Plasmid-Based One-Pot Saturation Mutagenesis and Robot-Based Automated Screening for Protein Engineering

Fumihiro Kawai,† Akihiko Nakamura,‡ Akasit Visootsat,† and Ryota Iino*†‡

†Institute for Molecular Science, National Institutes of Natural Sciences, S-1 Higashiyama Myodaijicho, Okazaki, Aichi 444-8787, Japan
‡The Graduate University for Advanced Studies (SOKENDAI), Shonan Village, Hayama, Kanagawa 240-0193, Japan

ABSTRACT: We evaluated a method for protein engineering using plasmid-based one-pot saturation mutagenesis and robot-based automated screening. When the biases in nucleotides and amino acids were assessed for a loss-of-function point mutation in green fluorescent protein, the ratios of gain-of-function mutants were not significantly different from the expected values for the primers among the three different suppliers. However, deep sequencing analysis revealed that the ratios of nucleotides in the primers were highly biased among the suppliers. Biases for NNB were less severe than for NNN. We applied this method to screen a fusion protein of two chitinases, ChiA and ChiB (ChiAB). Three NNB codons as well as tyrosine and serine (XYSX1X2X3) were inserted to modify the surface structure of ChiAB. We observed significant amino acid bias at the X3 position in water-soluble, active ChiAB-X1YSX2X3 mutants. Examination of the crystal structure of one active mutant, ChiAB-FYSFV, revealed that the X3 residue plays an important role in structure stabilization.

INTRODUCTION

Saturation mutagenesis is a basic protein engineering technique for increasing thermostability1,2 and substrate specificity3,4 of enzymes and antibody affinity5,6. To increase the efficiency of randomization in saturation mutagenesis, primer design plays a critical role. Several design methods have been developed, including Max randomization,7 Tang’s “small-intelligent”,8 and 22c-tricks.9 These designed oligonucleotide mixtures are especially suitable for single-site saturation mutagenesis. In contrast, conventional degenerate oligonucleotides, such as NNN, NNB, and NNK/S, are more suitable for multisite saturation mutagenesis in terms of cost, simplicity of primer design, and reduced number of PCR primers. In addition to the design of degenerate oligonucleotides for saturation mutagenesis, the development of molecular cloning techniques also plays an important role in facilitating the mutagenesis procedure. The most popular techniques are QuickChange mutagenesis (Agilent), Gibson assembly10 (NEB), and In-Fusion11 (Clontech). These commercially available kits are useful for ligation with single or multiple fragments in vitro without any restriction enzymes and ligase. A recent development in the field is plasmid-based one-pot saturation mutagenesis12 using a pair of endonucleases (Nt.BbvCI and Nb.BbfCI) for nicking mutagenesis, in vivo assembly for homologous recombination cloning (IVA),13 and seamless ligation cloning extract from Escherichia coli14–17 (SLiCE). SLiCE is notable for its simplicity and low cost.

In this study, we evaluated a method of plasmid-based one-pot saturation mutagenesis employing the three techniques, namely conventional degenerate NNN or NNB codons for single or multisite saturation mutagenesis, IVA cloning for primer design, and SLiCE for ligating both ends of the fully amplified linear plasmid with homologous recombination in vitro. We also performed a robot-based automated method for E. coli cell culture, protein purification, and activity measurement in 96-well plates using a liquid-handling robot. We assessed this combination of methods for protein engineering. First, to assess our one-pot saturation mutagenesis and to understand the efficiency of randomization, we applied our method to a nonfluorescent, loss-of-function green fluorescent protein (GFP) mutant, GFPmut3-Y66H,12 using NNN or NNB primers from three suppliers. The recovery rates (H66Y mutation) in both colony counting and deep sequencing were similar among the primer suppliers, but large nucleotide biases were observed in deep sequencing. The fractions of thymine (T) and guanine (G) from suppliers 1 and 2 were much higher than those of adenine (A) and cytosine (C) in NNN and NNB degenerate codons. In contrast, supplier 3 showed ratios of A, T, G, and C similar to the expected ratios in both NNN and NNB degenerate codons. Biases in amino acid composition given by NNN degenerate codons were also evident in our
deep sequencing results. Valine, phenylalanine, tryptophan, and glycine all showed higher-than-expected fractions, and aspartic acid and isoleucine were less abundant than expected. However, there were no significant biases in amino acids from NNB degenerate codons and the variances of observed/expected ratios were low. Therefore, on the basis of analyses of biases in nucleotides and amino acids, degenerate NNB codons were used for multisite saturation mutagenesis.

Next, we applied this method to ChiAB, an artificial fusion protein of two chitinases, ChiA and ChiB, from Serratia marcescens.18,19 S. marcescens ChiA and ChiB processively hydrolyze crystalline chitin from reducing and nonreducing ends, respectively, and are linear motor proteins that move in opposite directions. Both ChiA and ChiB have a catalytic domain (CD) and a carbohydrate-binding module (CBM) connected by a short linker. The CDs show almost identical folds (TIM barrel), whereas the positions of the CBM relative to the CD were different. We are currently attempting to engineer a bidirectional motor protein based on ChiAB that carries one CD from ChiA and two CBMs from ChiA and ChiB. Multisite saturation mutagenesis was carried out with degenerate NNB codons as well as tyrosine and serine residue insertions (X1YSX2X3) to modify surface properties of ChiAB.

The observed ratios of each amino acid at three positions (X1, X2, and X3) were similar when the sequences were analyzed irrespective of solubility of expressed proteins. On the other hand, among water-soluble and active 120 clones, we observed an amino acid bias at the third (X3) position in ChiAB-X1YSX2X3 mutants. In an attempt to understand the reason for the bias, we solved the crystal structure of one active mutant, ChiAB-FYSFV. The results showed that the valine residue at the X3 position was oriented toward the inside of the molecule, presumably contributing to the stability of the structure.

