Construction of stabilized proteins by combinatorial consensus by mutagenesis

N. Amin1, A. D. Liu1, S. Ramer1, W. Aehle2, D. Meijer2, M. Metin2, S. Wong1, P. Gualfetti1 and V. Schellenberger1,3

1Genencor International, 925 Page Mill Road, Palo Alto, CA 94304, USA and 2Genencor International BV, Leiden, The Netherlands
3To whom correspondence should be addressed.
E-mail: vschellenberger@genencor.com

We constructed stabilized variants of β-lactamase (BLA) from Enterobacter cloacae by combinatorial recruitment of consensus mutations. By aligning the sequences of 38 BLA homologs, we identified 29 positions where the E. cloacae gene differs from the consensus sequence of lactamases and constructed combinatorial libraries using mixtures of mutagenic oligonucleotides encompassing all 29 positions. Screening of 90 random isolates from these libraries identified 15 variants with significantly increased thermostability. The stability of these isolates suggests that all tested mutations make additive contributions to protein stability. A statistical analysis of sequence and stability data identified 11 mutations that made stabilizing contributions and eight mutations that destabilized the protein. A second-generation library recombining these 11 stabilizing mutations led to the identification of BLA variants that showed further stabilization. The most stable variant had a midpoint of thermal denaturation (Tm) that was 9.1°C higher than the starting molecule and contained eight consensus mutations. Incubation of three stabilized BLA variants with several proteases showed that all tested isolates have significantly increased resistance to proteolysis. Our data demonstrate that combinatorial consensus mutagenesis (CCM) allows the rapid generation of protein variants with improved thermal and proteolytic stability. Keywords: combinatorial mutagenesis/consensus mutation/ lactamase/proteolysis/thermostability

Introduction

The stabilization of proteins based on their crystal structure requires precise calculation of small contributions of mutations to protein folding and stability. This limits our ability to design stabilized proteins based on structural information alone (Nicholson et al., 1988; Braxton and Wells, 1992; Gallagher et al., 1993; Shi et al., 1993; Strausberg et al., 1993). An alternative strategy is the stabilization of proteins by directed evolution, which has been applied successfully to several proteins (Cunningham and Wells, 1987; Chen and Arnold, 1993; Matsumura et al., 1999; Miyazaki et al., 2000). In general, stabilizing mutations occur at low frequency and evolutionary approaches require the screening of large populations of random mutants. Alternatively, one can enrich stabilized variants of proteins by surface display (Kristensen and Winter, 1998; Sieber et al., 1998; Jung et al., 1999; Shusta et al., 1999), selection in thermostable hosts (Hoseki et al., 1999) or other selection protocols. The efficiency of directed evolution can be improved in some cases by constructing random variant libraries via recombination of homologous genes (Stemmer, 1995), by recruitment of individual mutations from homologs (Sandgren et al., 2003) or by constructing synthetic gene libraries (Ostermeier, 2003). However, these methods require the availability of natural isolates that share a suitable level of homology with the parent gene or they require extensive knowledge about structure–function relationships of the parent protein. Here we describe the construction of semi-random libraries based on multiple sequence alignments of a protein with its natural homologs. The resulting combinatorial libraries contain a very large fraction of stabilized mutants, which eliminates the need for selective enrichment or robotics-based techniques and allows for direct identification of stabilized variants by screening a moderate number of variants in microtiter plates.

