Systematically Scrutinizing the Impact of Substitution Sites on Thermostability and Detergent Tolerance for *Bacillus subtilis* Lipase A

Christina Nutschel, Alexander Fulton, Olav Zimmermann, Ulrich Schwaneberg, Karl-Erich Jaeger, and Holger Gohlke*

**ABSTRACT:** Improving an enzyme’s (thermo-)stability or tolerance against solvents and detergents is highly relevant in protein engineering and biotechnology. Recent developments have tended toward data-driven approaches, where available knowledge about the protein is used to identify substitution sites with high potential to yield protein variants with improved stability, and subsequently, substitutions are engineered by site-directed or site-saturation (SSM) mutagenesis. However, the development and validation of algorithms for data-driven approaches have been hampered by the lack of availability of large-scale data measured in a uniform way and being unbiased with respect to substitution types and locations. Here, we extend our knowledge on guidelines for protein engineering following a data-driven approach by scrutinizing the impact of substitution sites on thermostability or/detergent tolerance for *Bacillus subtilis* lipase A (BsLipA) at very large scale. We systematically analyze a complete experimental SSM library of BsLipA containing all 3439 possible single variants, which was evaluated as to thermostability and tolerances against four detergents under respectively uniform conditions. Our results provide systematic and unbiased reference data at unprecedented scale for a biotechnologically important protein, identify consistently defined hot spot types for evaluating the performance of data-driven protein-engineering approaches, and show that the rigidity theory and ensemble-based approach Constraint Network Analysis yields hot spot predictions with an up to ninefold gain in precision over random classification.

**1. INTRODUCTION**

Improving a protein’s (thermo-)stability1−9 or tolerance against solvents10−16 and detergents17−19 has become of utmost importance in protein engineering: Considering that enzymes are predominantly used as detergent additives20 and that the global industrial enzyme market has been forecast to reach $7.0 billion by 2023 from $5.5 billion in 2018 makes it clear that an increasing demand exists for enzymes that are adapted to harsh temperature, solvent, and detergent conditions.21−22

Modifying protein stability based on rational approaches has a long history,33,34 and a number of, usually, structure-based algorithms have been developed that estimate the effect of a substitution on the stability of a protein.23−28 However, despite successful applications in single cases (e.g., see Table 2 in ref 20), the general reliability of these approaches is still unsatisfactory.23,29−32 One reason is that multiple attempts to identify key features in protein sequences and/or structures associated with protein stability have failed to paint a clear picture, which makes it difficult to define rules of universal validity and general applicability.20,33 Another reason lies in the data used in the design and evaluation of rational design algorithms. The ProTherm database,34,35 which has been most frequently used for such endeavors, contains on average ~12 single, ~12 double, and ~1 multiple substitutions for each of the ~1000 proteins stored.33 Thus, while overall exhaustive, the data may not include a sufficient number of variants per protein to compensate for outliers and, therefore, may not allow a stratification of the data to derive a generally applicable set of rules. As such data, furthermore, originate from different experimental methods, it is not surprising that different changes in protein stability have been found associated with the same variant.36 In addition, the data are strongly biased toward substitutions to alanine, whereas it is very limited for some other substitutions.37 Recently, comprehensive mutagenesis data on a domain level associated with protein stabilities against a denaturing agent have been reported as a means to overcome these limitations.38

Following the principles of natural evolution, albeit on a reduced time scale, protein engineering by directed evolution has emerged as an attractive strategy to improve stability
through iterative cycles of mutagenesis and screening or selection.\textsuperscript{20,39} However, the highly labor-intensive method can become technically challenging if beneficial mutations need to be accumulated over generations of mutagenesis and screening or selection to reach a desired effect.\textsuperscript{40} After all, evolution is not good for problems that require multiple, simultaneous, low-probability events.\textsuperscript{41} To successfully investigate the then necessary large protein libraries, powerful automated techniques for rapid high-throughput screenings were established.\textsuperscript{20,39}

As an intermediate, third route recent developments have tended toward data-driven approaches,\textsuperscript{42} where available knowledge about the protein is used to first identify a substitution site with high potential to yield protein variants with improved stability, and second, substitutions are engineered by site-directed (SDM) or site-saturation (SSM) mutagenesis.\textsuperscript{73} The “knowledge” can arise from sequence information,\textsuperscript{42,43} structure information,\textsuperscript{44–46} or computational techniques.\textsuperscript{2,4,7,8,47,48} By such data-driven approaches, the challenge of accurately predicting the effect of a substitution on protein stability is circumvented, and substitution effects are guided to a few, distinguished sequence positions, making subsequent combinations feasible. However, even with high-throughput screening techniques, it is difficult to handle all variants based on combinations of the 20 proteinogenic AAs at more than six substitution sites (i.e., more than $20^6 = 6.4 \times 10^{10}$ variants).\textsuperscript{20,39,49,50}

Here, to extend our knowledge on guidelines for time- and cost-efficient protein engineering following a data-driven approach, we scrutinize the impact of substitution sites on thermostability or/and detergent tolerance for one protein at very large scale. To do so, we systematically analyze a complete experimental SSM library of BsLipA produced by us,\textsuperscript{15,16,19} which contains all 3439 theoretically possible single variants (181 substitution sites of BsLipA \times 19 naturally occurring AAs) and was evaluated as to different protein stabilities under respectively uniform conditions. Previously, the SSM library has been characterized regarding solvent and detergent tolerance (D) data.\textsuperscript{15,16,19} Here, we characterize the SSM library for the first time regarding thermostability ($T_{50}$) as well as combined $T_{50}$ and D data. BsLipA is a particularly interesting protein for such analysis because a high-resolution X-ray crystal structure (PDB ID: 1ISP, 1.3 Å) is known,\textsuperscript{51} which provides valuable insights in atomic details. Furthermore, the protein has considerable biotechnological importance,\textsuperscript{2,3} possesses an $\alpha$/\(\beta\)-hydrolase fold\textsuperscript{52} such that the impact of substitution sites at $\alpha$-helices, $\beta$-strands, and other secondary structure elements can be tested, and has been used frequently as a model system in related experimental and computational small-scale studies.\textsuperscript{2,3}

Our systematic large-scale analysis focuses on the following five aspects: (I) We determined the likelihoods to find substitution sites showing significantly increased $T_{50}$ or $D$ and investigated the frequencies and magnitudes of effects caused by single AA substitutions. (II) We analyzed at which substitution sites variants result with increased $T_{50}$ or/and $D$ across the protein and compared the findings to random mutagenesis. (III) From these results, we defined hot spot classes, i.e., classes of substitution sites particularly promising to increase $T_{50}$ or/and $D$. (IV) We probed to what extent hot spots can be predicted based on structure or sequence characteristics. (V) We tested the predictive power of the rigidity theory-based approach Constraint Network Analysis (CNA) previously applied in related scenarios.\textsuperscript{2,4–8} i.e., how accurately hot spots can be predicted as structural weak spots identified in a thermal unfolding simulation of the protein.

The main outcomes from our analyses are that we provide systematic and unbiased reference data at large scale for thermostability measured as $T_{50}$ values and detergent tolerance measured as $D$ for a biotechnologically important protein, we identify and consistently define hot spot types for evaluating the performance of data-driven protein-engineering approaches, and we show that CNA-based hot spot prediction can yield a gain in precision over random classification up to ninefold.

2. MATERIALS AND METHODS

2.1. Generation and Screening of the BsLipA SSM Library toward Changes in $T_{50}$ or $D$. The BsLipA library was constructed by site-saturation mutagenesis (SSM) and site-directed mutagenesis (SDM) as described by Frauenkron-Machedjou et al.\textsuperscript{15,16} and Fulton et al.\textsuperscript{19} In the present study, we defined all 3439 single variants (181 substitution sites of BsLipA \times 19 naturally occurring AAs) generated with SSM and SDM as the “SSM library”.

Previously, the SSM library has been screened toward its tolerance against four different classes of detergents: anionic (sodium dodecyl sulfate, SDS), cationic (cetyltrimethylammonium bromide, CTAB), zwitterionic (3-[hexadecyl(dimethyl)-ammonium]propane-1-sulfonate, SB3-16), and nonionic (polyoxyethylene sorbitan monoolesate, Tween 80) by Fulton et al.\textsuperscript{19} Residual activities of the variants after incubation in the presence of the respective detergent ($D$) were obtained as described in ref 19.