### RESULTS AND DISCUSSION

#### Fraction of Gain-of-Function GFP Mutants Analyzed by Colony Counting and Deep Sequencing.

We have designed a method to generate and screen a wide variety of mutant proteins. The bottleneck steps include the mutant construction and protein purification. To resolve these limiting steps, we combined one-pot saturation mutagenesis with robot-based small-scale purification of a large number of mutants. Our experimental system is shown in Figure 1. To assess nucleotide and amino acid biases, one-pot saturation mutagenesis was performed with degenerate NNN and NNB codons from three different suppliers. The fractions of gain-of-function mutations of GFPmut3 were analyzed by colony counting and deep sequencing.

Prior to comparison of NNN and NNB, we checked bias in DNA amplification efficiency of PCR depending on codons, using 64 primers encoding different codons (Figure 2). The amounts of PCR products (∼4500 bp) were basically similar among all codons except ATA, which did not show clear bands. This result indicates that the PCR step does not cause bias basically, at least when pEDAS-GFPmut3-Y66H is used as a template.

Then, to determine the recovery rate for each supplier’s degenerate codons by colony counting, the pEDAS-GFPmut3-Y66H plasmid was amplified by PCR with primers containing either NNN or NNB codons. The linear products were ligated with SLiCE, transformed into E. coli cells, and cultured on agar plates at 37 °C overnight. Colonies were counted under visible light, and the number of fluorescent colonies (gain-of-function, H66Y mutants) was counted under a blue/green light-emitting diode (LED) light. The fractions of gain-of-function colonies are shown in Table 1. For both NNN and NNB primers, suppliers 1−3 showed gain-of-function fractions with values that were slightly lower or higher than expected. However, the differences were minor, ranging from −0.9 to 2.0% (Table 1).
All experimental values were similar to expectation, showing no apparent biases. To investigate the potential biases in more detail, all colonies were collected and plasmids were extracted and analyzed by deep sequencing. First, to estimate fractions of gain-of-function GFP mutants, tyrosine codons (TAT, TAC) were counted (Table 1). Again, for both NNN and NNB primers, suppliers 1–3 showed fractions of gain-of-function similar to the expected values, with differences ranging from ~0.9 to 1.1% (Table 1). For each primer set and supplier, the value estimated by deep sequencing was similar to that from colony counting, indicating the validity of the deep sequencing (Table 1).

**Biases in Codons and Nucleotides Estimated by Deep Sequencing.** Next, we determined the fraction of each codon from each primer set and supplier. Figure 3 shows the fraction of each codon, ordered from largest to smallest. Expected fractions of each codon from NNN and NNB are 1.6% (1/64) and 2.1% (1/48), respectively. However, NNN from suppliers 1 and 2 showed significantly larger fractions of the top 10 codons, whereas this bias was slightly less severe in NNB (Table 2).

We then assessed the nucleotide bias more directly by estimating fractions of each nucleotide (Table 3). NNN primers from suppliers 1 and 2 had significantly higher fractions of T and G than those of A and C. Notably, NNN from supplier 1 showed an extremely low fraction of A. Proportions of A, T, G, and C in primers from supplier 3 were similar to the expected values (25%). In the case of NNB, the expected values were different because A was not present at the third position of NNB; therefore, the expected value of fraction of A was 16% and those of T, G, and C were 28%. In primers from supplier 1, the value of the fraction of T was highest and that of G was equal to the expected value. However, the value of the fraction of C was lower than expected. From supplier 2, the values of the fractions of T and G were higher than expected and that of C was lower than expected. In contrast, the fractions of A, T, G, and C were highly similar to expectations in both NNN and NNB primers from supplier 3. In the previous study, it has been reported that hand-mixed degenerate primers showed lower nucleotide bias than machine-mixed degenerate primers when used for PFunkel.20 In our study, we have not used hand-mixed degenerate primers for both NNN and NNB codons from supplier 1 to 3. The biases observed in our study may be improved if hand-mixed degenerate primers are used.

To examine the nucleotide bias in more detail, we analyzed the fractions of A, T, G, and C at each of the three positions of each codon (Table 4). NNN and NNB primers from suppliers 1 and 2 showed higher fractions of T and G than those of A and C at all three positions. In contrast, NNN from supplier 3 showed smaller biases than those from suppliers 1 and 2 at all three positions. This was also the case for NNB primers from supplier 3. Overall, codon and nucleotide biases were lower in NNB than in NNN, although the effects of different template sequences and DNA polymerases were not evaluated in our study. To draw general conclusion about the difference in the biases between NNN and NNB, further quantitative analysis of different target proteins will be required.

**Biases in Amino Acids.** We also translated our deep sequencing results into amino acids (Figures 4 and 5). We found obvious biases in amino acids translated from NNN (Figure 4, top panels). Several amino acids with significantly higher and lower fractions than expected were observed for suppliers 1–3. Indeed, values more than twice as high from the
expected values for valine, phenylalanine, and tryptophan were observed for both suppliers 1 and 2 (Figure 5, top panels). Furthermore, the fractions of lysine, histidine, glutamic acid, threonine, asparagine, and glutamine were less than half of that expected from supplier 1, whereas no amino acids with significantly low fractions were observed with supplier 2.

Codons from supplier 3 showed only one amino acid, glycine, with a ratio 2 times higher than expected, whereas three amino acids, lysine, arginine, and isoleucine, showed ratios significantly lower than expected. The amino acid analysis therefore revealed biases in NNN codons from all three suppliers (Table 5).

Biases were relatively less severe in amino acids translated from NNB (Figures 4 and 5, bottom panels). Except for phenylalanine from supplier 1, and tryptophan, histidine, and asparagine from supplier 2, no amino acid ratios from NNB were more than twice as high or low from the expected values (Figure 5). The observed/expected ratios were nearly equal to 1 for each amino acid for supplier 3. Variances from NNB were lower than those from NNN, and, notably, the variance was lowest (0.08, Table 5) for supplier 3.

Overall, the observed and expected fractions of each amino acid from NNN and NNB codons showed wide variations (Figure 6). Ratios of mutations were significantly different among amino acids. However, observed fractions were relatively similar to the expected values in NNB codons from all three suppliers. Moreover, the variance among amino acid fractions was much lower for NNB than for NNN (Tables 5 and 6). As a result, frequencies of mutations from NNB were more uniform than from NNN. In terms of nucleotide, amino acid, and mutational biases, NNB codons were superior to NNN under the experimental conditions we evaluated. This was especially true of NNB from supplier 3.

