Directed evolution and structural analysis of N-carbamoyl-d-amino acid amidohydrolase provide insights into recombinant protein solubility in *Escherichia coli*

Shimin JIANG*†, Chunhong LI*, Weiwen ZHANG†, Yuanheng CAI*, Yulinli YANG*, Sheng YANG* and Weihong JIANG*‡

*Laboratory of Molecular Microbiology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, People's Republic of China, †Microbiology Department, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, U.S.A., and ‡Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200025, People's Republic of China

One of the greatest bottlenecks in producing recombinant proteins in *Escherichia coli* is that over-expressed target proteins are mostly present in an insoluble form without any biological activity. DCase (N-carbamoyl-d-amino acid amidohydrolase) is an important enzyme involved in semi-synthesis of β-lactam antibiotics in industry. In the present study, in order to determine the amino acid sites responsible for solubility of DCase, error-prone PCR and DNA shuffling techniques were applied to randomly mutate its coding sequence, followed by an efficient screening based on structural complementation. Several mutants of DCase with reduced aggregation were isolated. Solubility tests of these and several other mutants generated by site-directed mutagenesis indicated that three amino acid residues of DCase (Ala18, Tyr30 and Lys34) are involved in its protein solubility. *In silico* structural modelling analyses suggest further that hydrophilicity and/or negative charge at these three residues may be responsible for the increased solubility of DCase proteins in *E. coli*. Based on this information, multiple engineering designated mutants were constructed by site-directed mutagenesis, among them a triple mutant A18T/Y30N/K34E (named DCase-M3) could be overexpressed in *E. coli* and up to 80% of it was soluble. DCase-M3 was purified to homogeneity and a comparative analysis with wild-type DCase demonstrated that DCase-M3 enzyme was similar to the native DCase in terms of its kinetic and thermodynamic properties. The present study provides new insights into recombinant protein solubility in *E. coli*.

Key words: *N*-carbamoyl-d-amino acid amidohydrolase, directed evolution, negative charge, protein folding, protein solubility.

**INTRODUCTION**

Although various bacterial expression systems have been established for fast and efficient expression of heterologous proteins [1], one common problem of most heterologous expression systems is the poor solubility/misfolding of recombinant proteins. In most cases, target proteins failed to fold into their native states and instead accumulated in inactive IBs (inclusion bodies) when expressed heterologously [2]. Current experimental strategies to form soluble and native protein molecules include: the co-expression of foldases or chaperones [3–5]; the fusion of tags that contain a highly soluble polypeptide or an increased negatively charge polypeptide [6,7]; the use of promoters with different strengths, in order to control the rate of protein synthesis; the expression of the protein at a lower temperature; and the optimization of growth medium and culturing conditions [8]. Although with some success, the major disadvantage of these strategies is that they are very time-consuming. Another alternative method is to use the directed evolution technique, which can generate a large number of mutants in a very short period of time. For application of directed evolution techniques, the establishment of an efficient screening method is a prerequisite. Some commonly used high-throughput screening methods use fusion reporter systems, such as the green fluorescent protein folding report method [11], the N-terminal fusion system with chloramphenical acetyltransferase [12] and LacZ (β-galactosidase α peptide) complementation solubility reporter assay [13].

DCase (N-carbamoyl-d-amino acid amidohydrolase; EC 3.5.1.77) is a rate-limiting enzyme in a two-step reaction system for producing D-HPG (D-p-hydroxyphenylglycine), an important intermediate for semi-synthesis of β-lactam antibiotics such as penicillins and cephalosporins [14]. In the process, a starting substrate DL-5-substituted hydantoin is asymmetrically hydrolysed to N-carbamoyl-d-p-hydroxyphenylglycine by D-hydantoinase (EC 3.5.2.2), followed by stereo-specific transformation of N-carbamoyl-d-p-hydroxyphenylglycine into its corresponding D-HPG by DCase. DCase normally has lower activity relative to that of D-hydantoinase [15,16]. During the past decade, several studies have been reported on the cloning and characterization of DCase encoding genes from a variety of micro-organisms and their overexpression in the *Escherichia coli* system [17–20]. However, overproduction of DCase proteins in *E. coli* often results in the formation of biologically inactive IBs. In our previous study...
a DCase gene was cloned from *Burkholderia pickettii*. The gene encodes a peptide of 304 amino acids with a calculated molecular mass of 34 334 Da. DCase forms a homotetramer and is an intracellular protein without disulfide bonds. Attempts were made to overexpress DCase in *E. coli*, and the results showed that its activity was very low, because nearly 80% of the recombinant DCase was partitioned into insoluble aggregates. In the present study, the DCase encoding gene from *B. pickettii* was subjected to random mutation by combining error-prone PCR and DNA shuffling, followed by a simple and efficient screening protocol based on structural complementation between the α- and ω-fragments of β-gal (β-galactosidase) [13]. Using this approach, we have identified three amino acid residues (Ala18, Tyr30 and Lys34) that are related to the protein solubility of DCase. Molecular modelling analysis suggested that enhanced solubility of the DCase proteins attributed mainly to the increases in hydrophilicity and/or negative charge of the substituted amino acids in *E. coli*. Multiple mutants of DCase were rationally designed and overexpressed in *E. coli*; among them, DCase-M3 (a triple DCase mutant A18T/Y30N/K34E) was found, which was up to 80% soluble.

