Biased mutation-assembling: an efficient method for rapid directed evolution through simultaneous mutation accumulation

Norio Hamamatsu1, Takuyo Aita2,3, Yukiko Nomiya1, Hidefumi Uchiyama1, Motowo Nakajima1, Yuzuru Husimi2 and Yasuhiko Shibanaka1,4

1 Tsukuba Research Institute, Novartis Pharma KK, Ohkubo 8, Tsukuba 300-2611 and 2 Department of Functional Materials Science, Saitama University, Saitama 338-8570, Japan
3 Present address: Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, 2-43 Aomi, Koto-ku, Tokyo 135-0064, Japan
4 To whom correspondence should be addressed.
E-mail: yasuhiko.shibanaka@pharma.novartis.com

We have developed an efficient optimization technique, ‘biased mutation-assembling’, for improving protein properties such as thermostability. In this strategy, a mutant library is constructed using the overlap extension polymerase chain reaction technique with DNA fragments from wild-type and phenotypically advantageous mutant genes, in which the number of mutations assembled in the wild-type gene is stochastically controlled by the mixing ratio of the mutant DNA fragments to wild-type fragments. A high mixing ratio results in a mutant composition biased to favor multiple-point mutants. We applied this strategy to improve the thermostability of prolyl endopeptidase from Flavobacterium meningosepticum as a case study and found that the proportion of thermostable mutants in a library increases as the mixing ratio was increased. If the proportion of thermostable mutants increases, the screening effort increased as the mixing ratio was increased. If the proportion of thermostable mutants in a library is constructed on the basis of the overlap extension PCR technique (Pogulis et al., 1996), the mutant library contains every conceivable combination of advantageous mutations at a stochastically even frequency of occurrence. Thus, the directed evolution strategies that have been developed recently focus on expanding mutant diversity in a library by exploring as large a region in sequence space as possible and with an equal probability of mutation incorporation (Zha et al., 2003).

Wells proposed that mutational effects on the properties of biopolymers are almost independent and additive (Wells, 1990). Dill called this concept the additivity principle (Dill, 1997). Many studies, including our own (Aita et al., 2000, 2002), support these principles (Takeda et al., 1989; Houghten et al., 1991; Lowman and Wells, 1993). Therefore, it is tempting to design mutants that incorporate all known advantageous mutations. However, certain mutations (which are advantageous if they are the only mutation present in a protein) diminish each other’s physiochemical effects (the epistatic effect; Kauffman and Macready, 1995; Aita et al., 2002). Therefore, it is desirable that many component mutations are assembled but a few combinations of the mutations are removed to avoid their epistatic effect. Here, we report an original directed evolution strategy that biases the creation of mutant libraries to favor the presence of multiple advantageous mutations within a mutant. This strategy, biased mutation-assembling, uses the technical framework of mutation scrambling (Uchiyama, 1999) and the theoretical one developed previously (Aita and Husimi, 2000).

A schematic of biased mutation-assembling is depicted in Figure 1. The method consists of the following steps:

1. Initially, for a given gene, advantageous single-point mutations are identified using a random mutagenesis technique. These are then combined to produce new mutants containing multiple mutations as described below.
2. The gene sequence is mapped into blocks, such that a block contains, by definition, one mutation. Blocks are also designed so that the ends of each block overlap the ends of adjacent blocks. DNA fragments, corresponding to the block sequences, are amplified from both wild-type and mutant genes. Note that each block ideally incorporates a single mutation; however, it is inevitable that more than one mutation will be present if the mutations are close in sequence.
3. All DNA fragments are mixed in a reaction tube and recombinant genes are produced using overlap extension PCR without primers. This reaction produces full-length

