Enhancing subtilisin thermostability through a modified normalized B-factor analysis and loop-grafting strategy

Rational design–guided improvement of protein thermostability typically requires identification of residues or regions contributing to instability and introduction of mutations into these residues or regions. One popular method, B-FIT, utilizes B-factors to identify unstable residues or regions and combines them with other strategies, such as directed evolution. Here, we performed structure-based engineering to improve the thermostability of the subtilisin E-S7 (SES7) peptidase. The B-value of each residue was redefined in a normalized B-factor calculation, which was implemented with a refined bioinformatics analysis strategy to identify the critical area (loop 158–162) related to flexibility and to screen for suitable thermostable motif sequences in the Protein Data Bank that can act as transplant loops. In total, we analyzed 445 structures and identified 29 thermostable motifs as candidates. Using these motifs as a starting point, we performed iterative homologous modeling to obtain a desirable chimera loop and introduced five different mutations into this loop to construct thermostable SES7 proteins. Differential scanning fluorimetry revealed increases of 7.3°C in the melting temperature of an SES7 variant designated M5 compared with the WT. The X-ray crystallographic structure of this variant was resolved at 1.96 Å resolution. The crystal structure disclosed that M5 forms more hydrogen bonds than the WT protein, consistent with design and molecular dynamics simulation results. In summary, the modified B-FIT strategy reported here has yielded a subtilisin variant with improved thermostability and promising industrial applications, supporting the notion that this modified method is a powerful tool for protein engineering.

This work was supported by National Key Research and Development Program of China 2017YFB0308401 (to J. Z.), Program for Changjiang Scholars and Innovative Research Team in University Grant IRT_15R26, National First-class Discipline Program of Light Industry Technology and Engineering Grant LITE2018-08, National Institutes of Health Grant R35 GM118047 (to H. A.), and the China Scholarship Council fund. The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This article contains Tables S1–S6, Figs. S1–S2, and supporting files. The atomic coordinates and structure factors (code 6PAK) have been deposited in the Protein Data Bank (http://wwpdb.org/).

1 To whom correspondence may be addressed. E-mail: aihar001@umn.edu.
2 To whom correspondence may be addressed. E-mail: zhangj@jiangnan.edu.cn.
3 To whom correspondence may be addressed. E-mail: gcdu@jiangnan.edu.cn.

This is an open access article under the CC BY license.

Typically, two steps are required to enhance protein thermostability through rational design. First, unstable residues or regions must be identified. Second, candidate mutations must be introduced to these residues or regions. The consensus concept is the most accepted method of developing thermostable proteins (1, 2). This approach identifies stable residues or regions at a given position in an alignment of homologous thermostable proteins, which accomplishes both steps simultaneously, as both the target position and candidates are identified. Another popular method, B-FIT, utilizes B-factors to identify unstable residues or regions and combines them with other strategies, such as directed evolution, to enhance protein thermostability (3).

A previously described B-FIT method used B-factors to identify and interpret the dynamic characteristics of proteins. Several early studies suggested that high B-factors indicate higher flexibility (instability) and low B-factors indicate rigidity (stability) (4, 5). However, B-values from different structures cannot be reasonably compared without some normalization (6, 7), leading to the use of normalized B-factors in a broad range of applications, such as molecular dynamics (8), crystallography analysis (9), and protein engineering (10). In many previous studies, the mean B-factor is calculated based on the Cα B-value (11), which is thought to represent backbone motion.

Many methods have been considered for choosing candidate mutations to enhance protein thermostability. Homologous alignment indicates evolutionary trends but is always limited by the number of homologous proteins (12). Directed evolution often yields better results but requires a large quantity of mutation data, and the results have little reference value for other studies (13).

Subtilisin E (EC 3.4.21.62), which was assigned to the peptidase family S8 and has great potential for eliminating the milk allergen β-lactoglobulin (14), was selected as the starting point for rational design. Many studies have successfully improved protein stability in the subtilisin family using traditional methods to expand its industrial applications (15, 16). In this study, we developed a more straightforward rational design method (with decent sample space and better efficiency) to enhance the thermostability of subtilisin E. This method can easily be applied to other proteins as well.
Improved thermostability of subtilisin E-S7

Figure 1. Schematic illustration of the overall procedure used to perform loop transplant.

Results

New B-FIT strategy

We designed a new B-FIT strategy that combines modified normalized B-factor analysis with a loop grafting strategy to enhance protein thermostability (Fig. 1). The normalized B-factors in a new algorithm were applied to screen for unstable regions and thermostable motifs. The motif was chosen from all peptidase data in the PDB4 (17), which gave sufficient sample space and could be used to engineer other proteins. Stable natural motifs were then transplanted into the target protein to substitute for unstable loops. After determining a skeleton sequence, we adjusted other residues in the chosen motif to fit the host structure in an iterative homology modeling to obtain a satisfying chimera candidate. These desirable loops were experimentally verified, and the structure of the protein with the favorable experimental result was solved, providing additional information for mechanism analysis and future rational design.

Unstable region identification

In this study, B-values based on $C_n$ and those based on all atoms in the residue were briefly compared (Table 1). The B-values from all atoms in the selected residue were extracted using our modified PyMOL plugin “Average b” (Table S1), and the average value was calculated. Using the $C_n$ B-value, flexible small side-chain residues were identified, whereas long side-chain residues were missed. Because of this limitation, we improved the algorithm of normalized B-factors (see “Bioinformatics analysis”), which is vital for identifying flexible and rigid regions.