**MATERIALS AND METHODS**

**Random mutagenesis combining error-prone PCR and DNA shuffling**

Primers, 5′-GGGAATTCATATGACACGTGATGATGATCTT-3′ and 5′-CCCAAGCTTGCCGCAAGACAGACGGGAGCC-AGGGATCCGGATTCGGCCGATGACACC-3′, were designed to PCR amplify the DCase gene from the plasmid pXZ-total [21]. Of note, we chose to use the DNA fragment 5′-GGATCC-GCTGGCTGCCGTGTCGGTCTCCGACAGTTT-3′ coding for an amino acid linker Gly-Ser-Ala-Gly-Ser-Ala-Gly-Ser-Gly-Ala-Ser between DCase and the α-fragment of β-gal in the latter primer [11]. To enhance the natural mutation rate, the DCase gene was initially amplified by error-prone PCR [22]. The reaction mixture for a 100 μl error-prone PCR sample contained: 7 mM MgCl2, 50 mM KCl, 10 mM Tris/HCl (pH 8.5), 0.1% gelatin, 0.2 mM dATP, 0.2 mM dGTP, 1.0 mM dCTP, 1.0 mM dTTP, 250 pmol of each primer, 500 ng of pXZ-total template DNA, MnCl2 (0.075–0.15 mM final concentration) and 5 units of Taq polymerase (Promega). After denaturation for 6 min at 94°C, the PCR was run for 30 cycles of 94°C, 30 s; 58°C, 30 s; 72°C, 30 s; with a final extension step at 72°C for 5 min. PCR products were checked by electrophoresis on a 1.0% (w/v) agarose gel and then purified using a DNA purification kit (Qiagen). Subsequently, the PCR product was randomly cleaved with DNaseI (New England Biolabs). Finally, DNA fragments were reassembled with a primerless PCR and an additional primer PCR as described by Stemmer [23].

**Construction and screening of mutant library**

The reassembled genes, double digested with NdeI and HindIII, were ligated into the pMAL-c2x vector (New England Biolabs) digested with Ndel and HindIII using the rapid DNA ligase to form an in-frame DCase–α-fragment fusion protein. The ligation mixture was transformed into *E. coli* DH5α competent cells (Novagen) yielding a library of over 10000 transformants on LB (Luria–Bertani) plates supplemented with 100 μg/ml ampicillin, 80 μg/ml X-gal (5-bromo-4-chloroindol-3-yl β-D-galactopyranoside) and 0.1 mM IPTG (isopropyl-β-D-thiogalactoside) [13]. After the plates were incubated at 37°C for 18–24 h, colonies showing darker blue than the WT (wild-type) were selected.

**Sequencing and analysis**

Plasmids isolated from the possible positive colonies were purified using a plasmid mini kit (Qiagen). DNA sequencing was carried out using fluorescent dye terminator sequencing chemistry on an ABI 3730 sequencer (Invitrogen). Nucleotide and amino acid sequence alignment analysis was performed using ClustalW [24]. The secondary structure prediction of protein was provided by ESPript [25].

**Measurement of β-gal activity *in vitro* and solubility analysis**

*E. coli* DH5α cells containing the WT DCase–α-fragment and mutated DCase–α-fragment fusion expression constructs were inoculated into a liquid LB medium supplemented with 100 mg/ml ampicillin and grown at 37°C until a D₅₀₀ of 0.6 was reached. The target proteins were induced with 0.3 mM IPTG for 3 h. A 1 ml aliquot of the culture was then transferred into a 1.5 ml Eppendorf tube. After adding 10 μl I of X-gal (100 μg/ml), the tube was shaken at 37°C for 30 min. β-Gal activity was then evaluated by the appearance of a blue colour. Moreover, solubility of the DCase–α-fragment fusion protein was monitored by native PAGE. A 1 ml aliquot of the cell culture was pelleted in a 1.5 ml Eppendorf tube and resuspended in 200 μl lysis buffer A (100 mM Tris/HCl, pH 7.5, 100 mM NaCl and 1 mM EDTA). After sonication on ice (six bursts, 5 s each at 250 W), the lysates were centrifuged at 15 000 g for 20 min at 4°C. The supernatant fractions were analysed using native-PAGE. The gel was stained with buffer Z (10 mM KCl, 2.0 mM MgSO₄, 60 mM Na₂HPO₄, 40 mM Na₃PO₄, pH 7.0) containing 80 μg/ml X-gal. The activity of β-gal was determined using ONPG (O-nitrophenyl-β-D-galactopyranoside) as a substrate as described previously [13].