As to the screening procedure regarding thermostability, the screening cultures were incubated as described in ref 19. The culture supernatant was collected by centrifugation (1500 g, 40 min) and diluted 2.5-fold with Sørensen buffer (42.5 mL of Na$_2$HPO$_4$ (8.9 g l$^{-1}$), 2.5 mL of KH$_2$PO$_4$ (6.8 g l$^{-1}$)) before screening. The protein-containing supernatant was incubated in a 0.2 mL PCR microtiter plate (MTP) in a programmable thermal cycler (Eppendorf Mastercycler Thermal Cycler PCR). The supernatant samples were incubated at temperatures between 40 and 60 °C for 20 min. A dry block incubator (MRK 23 Cooling-ThermoMixer, DITABIS) was equipped with a “15 and 50 mL falcon tube adaptor” (BT 03, DITABIS). Three falcon tubes with 19.8 mL of pNP solution A (19.8 mL of Sørensen buffer, 45.54 mg of sodium deoxycholate, 22 mg of gum arabic) were inserted into the falcon tube incubator. All dry block incubators were set to 40 °C, 30 min prior to the beginning of the experiment. Twenty seconds before the end of the incubation, 2.2 mL of pNP solution B (48 mg of pNP in 8 mL of 2-propanol) was added into prewarmed pNP solution A and briefly mixed. The substrate mixture was applied to the wells of the MTPs in 50 μL aliquots to start the measurement of thermostability and measured in a MTP reader (Molecular Devices Spectramax). The enzymatic activity in each sample was measured by the rate of increase in absorption at O.D. 410 nm. The residual activity in each sample was calculated from the slope of the change in absorption at O.D. 410 nm relative to the slope of the sample heated to 40 °C during a measurement time of 3 min. From that, $T_{50}$ was obtained from the inflection point of a sigmoid curve fit. Control experiments with just pNP, or pNP in the presence of BsLipA at temperatures up to 60.6 °C, that way leading to denaturation of BsLipA, show no change in the para-nitrophenolate (pNP) absorption over time, demonstrating that pNP is only produced in the presence of a functional enzyme (Figure 1569)}
2.2. Global Characterization of BsLipA Variants’ Changes in \( T_{50} \) or \( D \). For analyzing the changes in \( T_{50} \) (eq 1) or \( D \) (eq 2) of BsLipA variants, the values of wtBsLipA were used as references; i.e., the differences between the values of the variants and those of wtBsLipA were calculated. Positive (negative) \( \Delta \)-values indicate variants with increased (decreased) \( T_{50} \) or \( D \).

\[
\Delta T_{50} = T_{50}(\text{variant}) - T_{50}(\text{wtBsLipA})
\]

\[
\Delta D = D(\text{variant}) - D(\text{wtBsLipA})
\]

For the large-scale analysis, only \( \Delta T_{50} \) of variants higher (lower) than the experimental uncertainty, taken as the standard deviation \( \sigma \) for the respective variant determined from three screenings of \( T_{50} \) were considered significantly increased (decreased) in \( T_{50} \) compared to wtBsLipA. Furthermore, only \( \Delta D \) of variants higher (lower) than two times the experimental standard deviation (\( 2\sigma_D \)) of wtBsLipA determined from screenings of 2997 wtBsLipA replicates toward the respective detergent were considered significantly increased (decreased) in \( D \) compared to wtBsLipA. Here, \( \sigma_D \) of wtBsLipA was used as significance criterion, as the experimental standard deviation for each variant was not available. \( 2\sigma_D \) was chosen because it corresponds to a \( p \)-value below 0.05.

2.3. Definitions of Classes of BsLipA Substitution Sites. The different classes of substitution sites regarding significantly increased \( T_{50} \) or \( D \) were defined based on the set theory. Therefore, the following binary operations on sets were applied: The union of the sets \( A \) and \( B \) is the set of elements which are in \( A \), in \( B \), or in both \( A \) and \( B \) (eq 3).

\[
(A \cup B) = \{ x \in A \vee x \in B \}
\]

The intersection of the sets \( A \) and \( B \) is the set of elements which are in \( A \) and \( B \) (eq 4).

\[
(A \cap B) = \{ x \in A \wedge x \in B \}
\]

Finally, the Jaccard index (\( J \)) was used to compare the similarity of two sets \( A \) and \( B \), i.e., the cardinal number of the respective intersection divided by the cardinal number of the respective union (eq 5). The range of \( J \) is \([0, 1]\), with 1 indicating identical sets \( A \) and \( B \).

\[
J(A, B) = \frac{|A \cap B|}{|A \cup B|}
\]

Based on the different classes of substitution sites, we defined hot spots, which are substitution sites particularly promising to yield significantly increased \( T_{50} \) or \( D \).

2.4. Structural Determinants of BsLipA Hot Spots. Hot spots were assigned to groups according to their location in secondary structure elements (yielding 20 subgroups), solvent-accessible surface areas (SASAs) (yielding five subgroups), and physicochemical properties (yielding five subgroups). The secondary structure elements of the wtBsLipA crystal structure (PDB ID: 1ISP with highest resolution of 1.3 Å\(^{45}\)) were identified with the DSSP program.\(^{58}\) Subsequently, the SASAs of the wtBsLipA were analyzed with the DSSP program.\(^{59}\) The fractional solvent-accessible surface areas (fSASAs) were calculated with respect to the maximum solvent-accessible surface area of each hot spot (maxSASA) (eq 6).\(^{59}\)

\[
fSASA = \frac{100 \times \text{SASA}}{\text{maxSASA}}
\]

As the screening studies were performed at pH 8,\(^{49}\) hot spots were subgrouped by their physicochemical properties as follows: aliphatic (Ile, Ala, Val, Leu, Gly), aromatic (Phe, Tyr, Trp), neutral (Cys, Pro, Met, Ser, Thr, Asn, Gln), positively charged (His, Lys, Arg), and negatively charged (Asp, Glu).

2.5. Conservation of wtBsLipA Residues within Bacterial Lipases. Apart from the catalytic triad (S77, D133, and H156), also variants at conserved sequence positions were considered because the SSM library revealed significantly increased \( T_{50} \) or \( D \) at such positions. The conservation of wtBsLipA residues within the bacterial lipases was calculated using the available sequences from the Pfam database,\(^{60}\) for the lipase class 2 (PF01674). The sequences were limited to the bacterial sources, which contain 1138 sequences from 603 bacterial species. All sequences were aligned using Clustal Omega.\(^{61,62}\) For the alignment, the full-length sequence of wtBsLipA (UniProt ID: P37957) was used.\(^{63}\) The conservation was calculated using AAcOn Calculations,\(^{64}\) through Jalview.\(^{65}\) The conservation range is \([0, 10]\) with 10 showing no (high) conservation.

2.6. Constraint Network Analysis. The Constraint Network Analysis (CNA) aims at linking structural rigidity and flexibility to the biomolecule’s structure, (thermo)stability, and function.\(^{66–68}\) The CNA software acts as front- and back-end to the graph theory-based rigidity analysis software floppy Inclusions and Rigid Substructure Topography (FIRST).\(^{69}\) In FIRST, proteins are modeled as constraint networks in a body-and-bar representation, which has been described in detail by Hespenheide et al.\(^{70}\) Based on the modeled constraint network of the protein structure, a pebble game algorithm decomposes the network into flexible and rigid subparts.\(^{71,72}\) In order to monitor the decay of network rigidity and to identify the rigidity percolation threshold, CNA performs thermal unfolding simulations by consecutively removing noncovalent constraints (hydrogen bonds, including salt bridges) from a network in increasing order of their strength.\(^{73}\) For this, a hydrogen bond energy \( E_{Hb} \) is computed by a modified version of the potential by Mayo et al.\(^{73}\) During the thermal unfolding simulations, phase transitions can be identified where the network switches from overall rigid to flexible states. For a given network state \( \sigma = f(T) \), hydrogen bonds with an energy \( E_{Hb} > E_{cut}(\sigma) \) are removed from the network at temperature \( T \). In this study, the thermal unfolding simulation was carried out by decreasing \( E_{cut} \) from \(-0.1 \) to \(-6.0 \) kcal mol\(^{-1}\) with a step size of \(0.1 \) kcal mol\(^{-1}\). \( E_{cut} \) can be converted to a temperature \( T \) using the linear equation introduced by Radestock et al. (eq 7).\(^{74,75}\) The range of \( E_{cut} \) is equivalent to increasing the temperature from 302 to 420 K with a step size of 2 K. Because hydrophobic interactions remain constant or become even stronger as the temperature increases,\(^{74,75}\) the number of hydrophobic tethers was kept unchanged during the thermal unfolding simulation, as done previously.\(^{76}\)

\[
T = \frac{-20 K}{\text{kcal}-\text{mol}^{-1}}E_{cut} + 300 K
\]