In nature, protein families have developed as a result of the continuous process of random mutagenesis, recombination and selection, which tends to eliminate most destabilizing mutations (Kimura, 1991). As a result, residues that stabilize a protein tend to be more prevalent than other amino acids at any given position in a protein family. It has been shown for several proteins that consensus mutations, which replace a particular amino acid of a specific protein with the most common amino acid that is present in that position among the family members, frequently lead to stabilized protein variants (Steipe et al., 1994). A particularly striking example is the recently described synthesis of a phytase gene that was based on the consensus sequence of 13 homologous phytase sequences (Lehmann et al., 2000). This consensus gene differed in 20% of the encoded amino acids from the closest natural phytase gene and the encoded protein was more stable towards thermal denaturation than any of its parent proteins. However, in a subsequent study, the same group showed that only a fraction of the consensus mutations contributed to protein stability and that several of the consensus mutations actually destabilized the protein (Lehmann et al., 2002). These results suggest that some consensus mutations should be avoided. Therefore, we developed a method that can rapidly identify stabilizing consensus mutations for a protein of interest. In this paper, we demonstrate that combinatorial consensus mutagenesis (CCM) allows the rapid introduction of multiple stabilizing consensus mutations into a protein while avoiding the introduction of destabilizing mutations. We have applied the process to β-lactamase from Enterobacter cloacae and showed that one-third of all variants in the CCM library were more stable than the parent protein. Regression analysis of sequence and stability information allowed us to identify the most stabilizing consensus mutations, which were added to the first-generation stabilized variant in a subsequent round of combinatorial mutagenesis, resulting in variants with further stabilization.
**Materials and methods**

**Library construction**

All libraries were based on plasmid pCB04, which contains the gIII BLa sequence of phage M13 fused to the mature sequence of BLA from *E.coli* driven by a lac promoter. The amino acid sequence of BLA from *E.coli* (Figure 1) was obtained from Genbank. The gene encoding this sequence which was codon optimized for *Escherichia coli* was synthesized by Aptagen. The sequence GGGAETVEHHHHHH was added to the C-terminus of BLA as a purification handle. Plasmid pCB04 is based on pUC19 and contains a chloramphenicol resistance marker.

The QuikChange multi site-directed mutagenesis kit (QCMS) (Stratagene) was used to construct combinatorial libraries NA01, NA02 and NA03 using 29 mutagenic primers. The primers were designed so that they had 17 homologous bases flanking each side of the mutagenic codon. The mutagenic codon was chosen based on an *E.coli* codon usage table. All primers were designed to anneal to the same strand of the template DNA (i.e. all were forward primers). The QCMS reaction was carried out as recommended by the manufacturer with the exception of the primer concentration, which was lowered to 4 or 0.4 μM total primer concentration. The individual primer concentrations in each library were NA01, 0.14 μM or ~40 ng each of 29 primers; NA02 and NA03, 0.014 μM or ~4 ng of each primer.

For each library, the reactions contained 50–100 ng of template plasmid pCB04, 1 μl of primer mix (100 or 10 μM stock of combined primers to give the desired primer concentrations mentioned above), 1 μl of dNTPs, 2.5 μl QCMS reaction buffer, 18.5 μl of deionized water and 1 μl of enzyme blend (QCMS kit), for a total volume of 25 μl. The thermo-cycling program was 1 cycle at 95°C for 1 min, followed by 30 cycles of 95°C for 1 min, 55°C for 1 min and 65°C for 10 min. DpnI digestion was performed by adding 1 μl of DpnI (provided in the QCMS kit) incubation at 37°C for 2 h, addition of 0.5 μl of DpnI and incubation at 37°C for an additional 2 h. A 1 μl volume of each reaction was transformed into 50 μl of TOP10 electrocompetent cells (Invitrogen). A 250 μl volume of SOC was added after electroporation, followed by 1 h of incubation with shaking at 37°C. Following incubation, 10–50 μl of the transformation mix were plated on LA plates with 5 mg/l chloramphenicol (CMP) or LA plates with 5 mg/l CMP and 0.1 mg/l cefotaxime (CTX) for selection of active BLA clones. The number of colonies obtained on both types of plates was comparable for all libraries. For the preparation of libraries NA02 and NA03 we carried out a single QCMS reaction, which was subsequently split into two portions after DpnI digestion. One portion, NA02, was transformed directly into *E.coli* and the second portion, NA03, was heated at 95°C for 2 min before transformation into *E.coli*. This experiment was conducted to determine if denaturation of hemimethylated DNA after DpnI digestion would reduce the fraction of parent clones in the resulting library.