Saturation Mutagenesis of ChiAB. Multisite saturation mutagenesis using NNB codons and insertion of tyrosine and serine residues were performed simultaneously on ChiAB to generate ChiAB-X1YSX2X3. This incorporated the curved α-helix that stabilizes the CBM-linker domain from ChiB. When ratios of each amino acid at three positions (X1,X 2, and X3) were compared without screening by solubility, they were very close to the expected values for valine, phenylalanine, and tryptophan were observed for both suppliers 1 and 2 (Figure 5, top panels). Furthermore, the fractions of lysine, histidine, glutamic acid, threonine, asparagine, and glutamine were less than half of that expected from supplier 1, whereas no amino acids with significantly low fractions were observed with supplier 2.

Codons from supplier 3 showed only one amino acid, glycine, with a ratio 2 times higher than expected, whereas three amino acids, lysine, arginine, and isoleucine, showed ratios significantly lower than expected. The amino acid analysis therefore revealed biases in NNN codons from all three suppliers (Table 5).

Biases were relatively less severe in amino acids translated from NNB (Figures 4 and 5, bottom panels). Except for phenylalanine from supplier 1, and tryptophan, histidine, and asparagine from supplier 2, no amino acid ratios from NNB were more than twice as high or low from the expected values (Figure 5). The observed/expected ratios were nearly equal to 1 for each amino acid for supplier 3. Variances from NNB were lower than those from NNN, and, notably, the variance was lowest (0.08, Table 5) for supplier 3.

Overall, the observed and expected fractions of each amino acid from NNN and NNB codons showed wide variations (Figure 6). Ratios of mutations were significantly different among amino acids. However, observed fractions were relatively similar to the expected values in NNB codons from all three suppliers. Moreover, the variance among amino acid fractions was much lower for NNB than for NNN (Tables 5 and 6). As a result, frequencies of mutations from NNB were more uniform than from NNN. In terms of nucleotide, amino acid, and mutational biases, NNB codons were superior to NNN under the experimental conditions we evaluated. This was especially true of NNB from supplier 3.

**Saturation Mutagenesis of ChiAB.** Multisite saturation mutagenesis using NNB codons and insertion of tyrosine and serine residues were performed simultaneously on ChiAB to generate ChiAB-X1YSX2X3. This incorporated the curved α-helix that stabilizes the CBM-linker domain from ChiB. When ratios of each amino acid at three positions (X1, X2, and X3) were compared without screening by solubility, they were very close to the expected values for valine, phenylalanine, and tryptophan were observed for both suppliers 1 and 2 (Figure 5, top panels). Furthermore, the fractions of lysine, histidine, glutamic acid, threonine, asparagine, and glutamine were less than half of that expected from supplier 1, whereas no amino acids with significantly low fractions were observed with supplier 2.

Codons from supplier 3 showed only one amino acid, glycine, with a ratio 2 times higher than expected, whereas three amino acids, lysine, arginine, and isoleucine, showed ratios significantly lower than expected. The amino acid analysis therefore revealed biases in NNN codons from all three suppliers (Table 5).

Biases were relatively less severe in amino acids translated from NNB (Figures 4 and 5, bottom panels). Except for phenylalanine from supplier 1, and tryptophan, histidine, and asparagine from supplier 2, no amino acid ratios from NNB were more than twice as high or low from the expected values (Figure 5). The observed/expected ratios were nearly equal to 1 for each amino acid for supplier 3. Variances from NNB were lower than those from NNN, and, notably, the variance was lowest (0.08, Table 5) for supplier 3.

Overall, the observed and expected fractions of each amino acid from NNN and NNB codons showed wide variations (Figure 6). Ratios of mutations were significantly different among amino acids. However, observed fractions were relatively similar to the expected values in NNB codons from all three suppliers. Moreover, the variance among amino acid fractions was much lower for NNB than for NNN (Tables 5 and 6). As a result, frequencies of mutations from NNB were more uniform than from NNN. In terms of nucleotide, amino acid, and mutational biases, NNB codons were superior to NNN under the experimental conditions we evaluated. This was especially true of NNB from supplier 3.

**Saturation Mutagenesis of ChiAB.** Multisite saturation mutagenesis using NNB codons and insertion of tyrosine and serine residues were performed simultaneously on ChiAB to generate ChiAB-X1YSX2X3. This incorporated the curved α-helix that stabilizes the CBM-linker domain from ChiB. When ratios of each amino acid at three positions (X1, X2, and X3) were compared without screening by solubility, they were very close to the expected values for valine, phenylalanine, and tryptophan were observed for both suppliers 1 and 2 (Figure 5, top panels). Furthermore, the fractions of lysine, histidine, glutamic acid, threonine, asparagine, and glutamine were less than half of that expected from supplier 1, whereas no amino acids with significantly low fractions were observed with supplier 2. Codons from supplier 3 showed only one amino acid, glycine, with a ratio 2 times higher than expected, whereas three amino acids, lysine, arginine, and isoleucine, showed ratios significantly lower than expected. The amino acid analysis therefore revealed biases in NNN codons from all three suppliers (Table 5).

Biases were relatively less severe in amino acids translated from NNB (Figures 4 and 5, bottom panels). Except for phenylalanine from supplier 1, and tryptophan, histidine, and asparagine from supplier 2, no amino acid ratios from NNB were more than twice as high or low from the expected values (Figure 5). The observed/expected ratios were nearly equal to 1 for each amino acid for supplier 3. Variances from NNB were lower than those from NNN, and, notably, the variance was lowest (0.08, Table 5) for supplier 3.