The CNA software is available under academic licenses from http://cpclab.uni-duesseldorf.de/index.php/Software, and the CNA web server is accessible at http://cpclab.uni-duesseldorf.de/cna/.
2.7. Generation of a Structural Ensemble of wtBSLipA. MD simulations of wtBSLipA were carried out with the GPU-accelerated version of PMEMD\textsuperscript{79} of the AMBER14 suite of programs\textsuperscript{1} together with the ff14SB force field.\textsuperscript{78} As a starting structure, the X-ray crystal structure of wtBSLipA (PDB ID: 1ISP) was used.\textsuperscript{57} Hydrogens were added, and side-chain orientations ("flips") of Asn, Gln, and His were optimized by the REDUCE program\textsuperscript{79} based on suitable hydrogen-bonding geometries and avoiding potential steric clashes. This was done to take into account that O versus N or N versus C is difficult to distinguish in X-ray crystallography experiments.\textsuperscript{79} For neutralization of the system, sodium counterions were added. Subsequently, the system was solvated by a truncated octahedral box of TIP3P water\textsuperscript{80} such that a layer of water molecules of at least 11 Å widths covers the protein surface. The particle mesh Ewald method\textsuperscript{81} was used with a direct-space nonbonded cutoff of 8 Å. Bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm,\textsuperscript{82} and the time step for the simulation was 2 fs. As done before,\textsuperscript{78} a trajectory of 100 ns length was generated after thermalization, simulating in the canonical (NVT) ensemble at $T = 300$ K, with conformations extracted every 40 ps from the last 80 ns, resulting in a structural ensemble of 2000 conformations. We assessed the statistical independence of the extracted conformations by calculating the autocorrelation function of the cluster configuration entropy $H_{\text{type2}}$, the measure used to identify phase transitions in the constraint networks (see section 2.9 below) (Figure S2). Because fluctuations of $H_{\text{type2}}$ decorrelate already within the first two snapshots, the snapshots used for CNA, which were extracted at time intervals of 40 ps, are considered independent.

2.8. Thermal Unfolding Simulation of wtBSLipA. For analyzing the rigid cluster decomposition of wtBSLipA, a thermal unfolding simulation was performed by CNA on an ensemble of network topologies (ENT\textsuperscript{MD}) generated from a molecular dynamics (MD) trajectory. The ensemble-based CNA was pursued to increase the robustness of the rigidity analyses.\textsuperscript{58,83} Subsequently, the unfolding trajectory was visually inspected by VisualCNA\textsuperscript{64} for identifying secondary structure elements that segregate from the largest rigid cluster at each major phase transition. VisualCNA is an easy-to-use PyMOL plug-in that allows setting up CNA runs and analyzing CNA results linking data plots with molecular graphics representations.\textsuperscript{84} VisualCNA is available under an academic license from https://cpclab.uni-duesseldorf.de/index.php/Software.

2.9. Local and Global Indices for Analyzing Structural Rigidity of wtBSLipA. From the thermal unfolding simulation, CNA computes a comprehensive set of indices to quantify biologically relevant characteristics of the biomolecule’s stability.\textsuperscript{85} Global indices are used for determining the flexibility and rigidity at a macroscopic level. Local indices determine the flexibility and rigidity at a microscopic level of bonds.

The cluster configuration entropy $H_{\text{type2}}$ is a global index, which has been introduced by Radestock and Gohlke.\textsuperscript{4} $H_{\text{type2}}$ is used to identify the phase transition temperature $T_p$ at which a biomolecule switches from a rigid to a floppy state and the largest rigid cluster stops to dominate the whole protein network. As long as the largest rigid cluster dominates the whole protein network, $H_{\text{type2}}$ is low because of the limited number of possible ways to configure a system with a very large cluster. When the largest rigid cluster starts to decay or stops to dominate the network, $H_{\text{type2}}$ jumps. There, the network is in a partially flexible state with many ways to configure a system consisting of many small clusters. The percolation behavior of protein networks is usually complex, and multiple phase transitions can be observed.\textsuperscript{2,4,5,7,8} In order to identify $T_p$ a double sigmoid fit was applied to an $H_{\text{type2}}$ versus $T(E_{\text{cut}})$ curve as done previously,\textsuperscript{2,4,5,7,8} and $T_p$ taken as that $T$ value associated with the largest slope of the fit.

The stability map $r_{cij}$ is a local index, which has been introduced by Radestock and Gohlke.\textsuperscript{4} $r_{cij}$ represents the local stability within a protein structure for all residue pairs at which a rigid contact $rc$ between two residues $i$ and $j$ (represented by their $C_{\alpha}$ atoms) is lost during the thermal unfolding. $rc$ exists if $i$ and $j$ belong to the same rigid cluster $c$ of the set of rigid clusters $C^R$.\textsuperscript{4} Thus, $r_{cij}$ contains information cumulated over all network states along the unfolding trajectory as to which parts of the network are (locally) mechanically stable at a given $\sigma$ and which are not. This stability information is not only available in a qualitative manner but also quantitatively in that each $r_{cij}$ has been associated with $E_{\text{cut}}$ at which the rigid contact is lost. The sum over all entries in $r_{cij}$ represents the chemical potential energy due to noncovalent bonding, obtained from the coarse-grained, residue-wise network representation of the underlying protein structure. To focus only on the stability of $rc$ between structurally close residues, $r_{cij}$ was filtered such that only rigid contacts between two residues that are at most 5 Å apart from each other were considered (neighbor stability map $r_{cij{\text{neighbor}}}$).

Finally, CNA predicts unfolding nuclei as structural features from which macroscopic (in)stability originates.\textsuperscript{2} Unfolding nuclei are represented by residues that percolate from the largest rigid cluster at the latest phase transition. If such residues become flexible, it will have a detrimental effect on protein stability. Fringe residues of the unfolding nuclei percolate from the largest rigid cluster during earlier steps of the thermal unfolding. We follow the hypothesis that the more structurally stable the fringes of unfolding nuclei are, the more structurally stable will be those unfolding nuclei.\textsuperscript{7} Therefore, if such fringe residues (termed weak spots) are targeted by substitutions, the likelihood to stabilize the rigid core of a protein should be high. If two unfolding nuclei were only separated by one residue, this residue was also considered a weak spot. This procedure of identifying weak spots is in agreement with a previous study by us.\textsuperscript{2}

2.10. Statistical Evaluation of CNA as a Binary Classifier. The performance of CNA was investigated as a binary classifier with the following possible outcomes: true positives (TP) are predicted weak spots that are hot spots, whereas false positives (FP) are predicted weak spots that are non-hot spots. In turn, true negatives (TN) are predicted non-weak spots that are non-hot spots, whereas false negatives (FN) are predicted non-weak spots that are hot spots. Different metrics were then applied to evaluate CNA.

The recall ($r$) answers the question how many hot spots were predicted as weak spots (eq 8).\textsuperscript{86}

$$r = \frac{TP}{TP + FN} = \frac{\text{No. of predicted weak spots that are hot spots}}{\text{No. of hot spots}}$$

The precision ($p$) evaluates how many predicted weak spots are actually hot spots (eq 9).\textsuperscript{86}
Figure 1. Distribution of BsLipA variants’ changes in $T_{50}$ or $D$ toward one detergent. Distribution of BsLipA variants’ changes in (A) $T_{50}$ ($\Delta T_{50}$) or $D$ ($\Delta D$) with respect to (B) SDS, (C) CTAB, (D) SB3-16, and (E) Tween 80 at the indicated concentrations compared to wtBsLipA ($\Delta T_{50}/\Delta D = 0$). (A) Variants with $\Delta T_{50}$ lower than the experimental uncertainty (standard deviation $\sigma_T$ for the respective variant) were excluded from further analyses (gray). (B–E) Variants within $2\sigma_D$ of $\Delta D$ of wtBsLipA determined from screenings of 2997 wtBsLipA replicates toward the respective detergent were excluded from further analyses (gray). The insets show the numbers of variants which cause a significant increase or decrease in $T_{50}$ or $D$ toward one detergent. A red (blue) color indicates a significantly increased (decreased) $T_{50}$ or $D$ toward one detergent.

\[
p = \frac{TP}{TP + FP} = \frac{\text{No. of predicted weak spots that are hot spots}}{\text{No. of weak spots}} \tag{9}
\]

The precision in random classification ($p_{\text{random}}$) indicates how many of the 181 BsLipA residues are actually hot spots (eq 10).\textsuperscript{86}

\[
p_{\text{random}} = \frac{TP + FN}{TP + FP + TN + FN} = \frac{\text{No. of hot spots}}{181 \text{ residues of BsLipA}} \tag{10}
\]

The gain in precision over random classification (gip) represents how many predicted weak spots are actually hot spots in comparison to random classification (eq 11).\textsuperscript{86}
Table 1. Identified Classes of Substitution Sites