**Screening for BLA stability**

Libraries NA01, NA02 and NA03 were plated on to agar plates with LA medium containing 5 mg/l chloramphenicol. No CTX was added to these plates to allow sequencing of isolates without BLA activity. Thirty colonies from each library were transferred into the wells of one 96-well plate containing 200 μl of LB with 5 mg/l CMP. Four additional wells were inoculated with TOP10/pCB04, which served as the wild-type control during the assay. A master plate was generated by adding glycerol and was stored at ~80°C.

A 96-well plate containing 200 μl per well of LB with 5 mg/l CMP and 0.1 mg/l CTX was inoculated from the master plate using a repipette tool. The plate was incubated for 3 days at 25°C in a humidified incubator shaking at 225 r.p.m. The following operations were performed with each well of the cultured 96-well plate: 50 μl of culture were transferred into a plate that contained 50 μl of B-PER cell lysis reagent (Pierce). The suspension was incubated at room temperature for 90 min to lyse the cells and release BLA. The lysate was diluted 1000-fold and 10 000-fold into 100 mM citrate−200 mM phosphate buffer pH 7.0 containing 0.125% octyl glucosynside (Sigma). The diluted samples were heated at 56°C for 1 h with mixing at 650 r.p.m. Subsequently, 20 μl of the sample were transferred to 180 μl of nitrocefin assay buffer [0.1 mg/l nitrocefin (Oxoid) in 50 mM phosphate-buffered saline containing 0.125% octyl glucosynside] and the BLA activity was determined using a Spectramax Plus plate reader (Molecular Devices) at 490 nm. In parallel, a control sample was subjected to the same procedure but the heating step was omitted. Based on both activity readings the fraction of BLA activity that remained after the heat treatment was calculated for each of the 90 variants and four controls on the plate.

The screen of library NA04 was done following the protocol given above, except that the samples were incubated at 46°C in 50 mM imidazole pH 7.0 containing 10 mM CaCl₂, 0.005% Tween-20 and 0.1 mg/ml thermolysin (Sigma) to test the protease resistance of the variants. Proteolysis was quenched by transferring 40 μl of sample into a fresh plate containing 10 μl of 50 mM EDTA to inactivate thermolysin.

**Statistical analysis**

The contributions of all mutations to stability were calculated using MS Excel. A matrix, Mₖᵢ, was established containing one row for each isolate and one column for each of the 28 observed mutations. The matrix contained the value of one if the particular isolate carried the particular mutation. The logarithm of the remaining activity of each isolate was entered. In a separate
Combinatorial consensus mutagenesis

column, the theoretical stability of each isolate was calculated using Equation 2 and parameters $P_k$ and $C$. The sum of squares of the differences between the measured and the calculated values of $\log R_m$ was calculated for each isolate as the target cell. Using the solver function of Excel, we optimized the parameters $P_k$ and $C$ to minimize the value in the target cell. All parameters were set to zero as initial estimates.

**Purification of BLA variants**

Strains were inoculated into 1 l of Terrific Broth (TB) containing 5 mg/l CMP and incubated at 37°C overnight. Cells were harvested by centrifugation (6000 g for 15 min). The pellets were resuspended in 200 ml of phosphate-buffered B-PER solution (Pierce). The suspension was shaken for 1 h at room temperature until the pellets were solubilized. Cell debris and insoluble protein were removed by centrifugation (15 000 g for 15 min). The supernatant was stored at 4°C until purification.

Proteins were first purified using immobilized metal affinity chromatography (INAC), The purification was done using a Bio-Cat (Applied Biosystems) and a Waters column (22x95 mm) packed with POROS 20MC media (Applied Biosystems). The column was loaded with 250 mM NiCl, washed with water and equilibrated with 10 mM HEPES, 0.5 M NaCl, pH 8.4. Samples were loaded on the column, washed with equilibration buffer and eluted with 10 mM HEPES, 0.5 M NaCl and a gradient of 200 mM imidazole.

The proteins were further purified by affinity chromatography using m-aminophenylboronic acid (PBA) resin (Sigma). This purification was done by gravity flow. A 15 ml amount of PBA resin was packed in a disposable column (15x120 mm) (Bio-Rad Laboratories) and equilibrated with 20 mM triethanolamine, 0.5 M NaCl, pH 7.0. After loading the sample, the columns were washed with four column volumes of equilibration buffer and subsequently BLA was eluted with 0.5 M sodium borate, 0.5 M NaCl, pH 7.0. The resulting proteins had a purity of 99% as judged by SDS–PAGE.

