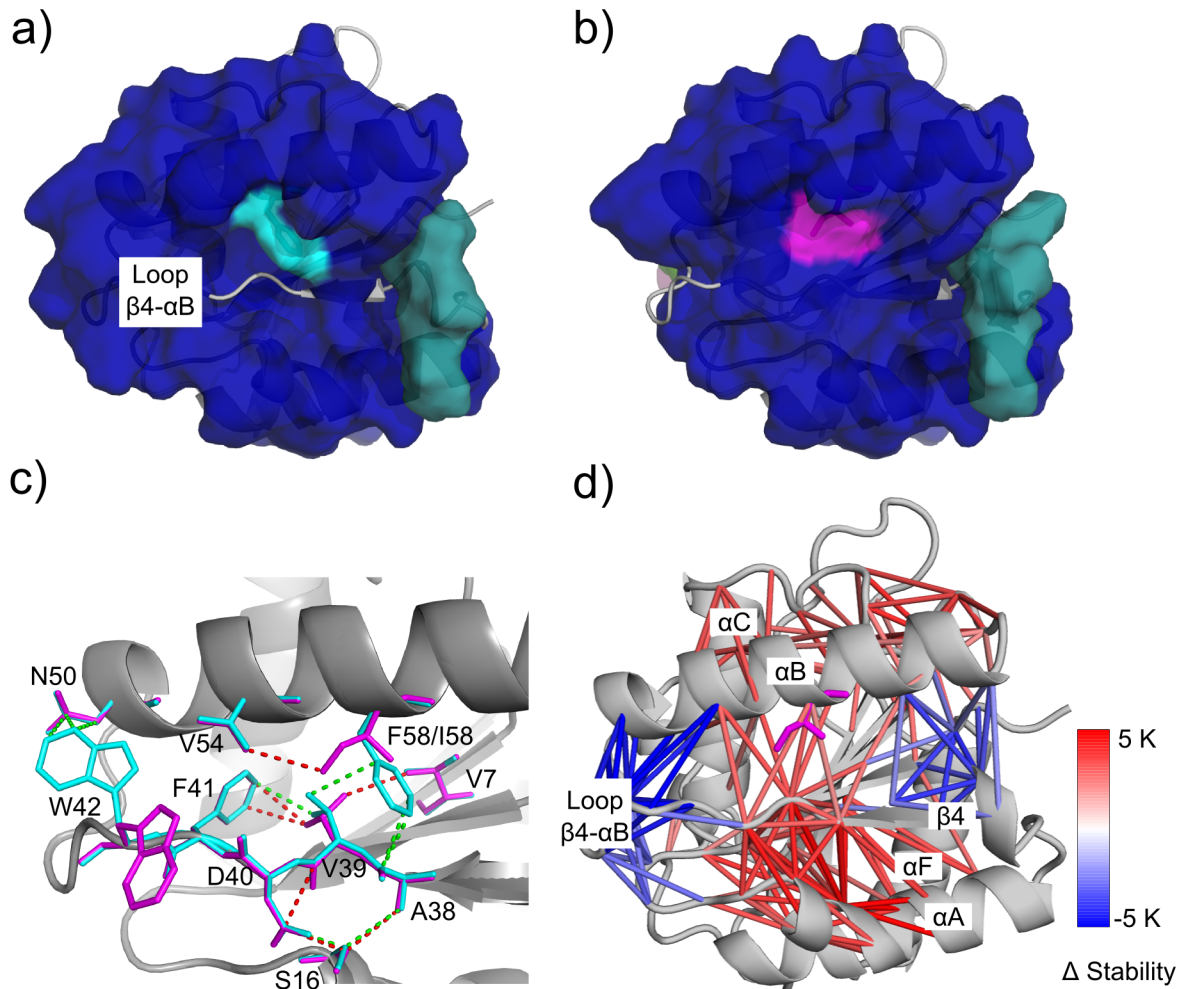
Fig 4. Structural origin of differences in the thermostability of WT and F58I, shown by a rigid cluster decomposition of WT (a) and F58I (b) at 316 K.
Rigid clusters are shown as uniformly colored bodies. The mutation site (residue 58), shown by a cyan (a) and a magenta (b) surface, is part of the largest rigid cluster (blue) in both WT and F58I. Hydrophobic contacts in the proximity of the mutation site between carbon atom pairs at most 3.8 Å apart are shown as green (WT) and red (F58I) dashed lines (c). Residues involved in making such contacts are shown as cyan (WT) and magenta (F58I) sticks. Differences in the stability of "rigid contacts" between variant F58I and WT are depicted on the variant structure (d). Two residues form a "rigid contact" if they belong to one rigid cluster. A red (blue) stick connecting C_α atoms of two residues indicates that a rigid contact in the variant is more (less) stable than in the WT (see color scale). Only those contacts of variant F58I that are stabilized or destabilized by $\geq$ 2 K are shown for clarity; for the same reason, contacts between two residues of the same secondary structure element are not shown. The mutated residue I58 is displayed by magenta sticks. Blow-ups of panel d showing the contact stability between secondary structure pairs mentioned in the main text can be found in S5 Fig.

doi:10.1371/journal.pcbi.1004754.g004

elements are involved in the thermal stabilization of the variant F58I (S4 Fig). Furthermore, the amino acids forming the catalytic triad in *Bs*LipA are S77 located between strand β5 and helix αC, D133 between strand β7 and helix αE, and H156 between strand β8 and helix αF [62]. Stabilization of these secondary structural elements due to introducing mutation F58I (S4 Fig; in particular, loops αB-β5 and αC-β6 (S5d Fig), loops αC-β6 and αD-β7 (S5e Fig), and helices αB and αC (S5c Fig)) may thus delay the unfolding of the active site.