Overall, the observed and expected fractions of each amino acid from NNN and NNB codons showed wide variations (Figure 6). Ratios of mutations were significantly different among amino acids. However, observed fractions were relatively similar to the expected values in NNB codons from all three suppliers. Moreover, the variance among amino acid fractions was much lower for NNB than for NNN (Tables 5 and 6). As a result, frequencies of mutations from NNB were more uniform than from NNN. In terms of nucleotide, amino acid, and mutational biases, NNB codons were superior to NNN under the experimental conditions we evaluated. This was especially true of NNB from supplier 3.
similar each other (Figure 7). Therefore, the amino acid ratio is not affected by the position and successive introduction of NNB. Then, we obtained 120 water-soluble and catalytically active samples and 24 insoluble and inactive samples. These 144 clones were sequenced, and amino acid residues in the X1, X2, and X3 positions were determined (Figure 8). The results showed no large biases at the X1 and X2 positions in either soluble or insoluble mutants or at the X3 position in insoluble mutants. However, in water-soluble mutants, there was a remarkable bias at the X3 position in favor of valine (27/120) and leucine (56/120) residues. Amino acid residues with basic (K, R, H), acidic (D, E), or uncharged polar (S, T, Y, N, Q) side chains were not introduced at the X3 position of the water-soluble mutants, presumably because these residues could not support correct folding and/or resulted in aggregation.

Structure Determination of ChiAB-FYSFV. To elucidate the reason for the bias at the X3 position of ChiAB-X1YSX2X3 mutants, we solved the crystal structure of the most active mutants, ChiAB-FYSFV, at 2.6 Å resolution (Figure 9a,b, Table 7). In the ChiAB-FYSFV structure, the ChiA moiety was not significantly different from the wild-type ChiA structure. A clear electron density was observed in the CBM domain of ChiB, and it was located on the bottom surface of ChiA.

Table 4. Fractions of A, T, G, and C at the First, Second, and Third Nucleotide Positions of the Codons in Each Primer Set and Supplier, Determined by Deep Sequencing

| primer set | supplier | first nucleotide (%) | second nucleotide (%) | third nucleotide (%) |
|------------|----------|----------------------|-----------------------|----------------------|
| NNN        |          | A | T | G | C | A | T | G | C | A | T | G | C | A | T | G | C |
| 1          | 1        | 9 | 42| 33| 16| 8 | 43| 33| 16| 9 | 39| 37| 15|    |    |    |    |
| 2          | 14       | 16| 38| 30| 16| 17| 38| 30| 15| 16| 39| 31| 14|    |    |    |    |
| 3          | 21       | 28| 32| 19|    | 28| 19| 26| 27| 22| 30| 27| 21|    |    |    |    |
| NNB        |          | 1 | 19| 36| 26| 19| 20| 36| 26| 18| 0 | 44| 33| 23|    |    |    |
| 2          | 18       | 34| 30| 17|    | 19| 33| 31| 17| 0 | 39| 41| 20|    |    |    |    |
| 3          | 21       | 26| 29| 24|    | 24| 24| 30| 22| 0 | 36| 36| 28|    |    |    |    |

*a* In NNN, the expected fractions of A, T, G, and C were all 25%. *b* In NNB, the expected fractions of T, G, and C at the third position were all 33% and the expected fraction of A at the third position was 0%.
However, the flexible linker region of the CBM-linker from ChiB did not show a clear electron density. In the wild-type ChiB structure, the curved \(\alpha\)-helix stabilizing with the CBM-linker domain is observed (Figure 9c). Wild-type ChiA does not have a similarly curved \(\alpha\)-helix, but there was an \(\alpha\)-helix containing a short loop in its ChiB counterpart (Figure 9d). In the present study, this \(\alpha\)-helix was mutated by multisite saturation mutagenesis and YS residue insertion. The results showed that the mutated part of the \(\alpha\)-helix in the ChiAB-FYSFV structure was similar to neither the ChiA nor the ChiB \(\alpha\)-helix. Instead, a straight \(\alpha\)-helix was observed (Figure 9e). The side chain of the valine residue at the X3 position of the X1YSX2X3 (FYSFV) moiety of the mutated \(\alpha\)-helix was oriented toward the inside of the protein core. These results suggested that the residue at the X3 position of X1YSX2X3 plays an important role in stabilizing the structure and preventing aggregation, supporting the abovementioned hypothesis that mutants with basic, acidic, uncharged polar amino acid residues at the X3 position cannot fold correctly, resulting in aggregation. Therefore, it seems very likely that the bias in favor of valine and leucine residues at the X3 position was indispensable for stabilization and solubilization of the ChiAB-X1YSX2X3 mutant.

### CONCLUSIONS

The method of protein engineering evaluated in this study was based on one-pot saturation mutagenesis and robot-based automated screening. According to the nucleotide and amino acid biases in our gain-of-function GFP mutant experiments, the NNB primer was much more useful than the NNN primers for multisite saturation mutagenesis. The least-biased NNB primer was obtained from supplier 3. In this study, we have not attempted to use other degenerate primers, such as NNK and NNS. The NNK and NNS primers contain only 32 codons, including one stop codon, therefore incorporating less redundancy than NNB. They may provide improved random mutagenesis by mitigating biases. Furthermore, as a proof-of-concept, we applied our methods to ChiAB with three degenerate NNB codons and the insertion of tyrosine and serine residues. After screening active 120 clones, we observed heavy amino acid biases at the X3 position of ChiAB-X1YSX2X3. This result clearly reflected the effects of amino acid bias in the mutant generation on the efficiency of protein
engineering. With our method, we will try to engineer non-natural motor proteins with novel functions.

**METHODS**

**Reagents.** All chemicals were purchased from Wako. The template plasmid pEDAS_GFPmut3_Y66H was a gift from Timothy Whitehead (Michigan State University, Addgene plasmid #80085). PrimeSTAR HS DNA polymerase was purchased from Takara. Plasmid extraction kit was purchased from NIPPON Genetics Co., Ltd. Wizard SV gel and PCR clean-up system was purchased from Promega. All primers (containing NNN or NNB, where B = T/G/C) were purchased from supplier 1 (Fasmac), supplier 2 (Eurofin), and supplier 3 (Integrated DNA Technologies).