**Thermostability analysis**

Circular dichroism (CD) experiments were performed on an Aviv 62ADS spectrophotometer, equipped with a five-position thermoelectric cell holder supplied by Aviv. Buffer conditions were 10 mM adjusted to pH 7.1 and 50 mM citrate–phosphate buffer conditions were 10 mM adjusted to pH 7.1 and 50 mM citrate–phosphate buffer. Buffer conditions were 10 mM adjusted to pH 7.1 and 50 mM citrate–phosphate buffer. The concentration and type of protein were of particular importance as expected for a mixed a-helix. The theoretical midpoint of the transition ($T_m$) is an apparent value because the thermal denaturation of all BLA variants studied was not reversible. Because of irreversibility and a slow kinetic component caused by aggregation, we chose sufficiently fast scan rates and experiments were carefully controlled for time spent in the thermal denaturation transition, so that small temporal variations did not lead to apparent $T_m$ differences.

**Protease sensitivity**

A 1 μg amount of purified protein was incubated with different concentrations of protease in 150 μl in 0.2 ml strip tubes (MJ Research), 100 mM Tris–HCl, 10 mM CaCl$_2$, 0.005% Tween-20, pH 7.9, for different time periods at 37°C in quadruplicate. Trypsin, chymotrypsin and thermolysin (Sigma) were added at 200–800 μg/ml concentration. The BLA activity was measured for samples incubated with and without protease by monitoring the hydrolysis of the chromogenic substrate nitrcefin (Oxoid). The activity of protease-treated sample relative to the untreated sample was calculated for each variant.

**Results**

**Construction of first-generation CCM libraries**

Using the sequence of the parent gene β-lactamase from *E. cloacae* as a query, we identified 37 homologous sequences in GenBank. These sequences were aligned using the clustalW algorithm in the Vector NTI software package (Informax). The resulting consensus sequence differed in 29 positions from the parent gene. The amino acid sequence of β-lactamase from *E. cloacae* and the 29 consensus mutations are depicted in Figure 1. Based on this information, we designed 29 mutagenic primers such that each primer encoded the replacement of one residue in the parent sequence with the corresponding consensus residue. Using these 29 mutagenic primers, we constructed combinatorial libraries using the QCMS method. To our knowledge, there are no prior reports of this method being applied with such a large number of mutagenic primers. Therefore, we decided to test several variations of the protocol as described in Materials and methods, yielding three libraries, NA01, NA02 and NA03.

**Screening first-generation libraries**

We chose 30 clones randomly from each library for characterization. Sequence analysis showed that the individual isolates contained up to 11 of the planned consensus mutations. Only 24 of the 90 isolates contained no mutations. Two isolates contained significant deletions in the BLA gene and eight isolates contained unplanned point mutations which must have been introduced by the QCMS approach. Twenty-eight of the planned 29 consensus mutations were observed in at least two of the isolates. The only mutation not found in any of the isolates was T1A. Subsequent sequence analysis revealed an error in the design of the primer encoding this particular mutation and we did not consider mutation T1A during further analysis.

The frequency distribution of mutations in the libraries is shown in Figure 2. Libraries NA02 and NA03 contained a large fraction of combinatorial variants whereas almost half the isolates from NA01 did not contain any mutations. We conclude that reducing the concentration of mutagenic primers below the manufacturer’s recommendation results in a more favorable distribution of mutations when constructing large combinatorial libraries. Subsequently, we measured the thermostability of BLA activity for each isolate. Figure 3 shows that about one-quarter of all isolates in the libraries were more stable than the parent protein. Hence the fraction of stabilized mutants in the CCM libraries is significantly higher than the fraction of stabilized variants in random libraries that can be obtained by PCR mutagenesis or similar methods. Such non-targeted
mutagenesis libraries typically require the screening of hundreds to thousands of variants in order to identify a few stabilized mutants (Matsumura and Aiba, 1985; Kuchner and Arnold, 1997; Cherry et al., 1999). Therefore, our data support the hypothesis that a large fraction of consensus mutations leads to the stabilization of a protein. The most stable variant, NA03.8, contained three mutations: Q95E, A153S and I334L. This variant was chosen as parent for a second-generation library.