The modified plugin enabled calculation of the B-factor of each residue or loop. When we refined the SES7 crystal structure (PDB entry 6O44) in our previous work, we obtained two molecules of SES7 in slightly different conformations that were connected to each other (14). Furthermore, we calculated the average B-factor based on these two different symmetries, chain A and chain B. The B-factors of all residues are listed in Table S2. The mobile residues and loops with normalized B-factors ($B_i^\prime$) greater than 1.75 were chosen and are listed in Table 2. The N terminus, C terminus, four residues, and three loops were identified as having relatively high $B_i^\prime$ values. Considering that each unstable loop contained 4–6 high $B_i^\prime$ residues, we predicted that a loop transplant strategy could provide better results than saturated mutation at a single residue. Therefore, we focused on the three loops.

The SES7 crystal structure does not contain the propeptide, and thus we aligned our structure with a highly homologous subtilisin E complex structure (PDB entry 1SCJ). We used the propeptide in the structure to dock our mature SES7. The all-atom root mean square deviation was 0.567 Å (1946 atoms) after superposing these two crystal structures using the PyMOL align function. Loops 98–103 and 127–132, which are involved in substrate binding, were identified based on the docking results. And any change in these two loop can dramatically alter the protein structure. All flexible residues and loops screened by B-factor analysis are labeled in Fig. 2. Based on the B-factor, loop 158–162 is the most flexible loop in SES7 and does not participate in substrate catalysis. Meanwhile, this loop has a proper distance to the catalytic cave, because the average distance between Ca atoms of all residues in loop 158–162 and the nucleophile Ser-221 is 20.56 Å; thus, it was chosen as the target loop for replacement.

Stable motif identification

To identify suitable candidate motifs for transplantation, we used the terms “StructTitleQuery: struct.title.comparator = contains struct.title.value = peptidase and StructTitleQuery: struct.title.comparator = 'contains struct.title.value = proteasome” to search the PDB database in March 2017. A total of 445 peptidase structures from different papers were downloaded and analyzed. The number of PDB structures we analyzed in the end is much lower than the number of search hits due to four principles. First, some papers reported different-resolution PDB structures for the same protein, and we only chose one of them (e.g. 6HC6 and 6HC7). Second, some papers reported many complex PDB structures for the same protein with different ligands, and we only chose one of them (e.g. 6P52, 6P53, 6P54, 6P55, and 6P56). Third, some papers reported many single-residue mutant PDB structures for one protein, and we only chose one of them (e.g. 5YOH, 5YOI, 5YPD, and 5YPI). Those mutation sites normally were at the active site and did not affect the motifs we wished to collect. Fourth, some papers reported ligand peptide or one-domain PDB structures (<100 amino acids) (e.g. 6R9Z and 4HE5). We did not analyze these structures because they are too small to offer useful motifs. All PDB codes are listed in Table S3. Using the PyMOL script “Screen stable loops” (Table S4), we located stable loops efficiently.

The abbreviations used are: PDB, Protein Data Bank; RMSF, root mean square fluctuation; pNA, p-nitroanilide; SASA, solvent-accessible surface area; DSF, differential scanning fluorimetry.
and visually. Further calculation based on $B_i'$ was applied to these loops to verify their $B$-values. Only 29 loops with low $B_i'$ or those forming a desirable hydrogen bond from 22 different peptides were selected, which are listed in Table 3 along with the calculated $B$-values ($B_i$, $B_{(i)}$, and $B_{(i)}'$) of all residues in these 22 structures (see supporting files).

**Table 1**

| Sequence number | Small side-chain residue | Long side-chain residue |
|-----------------|--------------------------|-------------------------|
| 1               | 159                      | 160 27 141 265 59       |
| 2               | 232                      | 219 26 04 18 24         |
| 3               | 333                      | 34 35 36 37 38 39       |
| 4               | 444                      | 45 46 47 48 49 50       |

**Table 2**

Normalized $B$-factor of mobile residues and loops in SES7

| Sequence number | N-terminal | Residue 27 | Residue 59 |
|-----------------|------------|------------|------------|
| Sequence        | $B_i'$     |            |            |
| 1               | 2.29       | 1.93       | 3.23       |
| 2               | 2.89       | 2.42       | 2.63       |
| 3               | 3.02       | 3.71       | 3.23       |
| 4               | 3.56       | 5.15       | 3.32       |

**Figure 2. B-factor demonstration of the SES7 crystal structure.** All mobile residues and loops are labeled. Residues 27, 59, 141, and 265, as well as the N terminus and C terminus of SES7, are displayed as spheres. Loop 127–132 and loop 158–162 show higher mobility (thicker red coils) than other parts of the protein. Loop 98–103 and loop 127–132 hold the propeptide in the middle of the structure and are related to substrate binding.