Five mutations at weak spots identified at transitions T4 and T5 resulted in lower T'_{50} values than that of WT *Bs*LipA (Table 2). This result appears to contradict our reasoning that mutations which strengthen connections of weak spot residues identified at late phase transitions should particularly improve thermostability. In each case, however, a small amino acid was substituted by a large amino acid, which likely could not be accommodated by the fold. This

calls for improved modeling considering backbone relaxation [63] for variant construction in future studies with the aim to improve discrimination between amino acid substitutions in already densely packed regions, which could not accommodate small-to-large residue mutations, and substitutions in the vicinity of a protein cavity, where small-to-large residue mutations are an established strategy to increase protein stability [39,64]. Along the same lines, the two variants G104I and G104L out of the three variants that showed a nearly complete loss of activity at room temperature, and no residual activity after 30 min incubation at temperatures between 40–60°C, involved a residue located in the active site. While at the opposite side of the catalytic triad, introducing larger residues may occlude the substrate binding region. Such weak spots can be filtered out in future studies based on their location in the protein [65].

Discussion

We developed a novel rational approach based on increasing structural rigidity for improving a protein's thermostability and applied it prospectively to *Bs*LipA. The approach combines ensemble- and rigidity theory-based weak spot prediction by CNA, filtering of weak spots according to sequence conservation, computational site saturation mutagenesis, assessment of variant structures with respect to their structural quality, and screening of the variants for increased structural rigidity by ensemble-based CNA. Two reasons account for its high computational efficiency: In the first step, the number of potential mutation sites is dramatically reduced due to concentrating only on structural weak spots. In the second step, the use of ensembles of network topologies, rather than structural ensembles, alleviates the need for costly conformation sampling. As a result, about one mutation per hour can be processed on one core once weak spots have been detected (Table 4); this task is trivially parallelizable for multiple mutations. From a methodological point of view, this majorly distinguishes our approach from other state-of-the art methods for predicting effects of mutations on protein stability [27–33] in that these methods would need to consider all potential mutation sites due to the lack of an equivalent "step one". Furthermore, these methods either do not consider ensemble

Table 4. Computing times for weak spot identification, site saturation mutagenesis, and screening for increased structural rigidity.

Step[a]	Time required[b]	Comment
a) MD simulation	~78 h	100 ns long MD simulation on a single GPU
b) Thermal unfolding simulation	4 h and 35 min	Structural ensemble of 2000 structures run on one CPU core
c) Weak spot detection	2 h	Manual identification by visual inspection
d) Filtering weak spots	Instantaneous	Highly conserved weak spots were discarded
e) Variant modeling by SCWRL	< 1 s	For a single mutation
f) ENT^FNC run	~ 1 h and 10 min	For a single mutation applying 1000 network topologies
Total	~700 h	The computations for 522 times step f) (~610 h) can be trivially parallelized.

[a] Steps are according to Fig 1.

[b] For the case study with *Bs*LipA described here. With respect to the protein size N, the times required for the steps scale as a) $N \log N$ [86], b) N [23], c) N (assuming that the number of weak spots scales linearly with the protein size), e) N, and f) N^2.

doi:10.1371/journal.pcbi.1004754.t004

representations of the protein [28–33] or use structural ensembles [27]. Finally, our approach does not require weighting or fitting parameters, in contrast to other methods [27,30,31,66].

As to the application to *Bs*LipA, our approach resulted in three out of twelve experimentally tested single-point mutations with significantly increased thermostability with respect to WT, yielding 6.6°C as the largest increase. This increase compares favorably to the median increase in the apparent melting temperature of 8°C found for 93 cases of engineered proteins, most of which contain more than one mutation [67]. Considering all tested single-point mutations, our approach yielded a success rate as to significantly increased thermostability of 25%, which raises to 60% if the five small-to-large residue mutations and the two mutations in the active site are excluded. These success rates are markedly higher than the 5% of mutations showing an increase in protein stability found within 1285 variants of ten different proteins [68,69]. It is also instructive to compare our results to those obtained by testing a complete site saturation mutagenesis library of *Bs*LipA for improved detergent tolerance, where the success rate amounts to 2% [57]. Furthermore, for state-of-the-art methods for predicting the sign of stability change due to a mutation, impressive accuracies of over 80% have been reported [28]. These values result from the methods being very good at predicting destabilizing mutations and the prevalence of such mutations in the investigated data sets [28]. In line with this, for our predicted negative controls, we found a success rate as to significantly decreased thermostability of 90%. In contrast, the methods' performances are much worse in predicting stabilizing mutations, yielding an average success rate for such mutations of 36% over 12 methods [28].

Evaluating the performance of our approach as a binary classifier [70] (S2 Table), our approach discriminates between mutations leading to increased thermostability *versus* those leading to decreased thermostability with a sensitivity of 83%, a specificity of 56%, and an accuracy of 63% considering all variants in Tables 2 and S1, and a sensitivity of 100%, a specificity of 77%, and an accuracy of 83% if the small-to-large residue mutations and the two mutations in the active site are excluded. In our view, this signifies that our approach provides for a robust binary classifier. Our approach has a precision (predictive value for increased thermostability) of 42% (60% if the small-to-large residue mutations and the two mutations in the active site are excluded) (S2 Table), which leads to a gain in precision with respect to a random classifier of a factor of 1.6 (2.4). Furthermore, a Mann–Whitney U test [71] demonstrates that predicted positive ΔT_p significantly points to increased experimental thermostability ($p < 0.05$) (see S2 Text).