**Primer Design for Saturation Mutagenesis of Loss-of-Function GFP.** All NNN degenerate codon primers contained all 64 codons, including three stop codons. The forward primer (5′-CAC TTGTCACTTTTCGGTGGTGTTCAATGCTTTGCG-3′), containing one degenerate NNN motif, had a melting temperature ($T_m$) of 68 °C. The NNB primers included 48 codons, including one stop codon. The forward primer (5′-CACTTGTCACCTTTTCGGTGGTGTTCAATGCTTTGCG-3′), containing one degenerate NNB motif, had a $T_m$ of 68 °C. The concentration of all primers was adjusted to 10 pmol/μL for PCR. The following reverse primer was used for all saturation mutagenesis experiments: 5′-ACCGAAAGTAGTGAACAGTGTTGGCCATGGAACAGGTAG-3′.

**PCR and DpnI Treatments.** All three-step PCRs were carried out with PrimeSTAR HS polymerase. The PCR mixture was as follows: 10 μL of 5X PS buffer, 4 μL of 2.5 mM dNTP, 0.5 μL of PrimeSTAR HS DNA polymerase, 2 μL of 10 pmol/μL primer mix, and 1 ng of template plasmid, made up to a volume of 50 μL with sterilized water. The thermocycling protocol was as follows: 30 cycles of 98 °C for 10 s, 55 °C for 5 s, and 72 °C for 4 min and 30 s. After amplification, all products were incubated with 1 μL of DpnI (NEB) for 15 min at 37 °C to digest the template plasmid in the reaction mixture. To increase efficiency of SLiCE cloning, 1% agarose electrophoresis was performed, then each fragment was extracted, and purified by gel clean-up system.

**Figure 6.** Observed fractions of each amino acid in NNN and NNB from three suppliers. All histograms show experimental amino acid fractions. Light blue: NNN (64 codons, top), green: NNB (48 codons, bottom). The thick horizontal lines (dark blue) on the histograms indicate expected amino acid fractions.

**Table 6.** Statistics of Each Experimental Fraction of 20 Amino Acids in Each Primer Set, from Each Supplier, Determined by Deep Sequencing

| primer set | supplier | mean (%) | variance | SD (%) | median (%) | stop codon (%) |
|------------|----------|----------|----------|--------|------------|----------------|
| NNN        | 1        | 4.93     | 21.4     | 4.63   | 2.9        | 1.49           |
|            | 2        | 4.84     | 10.7     | 3.27   | 3.79       | 3.29           |
|            | 3        | 4.74     | 13.9     | 3.73   | 3.50       | 5.19           |
| NNB        | 1        | 4.96     | 7.78     | 2.79   | 4.21       | 0.74           |
|            | 2        | 4.92     | 8.47     | 2.91   | 4.22       | 1.63           |
|            | 3        | 4.94     | 5.31     | 2.30   | 4.35       | 1.28           |
SLiCE Cloning and Transformation. SLiCE cloning uses homologous recombination to ligate DNA fragments from cellular extracts in vitro. SLiCE and 10× SLiCE reaction buffer (0.5 M Tris–HCl pH 7.5, 100 mM MgCl2, 10 mM ATP, and 10 mM DTT) were prepared as previously described. Briefly, the product (20 ng), 1 μL SLiCE, and 1 μL SLiCE reaction

Figure 7. Ratios of each amino acid at X1, X2, and X3 positions of ChiAB-X1YSX2X3 without screening by solubility. The 265 clones were sequenced, and coded amino acid residues were counted at each position, irrespective of solubility of expressed proteins. The thick horizontal lines (dark blue) on the histograms indicate expected amino acid fractions.

Figure 8. Ratios of each amino acid at X1, X2, and X3 positions for soluble and insoluble ChiAB-X1YSX2X3. The soluble 120 clones and insoluble 24 clones were sequenced, and coded amino acid residues were counted at each position.

ACS Omega
Article
DOI: 10.1021/acsomega.8b00663
ACS Omega 2018, 3, 7715−7726
buffer were mixed in a reaction volume of 10 μL. The mixture was incubated at 37 °C for 30 min. After the SLiCE reaction, the samples were immediately used for transformation or stored at −30 °C. SLiCE reactant (5 μL) was added to 50 μL of Tuner(DE3) competent cells on ice and mixed moderately. Within 1 min, these mixtures were transferred to a cold cuvette and transformation was carried out with a MicroPulser electroporator (Bio-Rad). Immediately after electroporation, 300 μL of iced super optimal broth with catabolite repression (SOC) was added. The transformed cells were incubated at 37 °C for 30 min. Finally, 150 μL of each transformant was spread on 15 cm LB-agar plates with 50 μg/mL ampicillin and incubated at 37 °C overnight.

**Colony Counting.** Colonies were counted under blue/green LED (Handy Blue/Green LED). The recovery rate was represented as illuminated colonies normalized to all colonies from each primer set and supplier [(number of illuminated colonies/total number of colonies) × 100], respectively. For calculating the fraction of gain-of-function by deep sequencing, the number of TAT and TAC codons was counted [(number of TAT and TAC/total number of counts) × 100]. For calculating the expected fraction of gain-of-function, the number of TAT and TAC codons were counted and divided by the number of codons in each primer set [(number of TAT and TAC/total number of codons) × 100].

**Comparison of PCR Products with 64 Primers Encoding Different Codons.** The pEDAS_GFPmut3_Y66H was amplified with 64 kinds of forward primers encoding different codons (5′- C A C T T G T C A C - T A C T T C G G T X X G G T T C A T G C T T G C G - 3′, where X is A, C, G, or T) and the same reverse primer as described above. Other conditions of PCR were same as the saturation mutagenesis experiments. Products (2 μL) were electrophoresed in 1% agarose.

**Deep Sequencing.** All colonies on each plate were collected with 10 mL of LB medium and centrifuged at 4000g for 10 min. The number of transformants used for deep sequencing were supplier 1 (NNN: 1453, NNB: 1618),

**Figure 9.** Crystal structure of ChiAB-FYSFV. Side (a) and top (b) views of ChiAB-FYSFV (PDB ID: 5ZL9). The ChiA-part is shown in green, and the CBM-linker from ChiB is shown in blue. (c) The original α-helix in the wild-type ChiB (PDB ID: 1E6N). (d) The original α-helix in the wild-type ChiA (PDB ID: 1EIB). (e) The α-helix in the crystal structure of ChiAB-FYSFV. Pink circles indicate the mutated α-helix.