Introduction

Directed evolution is a useful approach for improving a protein’s physicochemical fitness, e.g. its thermostability. Various currently used enzymatic industrial catalysts and pharmaceutical products were developed using directed evolution techniques (Neylon, 2004), with the DNA shuffling method finding widespread developmental applicability (Stemmer, 1994a,b). However, since wild-type fragments are more included than the mutant fragments in a recombination reaction, reconstituted genes in a DNA shuffling library have only a few mutations, which leads to only small diversity of the mutant library. Ness et al. (2002) developed the synthetic shuffling method, which increases mutant diversity in a DNA shuffling library. In this approach, instead of double-stranded DNA fragments produced by random DNase digestion, ingeniously designed single-stranded DNAs are used for the production of mutant genes. Kikuchi et al. (2000) also reported an effective family shuffling technique, with single-stranded DNAs randomly digested by DNase. We reported ‘mutation scrambling’ as our original optimization strategy (Uchiyama, 1999), which was constructed on the basis of the overlap extension PCR technique (Pogulis et al., 1996). The mutant library contains every conceivable combination of advantageous mutations at a stochastically even frequency of occurrence. Thus, the directed evolution strategies that have been developed recently focus on expanding mutant diversity in a library by exploring as large a region in sequence space as possible and with an equal probability of mutation incorporation (Zha et al., 2003).

Wells proposed that mutational effects on the properties of biopolymers are almost independent and additive (Wells, 1990). Dill called this concept the additivity principle (Dill, 1997). Many studies, including our own (Aita et al., 2000, 2002), support these principles (Takeda et al., 1989; Houghten et al., 1991; Lowman and Wells, 1993). Therefore, it is tempting to design mutants that incorporate all known advantageous mutations. However, certain mutations (which are advantageous if they are the only mutation present in a protein) diminish each other’s physiochemical effects (the epistatic effect; Kauffman and Macready, 1995; Aita et al., 2002). Therefore, it is desirable that many component mutations are assembled but a few combinations of the mutations are removed to avoid their epistatic effect. Here, we report an original directed evolution strategy that biases the creation of mutant libraries to favor the presence of multiple advantageous mutations within a mutant. This strategy, biased mutation-assembling, uses the technical framework of mutation scrambling (Uchiyama, 1999) and the theoretical one developed previously (Aita and Husimi, 2000).

A schematic of biased mutation-assembling is depicted in Figure 1. The method consists of the following steps:

1. Initially, for a given gene, advantageous single-point mutations are identified using a random mutagenesis technique. These are then combined to produce new mutants containing multiple mutations as described below.
2. The gene sequence is mapped into blocks, such that a block contains, by definition, one mutation. Blocks are also designed so that the ends of each block overlap the ends of adjacent blocks. DNA fragments, corresponding to the block sequences, are amplified from both wild-type and mutant genes. Note that each block ideally incorporates a single mutation; however, it is inevitable that more than one mutation will be present if the mutations are close in sequence.
3. All DNA fragments are mixed in a reaction tube and recombinant genes are produced using overlap extension PCR without primers. This reaction produces full-length
Fig. 1. A schematic example of biased mutation-assembling, assuming a basis set of three mutations. The circle, triangle and square each represent one mutation. A block represents a portion of the gene containing one mutation and represents a recombination unit. The double-headed arrows represent overlapping sequences between adjacent blocks and these overlapping sequences hybridize during PCR recombination.

4. Mutants with the desired properties, owing to incorporation of multiple mutations and without epistatic mutations, are then identified.

This protocol defines the optimization algorithm for biased mutation-assembling in binary n-dimensional sequence space. The probability of incorporating a mutation into a site is the bias, r, which is the fraction of all mutant DNA fragments containing the mutation that defines the block to the total DNA concentration present during recombination:

$$r = \frac{[\text{mutant DNA fragments}]}{[\text{mutant DNA fragments}] + [\text{wild-type DNA fragments}]}$$  (1)

where [mutant DNA fragments] and [wild-type DNA fragments] are the concentrations of the DNA fragments with and without the defining mutation in the recombination mixture, respectively. The extent to which mutations accumulate is stochastically controlled by the r value (0 ≤ r ≤ 1), because a complete gene will be reconstituted only at overlapping ends of adjacent DNA fragments. Let n be the number of component mutations. The library, constructed with an r value of 0 or 1.0 contains only wild-type gene or only mutant gene with all n mutations, respectively. Libraries constructed with intermediate r values (0 < r < 1) will contain mutant genes with an average of approximately rn mutations. The probability of the number of accumulated mutations, d, obeys the binomial distribution: $\binom{n}{d}p^d(1 - p)^{n - d}$. In constructing assembling libraries, the r value should be large (near one, but not equal to one), so that mutants that incorporate many advantageous mutations, but not epistatic mutations, could be found. Because libraries, with large r values also contain few mutants that have a small number of mutations, the necessary screening effort required to find mutants with the most desirable properties is reduced (Aita and Husimi, 2000).