**Oligonucleotide-directed mutagenesis**

Site-directed mutagenesis for 22 single point mutants of DCase, in the three mutation sites: Ala18, Tyr30 and Lys34 (A18L, A18Y, A18E, A18D, A18N, A18K, A18R, Y30L, Y30A, Y30G, Y30E, Y30D, Y30T, Y30C, Y30K, Y30S, K34L, K34A, K34G, K34H, K34D, K34R), three double mutants (A18T/Y30N, A18T/K34E, and Y30N/K34E) and one triple mutant (A18T/Y30N/K34E) were generated by overlap extension PCR [26]. All the mutations were confirmed by DNA sequencing.

**Protein expression and solubility test**

All DCase genes, including WT and the mutated DCase by directed evolution as well as mutants by site-directed mutagenesis, were PCR amplified using the forward primer: 5′-GGGAATTCATATGACACGTGATGATGATCTT-3′ and the reverse primer: 5′-CCCAAGCTTTCAGAGTTCCGCGATGACACC-3′; and the reverse primer: 5′-CCCAAGCTTTCAGAGTTCCGCGATGACACC-3′. After double digestion with Ndel and HindIII, the DCase genes were subcloned into the pET-28a expression vector (Novagen) and heterologously expressed in *E. coli* strain BL21 (DE3) (Novagen). Cells harbouring the plasmid with WT or the evolved DCase genes were grown to the mid-log phase at 37°C, induced with IPTG at a final concentration of 0.3 mM at 22°C. Cells from 1 ml of culture were harvested about 10 h post-induction by centrifugation at 15 000 g for 10 min at 4°C and resuspended in 400 μl lysis buffer A without 1 mM EDTA. The cell suspensions were lysed by sonication on ice (four bursts, 30 s). After removal of insoluble protein and cell debris by centrifugation of the lysates at 15 000 g for 10 min at 4°C, the supernatant fraction and precipitant fraction were subjected to SDS/PAGE analysis (10% gels). Protein solubility was estimated by scanning Coomassie Blue-stained gels in relation to a reference protein.
Directed evolution and structural analysis of DCase

RESULTS

Directed evolution of soluble DCase variants

The directed evolution technique involves two key steps: first, the construction of mutant library with enough variations; and secondly, high-throughput screening. In the present study, attempts were made to increase the mutation rate. A variety of DCase gene fragments were first created by error-prone PCR, the fragments were then randomly fragmentized with DNase I before the steps of primerless PCR and primer PCR of DNA shuffling. Through such modifications a higher frequency of mutations, 0.8 % at nucleotide level and approx. 3 amino acid changes per enzyme variant, was achieved (data based on sequence analysis of 15 randomly selected clones). Overall, a library of mutated DCase with over 10,000 colonies was obtained. The mutated DCase genes were cloned into the pMAL-c2x fusion expression vector (Figure 1A). In addition, we also introduced a colorimetric screening system based on structural complementation between the α- and ω-fragments of β-gal in E. coli to identify DCase variants with improved solubility. The solubility can be visualized by the intensity of blue colour on IPTG/X-gal-treated indicator plates. The method is very high throughput and reproducible. From the initial screening, four mutants (MU1–MU4) with increased solubility were isolated from the library. The relative intensities of blue colour in cells expressing WT DCase–α-fragment or mutated DCase–α-fragment were compared in liquid medium containing esence experiments were performed at 25 °C. The experimental data were used to calculate the thermodynamic parameters according to the two-state unfolding model [29].

Structural homology modelling

Structural models of the WT DCase and the DCase-M3 proteins of B. picketii were constructed based on the crystal structure of DCase from Agrobacterium sp. KNK712 from the Protein Data Bank (PDB accession code 1ERZ) [30]. The amino acid replacements were generated using the rotamer library approach of the SCWRL program [31]. The program RasMol was used for display of the protein structure [32]. Charge distribution of protein was generated with GRASP [33].