**Table 7. Statistics of Data Collection and Refinement of the Crystal Structure of ChiAB-FYSFV**

| data collection statistics | SmChiAB-FYSFV (PDB ID: 5ZL9) |
|----------------------------|-------------------------------|
| beam line                  | Aichi SR BL2S1                |
| wavelength (Å)             | 1.12                          |
| space group                | C222                          |
| unit cell parameters (Å)   | a = 71.8, b = 190.2, c = 132.9 |
| exposure time (frame/s)    | 10                            |
| number of frames           | 360 (ω = 1°)                  |
| resolution (Å)             | 47.5–2.6 (2.69–2.6)           |
| observed/unique reflections| 418 227/28 411                |
| multiplicity               | 14.7 (14.7)                   |
| completeness (%)           | 100 (100)                     |
| Rmerge (%)                 | 19.8 (101.4)                  |
| mean ⟨I/σ(I)⟩              | 14.5 (2.6)                    |

**refinement statistics**

| SmChiAB-FYSFV (PDB ID: 5ZL9) |
|-------------------------------|
| Rwork/Rfree                   | 18.8/20.5                     |
| rmsd bonds (Å)                | 0.004                         |
| rmsd angles (deg)             | 0.845                         |
| atoms                         | 4704                          |
| solvent molecules             | 292                           |
| Ramachandran plot             | SmChiAB-FYSFV (PDB ID: 5ZL9)  |
| favored (%)                   | 95.7                          |
| allowed                       | 3.3                           |
| outliers (%)                  | 1.0                           |
| rotamer outliers (%)          | 0.4                           |
suppliers 2 (NNN: 2048, NNB: 1573), and supplier 3 (NNN: 1527, NNB: 2128). Plasmid mixtures were extracted from each cell pellet by a plasmid extraction kit, diluted to 1 ng/μL, and amplified by PCR for deep sequencing. The forward primer was 5′-TCGTCGCTAAGCTTGCATTGTATAAGAACGACAG + ATATTAGGAGACACCAAGGTTC-3′, and the reverse primer was 5′-GTCGTCGTGGGCTCGGA-GATGTTGATATAGAGACAG + GTGTCATTAGTTCCCGTCATCTTGTG-3′. Both primers had additional sequences (underlined) for deep sequencing. The PCR was performed in two steps, each with 30 cycles of 94 °C for 10 s and 68 °C for 4 min and 30 s, with PrimeSTAR HS DNA polymerase. All products were electrophoresed in 1% agarose and purified with a gel clean-up system. All experiments described below were performed by Hokkaido System Science Co., Ltd. Additional adapter sequences were attached with KAPA HiFi DNA polymerase (HotStart Ready mix) by second index PCR (preincubation 95 °C for 3 min, 8 cycles of 95 °C for 10 s, 55 °C for 30 s, and 72 °C for 30 s, followed by 72 °C for 5 min). Nextera XT index primers N7xx and S5xx were used. All fragments were purified by AMPure XP beads (Beckman Coulter Genomics). Deep sequencing was performed with MiSeq system (Illumina) by Hokkaido System Science Co., Ltd. Processed codon counts data used for analysis were as follows: supplier 1 (NNN: 86 048, NNB: 82 657), supplier 2 (NNN: 83 927, NNB: 79 696), and supplier 3 (NNN: 80 325, NNB: 83 835). For calculation of the gain-of-function fraction from the results of deep sequencing, the numbers of TAT and TAC codons were counted and used [(number of TAT and TAC/total number of counts) × 100].

For calculation of the expected value of gain-of-function fraction, the number of TAT and TAC codons in each primer set were counted and divided by the total numbers of codons.

**Multisite Saturation Mutagenesis for ChiAB-X1YSX2X3.** To generate the curved α-helix in ChiA helix (Y373 A374 Y375), tyrosine and serine residues were inserted between V373 and A374. The three adjacent residues (373-375) were mutated with degenerate NNB codons. The synthetic ChiAB gene was cloned into pET-27b with NdeI and NotI restriction sites. To introduce three site saturation mutagenesis and two extra tyrosine and serine residues (X,YSX,X3), the following primers were used: forward primer 5′-GACAAAGTACGACAGAGNBNTATAGCTCNBN-BAACGTTGGCCGCAAGACTCGATGGAATCACATC-3′ (Tm = 70 °C) and reverse primer 5′-CTTGTCATCTGTGTCCT- TACCGGCGCTGTAGT-3′ (Tm = 62 °C). All reactions were carried out with three-step PCRs using PrimeSTAR HS DNA polymerase. The PCR mixture consisted of 10 μL of 5× PS buffer, 4 μL of 2.5 mM dNTP, 0.5 μL of PrimeSTAR HS DNA polymerase, 2 μL of 10 pmol/μL primer mix, and 1 ng of template plasmid, made up to a volume of 50 μL with sterilized water. The thermocycling protocol included 30 cycles at 98 °C for 10 s, 68 °C for 7 min. All products were incubated with 1 μL of DpnI (NEB) for 15 min at 37 °C to digest the template plasmid in the reaction mixture. To increase the efficiency of SLiCE cloning, 1% agarose electrophoresis was performed, then each amplified linear plasmid was extracted, and purified by a gel clean-up system. The fragments were ligated by SLiCE and transformed into E. coli cells by electroporation. Transformed cells were spread and incubated on agar plates with 25 μg/mL of kanamycin at 37 °C overnight. The SLiCE reaction and transformation protocols were same as described above.