Prolyl endopeptidase (PEP) (prolyl oligopeptidase, EC 3.4.21.26) is the only endopeptidase known to show substrate specificity for proline and therefore it is expected to be applicable as a catalyst for pharmaceutical production, e.g. coupling reactions of biologically active peptides, such as C-terminal amidation of luteinizing hormone-releasing hormone (LH-RH) (Togame et al., 1994). However, PEP is susceptible to inhibition and inactivation under conditions for coupling reactions, such as the presence of organic solvents and high temperatures (Kreig and Wolf, 1995). Uchiyama et al. (2000) suggested that the thermostabilization (prolongation of activity half-life at high temperature) had a beneficial effect on the stability in the coupling reaction. Here, we describe the biased mutation-assembling protocol and the application of the strategy to the prolongation of the activity half-life of PEP from Flavobacterium meningosepticum at high temperature (thermostabilization of PEP) as a case study. We developed a basis set of 14 mutations that individually improved the thermostability of PEP and used these mutations to prepare assembling libraries with r values of 0.3, 0.5, 0.7, 0.8, 0.9 and 1.0. We found that the r value of a library directly correlates with the number of thermostable mutants in the library. After assaying only 2000 mutants, we found one, with the library with r value of 0.9, that has a 1200-fold longer activity half-life than the wild-type enzyme at 60°C. This mutant contains 12 mutations and lacks an epistatic pair.

Materials and methods

Development of a mutation basis set to improve PEP thermostability

We previously constructed the vector pUK-FPEPb to express PEP in Escherichia coli (Uchiyama et al., 2000). This vector contains the PEP gene, which is inserted between EcoRI and PstI restriction sites (Figure 2). The gene was amplified using error-prone polymerase chain reaction (PCR) as described previously (Uchiyama et al., 2000), with experimental conditions optimized so that only one or at most a few nucleotide substitutions were introduced (Leung et al., 1989). Amplified DNAs were digested with EcoRI and PstI and then ligated into pUK-FPEPb vector, which was digested with the same enzymes. The vectors were transformed into E.coli JM109.
and randomly chosen transformants were screened for PEP thermostability using an active-staining method (Uchiyama et al., 2000). Mutations that were responsible for the thermostabilization were identified with DNA sequencing.

**Construction of PEP biased mutation-assembling libraries**

We found 14 single-point mutations, S19T, E67Q, F70L, S110F, N387K, A388V, S475G, I493V, E496K, N542I, K615R, G652V, S653A and Q656R, which individually improved PEP thermostability at 57.5 °C. These mutations form the basis set for assembling libraries reported here. Since E67Q–F70L, N387K–A388V, I493V–E496K and G652V–S653A–Q656R are each close in sequence, we constructed libraries by modifying the protocol shown in Figure 1. The PEP gene was divided into six blocks, labeled I–VI, and the corresponding segments contained in the primers; for example, 67Q/70L-f is the forward primer encoding the 67Q and 70L mutations.

**Preparation of periplasmic lysates containing PEP mutants**

In our expression system of PEP using pUK-FPEPb and E.coli JM109, PEP is found not only in the cytoplasmic compartment of E.coli, but also in the periplasmic compartment (Uchiyama et al., 2000). As shown by Beacham (1979) and Swamy and Goldberg (1982), the periplasmic space of E.coli has a limited

<table>
<thead>
<tr>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Template</th>
<th>Mutation on fragment I</th>
<th>i</th>
<th>Mixing fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>EcoRI-396-f</td>
<td>67Q/70F-r</td>
<td>Wild-type</td>
<td>E67Q, F70L</td>
<td>2</td>
<td>0.081</td>
</tr>
<tr>
<td>EcoRI-396-f</td>
<td>67Q/70L-r</td>
<td>S19T, E67Q, F70L</td>
<td>0.099</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRI-396-f</td>
<td>67Q/70F-r</td>
<td>None</td>
<td>1</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>EcoRI-396-f</td>
<td>67Q/70L-r</td>
<td>S19T, E67Q, F70L</td>
<td>0.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRI-396-f</td>
<td>67Q/70F-r</td>
<td>S19T, E67Q, F70L</td>
<td>0.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EcoRI-396-f</td>
<td>67Q/70L-r</td>
<td>S19T, E67Q, F70L</td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The mixing fraction is calculated as $r \times (1-r)^{i-1}$, where $r = 0.9$ and $i$ is the number of mutations contained in each DNA fragment.