<table>
<thead>
<tr>
<th>class</th>
<th>definition</th>
<th>no. of substitution sites</th>
<th>no. of weak spots</th>
<th>gip</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>{substitution site, 1 ≤ x ≤ 181, T_{SB}(x) is significantly increased}</td>
<td>69</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>II</td>
<td>{substitution site, 1 ≤ x ≤ 181, D_{SS}(x) is significantly increased}</td>
<td>74</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>III</td>
<td>{substitution site, 1 ≤ x ≤ 181, D_{CtAB}(x) is significantly increased}</td>
<td>42</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>IV</td>
<td>{substitution site, 1 ≤ x ≤ 181, D_{BS-16}(x) is significantly increased}</td>
<td>46</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>V</td>
<td>{substitution site, 1 ≤ x ≤ 181, D_{Tween80}(x) is significantly increased}</td>
<td>34</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>VI</td>
<td>II ∪ III ∪ IV ∪ V</td>
<td>109</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>VII</td>
<td>I ∪ VI</td>
<td>124</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>VIII</td>
<td>II ∩ III ∩ IV ∩ V</td>
<td>11</td>
<td>2</td>
<td>3.30</td>
</tr>
<tr>
<td>IX</td>
<td>I ∩ VIII</td>
<td>7</td>
<td>2</td>
<td>5.17</td>
</tr>
<tr>
<td>X</td>
<td>{substitution site, 1 ≤ x ≤ 181, six highest effects in significantly increased T_{SB}(x)}</td>
<td>6</td>
<td>1</td>
<td>3.02</td>
</tr>
<tr>
<td>XI</td>
<td>{substitution site, 1 ≤ x ≤ 181, six highest effects in significantly increased D_{SS}(x)}</td>
<td>6</td>
<td>1</td>
<td>3.02</td>
</tr>
<tr>
<td>XII</td>
<td>{substitution site, 1 ≤ x ≤ 181, six highest effects in significantly increased D_{CtAB}(x)}</td>
<td>6</td>
<td>3</td>
<td>9.05</td>
</tr>
<tr>
<td>XIII</td>
<td>{substitution site, 1 ≤ x ≤ 181, six highest effects in significantly increased D_{BS-16}(x)}</td>
<td>6</td>
<td>2</td>
<td>6.03</td>
</tr>
<tr>
<td>XIV</td>
<td>{substitution site, 1 ≤ x ≤ 181, six highest effects in significantly increased D_{Tween80}(x)}</td>
<td>6</td>
<td>0</td>
<td>—</td>
</tr>
<tr>
<td>XV</td>
<td>XI ∪ XII ∪ XIII ∪ XIV</td>
<td>20</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>XVI</td>
<td>X ∪ XV</td>
<td>24</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>XVII</td>
<td>XI ∩ XII ∩ XIII ∩ XIV</td>
<td>0</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>XVIII</td>
<td>X ∩ XVII</td>
<td>0</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
<td>nd&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Class of substitution sites; underlined classes represent hot spots.  
<sup>b</sup>Numbers of hot spots that are predicted as weak spots.  
<sup>c</sup>Gain in precision over random classification (eq 11).  
<sup>d</sup>Not determined.

range is [0, ∞], with values < 1 indicating a lower precision than obtained by random classification.

\[
gip = \frac{p}{p_{\text{random}}} \tag{11}
\]

The \( F_1 \)-score \( (F_1) \) is a measure of the test’s accuracy. It represents the harmonic mean of \( p \) and \( r \); i.e., if there is an uneven class distribution, it is used to seek a balance between \( p \) and \( r \) (eq 12).\(^\text{87}\) The \( F_1 \) range is \([0, 1]\), with 1 indicating perfect \( p \) and \( r \).

\[
F_1 = 2 \cdot \frac{p \cdot r}{p + r} \tag{12}
\]

2.11. Markov Chain Monte Carlo-Based Unfolding Simulations of wt8BSliP.

As an independent method to assess the order of unfolding of wt8BSliP, we used a Markov Chain Monte Carlo (MCMC) simulation with an all-atom model restricted to dihedral degrees of freedom.\(^\text{88}\) This method has been successfully used for protein-folding simulations\(^\text{89}\) and has been shown to reproduce the order of melting temperatures for a set of protein variants.\(^\text{90}\) In this MCMC model, implemented in the open source tool ProFASi (Protein Folding and Aggregation Simulator), the protein conformation is modified by changing one or few dihedral angles in each step. A step is accepted according to the Metropolis criterion, i.e., with a probability that depends on the absolute temperature and the resulting change of energy of the system. In ProFASi, the energy is calculated by an all-atom implicit solvent force field.\(^\text{90,91}\) While MCMC simulations allow arbitrarily large changes to the molecule, the unfolding simulations for this study have been restricted to side chain dihedral updates and small, locally correlated updates of main chain dihedral angles.\(^\text{92}\) To ensure adequate sampling, 96 MCMC simulations at 330 K were performed with a total of 3.05 × 10<sup>16</sup> elementary updates.

3. RESULTS

3.1. About One-Tenth of All Variants in the Complete SSM Library Show Significantly Increased \( T_{50} \) or \( D \) toward at Least One Detergent, and Such Variants Were Found at Two-Thirds of All Substitution Sites.

The BSLipA SSM library contained \( T_{50} \) as well as \( D \) data toward the four detergents SDS, CTAB, SB3-16, and Tween 80 for all 3439 single variants (181 substitution sites of BSLipA × 19 naturally occurring AAs), including also inactive variants (see section 2.1). Initially, the results of both experimental screening studies of the SSM library with respect to changes in \( T_{50} \) (\( \Delta T_{50} \)) or \( D \) toward one detergent (\( \Delta D \)) were assessed in terms of the variance of the data and its significance (see section 2.2).

As to the \( T_{50} \) data, only variants with \( \Delta T_{50} \) higher (lower) than the experimental uncertainty, taken as the standard deviation \( \sigma_T \) for the respective variant determined from three screenings of \( \Delta T_{50} \) were considered significantly increased (decreased) in \( T_{50} \) compared to wtBSliP (\( \Delta T_{50} = 0 \) K) (eq 1). The average \( \sigma_T \) is 0.44 K. In total, 1856 variants with significantly increased \( T_{50} \) were obtained, of which 214 (\( \sim 12\% \)) show an increase and 1642 (\( \sim 88\% \)) a decrease (Figure 1A, Table S1). This proportion represents what one would obtain in the case of random mutagenesis. The distribution of \( \Delta T_{50} \) is left-skewed, with extreme \( \Delta T_{50} \) values of −8.3 and +7.7 K, with the most frequent \( \Delta T_{50} \) range being −2 to −1.5 K (−12% out of 1856 variants), followed by \( \Delta T_{50} \) between −1.5 and −1 K (−10% out of 1856 variants) (Figure 1A). In turn, for each of 69 substitution sites (−38% out of 181 substitution sites) at least one variant with significantly increased \( T_{50} \) was found. These substitution sites are summarized in class I (I = {Substitution site, 1 ≤ x ≤ 181, \( T_{50}(x) \) is significantly increased}) (Tables 1 and S2).

Likewise, only variants with \( \Delta D \) higher (lower) than two times the experimental standard deviation (2\( \sigma_D \)) of wtBSliP determined from screenings of 2997 wtBSliP replicates\(^\text{19}\) toward the respective detergent were considered significantly increased (decreased) in \( D \) compared to wtBSliP (\( \Delta D = 0 \) K) (eq 2). The screening revealed the highest \( \sigma_D \) in the presence of SB3-16, followed by Tween 80, CTAB, and SDS (Table S1).\(^\text{19}\) This may be related to the fact that SB3-16 and Tween 80 were tested above the critical micelle concentration (cmc), while CTAB and SDS were tested below it.\(^\text{19,93}\) The respective detergent concentration had been chosen based on the inactivation of purified wtBSliP (Table S1).\(^\text{19}\) On average, 900 variants with...
significantly increased $D$ were obtained, of which 126 (~14%) show an increase and 774 (~86%) a decrease, on average across each detergent (Figures 1B–E, Table S1). This proportion represents what one would obtain in the case of random mutagenesis. The distribution of $\Delta D$ is left-skewed. The magnitude of the increase (decrease) in $\Delta D$ is between 1.6-fold and 2.4-fold (0.6-fold and 2.9-fold) of the residual activity of wtBsLipA. Furthermore, variants tested against SDS and SB3-16 showed an up to two times higher $\Delta D$ than against CTAB and Tween 80 (Figures 1B–E). This may be related to the different classes of the detergents. In turn, for each of 74, 42, 46, or 34 substitution sites at least one variant with significantly increased $D$ toward SDS, CTAB, SB3-16, or Tween 80 (~41, 23, 25, or 19% out of 181 substitution sites) was found. These substitution sites are summarized in classes $\{I \rightarrow V\}$ (Tables 1 and S2). The union of $\{II \rightarrow V\}$ contains 109 substitution sites (~60% out of 181 substitution sites) and is represented by class VI ($\{VI = I \cup II \cup III \cup IV \cup V\}$) (Tables 1 and S2, eq 3). For each of these substitution sites at least one variant shows significantly increased $D$ toward at least one detergent.

Finally, 124 substitution sites are summarized in the union of $I$ and $VI$ (~69% out of 181 substitution sites) ($\{VII = I \cup VI\}$ (Tables 1 and S2, eq 3). Thus, only for two-thirds of all substitution sites at least one variant with significantly increased $T_{S0}$ or $D$ toward at least one detergent was obtained.