**Identification of stabilizing mutations**

The CCM process allows one to add multiple stabilizing mutations to a protein in a single round of mutagenesis and screening. Because most variants contain multiple mutations, it is not immediately obvious which mutations are actually responsible for the increase in protein stability. However, we reasoned that it should be possible to estimate the stabilizing contribution of individual mutations by analyzing the correlation between the stability of a variant and its sequence. The set of 90 random isolates contained a total of 250 consensus mutations; thus each of the 28 mutations was on average observed nine times. We anticipated that stabilizing mutations should confer increased stability to most of the variants that contain them. Figure 4 shows the frequency of individual consensus mutations in the 20 most stable and the 20 least stable of the 90 random isolates from libraries NA01-3. It shows that many of the mutations are more prevalent in stabilized variants compared with less stable variants. The most prominent example is V284I, which occurred in 12/20 of the most stable variants but in only 1/20 of the least stable isolates. Other mutations that are prevalent in stabilized variants are Q95E, A153S, A208P, N232R and I262V. However, a few mutations occur more frequently in unstable isolates, Y170F, Q219E, V227A and P295A, which suggests that some of the consensus mutations may actually destabilize the protein.

We decided to estimate the contribution of each consensus mutation to BLA stability under the assumption that individual mutations have additive effects on protein stability. We assigned a parameter $P_k$ to each of the 28 mutations in the data set and assumed that the remaining activity $R_i$ of each variant can be calculated based on these parameters using the equation

$$\log R_i = \sum_{k=1}^{m} M_{ki}P_k + C$$

where $M_{ki} = 1$ if variant $i$ contains mutation $k$ and 0 if variant $i$ does not contain mutation $k$. $C$ is a constant that reflects the remaining activity of the parent enzyme. The parameters were determined by solving Equation 2 using Microsoft Excel.

$$\sum_{i=1}^{n} \left( \log R_i - \sum_{k=1}^{m} M_{ki}P_k - C \right) = \min$$

The calculation considered all 80 BLA variants which did not contain any unplanned mutations. Table I shows the calculated stabilization parameters for all 28 consensus mutations. Figure 5 shows that the model describes most of the measured variations in BLA stabilities. The analysis identified several mutations that appear to have significant stabilizing effect on BLA: V11I, V25I, R91K, Q95E, A153S, N232R, S247T and I262V. In addition, we identified mutations that reduce BLA stability: Q219E, V227A, P295A and Y170F.
Construction and screening of a second-generation CCM library

We chose variant NA03.8 as parent for a second-generation combinatorial library, NA04. Variant NA03.8 contains the two most stabilizing mutations, Q95E and A153S, in addition to one neutral mutation, I334L. Our initial statistical analysis, which was based on limited sequence information, had identified nine additional stabilizing mutations: V11I, V25I, R91K, N232R, S247T, I262V, V283L, V284I, T342K. We reused the nine mutagenic primers from round one to add these mutations to variant NA03.8. One primer was added to the mixture that encoded mutations V283L and V284I, which cannot be recombined using individual mutagenic primers. The mutagenesis was performed using the same protocol as for library NA02.

We screened library NA04 for increased stability of BLA in the presence of thermolysin to avoid the risk of isolating variants that re-fold after heat stress. Thermolysin is a thermostable protease with a broad specificity for hydrophobic amino acids. It has been reported that thermolysin selectively cleaves unfolded proteins (Arnold and Ulbrich-Hofmann, 1997). We tested 352 random isolates from library NA04, 332 of which expressed sufficient activity to allow measurement of protease stability. The screening results are shown in Figure 6. It is clear that a very significant number of isolates is more stable than the parent of the library, NA03.8. We chose 22 of the most stable variants for a repeat analysis and found all of them to be