**Mutant design and kinetic characterization**

Using SWISS-MODEL, we transplanted 29 candidate loops to the SES7 target position and built homologous modeled structures. Original and reverse sequences of each candidate loop were tested during homologous simulations, like original sequence SATCQ from 2AV9 and its reverse sequence QCTAS.
Improved thermostability of subtilisin E-S7

Table 3
Average and normalized B-factor of candidate motifs

| Organism(s)                              | PDB ID | Sequence | B_i  | B’  |
|------------------------------------------|--------|----------|------|-----|
| Thermoplasma acidophilum                 | 2A9V   | SATCQ    | 28.06| −0.61|
|                                           | 1W18   | RSETC    | 12.27| 0.33|
|                                           | 1WY2   | IPKIG    | 22.79| −0.41|
|                                           | 2D7J   | SETCP    | 12.22| −0.39|
| Escherichia coli (strain K12)            | 2QR5   | SPG MK   | 33.16| −0.73|
|                                           | 1KN9   | QQPQQ    | 40.10| −0.74|
|                                           | 1IZY   | DRKL     | 9.47 | −0.56|
| Homo sapiens                             | 2C1J   | LPDR     | 24.45| −0.48|
|                                           | 2ZED   | TPYG     | 26.53| −0.31|
|                                           | 3RBU   | LPDR     | 24.45| −0.40|
| Bacillus subtilis (strain 168)           | 1YHE   | TDRLGS   | 6.49 | −0.47|
|                                           |        | QEEVGLR  | 5.89 | −0.56|
|                                           |        | TANG     | 7.42 | −0.33|
|                                           |        | TKKG     | 13.69| 0.61|
|                                           |        | REDGY    | 26.17| −1.11|
|                                           |        | GIDGGS   | 12.47| −0.20|
|                                           |        | DVGF DKA | 12.40| −0.54|
|                                           |        | AEELGL   | 10.38| −0.11|
|                                           |        | GLND E   | 25.95| −0.18|
|                                           |        | RNFGLQ   | 24.90| −0.27|
|                                           |        | HPNY     | 22.92| −0.26|
|                                           |        | ITDYT    | 17.10| −0.62|

From our simulation analysis, the reverse sequences QCTAS from Thermoplasma acidophilum 2A9V and the original sequence PCTES from Pyrococcus horikoshii 2D7J worked well. Whereas WT sequence SSGSS forms no hydrogen bond outside the loop, Gln-158 in QCTAS sequences could form a hydrogen bond with Tyr-262, and Thr-160 in PCTES sequences could form a hydrogen bond with Tyr-262, which could stabilize the loop region. A few results indicated that some residues in certain positions could help the loop form extra hydrogen bonds. Based on these results, we kept these key residues in their original positions and tried to replace other residues in the loop. The key residues that could help the loop form extra hydrogen bonds in QQPQQ from 1KN9 were extracted to build a skeleton structure, QXPGQX. A similar skeleton structure, XXPXQX, was built based on VLPHQR from 1YLO. Another skeleton structure, XXPXXG, was built based on IPKIG from 1WY2. Based on all of the first-round simulation results, we built two combinations of skeleton structures, including either QXPGQX or PXPXGX as starting points for next-round simulation. Some residues at the X positions in the motifs were adjusted to charged residues to fit the SES7 structure in an iterative simulation, thus rendering the chimera loops. Charged residues, like His, Lys, Arg, Asp, and Glu, were used here because their side chains are more likely to form hydrogen bonds with other surface residues, and their hydrophilic properties are also helpful for the stability of the surface loop. Combined with the analysis described above, we identified two suitable loops that were ideal for transplantation, QRPQGD and QHPKEG, which were compatible with the local structure of the loop being replaced in SES7. All of the selected homologous modeled structures were then used as templates to build MDS structures. The root mean square fluctuations (RMSFs) were generated to evaluate the transplantation effect to the structure. The mutants that showed lower overall RMSF than that of WT were chosen for experimental verification.

The enzyme activities of the WT SES7 and all mutants were characterized by kinetic analysis of succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (AAPF-pNA) hydrolysis. A Lineweaver–Burk plot of SES7 and its mutants is shown in Fig. S1. All the kinetic parameters were measured based on the Lineweaver–Burk plot. As shown in Table 4, the $k_{cat}/K_m$ values of four transplant mutants (M1–M4) decreased to 61.7–88.2% of that obtained for WT SES7, whereas the fifth mutant (M5) retained a $k_{cat}/K_m$ value similar to that of WT, indicating that only the M5 loop transplant enhanced thermostability without compromising activity.

Thermostability analysis

SES7 and its five mutants were tested for thermostability. The residual activity after treatment at 55 °C was tested with an AAPF-pNA activity assay. Details are shown in Table S5. M5 is the best mutant, with the $T_{50}$ reaching 75.56 min. As shown in Fig. S2, the M1 and M2 showed decreased thermostability, whereas M3, M4, and M5 showed improved thermostability. Among the five mutants containing the transplant loop 158–162, M5 maintained the closest enzymatic activity to WT SES7. As shown in the differential scanning fluorimetry (DSF) results presented in Fig. 3, WT SES7 had relatively poor thermostability, with a melting temperature of 55.0 ± 0.0 °C, whereas that of M5 was 62.3 ± 0.2 °C. From the CD thermal denaturation experiment, we found that the melting temperature of M5 was 4.86 °C higher than that of the WT protein. Table 5 shows juxtaposed melting temperatures of SES7 and M5 obtained by this two methods. Therefore, transplantation of the QHPKEG sequence (the optimal sequence produced by the simulation) to loop 158–162 of SES7 greatly improved the thermostability of

J. Biol. Chem. (2019) 294(48) 18398–18407
the protein, indicating that this loop plays an essential role in enzyme stability.