**Ratio of Each Amino Acid at X1, X2, and X3 Positions of ChiAB-X1YSX2X3.** A 96-well deep well plate was used for small-scale cultivation. Super Broth (1 mL, 3.2% tryptone, 2% yeast extract, and 0.5% sodium chloride) with 25 μg/mL of kanamycin was added to each well and used to inoculate colonies from the plate. Master plates were prepared to store the cells. The 96-well deep well plate was cultured at 37 °C with shaking at 1300 rpm for overnight and centrifuged at 4400 rpm for 10 min. Harvested cells were used for plasmid extraction, and the obtained codons and amino acids at each position were analyzed.

**Small-Scale Culture, Purification, and Activity Measurement of ChiAB-X1YSX2X3 by Liquid-Handling Robot.** Small-scale cultivation was basically same as described above except that the 96-well deep well plate was cultured at 30 °C with shaking at 1300 rpm until the cells reached an OD600 ∼ 1. Then, 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and the cells were then cultured at 20 °C overnight. Cells were then harvested by centrifugation (4400 rpm at 10 °C for 10 min). To disrupt the cells, 300 μL of BugBuster (Novagen) containing 10 unit/mL of Benzonase (Novagen) was added to each well, followed by shaking at 1000 rpm for 20 min. The disrupted cells were centrifuged at 4400 rpm at 25 °C for 10 min. We used Beckman Coulter Biomek 4000 for purification of ChiAB-X1,YSX,X3 mutants. After centrifugation, the supernatant was transferred to a new 96-well deep well plate and 100 μL of 50% slurry Ni-NTA agarose (QIAGEN) was added to each well. Washing was then carried out at 1000 rpm for 5 min to facilitate binding with the target protein. Ni-NTA agarose was washed with 200 μL of buffer A (50 mM sodium phosphate, pH 7.0) twice, 200 μL buffer B (50 mM sodium phosphate, pH 7.5, 50 mM imidazole) thrice, and eluted with 150 μL buffer C (50 mM sodium phosphate, pH 7.5, 100 mM imidazole). Eluted solution (5 μL) was loaded on 12% acrylamide gel and purity was checked. Purified enzymes (18 μL) were incubated with 0.1% (w/v) crystalline β-chitin in 100 mM sodium phosphate buffer (pH 6.0 at 37 °C for 8 min) in reaction mixture volumes of 180 μL. Reactions were stopped by adding 240 μL of Schaeles reagent (500 mM sodium carbonate, 1.5 mM potassium ferricyanide), and insoluble chitin was separated on 96-well 0.45 μm PVDF filter plates. The filtered solution was heated at 95 °C for 15 min, and 200 μL of the samples were transferred to 96-well plates. Absorbance at 420 nm was measured, and amounts of soluble products were calculated from a standard curve with chitobiase.

**Plasmid Extraction and Sequencing of Active Mutants.** Each colony was inoculated in 10 mL of LB medium with 25 μg/mL of kanamycin from the master plate and cultured at 37 °C overnight. The cells were harvested by centrifugation (4000g at 4 °C for 10 min). Plasmid extractions were performed following the manufacturer’s protocol. Sequencing of all plasmids was carried out by FASMAC Co., Ltd.

**Large-Scale Culture and Purification of ChiAB-FYSFV.** ChiAB-FYSFV, one of the ChiAB-X1YSX2X3 mutants, was selected by its solubility and activity. Cells were inoculated from the master plate into 10 mL of LB medium and cultured overnight at 37 °C. Overnight culture (5 mL) was added to 1 L of LB medium. After the cells were grown in LB medium containing 25 μg/mL of kanamycin at 37 °C to an OD600 ∼ 1, protein overexpression was induced by adding 0.5 mM IPTG, followed by overnight incubation at 20 °C. After harvesting by
centrifugation at 8000g for 20 min, the cells were resuspended in 100 mL of 100 mM Tris–HCl (pH 8.0) and sonicated with ethylenediaminetetraacetic acid-free protease inhibitor cocktail (Complete Mini, Roche). Then, 2.5 mL of 4 M NaCl and 5 mL of Ni-NTA Superflow (50% Slurry, Qiagen) were added to the supernatant directly, followed by gentle mixing, and incubated for 5 min at 25 °C. The Ni-NTA resin was packed into an open column and washed with a buffer (50 mM sodium phosphate, pH 7.0, 100 mM NaCl) containing 50 mM imidazole. The sample was eluted with the same buffer containing 100 and 200 mM imidazole in a stepwise manner. Protein fractions were mixed and concentrated to 500 μL using a 10 kDa cut VIVASPIN Turbo 15 (Sartorius). The concentrated sample was further purified by Superdex 200 10/300GL (GE Healthcare) equilibrated with a buffer (50 mM sodium phosphate, pH 7.0, 100 mM NaCl). Peak fractions were combined and concentrated to 20 mg/mL for crystallization.

Crystallization, Data Collection, and Structure Determination. Protein concentration of ChiAB-FYSFV was adjusted to 20 mg/mL. All drops contained 1 μL of protein solution and 1 μL of reservoir solution. Plate-shaped crystals were obtained under a wide range of sodium citrate concentrations (0.4–0.7 M) at pH 6.4–7.4 using 5–10% MeOH at 20 °C, for a few days. These crystals were brought to BL2S1 at Aichi synchrotron. The crystals were washed briefly in mother liquid containing 30% (v/v) glycerol and flash cooled in a nitrogen gas stream. Data were collected using ADSC Q315r detectors (1.12 Å wavelength, 1° oscillation angle). All datasets were processed with XDS22 (XDSGUI and XDSDME23). ChiAB-FYSFV crystals were in C2221 (a = 71.8 Å, b = 190.2 Å, c = 133.0 Å, α = 90°, β = 90°, γ = 90°). The initial structure was solved by molecular replacement with Phaser-MR (PHENIX24) using ChiA (PDB ID: 1eib) and the chitin-binding domain of ChiB (PDB ID: 1e6n) as template structures. Further modeling and refinement were carried out with COOT25 and phenix.refine (PHENIX24), respectively. Finally, the ChiAB-FYSFV structure was determined at 2.6 Å.