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Parameter</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4T</td>
<td>0.05</td>
<td>4</td>
</tr>
<tr>
<td>V11I</td>
<td>0.38</td>
<td>4</td>
</tr>
<tr>
<td>A13D</td>
<td>0.04</td>
<td>24</td>
</tr>
<tr>
<td>V25I</td>
<td>0.19</td>
<td>7</td>
</tr>
<tr>
<td>Y41F</td>
<td>−0.23</td>
<td>5</td>
</tr>
<tr>
<td>F43Y</td>
<td>−0.09</td>
<td>20</td>
</tr>
<tr>
<td>I65V</td>
<td>−0.22</td>
<td>2</td>
</tr>
<tr>
<td>R91K</td>
<td>0.37</td>
<td>2</td>
</tr>
<tr>
<td>Q95E</td>
<td>0.76</td>
<td>9</td>
</tr>
<tr>
<td>R105T</td>
<td>−0.22</td>
<td>2</td>
</tr>
<tr>
<td>M106L</td>
<td>−0.18</td>
<td>3</td>
</tr>
<tr>
<td>S130A</td>
<td>−0.10</td>
<td>14</td>
</tr>
<tr>
<td>A153S</td>
<td>0.67</td>
<td>10</td>
</tr>
<tr>
<td>Y170F</td>
<td>−0.26</td>
<td>4</td>
</tr>
<tr>
<td>A194S</td>
<td>0.14</td>
<td>4</td>
</tr>
<tr>
<td>A208P</td>
<td>−0.02</td>
<td>8</td>
</tr>
<tr>
<td>Q219E</td>
<td>−0.45</td>
<td>8</td>
</tr>
<tr>
<td>T225S</td>
<td>0.17</td>
<td>3</td>
</tr>
<tr>
<td>N226T</td>
<td>−0.20</td>
<td>8</td>
</tr>
<tr>
<td>V227A</td>
<td>−0.31</td>
<td>6</td>
</tr>
<tr>
<td>N232R</td>
<td>0.44</td>
<td>6</td>
</tr>
<tr>
<td>S247T</td>
<td>0.18</td>
<td>4</td>
</tr>
<tr>
<td>I262V</td>
<td>0.29</td>
<td>15</td>
</tr>
<tr>
<td>V283L</td>
<td>0.33</td>
<td>3</td>
</tr>
<tr>
<td>V284I</td>
<td>0.38</td>
<td>19</td>
</tr>
<tr>
<td>P295A</td>
<td>−0.30</td>
<td>5</td>
</tr>
<tr>
<td>I334L</td>
<td>−0.06</td>
<td>6</td>
</tr>
<tr>
<td>T342K</td>
<td>0.11</td>
<td>9</td>
</tr>
<tr>
<td>C</td>
<td>0.47</td>
<td></td>
</tr>
</tbody>
</table>

*Frequency is the number of variants that contained the particular mutation.

Fig. 4. Frequency of mutations in the 20 most stable and the 20 least stable isolates from libraries NA01-3. The 20 most stable isolates contained a total of 107 consensus mutations whereas only 41 mutations were found in the 20 least stable isolates. Three of the least stable isolates had no mutations.

Fig. 5. Correlation between the measured stability and calculated stability of BLA isolates from NA01-3. The natural logarithm of the remaining activity is plotted. The activity was calculated based on parameters shown in Table I.
more stable than variant NA03.8. The stabilized variants contained between two and eight of the nine intended mutations. Mutation I262V was observed in 21 of the 22 isolates and mutation R91K was detected in 17 isolates. These two mutations appear to have the largest stabilizing effect. The most stable isolate, NA04.17, contained eight consensus mutations.

**Characterization of stabilized BLA variants**

We chose three BLA variants for purification and further characterization. Thermal denaturation was monitored at 220 nm (Figure 7). All variants showed irreversible denaturation. However, CCM variants showed an increase in apparent \( T_m \) between 4.8°C for NA03.8 and 9.1°C for NA04.17 (Table II). Studies on sensitivity of the variants towards three proteases, chymotrypsin, trypsin and thermolysin, which have very different primary specificities, showed that all variants were stabilized against inactivation by all three tested proteases (Figure 8). Furthermore, there was a very clear correlation between the stability towards proteolysis and the thermostability of these variants. These observations strongly suggest that the rate of proteolysis of BLA is mainly driven by the overall thermodynamic stability of the protein and not by the specificity of the attacking protease.