An approach related to CNA is the distance constraint model (DCM)[72], which reaches average percent errors of 1.1% (Pearson correlation coefficient $R = 0.72$) [73] and 4.3% ($R = 0.64$) [74] for melting point predictions of protein variants with single and multiple mutations, corresponding to an error of ~4 K [73] and ~14 K [74]. This model requires a system-specific fitting to experimental heat capacity curves from differential scanning calorimetry, however [73,74]. Over all variants predicted (including the negative controls but excluding the three variants for which no activity could be measured (Table 2)), our approach, which does not require fitting parameters, yields a significant ($R = 0.48$, $p = 0.02$) correlation between predicted and experimental thermostabilities; if small-to-large residue mutations and the two mutations in the active site are excluded, the correlation improves further ($R = 0.62$, $p = 0.02$; S6 Fig). These results show that our approach can reproduce experimental trends with sufficient accuracy.

The effectiveness of our approach is also demonstrated when comparing it to the study by Reetz and coworkers [15] applying iterative saturation mutagenesis to *Bs*LipA. The largest increase in T_{50} they have found for a variant containing a single point mutation in the first step was 4.3°C; our largest increase of 6.6°C compares favorably to this value. Four more steps of optimization and screening of about 8000 colonies then yielded two variants carrying five and

seven mutations that showed an increase of T_{50} by 45°C. The study of Reetz *et al.* also differs from ours in a fundamental aspect: in the former study, those residues that showed the highest crystallographic B-factors, i.e., were the most mobile, were chosen as weak spots. In our study, weak spots constitute residues that segregate from large rigid, i.e. internally immobile, clusters during thermal unfolding.

In summary, these results suggest that our approach is a valuable, orthogonal complement to existing methods for rational protein design aimed at improving thermostability. The more thermostable variants can serve as starting points for further engineering of substrate scope and/or enantioselectivity by directed evolution, exploiting that enhanced thermostability promotes the ease of evolvability [75].

Materials and Methods

Constraint Network Analysis: Thermostability prediction

Constraint Network Analysis (CNA) predicts rigid and flexible regions within a biomolecule, which allows linking these static characteristics to the molecule's stability and function [17,21]. CNA has been described in detail in refs. [17,21,35,76]. The approach has been used previously to predict the (thermodynamic) thermostability of proteins and to identify weak spot residues that, when mutated, are likely to improve thermostability [16,18,19].

In CNA, a protein is modeled as a *body-and-bar* network of bodies (atoms) and bars (covalent and noncovalent interactions). Each atom has six degrees of freedom, and each bar removes one degree of freedom [22]. An interaction between two atoms can be modeled as any number of bars between one and six depending on the strength of the interaction. Here, single covalent bonds (double and peptide bonds) were modeled as five (six) bars, hydrogen bonds and salt bridges (together referred to as "hydrogen bonds") as five bars, and hydrophobic interactions as two bars. For hydrogen bonds a hydrogen bond energy E_{HB} is computed by a modified version of the potential by Mayo and coworkers [77] as described in ref. [26]. By successively removing noncovalent constraints from a network, a thermal unfolding of the protein is simulated [16,18,19,26]. Hydrogen bonds are removed from the network in increasing order of their strength [77], i.e., hydrogen bonds with an energy $E_{HB} > E_{cut}(\sigma)$ are discarded from the network of state σ. In the present study, E_{cut} values ranging from −0.1 kcal mol^{-1} to −6.0 kcal mol^{-1} with a step size of 0.1 kcal mol^{-1} were used. E_{cut} can be converted to a temperature using a linear relation introduced by Radestock and Gohlke [16,18], according to which the range of E_{cut} used in this study is equivalent to increasing the temperature of the system from 302 K to 420 K with a step size of 2 K. The rigidity of each network state σ during the thermal unfolding simulation is analyzed by the pebble game algorithm [23,24] as implemented in the FIRST program [25]. From these analyses, the change in the global rigidity characteristics is monitored by the cluster configuration entropy H_{type2} [76]. Finally, a phase transition temperature T_p is identified as the temperature when a largely rigid network becomes largely flexible. We showed that T_p can be used for predicting the thermodynamic thermostability of and identifying structural weak spots in a protein [16,18,19]. Usually, multiple phase transitions occur during the thermal unfolding of a protein because of its modular architecture, i.e., secondary structure elements can segregate from the largest rigid cluster as a whole [18].

Constraint Network Analysis: Local rigidity indices

In contrast to global indices, local indices monitor rigidity at a residue level. One such index, the rigidity index r_i, is defined for each covalent bond i between two atoms as the E_{cut} value during the thermal unfolding simulation at which the bond changes from rigid to flexible [76]. For a C_α atom-based representation, the average of the two r_i values of the two backbone bonds

is taken. As a two-dimensional itemization of r_i, a stability map rc_{ij} indicates for all residue pairs the E_{cut} value at which a rigid contact between the two residues i, j is lost, i.e., when the two residues stop belonging to the same rigid cluster [76]. From rc_{ij}, a rigid cluster decomposition, i.e., a set of rigid clusters and flexible links in between, can be computed for each network state σ during the thermal unfolding simulation.

When the stability map rc_{ij} is filtered such that only rigid contacts between residues that are at most 5 Å apart from each other (measured as the distance between the closest atom pair of the two residues) are considered, a neighbor stability map results. This map helps focusing on short-range rigid contacts that can be directly modulated by mutagenesis with the aim to stabilize them for improving the overall stability of a protein.

In this study we use neighbor stability maps to analyze the (local) effect of mutations on the stability of rigid contacts of neighboring residues (S4 Fig). The increase in the strength of rigid contacts is calculated as the average over differences in rc_{ij} of the variant *versus* WT for all neighboring residue pairs (lower triangles in S4 Fig). The increase in the strength is measured in K.