It is worth notice that experimental variability, like different scanning rates, buffering chemicals, ionic agents, and pH range, can result in diverse melting temperature measurements (18, 19). In some cases, $T_m$ values could cover a wide range of temperatures from 39 to 48 °C (20). However, the melting temperatures measured under the same conditions by the same method are comparable.

**Structure analysis**

Far-UV CD spectra were recorded from 200 to 260 nm for WT SES7 and the M5 mutant to characterize their secondary structures. As shown in the CD curves in Fig. 4, each protein contained a similar degree of $\alpha$-helical (31–29%) and $\beta$-sheet (12–11%) structure, indicating that the loop transplant of M5 did not significant change the secondary structure of SES7.

MDS was performed to interpret the molecular stabilization mechanism. RMSF could reflect the stability of individual residue of protein, and a low RMSF denoted the rigidity of a given residue. The RMSF values of SES7 and M5 are shown in Fig. 5. The majority of M5 showed decreased RMSF values, and the transplanted region shows similar RMSF values compared with WT.

The M5 mutant protein was crystallized, and its crystal structure 6PAK was solved. After comparing loop SSGSS in 6O44 (WT) and the chimera loop QHPKEG in 6PAK, we found that the $B_J$ values were similar (Table 6), also indicating that the loop QHPKEG was not rigid.

---

**Table 4**

Kinetic parameters of SES7 wild type and mutants

Uncertainties are denoted as the average ± S.D.

|        | WT   | M1   | M2   | M3   | M4   | M5   |
|--------|------|------|------|------|------|------|
| Loop   | 158–162 | SSGSS | PCTES | QCTAS | QQPQQ | QRPGQ | QHPKEG |
| $K_m$ (μM) | 23.40 ± 0.46 | 25.69 ± 1.39 | 20.01 ± 0.67 | 22.26 ± 0.05 | 22.85 ± 0.29 | 49.09 ± 1.71 |
| $k_{cat}$ (min$^{-1}$) | 0.80 ± 0.04 | 0.54 ± 0.02 | 0.53 ± 0.04 | 0.52 ± 0.01 | 0.68 ± 0.03 | 1.76 ± 0.10 |
| $k_{cat}/K_m$ (min$^{-1}$ μM$^{-1}$) | 0.034 | 0.021 | 0.026 | 0.023 | 0.030 | 0.036 |

---

**Table 5**

Comparison of experimentally obtained melting temperatures of SES7 and M5 using DSF and CD

|        | SES7   | M5   |
|--------|--------|------|
| DSF (°C) | 55.0 ± 0.0 | 62.3 ± 0.2 |
| CD (°C)   | 53.31 | 58.17 |

---

**Figure 3.** Melt curves from differential scanning fluorimetry for SES7 and M5. A, the plots show the fluorescence response (relative fluorescence units; RFU). B, the first derivative of the melt curve ($-d$(RFU)/dT), SES7 (red curve), and M5 (blue curve). C, the melting curve monitored by far-UV CD, where the J-815 CD spectropolarimeter was set to record three times for each degree and generate the average results, which do not give a margin of error.

**Figure 4.** CD spectra of SES7 WT and M5. The M5 mutant protein was crystallized, and its crystal structure 6PAK was solved. After comparing loop SSGSS in 6O44 (WT) and the chimera loop QHPKEG in 6PAK, we found that the $B_J$ values were similar (Table 6), also indicating that the loop QHPKEG was not rigid.
The solvent-accessible surface area (SASA) of the two symmetries and whole molecule were calculated using the modified PyMOL script “Buried surface area” (Table 7). The SASA of M5 was more prominent than in the WT because of the longer loop QHPKEG. The buried surface area was also generated, which revealed that M5 has a more significant buried surface area, indicating that M5 forms a more compact dimer than the WT protein.

Fig. 6 shows a protein surface on which the surface residues in the WT protein were changed to residues in the transplant loop. These changes made the two symmetries, chain A and chain B, more compact (Fig. 6, A and B). Examination of the two crystal structures in detail revealed that QHPKEG in M5 molecule B (Fig. 6D) formed an intramolecular hydrogen bond involving Ser-189 and QHPKEG in SES7 (Fig. 6C) did not interact with any other residue. This indicates that the sequence QHPKEG is more likely to form hydrogen bonds and that these interactions enhance the thermostability of the M5 protein. The sequence QHPKEG has a somewhat β-strand–like (extended) polypeptide conformation that may contribute to the slight difference observed in the CD curve shown in Fig. 4. The details for all two-dimensional loop residue interactions generated in LigPlot+ are shown in Fig. 6 (E and F).

Discussion

A new B-FIT strategy combining modified normalized B-factor analysis with a loop-grafting strategy was established to enhance the thermostability of SES7 protein. We redefined the mean B-factor for normalized B-factor analysis, which allows us to locate unstable and replaceable regions more accurately and screen stable motifs more efficiently. A loop transplant strategy was used to obtain candidate loops from different domains of other peptidases, providing a larger evolutionary scale.

Many studies have verified that a mutation on the catalytic pocket could dramatically affect the enzyme activity (21, 22). It is unwise to pick a target in the catalytic pocket if we only want to enhance the thermostability. On the other hand, the purpose of thermostability improvement is to help the catalytic pocket maintain its functional conformation at higher temperatures; thus, the mutation target should not be too far from the catalytic pocket. In the end, loop 158–162 SSGSS was chosen as the target from the three most unstable loop regions (with high normalized B-factor) in SES7 because it does not belong to but is not far (~20 Å) from the catalytic pocket.