To conclude, for the first time, we performed a systematic large-scale analysis of a complete experimental SSM library toward two types of stabilities of one protein containing all single variants. The likelihoods to generate variants with significantly increased $T_{S0}$ (~12%) or $D$ toward one detergent (~14% on average across all detergents) by random mutagenesis ($I \rightarrow V$) are similar. Variants with significantly increased $T_{S0}$ or $D$ toward at least one detergent were obtained at only two-thirds of all substitution sites ($VII$), and this value falls to about one-third or below if $T_{S0}$ and $D$ toward one detergent are considered separately ($I \rightarrow V$). Hence, such substitution sites are not uniformly distributed across the protein. For the following analyses, only substitution sites with at least one variant yielding significantly increased $T_{S0}$ or $D$ toward at least one detergent were considered.

3.2. The Higher the Frequency of Substitution Occurrences That Lead to Significantly Increased $T_{S0}$ or $D$

![Figure 2](https://dx.doi.org/10.1021/acs.jcim.9b00954)
D toward One Detergent, the More Pronounced the Highest Effect, and Vice Versa. Next, we investigated the BsLipA SSM library regarding the respective frequency of substitution occurrences at substitution sites that lead to significantly increased $T_{50}$ ($N_{\text{BsLipA}}$) or $D$ ($\Delta N_{\text{BsLipA}}$) toward one detergent. Additionally, we analyzed the respective highest effects in significantly increased $T_{50}$ ($\Delta T_{50\text{max}}$) or $D$ ($\Delta D_{\text{max}}$) toward one detergent at substitution sites. Finally, we address the question if the frequency of substitution occurrences and the highest effects per substitution site are related to each other.

The highest $N_{\text{BsLipA}}$ of I was 12 (F17) (Figure 2A), whereas the highest $N_{\text{BsLipA}}$ of II–V were 14 (E65), 6 (1135 and D144), 11 (G46), and 5 (V99) (Figure 2B, Table S14), respectively, indicating that up to ~60% and more of the variants for some substitution sites yield significantly increased $T_{50}$ or $D$ toward one detergent. Correlations between $N_{\text{BsLipA}}$ of I and $N_{\text{BsLipA}}$ of II–V yielded, on average, $R^2 = 0.03; p > 0.1$ (Figure 2C, Table S3). The highest correlation was found between $N_{\text{BsLipA}}$ of I and $N_{\text{BsLipA}}$ of II–V, overall very weak to weak but mostly significant correlations were obtained (on average: $R^2 = 0.11; p < 0.01$) (Figure 2C, Table S3). The highest correlation was observed between $N_{\text{BsLipA}}$ of III and IV ($R^2 = 0.26; p < 0.001$).

The highest $\Delta T_{50\text{max}}$ of I was 7.7 K (M137), whereas the highest $\Delta T_{50\text{max}}$ of II–V were 1.49 (M137), 1.63 (T110), 2.41 (G46), and 2.29 (S127), respectively (Table S9), indicating that specific single AA substitutions have a great impact on the magnitudes of the effects. Correlations between $\Delta T_{50\text{max}}$ of I and $\Delta T_{50\text{max}}$ of II–V shown, on average, $R^2 = 0.06; p > 0.1$ (Figure 2D, Table S4). The highest correlation was observed between $\Delta T_{50\text{max}}$ of I and $\Delta D_{\text{max}}$ of IV ($R^2 = 0.13; p < 0.1$). With respect to $\Delta D_{\text{max}}$ of II–V, overall very weak to weak and mostly insignificant correlations were obtained (on average: $R^2 = 0.08; p > 0.1$) (Figure 2D, Table S4). The highest correlations were observed between $\Delta D_{\text{max}}$ of II and V ($R^2 = 0.24; p < 0.05$) as well as $\Delta D_{\text{max}}$ of III and IV ($R^2 = 0.13; p < 0.1$).

Finally, mostly good to fair and significant correlations between $N_{\text{BsLipA}}$ of I and $\Delta T_{50\text{max}}$ of I as well as $N_{\text{BsLipA}}$ of I and $\Delta D_{\text{max}}$ of II–V were found (on average for increase: $R^2 = 0.27; p < 0.01$) (Figure 2E, Table S5).

To conclude, these findings indicate that the relation “the higher the frequency of substitution occurrences that lead to significantly increased $T_{50}$ or $D$ toward one detergent, the more pronounced the highest effect, and vice versa” holds for substitution sites at which at least one variant shows significantly increased $T_{50}$ or $D$ toward one detergent (I–V). Together with the results from the previous chapter, this result suggests that identifying a priori substitution sites with a high likelihood for significantly increased $T_{50}$ or $D$ toward one detergent will also be beneficial with respect to the magnitude of effects that can be achieved there by substitutions.

3.3. Eleven Substitution Sites Yield a ~4.6-fold Higher Likelihood To Find for Each Detergent Variants with Significantly Increased $D$ than Random Mutagenesis. Next, we focused on pairwise intersections of II–V to investigate if there are substitution sites at which for two detergents at least one variant shows significantly increased $D$, regardless of the magnitude of the single effect (see section 2.3). We compared the pairwise similarities between II–V by calculating the Jaccard index (J), i.e., the cardinal number of the respective intersection divided by the cardinal number of the respective union (Table S6, eq 5). The highest similarity was found between III and IV with $J(\text{III}, \text{IV}) = 0.47$, whereas the lowest similarity was observed between II and V with $J(\text{II}, \text{V}) = 0.23$. This may be related to the different classes of the detergents.

Encouraged by the findings of overlapping II–V, we also looked at the overall intersection of II–V (III = I ∩ VIII), i.e., substitution sites at which for each detergent at least one variant shows significantly increased $D$, regardless of the magnitude of the single effect (Tables 1 and S2, eq 4). VIII contains the 11 substitution sites E2, G13, D43, T45, Y49, N51, V54, E65, N98, M134, and M137 (~6% out of 181 substitution sites) (Tables 1, S2, and S14). These substitution sites are associated with 50 variants causing a significant change in $D$, of which 32 (~64%) show a significant increase, on average across all detergents (Table S7). Thus, this likelihood is ~4.6-fold higher in comparison to random mutagenesis. The most promising substitution sites of VIII are M134, N51, and T45 with variants showing increased $\Delta D_{\text{max}}$ of 2.25, 2.10, and 1.90, respectively.

To conclude, a dramatically reduced number of 11 substitution sites (VIII) yield a ~4.6-fold higher likelihood to find for each detergent variants with significantly increased $D$ compared to random mutagenesis. These findings indicate that if a protein-engineering study aims at identifying variants showing significantly increased $D$ toward each detergent, such substitution sites (VIII) should be identified prior to SDM.
significantly increased $D$ compared to random mutagenesis. These findings indicate that if a protein-engineering study aims at identifying variants showing significantly increased $T_{50}$ and $D$ toward each detergent, such substitution sites (IX) should be identified prior to SDM.

3.5. Six Substitution Sites with Highest $\Delta T_{50;\text{max}}$ ($\Delta D_{\text{max}}$) Yield a ~5.3-fold (~4.5-fold) Higher Likelihood To Find Variants with Significantly Increased $T_{50}$ ($D$) than Random Mutagenesis. The above analyses focused on substitution sites at which significantly increased $T_{50}$ or $D$ toward one detergent (I–V), significantly increased $D$ toward each detergent (VIII), as well as significantly increased $T_{50}$ and $D$ toward each detergent (IX) were observed, regardless of the magnitude of the effect. Now, we identified those six substitution sites for which the respective highest effects ($\Delta T_{50;\text{max}}$ or $\Delta D_{\text{max}}$) were found. The number of 6 is motivated by the current technical limitation to screen more than $20^6$ variants.

The six substitution sites M137, M134, G155, F17, I157, and Y139 yield variants with the highest $\Delta T_{50;\text{max}}$ of 7.7, 5.6, 4.5, 3.8, 3.6, and 3.2 K, respectively, and constitute class $X$ ($X = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } T_{50}(x)\}$. Tables 1, S2, and S9). The substitution sites of $X$ are associated with 68 variants causing a significant change in $T_{50}$, of which 43 (~63%) yield a significantly increased $T_{50}$ (Table S10). Thus, this likelihood is ~5.3-fold higher in comparison to random mutagenesis.

The most promising substitution sites exhibiting variants with the highest $\Delta D_{\text{max}}$ toward one detergent ($XI = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } D_{\text{SDS/CTAB/SB}3/16/\text{Tween 80}(x)\}$. Tables 1, S2, S9) are M137 (XI), T110 (XII), G46 (XIII), and S127 (XIV) with variants showing highest $\Delta D_{\text{max}}$ of 1.49, 1.63, 2.41, and 2.29, respectively (Tables 1, S2, and S9). With these substitution sites, 43 variants are associated causing a significant change in $D$, of which 27 (~63%) cause significantly increased $D$, on average across all detergents (Table S10). Thus, this likelihood is ~4.5-fold higher in comparison to random mutagenesis.