Generation of a structural ensemble of wild-type *Bs*LipA

Rigidity analyses using CNA are sensitive with respect to the input structure [45,78]. One way to improve the robustness is to carry out CNA on a structural ensemble derived from molecular dynamics (MD) simulations; then results (T_p values and stability maps) are averaged [19]. In the present study, MD simulations of WT *Bs*LipA were performed using the GPU accelerated version of PMEMD [79] of the AMBER 11 suite of programs [80,81] together with the ff99SB force field [82]. The X-ray crystal structure of *Bs*LipA with the highest resolution (PDB ID: 1ISP; resolution 1.3 Å) was used as input structure [83]. Hydrogen atoms were added using the REDUCE program [84] during which side-chains of Asn, Gln, and His were flipped if necessary to optimize the hydrogen bond network. Then, the system, neutralized by addition of sodium counter-ions, was solvated by a truncated octahedral box of TIP3P [85] water such that a layer of water molecules of at least 11 Å width covers the protein surface. The particle mesh Ewald method [86] was used with a direct-space non-bonded cutoff of 8 Å. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm [87], and the time step for the simulation was 2 fs. After equilibration, a production run of unrestrained MD in the canonical ensemble (NVT) was performed to generate a trajectory of 100 ns length, with conformations extracted every 40 ps from the last 80 ns resulting in a structural ensemble of 2000 conformations. The ensemble was used to predict weak spot residues on *Bs*LipA.

Strategy for predicting thermostabilzing mutations on *Bs*LipA

According to our strategy (Fig 1), the structural ensemble of 2000 conformations of WT *Bs*LipA (Fig 1a) was initially submitted to CNA for weak spot identification and prioritization. An average stability map was generated from individual stability maps for each conformation in the ensemble. A thermal unfolding trajectory showing average rigid cluster decompositions during the thermal unfolding simulation was reconstructed from the average stability map (Figs 1b and S7). For this we exploited that rigid cluster decompositions can be reconstructed from stability maps as the latter store E_{cut} (or temperature, according to the above mentioned linear relation) values for all residue pairs at which these residues cease to be in one rigid cluster during the thermal unfolding. The thermal unfolding trajectory was visually inspected for identifying transitions at which the rigidity of WT *Bs*LipA is substantially reduced using VisualCNA [88]. The inspection was done with a view that rigidifying contacts between the largest rigid cluster and residues that segregate at these substantial phase transitions should

improve the thermostability of the protein by delaying the disintegration of the largest rigid cluster. Accordingly, at every transition, residues that are in the neighborhod of, and whose side-chains point towards the largest rigid cluster from which they segregated, were identified as potential weak spots (Fig 1c). Weak spot residues that showed a high sequence conservation ($\geq 80\%$ identity) in a multiple sequence alignment of 296 lipase class 2 sequences obtained from the Pfam database [89] were not considered further (Fig 1d).

Next, for modeling single-point site-saturation mutations, structures of all possible mutations at each weak spot residue were generated by the SCWRL program [50] using WT BsLipA (PDB ID: 1ISP) as a template (Fig 1e). Conformations of side-chains of all residues within 8 Å of a mutated residue were re-predicted in order to allow for a local structural relaxation. The goodness of fit of the mutated side-chain in its environment was assessed using the ANOLEA server [51,52]. A variant was discarded if its average ANOLEA energy of the neighboring residues (≤ 5 Å of the mutation) is higher than the average energy of the same residues in WT by ≥ 2 kcal mol^{-1}. For all variant structures, hydrogen atoms were added using REDUCE [84] in an identical way as done for WT BsLipA (see section "Generation of a structural ensemble of WT BsLipA" for details). The structures were minimized by 2000 steps of conjugate gradient minimization (including an initial steepest descent minimization for 100 steps) or until the root mean-square gradient of the energy was $< 1.0^*10^{-4}$ kcal mol^{-1} Å^{-1}. The energy minimization was carried out with Amber11 [80] using the ff99SB force field [82] and the GB^{OBC} generalized Born model [90].

Finally, the generated variant structures were subjected to thermostability prediction and prioritization. In order to circumvent compute-intensive MD simulations for generating structural ensembles of each of the BsLipA variants, the more efficient ENT^{FNC} approach [35] was used in connection with thermal unfolding simulations by CNA [21]. Ensembles of 1000 network topologies of all single point variants of BsLipA were analyzed; for consistency, the WT BsLipA structure was treated in the same way (including an energy minimization as described above). For each variant and WT, T_p was identified as the inflection point of the sigmoid with the larger change in H_{type2} using a double sigmoid function [19] fitted to H_{type2} vs. T curves. That way, in most cases, a late transition involving the final decay of the largest rigid cluster is identified as T_p [18] except when a very large loss of rigidity occurs during an early transition. Based on ensemble-averaged T_p (Fig 1f), variants were selected for experimental characterization of their thermostability and Michaelis-Menten kinetics. See S1 Text for details. Table 4 summarizes the required computing times.