![Fig. 2. Block and fragment design for the PEP assembling library. The open circles represent the I4 mutations that form the basis set. The PEP gene is divided into six blocks and the corresponding sets of DNA fragments are numbered I–VI. Fragments I–V include the mutants S19, S110, S475, N542 and K615 near the sequence midpoint, respectively. Arrows represent the primers used to prepare the fragments. The open triangles on the arrows represent mutations located on the sequence midpoint, respectively. Arrows represent the primers used to prepare each DNA fragment, and also reconstitute genes. For example, primer 67Q/70L-r is the reverse primer coding both E67Q and F70L mutations.](Image)
number of proteases, only seven out of the 25 known cellular proteases, and comprises only 4–8% of the total cell protein. Also, we have reported the isolation of thermostable PEP mutants using whole cell lysate of E.coli JM109 (Uchiyama et al., 2000). In addition, we could not find any remarkable decrease in the PEP activity in the periplasmic lysate over 15 min even at 50°C. For these reasons, we considered that periplasmic lysates could be used for the evaluation of PEP thermostability with less proteolytic degradation effect. We prepared periplasmic lysates using a modified method reported by French et al. (1996) as follows: transformants, which formed colonies on LB agar, were randomly selected and then cultured in 200 μl of LB liquid medium containing 100 μg/ml ampicillin in 96 deep-well culture plates at 37°C for 14 h with shaking at 200 r.p.m. The transformants were harvested by centrifugation at 10,000 g for 5 min and washed with 200 μl of 20 mM potassium phosphate buffer (pH 7.0). The cells were centrifuged again, resuspended in 75 μl of osmotic shock buffer [100 mM Tris–HCl (pH 8.0), 0.5 M sucrose, 5 mM EDTA, 0.16 mg/ml lysozyme] and incubated on ice for 2 min. Next, 75 μl of cold distilled water were added and the incubation was continued on ice for 30 min. After centrifugation at 10,000 g for 5 min, the supernatant was used as the periplasmic lysates for evaluation of PEP activity half-life.

Evaluation of PEP mutant thermostability

Volumes of 50 μl of each periplasmic lysate were kept at 57.5 or 60°C for 5, 10 or 15 min, then PEP activity was assayed and, assuming first-order kinetics, a PEP activity half-life (t½) was calculated. The screening of mutants was conducted at 57.5°C, at which nearly 90% of the wild-type activity was lost during a 15 min heat treatment. Since highly thermostable mutants showed the great thermostability and showed very long half-lives with a large measurement error at 57.5°C, we evaluated such mutants by adopting higher temperature, 60°C, for precise measurement of their half-lives. For the PEP activity assay, 10 μl of a periplasmic lysate were added to 140 μl of PEP assay buffer [100 mM potassium phosphate (pH 7.0), 100 μg/ml bovine serum albumin, 1 mM dithiothreitol, 0.2 mM Z-Gly-Pro-pNA; Bachem] and the absorbance at 410 nm was monitored at 30°C. The PEP activity half-life is independent of total protein concentration of lysate and the initial lysate PEP activity (data not shown); consequently, it was not necessary to standardize or correct for protein concentration in the assay mixtures.

Results

Identification of advantageous mutations that improved PEP thermostability

Initially, using an error-prone PCR technique and active-staining method (Uchiyama et al., 2000), we found 14 mutations that individually improved PEP thermostability, namely S19T, E67Q, F70L, S110F, N387K, A388V, S475G, I493V, E496K, N542I, K615R, G652V, S653A and Q656R. The PEP activity half-lives at 57.5°C for mutants which have one of these mutations are longer than that of the wild-type enzyme (Aita et al., 2002). These 14 mutations form the basis set for the biased mutation-assembling libraries. As shown in Figure 3, the mutations were mapped to the corresponding positions of the three-dimensional porcine muscle PEP structure (Fülöp et al., 1998). With the exception of S19T, for which a homologous site does not exist, many of these mutations are spatially located near the N- and C-termini. Mutations near the termini might decrease thermal motion, thereby stabilizing the catalytic domain.