Furthermore, we determined the union of $XI = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } T_{50}(x)\}$ that yield variants showing the respective highest $\Delta D_{\text{max}}$ toward at least one detergent ($XV = XI \cup XII \cup XIII \cup XIV$) (Tables 1 and S2, eq 3). Additionally, the union of $X$ and XV was defined as the set of 24 substitution sites (~13% out of 181 substitution sites), which exhibit variants showing the respective highest $\Delta T_{50;\text{max}}$ or $\Delta D_{\text{max}}$ toward at least one detergent ($XVI = X \cup XV$) (Tables 1 and S2, eq 3).

The intersection between $XI = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } T_{50}(x)\}$ and $XVII = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } D_{\text{SDS/CTAB/SB}3/16/\text{Tween 80}(x)\}$ is empty; i.e., there are no common substitution sites among those six at which variants with highest $\Delta T_{50;\text{max}}$ and $\Delta D_{\text{max}}$ for each detergent were found (Tables 1 and S2, eq 4). Thus, $XVII$ and $XVIII$ were not considered for the following analyses.

Additionally, we compared the pairwise similarities between $X = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } T_{50}(x)\}$ and $XVII = \{\text{Substitution sites}, \, 1 \leq x \leq 181, \text{six highest effects in significantly increased } D_{\text{SDS/CTAB/SB}3/16/\text{Tween 80}(x)\}$ by calculating $J$ (eq 5). Regarding the highest $\Delta D_{\text{max}}$ only XII and XIII overlap to some extent ($J(XII, XIII) \simeq 0.2$) (Table S6). Regarding the highest $\Delta T_{50;\text{max}}$ and $\Delta D_{\text{max}}$ only $X$ and XI, XII, or XIII, respectively, slightly overlap ($J(X, XI) \approx J(X, XII) \approx J(X, XIII) \simeq 0.1$) (Table S6).

To conclude, a highest $\Delta T_{50;\text{max}}$ of 7.7 K and a highest $\Delta D_{\text{max}}$ of 2.41 were found. The six substitution sites with highest $\Delta T_{50;\text{max}}$ yield a ~5.3-fold higher likelihood to find variants with significantly increased $T_{50}$ ($X$); the six substitution sites with highest $\Delta D_{\text{max}}$ yield a ~4.5-fold higher likelihood to find variants with significantly increased $D$ ($XI \cup XIV$). There are no common substitution sites among those six at which for each detergent variants with highest $\Delta D_{\text{max}}$ were found (XVII). Neither are there common substitution sites among those six at which variants with highest $\Delta T_{50;\text{max}}$ and $\Delta D_{\text{max}}$ for each detergent were found (XVIII).

3.6. Definition of Hot Spots. Based on these results, we defined seven types of hot spots, i.e., substitution sites particularly promising to cause a significant increase in $T_{50}$ or/and $D$. First, the respective six substitution sites of $X \cup XIV$ are considered hot spots because variants yield the respective highest $\Delta T_{50;\text{max}}$ or $\Delta D_{\text{max}}$ toward one detergent for these substitution sites (Tables 1, S2, and S9). Furthermore, we showed that there is a correlation between the magnitude of an effect found at a substitution site and the frequency of substitution occurrences that lead to significantly increased $T_{50}$ or $D$ toward one detergent (see section 3.2). Finally, generating and evaluating variants based on combinations of all 20 AAs at six substitution sites is still manageable with current protein-engineering techniques.

As shown above, XVII and XVIII, which would constitute the substitution sites with the broadest impact on $\Delta D_{\text{max}}$ or $\Delta T_{50;\text{max}}$ and $\Delta D_{\text{max}}$ are empty (see section 3.5). Hence, we resorted to defining, second, the 11 substitution sites of VIII showing significantly increased $D$ toward each detergent, regardless of the magnitude of the single effect (see section 3.3) and, third, the seven substitution sites of IX showing significantly increased $T_{50}$ and $D$ toward each detergent, regardless of the magnitude of the single effect (see section 3.4) as hot spots (Tables 1 and S2). With 11 and 7 substitution sites, these classes are also the smallest besides $X \cup XIV$.

3.7. Hot Spots Are Diverse in Terms of Localization in Secondary Structure Elements, Degree of Burial, and Sequence-Based Characteristics of the Substituted AAs. Ideally, one would identify such hot spots based on structural or sequence characteristics of the protein (see sections 2.4 and 2.5) prior to performing experiments. Suitable structure-based characteristics are localization in secondary structure elements (Table S11) and the degree of burial as measured by fSASAs (Table S12, eq 6).

As to localization in secondary structure elements (Table S11), hot spots are rarely found in $\beta$-helices and $\beta$-strands. Exceptions are hot spots of class XIV, which are enriched in strand $\beta$. With respect to $\alpha$-helices, at least one and at most four hot spot(s) of each class is (are) found in that secondary structure class, mainly in helices $\alpha B$ and $\alpha E$. However, without further information, one would not know which particular secondary structure element to choose for hot spot prediction. Hence, if all sites of a certain secondary structure class were chosen as hot spots, in the best case, a gain in precision (gip, eq 11) over random classification of 4.71 is found for $\beta$-strands, albeit at the expense of predicting 32 substitution sites (~18% of 181 AAs). Far more than the 6 sought. As to bridges, turns, loops, and bends defined by DSSP,38 no hot spot is found in the first secondary structure type. At most three hot spots are found in any of the other three types, but only for hot spots of class XI and VIII. These cases are related to a maximal gip of 1.93, albeit at the expense of predicting 47 substitution sites (~26% of 181 AAs). Thus, in our study, identifying hot spots based on this secondary structure type results in a low precision.
As to the degree of burial (Table S12), the least hot spots are associated with substitution sites that are mostly solvent-exposed (0.8 < fSASA ≤ 1.0). By contrast, the most hot spots are associated with substitution sites that are partially solvent-exposed (0.6 < fSASA ≤ 0.8), although this statement does not hold for hot spots of class XIV. This case is related to a maximal gip of 6.70, albeit at the expense of predicting 18 substitution sites (∼10% of 181 AAs).

Suitable sequence-based characteristics are physicochemical properties of the substituted AAs (Table S13) and the degree of AA conservation (Table S14). As to the physicochemical properties of the substituted AAs (Table S13), the distribution of hot spots over the classes is generally broad. Exceptions are hot spots of classes XIII and XIV (in both cases preferentially found at aliphatic and neutral AAs (Table S15)) and class X (preferentially found at aliphatic, aromatic, and neutral AAs (Table S15)). Therefore, gip values are generally low, with the largest one being 4.02 for the case of hot spots of class X at aromatic AAs, albeit at the expense of predicting 15 substitution sites (∼8% of 181 AAs). As to the degree of AA conservation, hot spots are located at nonconserved and semiconserved positions (conservation in the range of 0–6) (Table S14). The highest conservations were found for I128 (conservation = 6) and V99, T126, and I128 (conservation = 5).

To conclude, while predicting hot spots based on structural characteristics can lead to marked gip values, usually many predicted hot spots result, which would require considerable experimental efforts. Still, if a higher number of predicted hot spots is acceptable, partially solvent-exposed residues are good hot spot candidates. Applying sequence-based characteristics, substituting aliphatic and neutral residues should more likely improve T50 or/and D. Additionally, nonconserved and semiconserved regions preferentially contain hot spots.

3.8. Rigidity Theory-Based (CNA) and Markov Chain Monte Carlo Simulation-Based (ProFASi) Approaches Predict Similar Thermal Unfolding Pathways of wtBsLipA

We intend to test if hot spots can be predicted as structural weak spots by our rigidity theory-based approach CNA66–68 (see section 2.6). As a prerequisite, information on the hierarchy of rigid and flexible regions in a protein structure is required. Therefore, a thermal unfolding simulation of wtBsLipA was carried out with CNA as done previously to predict major rigidity elements in the protein. The obtained thermal unfolding pathway of wtBsLipA (PDB ID: 1ISP) is shown in Figure 3A. The early phase transitions (T1–T2) and late phase transitions (T3–T5) are indicated. The rigid clusters are represented as uniformly colored blue, green, magenta, and cyan bodies in the descending order of their sizes.

To predict hot spots, we used the CNA-based stability map (Figure 3B) and the neighbor stability map (Figure 3C). The CNA-based stability map shows the RCij values at which a rigid contact between two residues is lost for all residue pairs during the thermal unfolding simulation (upper triangle). The neighbor stability map RCij,neighbor considering only the rigid contacts between two residues that are at most 5 Å apart from each other, with values for all other residue pairs colored gray (lower triangle). The RCij values are calculated with CNA based on a structural ensemble (ENTMD). A red (blue) color indicates that contacts between residue pairs are more (less) rigid.