Supporting Information

S1 Fig. Map of $\Delta T_p = T_p$ (variant) $- T_p$ (wt) values for each mutation (abscissa) at each weak spot residue (ordinate) identified by CNA in WT BsLipA. Weak spot residues are grouped by the major transition at which they are identified (Table 1 and Fig 2 in the main text). Mutations are colored according to thermostabilizing (red) or thermodestabilizing (blue) effects: Of the 239 single point mutations at the 13 weak spots identified from early transitions at low temperatures (T1 and T2), only four resulted in a higher T_p than WT BsLipA; two of these increases were statistically significant ($p < 0.05$ according to Welch's t-test [91]). At five weak spots identified at transition T3, seven mutations resulted in higher T_p values than WT BsLipA; three of these increases were statistically significant. The most pronounced predicted thermostabilization both in terms of the number of variants showing increased T_p values (55, of which 27 were significant) and the magnitude of the T_p increase was observed for mutations at the nine weak spots identified at transition T4. Finally, nine mutations at four weak spot residues identified at the last transition T5 resulted in an increase in T_p compared to WT BsLipA;

six of these increases were significant. Weak spot residues that are highly conserved in the multiple sequence alignment of the lipase family (see the main text) are shown in gray. Mutations that led to energetically unfavorable structures as calculated by the ANOLEA server are shown as white stripes on gray color. Experimentally tested variants are marked by a black box.
(TIF)

S2 Fig. Specific activities of *Bs*LipA variants between temperatures 40 and 60°C. The *Bs*LipA variants and the *p*NPP substrate solutions were incubated for 5 min at the indicated temperatures, and then the activity was measured at these temperatures. Variants G104I and I87W were inactive at these temperatures.
(TIF)

S3 Fig. Specific activities of *Bs*LipA variants between temperatures 40 and 60°C. The *Bs*LipA variants and the *p*NPP substrate solutions were incubated for 30 min at the indicated temperatures, and then the activity was measured at these temperatures. Variants G104I, G104L, and I87W were inactive at these temperatures.
(TIF)

S4 Fig. Differences in the stability of rigid contacts between WT and variants of *Bs*LipA: V54H (a), F58I (b), V96S (b). In the map, a red (blue) color indicates that a rigid contact in the variant is more (less) stable than in the WT (see color scale). The upper triangle shows differences in the stability values for all residue pairs; the lower triangle shows differences in the stability values only for residue pairs that are within 5 Å of each other, with values for all other residue pairs colored grey. Secondary structure elements as computed by the DSSP program [92,93] are indicated on both abscissa and ordinate and are labeled: α-helix (red rectangle), β-strand (green rectangle), loop (black line). Mutated residues are indicated by arrows. Blow-ups for secondary structure pairs of F58I described in the main text are shown.
(TIF)

S5 Fig. Blow-ups of Fig 4d in the main text showing differences in the stability of "rigid contacts" between variant F58I and WT depicted on the variant structure: αB/loop β4-αB (a); αA/β strands 3,4,5 (b); αB/αC (c); loop αB-β5/loop αC-β6 (d); loop αC-β6/loop αD-β7 (e). Two residues form a "rigid contact" if they belong to one rigid cluster. A red (blue) stick connecting C_α atoms of two residues indicates that a rigid contact in the variant is more (less) stable than in the WT (see color scale). Only those contacts of variant F58I that are stabilized or destabilized by ≥ 2 K are shown for clarity; for the same reason, contacts between two residues of the same secondary structure element are not shown. Note that the blow-ups shown here are related to the blow-ups shown in S4b Fig.
(TIF)

S6 Fig. Scatter plot of predicted ΔT_p values for all variants except the eight small-to-large residue mutations and the two mutations in the active site versus the experimental change in thermostability.
(TIF)

S7 Fig. Construction of average rigid cluster decompositions from using a structural ensemble. The input structure (1) is subjected to MD simulations for generating a structural ensemble (2). Stability maps are generated for each conformation in the ensemble using CNA (3). From all individual maps, an average stability map is generated (4). A trajectory showing the average rigid cluster decomposition during the thermal unfolding is then reconstructed from the average stability map (5).
(TIF)

S8 Fig. SDS-PAGE of all variants and the WT that were purified using a N-terminal his-tag and the Ni-NTA purification method. After purification the samples were desalted and stored in 10 mM glycine buffer pH 11. The variant I87W was in all biological replicates not expressed properly and could only be purified in small amounts.
(TIF)

S1 Table. *Bs*LipA variants with negative ΔT_{p} characterized experimentally.
(PDF)

S2 Table. Confusion matrix for two possible outcomes $\Delta T > 0$ and $\Delta T < 0$ for classifying 22 (15) *Bs*LipA variants with respect to predicted and experimental changes in thermostability related to WT *Bs*LipA.
(PDF)

S3 Table. *Bs*LipA variants and mutagenesis primer sequences.
(PDF)

S1 Text. Experimental characterization of thermostability and Michaelis-Menten kinetics.
(PDF)

S2 Text. Mann–Whitney U test.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: HG, KEJ, PCR, AF. Performed the experiments: PCR, AF. Analyzed the data: PCR, AF, KEJ, HG. Wrote the paper: PCR, AF, KEJ, HG.

References

1. Demirjian DC, Moris-Varas F, Cassidy CS (2001) Enzymes from extremophiles. Current Opinion in Chemical Biology 5: 144–151. PMID: 11282340

2. Van den Burg B (2003) Extremophiles as a source for novel enzymes. Current Opinion in Microbiology 6: 213–218. PMID: 12831896

3. Polizzi KM, Bommarius AS, Broering JM, Chaparro-Riggers JF (2007) Stability of biocatalysts. Current Opinion in Chemical Biology 11: 220–225. PMID: 17307381

4. Haki GD, Rakshit SK (2003) Developments in industrially important thermostable enzymes: a review. Bioresource Technology 89: 17–34. PMID: 12676497

5. Eijsink VGH, Gaseidnes S, Borchert TV, van den Burg B (2005) Directed evolution of enzyme stability. Biomolecular engineering 22: 21–30. PMID: 15857780

6. Ahmad S, Kamal MZ, Sankaranarayanan R, Rao NM (2008) Thermostable *Bacillus subtilis* lipases: In vitro evolution and structural insight. Journal of Molecular Biology 381: 324–340. doi: 10.1016/j.jmb.2008.05.063 PMID: 18599073

7. Jochens H, Aerts D, Bornscheuer UT (2010) Thermostabilization of an esterase by alignment-guided focussed directed evolution. Protein Engineering Design & Selection 23: 903–909.