Evaluation of PEP thermostability for assembling library mutants

Approximately 150–300 transformants were randomly picked from each assembling library prepared with individual r values of 0.3, 0.5, 0.7, 0.8 and 0.9. The PEP activity half-lives at 57.5°C for the periplasmic lysates of these transformants were determined. A proportion of transformants that showed markedly low PEP activity ranged from 10 to 20%, with these transformants eliminated from the evaluation. The distribution of the PEP activity half-lives for each library is presented as scatter plots in Figure 4a, in which each individual plot in a column represents a mutant in each assembling library. Figure 4b shows the maximum activity half-lives for the mutants in each library. The half-lives of wild-type PEP and the mutant with all 14 mutations are included in Figure 4a and b. Clearly, a large r value correlates directly with increased production of mutants with greatly improved thermostability. The most thermostable mutant among the 264 samples in the library prepared with r = 0.9 shows an ~1600-fold longer activity half-life at 57.5°C than wild-type PEP. However, when r = 1.0, the mutant with all 14 mutations is less thermostable than the most thermostable one and shows at most a 38-fold longer half-life than wild-type PEP. That the mutant containing all 14 mutations does not have a greatly increased thermostability is probably a consequence of the epistatic mutation pair. As the r value increases, the activity half-lives of the majority of mutants converge on approximately that of the mutant with
14 mutations. These results suggest that the mutational effect among the 14 mutations is almost additive, but that there are a few significantly epistatic pairs that cause a diminishing return in PEP thermostability.

**Estimation of screening size required to find the fittest mutant**

Figure 4 indicates that a biased library with a large $r$ value should reduce the number of samples that must be assayed when the fittest mutants in binary $n$-dimensional sequence space are to be found. Since assembling libraries have a binomial distribution of mutant genes, we can estimate the minimum number of assays required to find the mutant with the best mutation combination. The probability, $P$, that a particular mutant with $d$ point mutations exists among $N$ samples picked randomly from an assembling library with a given $r$ value is

$$P = 1 - \left[ 1 - r^d (1 - r)^{n - d} \right]^N$$

where $n$ is the number of mutations in the basis set and is 14 for the PEP assembling libraries. If a numerical value is assigned to $P$, $N$ can be calculated according to this equation. When $d = 12$ and $P = 0.999$, $N$ is determined as 113 200, 5500, 2500 and 2400 for $r$ values of 0.5, 0.7, 0.8 and 0.9, respectively. For these calculations, we assumed that the most thermostable mutant would have 12 mutations because our previous studies indicated that the fittest mutant contained ~80–90% of the basis set mutations. Our calculations suggest that the number of assays needed to find mutants with maximum thermostability is dramatically decreased as the $r$ value increases. To confirm this prediction, we screened 2000 and 40 000 transformants selected randomly from assembling libraries with $r$ values of 0.9 and 0.5, respectively. Table II lists the top four thermostable PEP mutants found in the library with an $r$ value of 0.9 and the top two thermostable mutants found in that with an $r$ value of 0.5. The PEP activities of these mutants under physiological conditions (37°C) are similar to that of wild-type PEP. Mutant 22, the most thermostable mutant found in the library with an $r$ value of 0.9, shows a 1200-fold longer activity half-life at 60°C than wild-type PEP. The activity half-life of mutant 22 is significantly longer than that of mutant 7, which is the most thermostable mutant found in the library with an $r$ value of 0.5, in spite of the fact that 38 000 more transformants were screened in the latter library. The top four mutants found in the library with an $r$ value of 0.9 contain almost all component mutations, but do not contain Q656R. It is known that if E67Q and Q656R are simultaneously incorporated in a sequence, this pair, at least, causes an epistatic effect and reduces the thermostability drastically.