All thermostable motifs screened from the database may have been evolving for 10 thousand years to form their stable conformations. Choosing these motifs as the starting point for iterative simulation may reveal a proper distance for each mutation and is more likely to yield a desirable result. The RMSF analysis was used as a filter, and only those results with lower RMSF than WT were picked out for experimental verification.

The WT loop sequence SSGSS is simple (serine-rich), whereas the best transplanted sequence, QHPKEG, has a greater charge and residues with longer side chains. These types of sequences are more likely to form hydrogen bonds with other residues, which make the whole structure more stable by contributing to the overall negative enthalpy (ΔH). On the other hand, our X-ray structure suggested that the loop retained considerable flexibility, which could be important in alleviating the entropic penalty of protein folding. That should be on account of the exposed loop position and the longer loop length. Loop 158–162 is on the surface of the protein and hardly forms a connection with other residues (Fig. 6C). Moreover, the sequence QHPKEG is longer than the WT sequence SSGSS, which makes the last three residues, KEG, exposed on the surface (Fig. 6D). However, this sacrifice is not in vain. As shown in Fig. 6D, Gln-158 and His-159 were lifted to a closer position to another surface area and even formed hydrogen bonds with surface residues. Also, it has been reported that proline could decrease the conformational freedom of Cα–N rotation (23, 24), so the introduction of proline in the middle of the loop 158–162 stabilizes the connection between loop residues and other surface residues. Although the value of QHPKEG in 6PAK is similar to that of SSGSS in 6O44, the conformation change in this loop enhances the overall structure thermostability.

**Experimental procedures**

**Materials**

*Escherichia coli* DH5α competent cells were obtained from New England Biolabs (Ipswich, MA), and *E. coli* BL21(DE3)
Competent cells were purchased from Lucigen (Middleton, WI). Vector pET-24a(/H11001) was obtained from Novagen (Madison, WI). Pfu Turbo polymerase was purchased from Agilent Technologies (Santa Clara, CA), and DpnI was purchased from New England Biolabs. AAPF-pNA was purchased from Bachem (Bubendorf, Switzerland). All other chemicals were obtained from standard commercial sources and were of the highest grade available.

**Bioinformatics analysis**

The B-factor was calculated with a modified PyMOL Plugin “Average b” (Table S1). For further calculation, the following quantities were computed: \( \langle B \rangle \) is the average B-factor of all residues in the protein; \( B_i \) is the average B-value of not only C\(_\alpha\) atoms, but all atoms in the \( i \)th residue; \( N \) represents the total number of residues; \( \sigma(B) \) represents the standard deviation; and \( B'_i \) represents the normalized B-factor of selected residues or loops.

\[
\langle B \rangle = \frac{\sum B_i}{N} \quad \text{(Eq. 1)}
\]

\[
\sigma^2(B) = \frac{\sum (B_i - \langle B \rangle)^2}{N} \quad \text{(Eq. 2)}
\]

\[
B'_i = \frac{B_i - \langle B \rangle}{\sigma(B)} \quad \text{(Eq. 3)}
\]

After normalization, the contribution of each residue to the stability was observed through visualization. Values over 0 indicate that the residue makes the whole structure unstable. The unstable loops in SES7 were identified in this manner. In contrast, \( B'_i \) values below 0 mean the residue contributes to the stability of the structure. Stable motifs were obtained from the PDB with the help of a script, “Screen stable loops” (Table S4), and then a detailed normalized B-factor was calculated based on “Average b.”

A bioinformatics analysis strategy was designed (Fig. 1). Stable motifs were transplanted into unstable loops. The mutants’ structures were constructed by homology modeling with SWISS-MODEL (25). Molecular dynamics simulations were then conducted using the AMBER16 (26) implementing ff14SB field force (27). The models solvated inside a box containing water molecules, with a side length chosen to ensure a 15-Å minimum distance between any protein atom and the box boundary. In total, 150 mM NaCl was used to balance the negative charges. The process of MDS included water equilibration, side-chain equilibration, minimization, 50 ps of heating from 0 K up to 300 K, and 30-ns dynamics simulations. RMSF values reflecting the variation of individual residues at 300 K were calculated after MDS. The structures were displayed and analyzed by the PyMOL Molecular Graphics System, version 1.6.x (Schrödinger, LLC). Moreover, the key residues in motifs that could form hydrogen bonds are identified as skeleton structures. These skeleton structures were then set as starting points for the iterative simulations. Normally, long side chain residues were used randomly to replace other residues in the skeleton structure. Any change that could help the loop area gain more hydrogen bonds between the loop and other residues
remained as new starting points for the next round of simulation. The mutants that could form a satisfactory structure were subjected to further experimental verification.

The modified PyMOL script “Buried surface area” (Table S6) was applied to calculate the SASA of SES7 and M5. Water and other small molecules were deleted before analysis. The buried surface area was defined as follows: ((chain A SASA + chain B SASA) – whole molecular SASA)/2.