8. Kotzia GA, Labrou NE (2009) Engineering thermal stability of L-asparaginase by in vitro directed evolution. FEBS Journal 276: 1750–1761. doi: 10.1111/j.1742-4658.2009.06910.x PMID: 19220855

9. Lehmann M, Wyss M (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Current Opinion in Biotechnology 12: 371–375. PMID: 11551465

10. Vazquez-Figueroa E, Chaparro-Riggers J, Bommarius AS (2007) Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept. Chembiochem 8: 2295–2301. PMID: 17990264

11. Tian JA, Wang P, Gao S, Chu XY, Wu NF, et al. (2010) Enhanced thermostability of methyl parathion hydrolase from *Ochrobactrum* sp. M231 by rational engineering of a glycine to proline mutation. FEBS Journal 277: 4901–4908. doi: 10.1111/j.1742-4658.2010.07895.x PMID: 20977676

12. Leemhuis H, Rozeboom HJ, Dijkstra BW, Dijkhuizen L (2004) Improved thermostability of *Bacillus circulans* cyclodextrin glycosyltransferase by the introduction of a salt bridge. Proteins-Structure Function and Genetics 54: 128–134.

13. Kaneko H, Minagawa H, Shimada J (2005) Rational design of thermostable lactate oxidase by analyzing quaternary structure and prevention of deamidation. Biotechnology Letters 27: 1777–1784. PMID: 16314970

14. Eijsink VGH, Bjørk A, Gåseidnes S, Sirevåg R, Synstad B, et al. (2004) Rational engineering of enzyme stability. Journal of Biotechnology 113: 105–120. PMID: 15380651

15. Reetz MT, Carballeira JD, Vogel A (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. Angewandte Chemie International Edition 45: 7745–7751.

16. Radestock S, Gohlke H (2008) Exploiting the link between protein rigidity and thermostability for data-driven protein engineering. Engineering in Life Sciences 8: 507–522.

17. Rathi PC, Pfleger C, Fulle S, Klein DL, Gohlke H (2011) Statics of biomacromolecules. In: Comba P, editor. Molecular Modeling. Weinheim: Wiley-VCH. pp. 281–299.

18. Radestock S, Gohlke H (2011) Protein rigidity and thermophilic adaptation. Proteins: Struct, Funct, Bioinf 79: 1089–1108.

19. Rathi PC, Radestock S, Gohlke H (2012) Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio. Journal of Biotechnology 159: 135–144. doi: 10.1016/j.jbiotec.2012.01.027 PMID: 22326626

20. Krüger DM, Rathi PC, Pfleger C, Gohlke H (2013) CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo-)stability, and function. Nucleic Acids Research 41: W340–W348. doi: 10.1093/nar/gkt292 PMID: 23609541

21. Pfleger C, Rathi PC, Klein DL, Radestock S, Gohlke H (2013) Constraint Network Analysis (CNA): A Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function. Journal of Chemical Information and Modeling 53: 1007–1015. doi: 10.1021/ci400044m PMID: 23517329

22. Hespenheide BM, Jacobs DJ, Thorpe MF (2004) Structural rigidity in the capsid assembly of cowpea chlorotic mottle virus. Journal of Physics: Condensed Matter 16: S5055–S5064.

23. Jacobs DJ, Thorpe MF (1995) Generic rigidity percolation: the pebble game. Physical Review Letters 75: 4051–4054. PMID: 10059802

24. Jacobs DJ, Hendrickson B (1997) An algorithm for two-dimensional rigidity percolation: the pebble game. Journal of Computational Physics 137: 346–365.

25. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF (2001) Protein flexibility predictions using graph theory. Proteins: Struct, Funct, Bioinf 44: 150–165.

26. Rader AJ, Hespenheide BM, Kuhn LA, Thorpe MF (2002) Protein unfolding: rigidity lost. Proceedings of the National Academy of Sciences of the United States of America 99: 3540–3545. PMID: 11891336

27. Benedix A, Becker CM, de Groot BL, Caflisch A, Bockmann RA (2009) Predicting free energy changes using structural ensembles. Nature Methods 6: 3–4. doi: 10.1038/nmeth0109-3 PMID: 19116609

28. Worth CL, Preissner R, Blundell TL (2011) SDM—a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Research 39: W215–W222. doi: 10.1093/nar/gkr363 PMID: 21593128

29. Pokala N, Handel TM (2005) Energy functions for protein design: Adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. Journal of Molecular Biology 347: 203–227. PMID: 15733929

30. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. Journal of Molecular Biology 320: 369–387. PMID: 12079393

31. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Research 33: W306–W310. PMID: 15980478

32. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. Methods in Enzymology 383: 66–93. PMID: 15063647

33. Potapov V, Cohen M, Inbar Y, Schreiber G (2010) Protein structure modelling and evaluation based on a 4-distance description of side-chain interactions. BMC Bioinformatics 11.