The aforementioned RCij,neighbor (lower triangle) was compared with a contact map simulated by ProFASi (upper triangle). A red (blue) color indicates contacts between residue pairs that have a longer (shorter) lifetime (in MC sweeps) than the contacts of the residue pairs of the initial protein structure. 3_10-helices are represented as G-helices.
phase transitions at which the network switches from overall rigid to flexible states (see sections 2.7, 2.8, and 2.9).

From the thermal unfolding pathway of wtBsLipA, five major phase transitions, T1−T5, were predicted based on the global index $H_{type2}$ (Figure 3A). Depending on the energy cutoff $E_{cut}$, the phase transitions were characterized as either early (T1−T2; with $-0.8 \text{kcal mol}^{-1} \leq E_{cut} \leq -0.9 \text{kcal mol}^{-1}$) or late (T3−T5; with $-1.7 \text{kcal mol}^{-1} \leq E_{cut} \leq -1.9 \text{kcal mol}^{-1}$). $E_{cut}$ can be converted to a temperature $T$ using a linear equation (eq 7), according to which the ranges of $E_{cut}$ in this study are equivalent to $316\leq T \leq 318\text{K}$ for T1−T2, and $334 \leq T \leq 338\text{K}$ for T3−T5. During the early phase transitions $\alpha A, 3_{\alpha}-1, \alpha F$, and $3_{\alpha}-5$ segregate from the largest rigid cluster. $\alpha D, \alpha E, \beta B, \beta C$, and $\beta$-strands segregate from the largest rigid cluster during the late phase transitions. Afterward, the $\beta$-sheet becomes sequentially flexible, beginning with $\beta 4$ and $\beta 8$, followed by $\beta 3, \beta 7, \beta 5$, and $\beta 6$. For the analysis, $\sim 3$ h of computational time on a single GPU is required to generate a 100 ns long MD trajectory as well as $\sim 4$ h of computational time on a single core for the thermal unfolding simulation.

Since the percolation behavior of a protein network is complex due to the protein’s structural hierarchy and composition of different modules, it is often challenging to assign a phase transition with $H_{type2}$. Thus, in addition to using $H_{type2}$, we also characterized the hierarchy of rigid and flexible regions of wtBsLipA at a local level by computing $r_{c_{i,j}}$neighbor (lower triangle in Figure 3B) based on $r_{c_{i}}$ upper triangle in Figure 3B). $r_{c_{i,j}}$neighbor showed that residue pairs at the N-terminus revealed higher $E_{cut}$ values than residue pairs at the C-terminus. Thus, $r_{c_{i,j}}$neighbor demonstrates that the rigid contacts between neighboring residues are stronger at the N-terminus than at the C-terminus along the thermal unfolding simulation, i.e., the C-terminus of wtBsLipA starts to unfold first.

As an independent approach to assess the order of unfolding of wtBsLipA, we used the Markov Chain Monte Carlo (MCMC) simulation software ProFASi (Protein Folding and Aggregation Simulator) (see section 2.11).$^{85}$ The results of the simulation were represented in a contact map (upper triangle in Figure 3C). They reveal that the contacts between the residue pairs of the N-terminus have a longer lifetime (in terms of MC sweeps) than the contacts of the residue pairs of the C-terminus compared to the initial structure. Thus, although methodologically different, ProFASi predicts a very similar unfolding pathway of wtBsLipA with respect to CNA.

To conclude, five major phase transitions, T1−T5, were predicted by thermal unfolding simulations using CNA at which first the different helices and, finally, the $\beta$-strands segregate from the largest rigid cluster during thermal unfolding simulations of wtBsLipA by CNA. Structural rigidity is initially lost at the C-terminus, which is uniformly revealed by the global index $H_{type2}$ and the local index $r_{c_{i,j}}$neighbor. Finally, the two independent approaches CNA and ProFASi predict very similar unfolding pathways of wtBsLipA. The results suggest that the loss of rigidity predicted by CNA along the thermal unfolding simulation closely mimics the temperature-induced unfolding of wtBsLipA.

3.9. Unfolding Nuclei and Major Phase Transitions Are Predictive Markers of Structural Weak Spots. We next probed to what extent structural weak spots predicted by CNA agree with the above-defined hot spots. Following previous work,$^2$ weak spots are fringe residues of unfolding nuclei that percolate from the largest rigid cluster during earlier steps of the thermal unfolding (see section 2.9). In total, we predicted 10 weak spots (~6% out of 181 substitution sites), i.e., I12, G13, G46, G52, P53, T66, M134, I135, V136, and H152 (Figure 4A, Tables 1, 2, and S2). Three weak spots each segregate from the largest rigid cluster at T1 or T2, and four from the largest rigid cluster at T4 (Table 2).

The performance of predicting hot spots as weak spots by CNA was evaluated in terms of a binary classification, considering predicted weak spots at hot spots true positives (TP) and predicted weak spots at not-hot spots false positives (FP) (see section 2.10). In particular, the gain in precision over random classification (gip) (eq 11) and the $F_{2}$-score ($F_{2}$) (eq 12), a measure of a classifier’s accuracy, were used as performance measures. Over all seven classes of hot spots, between one and three of the predicted weak spots are hot spots (except for XIV, where no weak spot was met), resulting in gip
Table 2. CNA-Predicted Weak Spots of BsLipA

<table>
<thead>
<tr>
<th>weak spot</th>
<th>location at secondary structure elements</th>
<th>phase transition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I12</td>
<td>turn</td>
<td>T1</td>
</tr>
<tr>
<td>G13</td>
<td>turn</td>
<td>T1</td>
</tr>
<tr>
<td>G46</td>
<td>loop</td>
<td>T4</td>
</tr>
<tr>
<td>G52</td>
<td>αB</td>
<td>T4</td>
</tr>
<tr>
<td>P53</td>
<td>αB</td>
<td>T4</td>
</tr>
<tr>
<td>T66</td>
<td>αB</td>
<td>T4</td>
</tr>
<tr>
<td>M134</td>
<td>bend</td>
<td>T2</td>
</tr>
<tr>
<td>I135</td>
<td>bend</td>
<td>T2</td>
</tr>
<tr>
<td>V136</td>
<td>bend</td>
<td>T2</td>
</tr>
<tr>
<td>H152</td>
<td>bend</td>
<td>T1</td>
</tr>
</tbody>
</table>

values between 3.02 and 9.05 (Tables 1 and S2). Note that these results are associated with only 10 predicted weak spots, about half as many predictions than in the case of identifying hot spots as partially solvent-exposed residues (Table S12). As the numbers of hot spots in VIII–XIV are of a very similar magnitude, the CNA predictions are also associated with similar recall (r) (eq 8) and precision (p) values (eq 9) in each case (Table S2), indicating a well-balanced classifier. In the case of XII, the CNA predictions yield an F1-score of 0.38, higher than any F1-score associated with hot spot predictions based on structure or sequence characteristics (Tables S2, S11, S12, S13, and S14), and the F1-score for IX is 0.24, generally higher than F1-scores associated with structure- or sequence-based predictions for this class and on par with the result obtained for identifying these hot spots as partially solvent-exposed residues (Tables S2, S11, S12, S13, and S14).

To conclude, predicting hot spots as weak spots by CNA results in several cases in very good to good gip values and good to fair accuracies and is associated with a very low number of predicted weak spots, such that also only few experimental efforts are required later. Considering the low computing time required to perform a CNA analysis, these results indicate that applying CNA-based weak spot prediction before experimental engineering is beneficial, in particular if the number of substitution sites that can be dealt with in experiment is low.

4. DISCUSSION

In this study, for the first time, we performed a systematic large-scale analysis of a complete experimental SSM library of a biotechnologically highly relevant protein, BsLipA,52,53 with respect to two types of protein stability. The library covers all 181 residues of BsLipA and results in 3439 variants, each with a single AA substitution as confirmed by DNA sequencing. Considering the screening results of the library toward thermostability and detergent tolerance together is unique compared to related studies8,4,17–19 and important in view of the challenges of multidimensional property optimization of modern biocatalysts.104–108 The measured T50 and D values provide valuable reference data for future analyses because, in contrast to other data sources,34–37 the different protein stabilities were measured under respectively uniform conditions, and there is no bias toward any particular substitution type or site. Note, though, that other factors than protein stability may influence T50 or D values measured here,32 including that substitutions can directly impact BsLipA function, e.g., when occurring in the vicinity of the active site.8 Moreover, the measured T50 and D values may be influenced by thermodynamic or kinetic factors.7,24 Therefore, in our analysis, we focused on scrutinizing the impact of substitution sites on thermo-stability or/and detergent tolerance to gain generally applicable rules for data-driven protein engineering. The following results stand out:

First, across the library, the likelihoods to find variants with significantly increased T50 (~12%) or D toward one detergent (~14%) are almost identical and small. The finding that the overwhelming number of single AA substitutions introduced by random mutagenesis causes a destabilizing effect is in agreement with previous studies.33,107–110 This finding becomes even more statistically relevant if multiple mutations need to be accumulated over generations to reach a desired effect because frequently a single, yet rather likely, destabilizing mutation is sufficient to annihilate the effect of several stabilizing ones.20 The proportions of variants with increased T50 or D found here are in line with the composition of databases such as ProTherm82 but markedly larger than the success rate of ~2% used as a reference to evaluate the performance of FoldX.111 Hence, beyond the single T50 and D data, due to the completeness of our library and the model character of our protein, our results also constitute unbiased reference data as to what efficiency can be expected for a protein system when optimizing thermostability or detergent tolerance by random mutagenesis. In turn, largest increases in T50 of 7.7 K and D of 2.4 found demonstrate that considerable improvements of protein stability can already be achieved by single AA substitutions. In that respect, previous studies on BsLipA applying either directed evolution44 or rational design7,8 already yielded close-to-optimal results in terms of increased thermostability.