**Discussion**

Our data show that CCM is a very effective method for the identification of stabilized variants of a protein of interest. Significantly stabilized isolates can be identified by screening very small random libraries. Using this approach, we generated mutants that were stabilized against both thermal denaturation and proteolysis and we were able to increase the apparent \( T_m \) of BLA by 9.1°C. The most stable variant, NA04.17, contained eight mutations.

To identify stabilizing consensus mutations, we assumed that the effects of all mutations on protein stability are additive. We are aware that mutations can have complex and non-additive interactions between each other. It appears, however, that the effect of small changes that result in subtle perturbations of protein structure can be described frequently by assuming simple additivity (Wells, 1990; Gregoret and Sauer, 1993; Sandberg and Terwilliger, 1993). The purpose of our analysis was not to calculate exact thermodynamic contributions of each mutation to the stability of the protein. Instead, we focused on the rapid and efficient generation of a data set that allows one to identify stabilizing mutations based on screening data. The validity of this approach is supported by the identification of stabilized variants containing multiple consensus mutations from our second-generation library.

Our data suggest that stabilization of BLA was achieved by adding multiple consensus mutations and that most of these mutations make small but positive contributions to stability. Visual inspection of the crystal structure of BLA (Lobkovsky et al., 1994) allowed us to divide the consensus mutations into

---

**Table II. Thermostability of selected BLA variants**

<table>
<thead>
<tr>
<th>Variant</th>
<th>Mutations</th>
<th>( T_m (°C) )</th>
<th>( \Delta T_m (°C) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLA</td>
<td>None</td>
<td>57.1</td>
<td>–</td>
</tr>
<tr>
<td>NA03.8</td>
<td>Q95E, A153S, I334L</td>
<td>61.9</td>
<td>4.8</td>
</tr>
<tr>
<td>NA04.2</td>
<td>R91K, Q95E, A153S, S247T, I262V, I334L</td>
<td>64.9</td>
<td>7.8</td>
</tr>
</tbody>
</table>

*aMutations that resulted from the first round of CCM are underlined.*
three groups: mostly solvent exposed (15 mutations), partially buried (11 mutations) and completely buried (three mutations).

Both stabilizing and destabilizing mutations were observed in all three groups and no correlation between surface exposure and effect on stability could be detected. Of the two mutations that appear to make the largest stabilizing contributions, one is located on the surface (Q95E) and the other buried within the protein (A153S).

The literature contains several reports demonstrating that the proteolytic susceptibility of a protein depends largely on the conformational stability of the protein and to a lesser degree on the actual amino acid sequence (Fontana, 1988; Zappacosta et al., 1996; Arnold and Ulbrich-Hofmann, 1997). Our results follow that trend and show that the stability of BLA variants depends mainly on their thermodynamic stability and not on the specificity of the attacking protease. This observation can have significant implications for protein design. For example, it should be possible to identify mutants of proteins that resist proteolysis during in vivo therapeutic or analytical applications, without exact knowledge of the enzyme(s) responsible for proteolysis.

In this study, we limited mutagenesis to those amino acid changes that occurred in the consensus sequence of the protein, i.e. the most abundant amino acid for a given position. However, our method can be expanded easily to include other amino acid changes that occur less frequently in a particular position of the multiple sequence alignment. This would enable one to generate larger first-generation libraries and subsequent analysis of sequence and performance of isolates would allow one to identify rapidly the most valuable individual mutations. Such a statistical analysis is less informative for more traditional random libraries, which contain a large number of different mutations, but in which most individual mutations are too rare to allow multiple combinatorial observations.

There are currently multiple genome sequencing projects, which have led to a rapidly growing database of available sequences. Such databases will greatly facilitate the identification of protein homologs and thus the application of CCM to protein engineering.

References

Received August 30, 2004; revised November 16, 2004; accepted November 19, 2004

Edited by Jacques Fastrez