**Mutagenesis**

The WT subtilisin E-S7 (SES7) gene was from *Bacillus subtilis* (JN-S7) and cloned into pET-24a (+) and used as the template for loop-grafting mutagenesis. All primers used in this study are listed in Table 8. The entire plasmid was produced and amplified by non-strand-displacing synthesis from complementary mutagenic primers. The remaining nonmethylated DNA was transformed into *E. coli* DH5α competent cells after digestion by DpnI at 37 °C overnight. Sanger DNA sequencing was performed by Sequetech Corp. (Mountain View, CA) to verify the correct clones. The resulting plasmid was then transformed into *E. coli* BL21 (DE3) competent cells for enzyme expression.

**Expression and purification**

The transformed *E. coli* cells were precultured overnight and then inoculated into fresh lysogeny broth medium. The precultured cells (1%) were grown at 37 °C in lysogeny broth medium supplemented with kanamycin (40 mg/liter) until the A600 reached 1.0, after which the expression of SES7 was induced by the addition of isopropyl β-D-1-thiogalactopyranoside to a final concentration of 50 μM in the medium. The induced bacteria were further incubated overnight at 18 °C, pelleted by centrifugation, resuspended in Buffer A (200 mM Tris-HCl (pH 7.4), 500 mM NaCl, 5 mM β-mercaptoethanol), and stored in a −20 °C freezer. To purify SES7, we disrupted the cells by sonication with a Branson Sonifier 450. The lysate was centrifuged at 48,384 × g (4 °C) for 1 h (Beckman Avanti J-25 I, JA-25.50, Brea, CA). Following centrifugation, the supernatant was applied to a HisPurTM nickel-nitrilotriacetic acid resin (GE Healthcare, Little Chalfont, UK) with 30 mM Tris-HCl (pH 9.0), 40% (v/v) DMSO, and 1 mM CaCl2) at 37 °C for 1 h. The reaction was stopped by adding 10 μl of 10 mM phenylmethylsulfonyl fluoride. The activity was determined by monitoring the release of pNA through changes in absorption at 410 nm using a SparkTM 10M multimode microplate reader (Tecan, Männedorf, Switzerland). SES7 consumed less than 10% of the substrate under these conditions. The

### Table 8

| Mutant | Positions | Primers |
|--------|-----------|---------|
| M1     | Loop 158–162 | Forward: GAAACGaAGGTCCCGTGCAACCAGAACATGCGGCTACCTGC |
|        |            | Reverse: CAAGCTGTCGTTTTCCGTCGGGACGACGACAGTCGGGCTACCTGC |
| M2     | Loop 158–162 | Forward: GAAACGaAGGTCCCGTGCAACCAGAACATGCGGCTACCTGC |
|        |            | Reverse: CAAGCTGTCGTTTTCCGTCGGGACGACGACAGTCGGGCTACCTGC |
| M3     | Loop 158–162 | Forward: GAAACGaAGGTCCCGTGCAACCAGAACATGCGGCTACCTGC |
|        |            | Reverse: CAAGCTGTCGTTTTCCGTCGGGACGACGACAGTCGGGCTACCTGC |
| M4     | Loop 158–162 | Forward: GAAACGaAGGTCCCGTGCAACCAGAACATGCGGCTACCTGC |
|        |            | Reverse: CAAGCTGTCGTTTTCCGTCGGGACGACGACAGTCGGGCTACCTGC |
| M5     | Loop 158–162 | Forward: GAAACGaAGGTCCCGTGCAACCAGAACATGCGGCTACCTGC |
|        |            | Reverse: CAAGCTGTCGTTTTCCGTCGGGACGACGACAGTCGGGCTACCTGC |

**Circular dichroism**

CD measurements were performed with a J-815 CD spectropolarimeter (Jasco, Tokyo, Japan) at 25 °C. The enzyme sample (0.5 mg/ml) was prepared in 10 mM phosphate buffer (pH 7.0) and loaded into a quartz cuvette with a path length of 0.1 cm. The far-UV CD spectra were recorded from 200 to 260 nm with a 1-nm step resolution and scanning speed of 50 nm/min. The spectra from the buffer blanks were subtracted, and the secondary structure composition was analyzed with the K2D Programme on the online server DichroWeb.

**Thermal shift assay**

First, the melting temperatures of SES7 and M5 were determined by DSF. Aliquots of each mixture (40 μl) in final concentrations of 1.6% DMSO, 80x SYPRO orange dye (Life Technologies, Inc.), and 0.5 mg/ml SES7 were pre-incubated at 4 °C for 20 min. The DSF protocol consisted of an initial temperature of 20 °C, after which the temperature was increased to 95 °C at a rate of 1 °C/s on a C1000 Touch Thermal Cycler (Bio-Rad). Fluorescence was measured using the FRET channel of a CFX96TM real-time system (Bio-Rad), and data were collected in 0.5 °C increments. Melting curves were analyzed using Bio-Rad CFX Manager software.

Thermal denaturation was also monitored by far-UV CD. Each 200-μl enzyme sample (0.5 mg/ml) was heated from 20 to 90 °C. The far-UV CD spectra were recorded at 222 nm. The step size of the temperature ramp was 1 °C. To determine the melting temperature values, we fit the data to a melting curve using GraphPad software (GraphPad, Inc., La Jolla, CA).