Second, in the context of data-driven protein engineering, we identified substitution sites for which variants yield significantly increased T50 or/and D. Not considering the magnitude of the increase, only about one-third or below of all BsLipA residues constitute such favorable substitution sites if T50 and D are considered separately, demonstrating that the location of a residue within a protein structure matters with respect to a substitution effect. This result corroborates previous studies.5,7,8 In addition, our complete SSM library allowed us to reveal for such substitution sites a significant and fair correlation between the frequency of T50 or/and D-increasing substitutions and the magnitude of the maximum effect. Together, these results show that addressing all substitution sites in an unbiased manner by random mutagenesis results in a considerable experimental effort coupled to low efficiency. In turn, approaches that can identify substitution sites with a high likelihood for significantly increased T50 or D prior to doing experiments will be of great value in protein engineering studies.

Third, notably, the conclusions from the last paragraph also hold if more than one protein property is considered at a time. As such, we showed that at 11 substitution sites a ~4.6-fold higher likelihood to find for each detergent variants with significantly increased D compared to random mutagenesis is found. Additionally, seven substitution sites yield a ~3.4-fold higher likelihood to find significantly increased T50 and a ~4.7-fold higher likelihood to find for each detergent variants with significantly increased D compared to random mutagenesis. The latter finding suggests that approaches that can identify substitution sites with a high likelihood for significantly increased T50 should also be beneficial for identifying substitution sites with a high likelihood for significantly increased D, or vice versa. This is an important finding for practical applications as many more algorithms have been devised that address thermostability than detergent tolerance.
Fourth, as another set of reference data, we defined hot spot types together with the associated substitution sites to provide benchmark data for evaluating the performance of data-driven approaches. The first five classes follow the strict criterion that only the six substitution sites with the respective highest $\Delta T_{50\text{max}}$ or $\Delta D_{\text{max}}$ are considered, according to that all combinations of the 20 proteinogenic AAs at such sites could still be experimentally investigated. The intersections comprising the substitution sites with the broadest impact on $\Delta D_{\text{max}}$ or $\Delta T_{50\text{max}}$ and $\Delta D_{\text{max}}$ are empty. Thus, we resorted to defining two further classes with the somewhat relaxed criterion that the comprised substitution sites show significantly increased $D$ toward each detergent, or significantly increased $T_{50}$ and $D$ toward each detergent, regardless of the magnitude of the single effect.

Fifth, we used the complete, unbiased, and uniformly generated $T_{50}$ and $D$ data to probe if universal rules for modulating thermostability or detergent tolerance can be identified. We thereby focused on “one-dimensional” descriptors in terms of location in secondary structure elements, degree of burial, and physicochemical properties and conservation degree of substituted AA. Such descriptors have been widely analyzed before and play a role in data-driven consensus approaches. Analyzing “two- or higher dimensional” descriptors in terms of residue–residue interactions, entropic contributions or other collective phenomena, or cross-correlations of “one-dimensional” descriptors remains for future work. Notably, considering our descriptors, many (up to 98 substitution sites) predicted hot spots result, which would require considerable experimental efforts particularly if beneficial substitutions need to be accumulated to reach a desired effect. This finding demonstrates on a single protein level that, with these descriptors, no universal and sufficiently discriminating rule(s) can be identified, a finding that is mirrored in studies across protein families and with respect to low successes in assessing thermostabilities. Still, if a higher number of predicted hot spots is acceptable, partially solvent-exposed residues are good hot spot candidates. This result differs from previous experimental studies showing that especially surface remodeling emerged as an effective strategy to improve protein stability. Furthermore, loop positions, which have elsewhere been identified to preferentially carry favorable substitution sites, show mostly destabilizing effects. Finally, and likely surprisingly, hot spots were preferentially found at nonconserved and semiconserved position, a finding that may help refine future consensus concepts where multiple sequence alignments are used to substitute nonconsensus residues by consensus ones.

Sixth, we made use of the reference data to unequivocally benchmark our ensemble- and rigidity theory-based CNA approach with respect to predicting hot spots as structural weak spots of the protein. In contrast to previous studies on much smaller data sets, the present work allows to systematically assess the quality of our predictions. To do so, and in contrast to other assessments of protein stability predictors, we apply recall precision as basic statistical measures, rather than sensitivity and specificity, because we are interested in the accuracy of predicting hot spots and not not-hot spots, the latter of which furthermore clearly dominate the data set in terms of occurrence frequency. Methodologically, CNA differs from other state-of-the-art methods that do not consider ensemble representations of the protein. Furthermore, CNA does not require system-specific weighting or fitting parameters. This should make the results obtained here with CNA transferable to other protein systems. Weak spot prediction by CNA relies on a realistic modeling of the thermal unfolding of a protein. The predicted major phase transitions and the order of the segregating secondary structure elements are in agreement with previous computational studies and experimental observations on other proteins with an $\alpha/\beta$ hydrolyase fold. Furthermore, we confirmed the unfolding pathway of wtBslIP predicted by CNA with the independent MCMC-based ProFAsi approach. From a practical point of view, it is relevant that CNA predicted only 10 weak spots, allowing to focus subsequent substitution efforts on only $\sim$6% of the protein residues. Furthermore, the gain in precision over random classification is between $\sim$3 and $\sim$9, depending on the hot spot class. Considering the properties of the majority of predicted weak spots, i.e., a location in a loop, turn, or bend and a neutral or aliphatic amino acid type (Table 2), the notion may arise that these two properties, when correlated, characterize hot spots. The gain in precision over random classification is only between $\sim$0.7 and $\sim$2.1, however, depending on the hot spot class (Table S16), and, hence, more than fourfold lower than when hot spots are predicted as weak spots by CNA (Table 1). Together with the low computational demand on the order of hours only, these results lead to the strong recommendation to apply CNA-based weak spot prediction for data-driven protein engineering toward increased $T_{50}$ or $D$.

In summary, we provide systematic and unbiased reference data at large scale for thermostability measured as $T_{50}$ values and detergent tolerance measured as $D$ for a biotechnologically important protein, identified consistently defined hot spot types for evaluating the performance of data-driven protein-engineering approaches, and showed that CNA-based hot spot prediction can yield a gain in precision over random classification up to ninefold.

ASSOCIATED CONTENT

Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00954.

Tables S1–S16 and Figures S1–S2 as described in text; supplemental references (PDF)
$T_{50}$ values (XLSX)

AUTHOR INFORMATION

Corresponding Author
Holger Gohlke – Forschungszentrum Jülich GmbH, Jülich, Germany, and Heinrich Heine University Düsseldorf, Düsseldorf, Germany; orcid.org/0000-0001-8613-1447; Email: gohlke@uni-duesseldorf.de, h.gohlke@fz-juelich.de

Other Authors
Christina Nutschel – Forschungszentrum Jülich GmbH, Jülich, Germany
Alexander Fulton – Heinrich Heine University Düsseldorf, Jülich, Germany
Olav Zimmermann – Forschungszentrum Jülich GmbH, Jülich, Germany
Ulrich Schwaneberg – RWTH Aachen University, Aachen, Germany, and DWI-Leibniz-Institute for Interactive
Materials, Aachen, Germany;  orcid.org/0000-0003-4026-701X
Karl-Erich Jaeger – Heinrich Heine University Düsseldorf, Jülich, Germany, and Forschungszentrum Jülich GmbH, Jülich, Germany;  orcid.org/0000-0002-6036-0708

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.9b00954

Author Contributions
H.G., K.-E.J., and U.S. conceived the study. C.N. analyzed the data, performed MD simulations and CNA computations, analyzed the results, and wrote the manuscript together with H.G. A.F. performed experimental work. O.Z. performed ProFASi simulations and analyzed the results. H.G. supervised and managed the project. All authors reviewed and approved the manuscript.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors declare no competing financial interest.

REFERENCES


(132) Krause, D. JUWELS: Modular Tier-0/1 Supercomputer at the Jülich Supercomputing Centre. JLSRF 2019, 5, A135.