**Table II.** The mutation composition and PEP activity half-life at 60°C for the most thermostable PEP mutant found in assembling libraries with $r$-values of 0.9 and 0.5

<table>
<thead>
<tr>
<th>Bias $r$</th>
<th>Mutant</th>
<th>S19T</th>
<th>E67Q</th>
<th>F70L</th>
<th>S110F</th>
<th>N387K</th>
<th>A388V</th>
<th>S475G</th>
<th>I493V</th>
<th>E496K</th>
<th>N542I</th>
<th>K615R</th>
<th>G652V</th>
<th>S653A</th>
<th>Q656R</th>
<th>$t_{1/2}$ (min)$^a$</th>
<th>Relative stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 0.9$</td>
<td>Mutant 22$^b$</td>
<td>×</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>170 ± 28**</td>
<td>1200</td>
</tr>
<tr>
<td>(2000 clones)</td>
<td>Mutant 57</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>120 ± 24</td>
<td>880</td>
</tr>
<tr>
<td>Mutant 9</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>73 ± 12</td>
<td>520</td>
<td></td>
</tr>
<tr>
<td>Mutant 28</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>96 ± 24</td>
<td>690</td>
</tr>
<tr>
<td>$r = 0.5$</td>
<td>Mutant 7</td>
<td>○</td>
<td>×</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>70 ± 10</td>
<td>500</td>
</tr>
<tr>
<td>(40 000 clones)</td>
<td>Mutant 45</td>
<td>○</td>
<td>×</td>
<td>×</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>70 ± 10</td>
<td>500</td>
</tr>
<tr>
<td>Reference</td>
<td>Wild-type</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>0.14 ± 0.01</td>
<td>1</td>
</tr>
<tr>
<td>14 mutations$^c$</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>12 ± 0.8</td>
<td>88</td>
</tr>
</tbody>
</table>

$^a$The reported half-lives are the mean value ± standard deviation, $n = 7$.

$^b$Mutant 22 also contains a neutral mutation V64I, which may have been introduced during PCR.

$^c$The properties of the mutant with all 14 mutations are given in the last line of the table.

$**P < 0.01$ when comparing mutant 22 with mutant 7.
(Uchiyama et al., 2000; Aita et al., 2002). This result strongly demonstrated that our strategy both reduced the required screening effort and found what may be the best combination of mutations, while avoiding epistatic effects. Especially the reduced screening effort is beneficial in cases where the number of conceivable combinations of mutations is large.

Discussion

Biased mutation-assembling and other methods, e.g. DNA shuffling (Stemmer, 1994a,b), are similar in concept: novel mutants, containing several mutations, are produced by recombination of single- or double-stranded DNA fragments derived from wild-type and mutant or homologous genes. DNA shuffling is a simple and effective method for preparing a chimera gene library; however, mutant diversity in shuffling libraries tends to be small. Various modified methods expand the mutant population, while leveling the concentrations of individual mutants by more thoroughly exploring sequence space. Our method of biased mutation-assembling also produces a library with a large diversity; however, the method biases the content to favor multiple-point mutants. Biased mutation-assembling depends on the assumption that additivity principles are valid and the results in Table II are consistent with this assumption.