**Kinetics characterization**

The model substrate AAPF-pNA was used to evaluate the Michaelis–Menten kinetics of SES7 and its mutant. Samples containing 100 μl of 0.5 mg/ml peptidase and 900 μl of the AAPF-pNA substrate at 10 different concentrations (0.5–5 mg/ml) were incubated in reaction buffer (100 mM Tris-HCl (pH 9.0), 40% (v/v) DMSO, and 1 mM CaCl2) at 37 °C for 1 h. The reaction was stopped by adding 10 μl of 10 mM phenylmethylsulfonyl fluoride. The activity was determined by monitoring the release of pNA through changes in absorption at 410 nm using a SparkTM 10M multimode microplate reader (Tecan, Männedorf, Switzerland). SES7 consumed less than 10% of the substrate under these conditions. The
Improved thermostability of subtilisin E-S7

Table 9
M5 data collection and refinement statistics
Statistics for the highest-resolution shell are shown in parentheses.

| Data collection                              |                   |                   |
|----------------------------------------------|-------------------|-------------------|
| Resolution range                             | 58.91–1.90 (1.96–1.90) |
| Space group                                  | P2₁,2₁,2₁         |
| Unit cell                                     | 74,045, 80,282, 86,713 |
| Total reflections                             | 244,062 (23,845)  |
| Unique reflections                            | 41,625 (3912)     |
| Multiplicity                                 | 5.9 (5.8)         |
| Completeness (%)                             | 98.94 (95.00)     |
| Mean I/σ(I)                                  | 5.93 (0.99)       |
| Rmerge                                       | 0.3124 (1.546)    |
| Rfree                                        | 0.3425 (1.702)    |
| Rwork                                        | 0.1835 (0.7029)   |
| CC(100)                                      | 0.986 (0.328)     |

| Refinement                                   |                   |                   |
|----------------------------------------------|-------------------|-------------------|
| Reflections                                  | 41,292 (3911)     |
| No. for Rmerge                               | 2037 (210)        |
| Rmerge (%)                                   | 23.88 (35.72)     |
| Rfree (%)                                    | 28.26 (38.08)     |
| No. of non-hydrogen atoms                    | 4243              |
| Macromolecules                               | 3948              |
| Ligands                                      | 30                |
| Solvent                                      | 265               |
| Protein residues                             | 554               |
| Root mean square deviation (Å)               | 0.002             |
| Bond angles (degrees)                        | 0.51              |
| Ramachandran plot                            |                   |                   |
| Favor (%)                                    | 97.24             |
| Allowed (%)                                  | 2.76              |
| Outliers (%)                                 | 0.00              |
| Average B-factor                             | 32.63             |
| Macromolecules                               | 31.98             |
| Ligands                                      | 72.83             |
| Solvent                                      | 37.79             |

kinetic constants were evaluated by fitting the experimental data in a Lineweaver–Burk plot through linear regression.

Crystalization condition screening

The purified SES7-M5 protein was concentrated to ~30 mg/ml. Initial crystallization screening was conducted using the CrystalMation system (Rigaku, Carlsbad, CA). In a typical robotic screening, 0.2 μl of the protein solution was mixed with 0.2 μl of reservoir solution, and the drop was incubated at 20 °C. Plate-shaped crystals grew under several conditions, with the best crystal growth observed in the solution of 0.2 M sodium bromide, 20% (w/v) PEG 3350.

Diffraction data collection and structure model building

All data sets used for model building and refinement were collected at the Advanced Photon Source at Argonne National Laboratory on beamlines 24-ID-C and 24-ID-E at 100 K using the X-ray wavelength of 0.979 Å. The data were indexed, integrated, and scaled by the Rapid Automated Processing of Data beamline software (XDS (28) and CCP4 (29)) or in HKL2000 (30). The PHENIX program suite was utilized for initial molecular replacement phasing, model building, and refinement. Manual model building and further refinement were performed in Coot (31) and PHENIX (32) Refine. The summary of data collection and model refinement statistics for M5 is shown in Table 9.

Author contributions—H. T. and G. D. conceptualization; H. T. and K. S. data curation; H. T., K. S., and C. S. software; H. T. formal analysis; H. T. investigation; H. T. visualization; H. T. methodology; H. T. writing—original draft; H. T., H. A., and G. D. writing—review and editing; H. A. and J. Z. supervision; H. A., J. Z., and G. D. funding acquisition.