34. Razvi A, Scholtz JM (2006) Lessons in stability from thermophilic proteins. Protein Science 15: 1569–1578. PMID: 16815912

35. Pfleger C, Gohke H (2013) Efficient and robust analysis of biomacromolecular flexibility using ensembles of network topologies based on fuzzy noncovalent constraints. Structure 21: 1725–1734. doi: 10.1016/j.str.2013.07.012 PMID: 23994009

36. Jaeger KE, Eggert T (2002) Lipases for biotechnology. Current Opinion in Biotechnology 13: 390–397. PMID: 12323363

37. Jaeger KE, Ransac S, Dijkstra BW, Colson C, Vanheuvel M, et al. (1994) Bacterial lipases. FEMS Microbiology Reviews 15: 29–63. PMID: 7946464

38. Tian F, Yang C, Wang C, Guo T, Zhou P (2014) Mutatomics analysis of the systematic thermostability profile of *Bacillus subtilis* lipase A. Journal of Molecular Modeling 20: 2257. doi: 10.1007/s00894-014-2257-x PMID: 24827611

39. Abraham T, Pack SP, Yoo YJ (2005) Stabilization of *Bacillus subtilis* Lipase A by increasing the residual packing. Biocatalysis and Biotransformation 23: 217–224.

40. Ahmad S, Rao NM (2009) Thermally denatured state determines refolding in lipase: Mutational analysis. Protein Science 18: 1183–1196. doi: 10.1002/pro.126 PMID: 19472328

41. Acharya P, Rajakumara E, Sankaranarayanan R, Rao NM (2004) Structural basis of selection and thermostability of laboratory evolved *Bacillus subtilis* lipase. Journal of Molecular Biology 341: 1271–1281. PMID: 15321721

42. Kamal MZ, Ahmad S, Molugu TR, Vijayalakshmi A, Deshmukh MV, et al. (2011) In vitro evolved non-aggregating and thermostable lipase: Structural and thermodynamic investigation. Journal of Molecular Biology 413: 726–741. doi: 10.1016/j.jmb.2011.09.002 PMID: 21925508

43. Kamal MZ, Ahmad S, Yedavalli P, Rao NM (2010) Stability curves of laboratory evolved thermostable mutants of a *Bacillus subtilis* lipase. Biochimica et Biophysica Acta-Proteins and Proteomics 1804: 1850–1856.

44. Jaeger KE, Dijkstra BW, Reetz MT (1999) Bacterial biocatalysts: Molecular biology, three-dimensional structures, and biotechnological applications of lipases. Annual Review of Microbiology 53: 315–351. PMID: 10547694

45. Gohlke H, Kuhn LA, Case DA (2004) Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. Proteins-Structure Function and Bioinformatics 56: 322–337.

46. Beermann B, Guddorf J, Boehm K, Albers A, Kolkenbrock S, et al. (2007) Stability, unfolding, and structural changes of cofactor-free 1H-3-hydroxy-4-oxoquinaldine 2,4-dioxygenase. Biochemistry 46: 4241–4249. PMID: 17371045

47. Hung HC, Chang GG (2001) Multiple unfolding intermediates of human placental alkaline phosphatase in equilibrium urea denaturation. Biophysical Journal 81: 3456–3471. PMID: 11721007

48. Lehmann M, Pasamontes L, Lassen SF, Wyss M (2000) The consensus concept for thermostability engineering of proteins. Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology 1543: 408–415.

49. Steipe B, Schiller B, Pluckthun A, Steinbacher S (1994) Sequence statistics reliably predict stabilizing mutations in a protein domain. Journal of Molecular Biology 240: 188–192. PMID: 8028003

50. Krivov GG, Shapovalov MV, Dunbrack RL (2009) Improved prediction of protein side-chain conformations with SCWRL4. Proteins: Structure, Function, Bioinformatics 77: 778–795.

51. Melo F, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: a www server to assess protein structures. Proceedings, International Conference on Intelligent Systems for Molecular Biology 5: 187–190.

52. Melo F, Feytmans E (1997) Novel knowledge-based mean force potential at atomic level. Journal of Molecular Biology 267: 207–222. PMID: 9096219

53. Dick M, Weiergraber OH, Classen T, Bisterfeld C, Bramski J, et al. (2016) Trading off stability against activity in extremophilic aldolases. Sci Rep 6: 17908. doi: 10.1038/srep17908 PMID: 26783049

54. Tielen P, Kuhn H, Rosenau F, Jaeger KE, Flemming HC, et al. (2013) Interaction between extracellular lipase LipA and the polysaccharide alginate of *Pseudomonas aeruginosa*. BMC Microbiology 13.

55. Eijsink VGH, Vriend G, Vandervinne B, Hazes B, Vandenburg B, et al. (1992) Effects of changing the interaction between subdomains on the thermostability of *Bacillus* neutral proteases. Proteins-Structure Function and Genetics 14: 224–236.

56. Cho AR, Yoo SK, Kim EJ (2000) Cloning, sequencing and expression in *Escherichia coli* of a thermophilic lipase from *Bacillus thermoleovorans* ID-1. FEMS Microbiology Letters 186: 235–238. PMID: 10802177

57. Fulton A, Frauenkron-Machedjou VJ, Skoczinski P, Wilhelm S, Zhu L, et al. (2015) Exploring the protein stability landscape: *Bacillus subtilis* lipase A as a model for detergent tolerance. Chembiochem 16: 930–936. doi: 10.1002/cbic.201402664 PMID: 25773356

58. Yilmaz LS, Atilgan AR (2000) Identifying the adaptive mechanism in globular proteins: Fluctuations in densely packed regions manipulate flexible parts. The Journal of Chemical Physics 113: 4454–4464.