For biased mutation-assembling, the incorporation frequency for each mutation is stochastically controlled by the mixing fraction (which is determined by the $r$ value) of the mutant DNA fragments present during recombination. Our previously reported mutation scrambling corresponds to the case with an $r$ value of 0.5, not importing the additivity principles. Therefore, the library contains every conceivable combination of advantageous mutations at a stochastically even frequency of occurrence (Uchiyama, 1999). In this study, the library of PEP mutants, for which all mixing fractions were calculated using an $r$ value of 0.9, contained many thermostable PEP mutants and consequently the search for such mutants was significantly reduced. For other methods, such as DNA shuffling, assuming that all mutations are introduced independently in a binary manner, the $r$ value is the reciprocal of the number of mutations, $r = 1/n$. If we had used such a method to find thermostable PEP mutants, the $r$ value would have been, at most, 0.07 (1/14) and the screening size $4.5 \times 10^{14}$. However, when DNA fragments are prepared by DNase digestion, mutations, especially close in sequence, usually remain linked and recombine together; therefore, the $r$ value would almost certainly have been smaller and the number of assays required larger. Synthetic shuffling (Ness et al., 2002), which is one of modified DNA shuffling methods, can also produce a biased library, but each mutation was introduced with an approximately equal probability of incorporation. It will be interesting to integrate additivity principles into this method. Additionally, for biased mutation-assembling, the incorporation probability of an individual mutation can be controlled by adjusting the value of the mixing fraction for each fragment containing a given mutation, although, in this study, the same $r$ value was used for all 14 mutations. For example, if a mutation has a large (or a small) beneficial effect on a desired property, the $r$ value in the equation mixing fraction $r = (1 - r)^{m-1}$ should be set near 1.0 (or near 0.5). We have used this strategy to improve the substrate specificity of glucose dehydrogenase-B (GDHB) (EC 1.1.5.2) from *Acinetobacter calcoaceticus* for glucose. This enzyme shows broad substrate specificity that, especially the activity against maltose, greatly diminishes its utility. A set of mutations was identified that decreased the activity against maltose. Some of these mutations also decreased that against glucose and some of the mutations did not. Using an $r$ value of 0.5 for the first subset of mutations and an 0.9 for the second subset, we were able to decrease specifically the activity against maltose by 4% of wild-type GDHB without significantly affecting the activity against glucose (retaining 40% of wild-type GDHB). Consequently, we found what is probably the best combination of mutations resulting in improved specificity for glucose (manuscript in preparation).

Meanwhile, biased mutation-assembling requires a lot of primers and PCR reaction for the preparation of fragments when a lot of mutations are closely located as described here, requiring a lot of work compared with other methods, such as DNA digestion. However, in cases where component mutations are scattered enough to design individual blocks to have only one mutation as shown in Figure 1, biased mutation-assembling is a very simple method like overlap extension PCR.

A biased library is also produced if the wild-type backcross strategy, for which a mutant with all mutations is recombined with wild-type DNA, is used (Zhao and Arnold, 1997). However, as discussed above, mutations, especially close in sequence, are usually linked during recombination, which makes the control of the incorporation probability of individual mutations virtually impossible. Additionally, this strategy requires the advanced preparation of a gene containing all desired mutations. For these reasons, the wild-type backcross strategy is a less efficient way to prepare a biased library than is biased mutation-assembling.

A crucial element for the successful design of biased mutation-assembling libraries is the composition of the mutation basis set. We found the basis set for the PEP assembling library using error-prone PCR. However, when error-prone PCR is the chosen method, codon degeneracy limits the extent to which the amino acid substitution occurs and, therefore, all desirable mutations might not be found. This limitation can be overcome if other methods are used to produce point mutations. For example, Murakami et al. developed a method that introduces mutations into a gene without the limitations imposed by codon degeneracy (Murakami et al., 2002). This powerful method exhaustively identifies single-point mutations that improve the physiochemical properties of a protein. Combining this method with biased mutation-assembling may make directed evolution experiments more efficient.

In our study, an $r$ value of 0.9 most effectively improved the thermostability of PEP; however, in general, the optimal $r$ value depends on how rugged the fitness landscape is for a particular protein (Aita et al., 2000). Previously, we roughly estimated that the optimal $r$ value is between 0.8 and 0.9 for most proteins (Aita and Hsusi, 2000; Aita et al., 2000, 2001, 2002). Additional characterization of protein fitness landscapes should improve the estimate of the statistically optimal $r$ value.

In summary, we have shown that biased mutation-assembling is an efficient optimization technique for improving protein properties, through applying this strategy to the improvement of the thermostability of PEP as a case study. To show that biased mutation-assembling is a generally applicable method, we have produced libraries for several other
enzymes and found mutants with improved thermostability and substrate specificity. Additional studies of other proteins will be needed to validate this method further. The statistical and theoretical nature of protein fitness landscapes is the foundation for the concept of biased mutation-assembling. We believe that if the knowledge gained from protein fitness landscape calculations is incorporated into new, rapid protein evolution methods, such methods will be robust and reliable.

Acknowledgements

This work was performed as part of the R&D Project of the Industrial Science and Technology Frontier Program supported by the New Energy and Industrial Technology Development Organization, Japan.

References


Received October 22, 2004; revised March 30, 2005; accepted April 11, 2005

Edited by Alan Fersht