References
1. Lehmann, M., Pasamontes, L., Lassen, S. F., and Wyss, M. (2000) The consensus concept for thermostability engineering of proteins. Biochim. Biophys. Acta 1543, 408–415 CrossRef Medline
2. Lehmann, M., Loch, C., Hindendorf, A., Studer, D., Lassen, S. F., Pasamontes, L., van Loon, A. P., and Wyss, M. (2002) The consensus concept for thermostability engineering of proteins: further proof of concept. Protein Eng. 15, 403–411 CrossRef Medline
3. Reetz, M. T., Soni, P., Fernández, L., Gumulya, Y., and Carballera, J. D. (2010) Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. Chem. Commun. 46, 8657–8658 CrossRef Medline
4. Karplus, P., and Schulz, G. (1985) Prediction of chain flexibility in proteins. Naturwissenschaften 72, 212–213 CrossRef
5. Radiovojac, P., Obradovic, Z., Smith, D. K., Zhu, G., Vucetic, S., Brown, C. J., Lawson, J. D., and Dunker, A. K. (2004) Protein flexibility and intrinsic disorder. Protein Sci. 13, 71–80 CrossRef Medline
6. Smith, D. K., Radiovojac, P., Obradovic, Z., Dunker, A. K., and Zhu, G. (2003) Improved amino acid flexibility parameters. Protein Sci. 12, 1060–1072 CrossRef Medline
7. Tronrud, D. E. (1996) Knowledge-based B-factor restraints for the refinement of proteins. J. Appl. Crystallogr. 29, 100–104 CrossRef
8. Mou, Y., Huang, P.-S., Thomas, L. M., and Mayo, S. L. (2015) Using molecular dynamics simulations as an aid in the prediction of domain swapping of computationally designed protein variants. J. Mol. Biol. 427, 2697–2706 CrossRef Medline
9. Reetz, M. T., Carballeira, J. D., and Vogel, A. (2006) Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. Angew. Chem. Int. Ed. Engl. 45, 7745–7751 CrossRef Medline
10. Korkegian, A., Black, M. E., Baker, D., and Stoddard, B. L. (2005) Computational thermostabilization of an enzyme. Science 308, 857–860 CrossRef Medline
11. Sun, Z., Liu, Q., Qu, G., Feng, Y., and Reetz, M. T. (2019) Utility of B-factors in protein science: interpreting rigidity, flexibility, and internal motion and engineering thermostability. Chem. Rev. 119, 1626–1665 CrossRef Medline
12. Lehmann, M., and Wyss, M. (2001) Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Curr. Opin. Biotechnol. 12, 371–375 CrossRef Medline
13. Arnold, F. H., Wintrobe, P. L., Miyazaki, K., and Gershenson, A. (2001) How enzymes adapt: lessons from directed evolution. Trends Biochem. Sci. 26, 100–106 CrossRef Medline
14. Tang, H., Zhang, J., Shi, K., Aihara, H., and Du, G. (2019) Insight into subtilisin E-S7 cleavage pattern based on crystal structure and hydrolysis peptide analysis. Biochem. Biophys. Res. Commun. 512, 623–628 CrossRef Medline
15. Estell, D. A., Graycar, T. P., and Wells, J. A. (1985) Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. J. Biol. Chem. 260, 6518–6521 Medline
16. Wells, J. A., and Powers, D. (1986) In vivo formation and stability of engineered disulfide bonds in subtilisin. J. Biol. Chem. 261, 6564–6570 Medline

Acknowledgments—This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by National Institutes of Health (NIH) Grant GM124165. The Pilatus 6M detector on the 24-ID-C beamline is funded by NIH Office of Research Infrastructure Program (ORIP) High-End Instrumentation (HEI) Grant S10 RR029205. This research used resources of the Advanced Photon Source, a United States Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract DE-AC02-06CH11357. We thank Editage for English language editing.
17. Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., and Abola, E. E. (1998) Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 1078–1084 CrossRef Medline

18. Seabrook, S. A., and Newman, J. (2013) High-throughput thermal scanning for protein stability: making a good technique more robust. *ACS Comb. Sci.* **15**, 387–392 CrossRef Medline

19. Uniewicz, K. A., Ori, A., Xu, R., Ahmed, Y., Wilkinson, M. C., Fernig, D. G., and Yates, E. A. (2010) Differential scanning fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans: a screening test for binding specificity. *Anal. Chem.* **82**, 3796–3802 CrossRef Medline

20. Culajay, J. F., Blaber, S. I., Khurana, A., and Blaber, M. (2000) Thermodynamic characterization of mutants of human fibroblast growth factor 1 with an increased physiological half-life. *Biochemistry* **39**, 7153–7158 CrossRef Medline

21. Gora, A., Brezovsky, J., and Damborsky, J. (2013) Gates of enzymes. *Chem. Rev.* **113**, 5871–5923 CrossRef Medline

22. Muller, B. H., Lamoure, C., Le Du, M. H., Cattolico, L., Lajeunesse, E., Lemaitre, F., Pearson, A., Ducancel, F., Ménez, A., and Boullain, J. C. (2001) Improving *Escherichia coli* alkaline phosphatase efficacy by additional mutations inside and outside the catalytic pocket. *ChemBioChem* **2**, 517–523 CrossRef Medline

23. Boone, C. D., Rasi, V., Tu, C., and McKenna, R. (2015) Structural and catalytic effects of proline substitution and surface loop deletion in the extended active site of human carbonic anhydrase II. *FEBS J.* **282**, 1445–1457 CrossRef Medline

24. Yu, H., Zhao, Y., Guo, C., Gan, Y., and Huang, H. (2015) The role of proline substitutions within flexible regions on thermostability of luciferase. *Biochim. Biophys. Acta* **1854**, 65–72 CrossRef Medline

25. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 CrossRef Medline

26. Case, D. A., Betz, R., Cerutti, D., Cheatham, T., 3rd, Darden, T., Duke, R., Giese, T., Gohlke, H., Goetz, A., and Homeyer, N. (2016) AMBER 2016 Reference Manual, University of California, San Francisco, CA

27. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 CrossRef Medline

28. Kabsch, W. (2010) XDS. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 CrossRef Medline

29. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242 CrossRef Medline

30. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 CrossRef Medline

31. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 CrossRef Medline

32. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 CrossRef Medline

---

**Improved thermostability of subtilisin E-S7**

---

J. Biol. Chem. (2019) 294(48) 18398–18407

18407