Author Contributions
R.I. conceived and supervised the project. F.K., A.N., and A.V. performed one-pot saturation mutagenesis and deep sequencing data analysis of GFP mutants. A.N. and F.K. performed one-pot saturation mutagenesis, robot-based automated screening, and structural analysis of ChiAB mutants. F.K., A.N., and R.I. wrote the manuscript.

Notes
The authors declare no competing financial interest. The coordinates of ChiAB-FYSFV have been deposited in the protein data bank (PDB) with the PDB ID: 5ZL9.

Acknowledgments
We would like to thank Mayuko Yamamoto and Yasuko Okuni for assisting the experiments. This work was supported in part by JSPS KAKENHI, JP17K19213 (to R.I.) and by the grant of Joint Research by the National Institutes of Natural Sciences (NINS; NINS program No. 01111704) (to R.I.).

Abbreviations
CBM, carbohydrate-binding module; CD, catalytic domain; ChiAB, fusion protein of ChiA and ChiB; GFP, green fluorescent protein; IVA, in vivo assembly for homologous recombination cloning; SLiCE, seamless ligation cloning extract from Escherichia coli

References
(1) Buettner, K.; Hertel, T. C.; Pietzsch, M. Increased thermostability of microbial transglutaminase by combination of several hot spots evolved by random and saturation mutagenesis. *Amino Acids* 2012, 42, 987–996.
(2) Zhang, X.-F.; Yang, G.-Y.; Zhang, Y.; Xie, Y.; Witters, S. G.; Feng, Y. A general and efficient strategy for generating the stable enzymes. *Sci. Rep.* 2016, 6, No. 33797.
(3) Reetz, M. T. Biocatalysis in organic chemistry and biotechnology: past, present, and future. *J. Am. Chem. Soc.* 2013, 135, 12480–12496.
(4) Yin, H.; Pijning, T.; Meng, X.; Dijkhuizen, L.; van Leeuwen, S. S. Engineering of the Bacillus circulans β-Galactosidase Product Specificity. *Biochemistry* 2017, 56, 704–711.
(5) Wu, N. C.; Grande, G.; Turner, H. L.; Ward, A. B.; Xie, J.; Lerner, R. A.; Wilson, I. A. In vitro evolution of an influenza broadly neutralizing antibody is modulated by hemagglutinin receptor specificity. *Nat. Commun.* 2017, 8, No. 15371.
(6) Jiao, L.; Liu, Y.; Zhang, X.; Liu, B.; Zhang, C.; Liu, X. Site-saturation mutagenesis library construction and screening for specific broad-spectrum single-domain antibodies against multiple Cry1 toxins. *Appl. Microbiol. Biotechnol.* 2017, 101, 6071–6082.
(7) Hughes, M. D.; Nagel, D. A.; Santos, A. F.; Sutherland, A. J.; Hine, A. V. Removing the Redundancy From Randomised Gene Libraries. *J. Mol. Biol.* 2003, 331, 973–979.
(8) Tang, L.; Gao, H.; Zhu, X.; Wang, X.; Zhou, M.; Jiang, R. Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques* 2012, 52, 149–158.
(9) Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* 2013, 2, 83–92.
(10) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Aebersold, R. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 2009, 6, 343–345.
(11) Zhu, B.; Cai, G.; Hall, E. O.; Freeman, G. J. In-fusion assembly: seamless engineering of multidomain fusion proteins, modular vectors, and mutations. *BioTechniques* 2007, 354–359.
(12) Wrenbeck, E. E.; Klessmit, J. R.; Stapleton, J. A.; Adeniran, A. T.; Tyo, K. E. J.; Whitehead, T. A. Plasmid-based one-pot saturation mutagenesis. *Nat. Methods* 2016, 13, 928–930.
(13) García-Nafría, J.; Watson, J. F.; Greger, I. H. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci. Rep.* 2016, 6, No. 27459.
(14) Zhang, Y.; Werling, U.; Edelmann, W. SLiCE: a novel bacterial cell extract-based DNA cloning method. *Nucleic Acids Res.* 2012, 40, No. e55.
(15) Motohashi, K. A simple and efficient seamless DNA cloning method using SLiCE from *Escherichia coli* laboratory strains and its application to SLiP site-directed mutagenesis. *BMC Biotechnol.* 2015, 15, 47.
(16) Okegawa, Y.; Motohashi, K. Evaluation of seamless ligation cloning extract preparation methods from an *Escherichia coli* laboratory strain. *Anal. Biochem.* 2015, 486, 51–53.
(17) Motohashi, K. Evaluation of the efficiency and utility of recombinant enzyme-free seamless DNA cloning methods. *Biochem. Biophys. Rep.* 2017, 9, 310−315.

(18) Nakamura, A.; Tasaki, T.; Okuni, Y.; Song, C.; Murata, K.; Kozai, T.; Hara, M.; Sugimoto, H.; Suzuki, K.; Watanabe, T.; Uchihashi, T.; Noji, H.; Iino, R. Rate constants, processivity, and productive binding ratio of chitinase A revealed by single-molecule analysis. *Phys. Chem. Chem. Phys.* 2018, 20, 3010−3018.

(19) Igarashi, K.; Uchihashi, T.; Uchiyama, T.; Sugimoto, H.; Wada, M.; Suzuki, K.; Sakuda, S.; Ando, T.; Watanabe, T.; Samejima, M. Two-way traffic of glycoside hydrolase family 18 processive chitinases on crystalline chitin. *Nat. Commun.* 2014, 5, No. 3975.

(20) Firnberg, E.; Ostermeier, M. PFunkel: efficient, expansive, user-defined mutagenesis. *PLoS One* 2012, 7, No. e52031.

(21) Watanabe, N.; Nagae, T.; Yamada, Y.; Tomita, A.; Matsugaki, N.; Tabuchi, M. Protein crystallography beamline BL2S1 at the Aichi synchrotron. *J. Synchrotron Radiat.* 2017, 24, 338−343.

(22) Kabsch, W. XDS. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 125−132.

(23) Legrand, P. XDSME: XDS Made Easier. *GitHub repository.* 2017, DOI: 10.5281/zenodo.837885.

(24) 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.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 213−221.

(25) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 2010, 66, 486−501.