59. Rader AJ, Yennamalli RM, Harter AK, Sen TZ (2012) A rigid network of long-range contacts increases thermostability in a mutant endoglucanase. Journal of Biomolecular Structure and Dynamics 30: 628–637. PMID: 22731517

60. Verma D, Jacobs DJ, Livesay DR (2012) Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. PLoS Comput Biol 8: e1002409. doi: 10.1371/journal.pcbi.1002409 PMID: 22396637

61. Rathi PC, Jaeger KE, Gohlke H (2015) Structural Rigidity and Protein Thermostability in Variants of Lipase A from *Bacillus subtilis*. PLoS One 10: e0130289. doi: 10.1371/journal.pone.0130289 PMID: 26147762

62. van Pouderoyen G, Eggert T, Jaeger KE, Dijkstra BW (2001) The crystal structure of *Bacillus subtilis* lipase: a minimal alpha/beta hydrolase fold enzyme. Journal of Molecular Biology 309: 215–226. PMID: 11491291

63. Yin S, Ding F, Dokholyan NV (2007) Eris: an automated estimator of protein stability. Nature Methods 4: 466–467. PMID: 17538626

64. Ohmura T, Ueda T, Ootsuka K, Saito M, Imoto T (2001) Stabilization of hen egg white lysozyme by a cavity-filling mutation. Protein Science 10: 313–320. PMID: 11266617

65. Korkegian A, Black ME, Baker D, Stoddard BL (2005) Computational thermostabilization of an enzyme. Science 308: 857–860. PMID: 15879217

66. Bordner AJ, Abagyan RA (2004) Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. Proteins-Structure Function and Bioinformatics 57: 400–413.

67. Wijma HJ, Floor RJ, Janssen DB (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. Current Opinion in Structural Biology 23: 588–594. doi: 10.1016/j.sbi.2013.04.008 PMID: 23683520

68. Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. Current Opinion in Structural Biology 19: 596–604. doi: 10.1016/j.sbi.2009.08.003 PMID: 19765975

69. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. Journal of Molecular Biology 369: 1318–1332. PMID: 17482644

70. Fawcett T (2006) An introduction to ROC analysis. Patt Recognit Lett 27: 861–874.

71. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics 18: 50–60.

72. Jacobs DJ, Dallakyan S, Wood GG, Heckarthorne A (2003) Network rigidity at finite temperature: Relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems. Phys Rev E 68: 061109.

73. Li T, Tracka MB, Uddin S, Casas-Finet J, Jacobs DJ, et al. (2014) Redistribution of flexibility in stabilizing antibody fragment mutants follows Le Chatelier's principle. Plos One 9.

74. Verma D, Jacobs DJ, Livesay DR (2010) Predicting the melting point of human C-type lysozyme mutants. Current Protein & Peptide Science 11: 562–572.

75. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. Proceedings of the National Academy of Sciences of the United States of America 103: 5869–5874. PMID: 16581913

76. Pfleger C, Radestock S, Schmidt E, Gohlke H (2013) Global and local indices for characterizing biomolecular flexibility and rigidity. Journal of Computational Chemistry 34: 220–233. doi: 10.1002/jcc.23122 PMID: 23007873

77. Dahiyat BI, Gordon DB, Mayo SL (1997) Automated design of the surface positions of protein helices. Protein Science 6: 1333–1337. PMID: 9194194

78. Mamonova T, Hespenheide B, Straub R, Thorpe MF, Kurnikova M (2005) Protein flexibility using constraints from molecular dynamics simulations. Physical Biology 2: S137–S147. PMID: 16280619

79. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC (2013) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. Journal of Chemical Theory and Computation 9: 3878–3888. doi: 10.1021/ct400314y PMID: 26592383

80. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, et al. (2005) The Amber biomolecular simulation programs. Journal of Computational Chemistry 26: 1668–1688. PMID: 16200636

81. Case D.A. D TA, Cheatham T.E. III, Simmerling C.L, Wang J., Duke R.E., Luo R., Walker R.C., Zhang W., Merz K.M., Roberts B., Wang B., Hayik S., Roitberg A., Seabra G., Kolossváry I., Wong K.F., Paesani F., Vanicek J., Liu J., Wu X., Brozell S.R., Steinbrecher T., Gohlke H., Cai Q., Ye X., Wang J., Hsieh M.-J., Cui G., Roe D.R., Mathews D.H., Seetin M.G., Sagui C., Babin V., Luchko T., Gusarov S., Kovalenko A., Kollman P.A. (2010) AMBER 11. University of California, San Francisco.

82. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins: Struct, Funct, Bioinf 65: 712–725.

83. Kawasaki K, Kondo H, Suzuki M, Ohgiya S, Tsuda S (2002) Alternate conformations observed in catalytic serine of *Bacillus subtilis* lipase determined at 1.3 A resolution. Acta Crystallographica Section D, Biological Crystallography 58: 1168–1174. PMID: 12077437

84. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. Journal of Molecular Biology 285: 1735–1747. PMID: 9917408

85. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. The Journal of Chemical Physics 79: 926–935.

86. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N.log (N) method for Ewald sums in large systems. The Journal of Chemical Physics 98: 10089–10092.

87. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics 23: 327–341.

88. Rathi PC, Mulnaes D, Gohlke H (2015) VisualCNA: a GUI for interactive constraint network analysis and protein engineering for improving thermostability. Bioinformatics 31: 2394–2396. doi: 10.1093/bioinformatics/btv139 PMID: 25770091

89. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. (2012) The Pfam protein families database. Nucleic Acids Research 40: D290–D301. doi: 10.1093/nar/gkr1065 PMID: 22127870

90. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins: Struct, Funct, Bioinf 55: 383–394.

91. Welch BL (1947) The generalization of Student's problem when several different population variances are involved. Biometrika: 28–35. PMID: 20287819

92. Kabsch W, Sander C (1983) Dictionary of Protein Secondary Structure—Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers 22: 2577–2637. PMID: 6667333

93. Joosten RP, Beek TAHT, Krieger E, Hekkelman ML, Hooft RWW, et al. (2011) A series of PDB related databases for everyday needs. Nucleic Acids Research 39: D411–D419. doi: 10.1093/nar/gkq1105 PMID: 21071423