Purification of WT and mutant DCase-M3

Recombinant WT and the evolved DCase-M3 fused to a His6 tag were purified from E. coli by a Ni2+-affinity column (Qiagen). Briefly, the procedure of protein purification was as follows. (i) The cell pellets from 200 ml of cell culture medium were harvested by centrifugation and resuspended in 6 ml of lysis buffer A containing 10 mM imidazole. After sonication at 4 °C (for 4 min in 5 s pulses), the lysates were centrifuged at 15 000 g for 20 min at 4 °C. (ii) Supernatant fractions of the cleared lysates were mixed completely with 1 ml of Ni-NTA (Ni2+-nitrilotriacetate)–agarose on ice by shaking for 1 h, and were loaded on to a 5 ml column. (iii) The Ni-NTA–agarose was cleared using 10 ml of wash buffer (50 mM Tris/HCl, pH 7.5, 100 mM NaCl and 20 mM imidazole). (iv) The target proteins were eluted by elution buffer (50 mM Tris/HCl, pH 7.5, 100 mM NaCl and 250 mM imidazole). Protein purity was assessed by SDS/PAGE followed by staining of the gel with Coomassie Blue. Protein concentration was determined by the method of Bradford [27], using BSA as a standard.

Kinetic analysis of WT DCase and DCase-M3

The kinetic parameters of WT DCase and DCase-M3 were determined as described previously [28]. The substrate concentration of N-carbamoyl-D-β-hydroxyphenylglycine was varied from 3.3 to 20 mM. One unit of enzyme activity was defined as the amount of enzyme that catalyses the formation of D-HPG at the rate of 1 µmol/min under the assay conditions.

In vitro equilibrium denaturation measurement

Native proteins (6 µg of purified WT DCase or DCase-M3) were added into 1 ml of a Tris buffer (20 mM Tris/HCl, pH 7.5 and 50 mM NaCl) with varying concentrations of urea, and the fluorescence intensity of each sample was recorded after equilibrium. The urea-induced unfolding or refolding of the target proteins was monitored by the increase/decrease of the intrinsic fluorescence of tryptophan. Fluorescence measurements were performed using a Hitachi F-4010 fluorescence spectrophotometer with an excitation wavelength of 295 nm and an emission wavelength from 300–400 nm. The excitation and emission bandwidths were 3 nm and 5 nm respectively. All fluor-
Figure 2  Sequence alignment of six homologous amidohydrolases

1, DCase from B. pickettii; 2, DCase from Agrobacterium sp. KNK712; 3, DCase from Pseudomonas sp. KNK03A; 4, nitriase from Polaronomas sp. JS666; 5, aliphatic amidase from Saccharopolyspora spinosa; 6, β-alanine synthase from Brevibacillus agri. The Figure was generated using the ClustalW program and drawn with ESPript. The secondary structure elements of DCase are shown above the sequences. Identical residues are boxed in black. Glu47, Lys127 and Cys172, the catalytically important residues, are indicated by stars. The three mutations in DCase, A18T, Y30N and K34E, are indicated by closed circles (•).

X-gal (Figure 1B). The results showed that cells harbouring four mutated DCase–α-fragments displayed blue colouration that was much darker than that of WT DCase–α-fragment. Increased β-gal activities were also observed in MU1–MU4, especially in MU1 where it had a more than 3-fold increase of the β-gal activity compared with that of WT (Figure 1C). Supernatant fractions of these fusion proteins were analysed on native-PAGE to compare further the solubility of WT and MU1–MU4. As expected, more soluble proteins were observed for MU1–MU4 than for WT (Figure 1D).

**Determination of amino acid sites related to solubility in DCase**

To identify the molecular basis of protein solubility change, plasmids harbouring evolved DCase genes were extracted for DNA sequencing. Only one substitution was found in MU1 and MU2, K34E and A18T respectively. Whereas MU3 and MU4 had the same substitution of Y30N, additional analysis showed that three other substitutions in MU3 or MU4 (Q23H, R278C and L300Q) did not affect the solubility of DCase (results not shown). The results suggested that the three amino acid residues (Ala18, Tyr30 and Lys34) were related to protein solubility of DCase. The sequence alignment of six amidohydrolases is listed in Figure 2 and the characteristics related to the three mutations are summarized in Table 1.

**Effects of site mutations on DCase solubility**

Due to the low expression level of DCase from the pMAL-c2x vector and the purification problems, the pET system was used to
overexpress the target proteins. SDS/PAGE analysis of proteins from whole cells showed that overall expression levels of the WT DCase and the evolved DCase proteins were at the same level, approx. 70% of total protein (results not shown). Fractionation of the cells by sonication and centrifugation showed that more of the evolved DCase proteins remained soluble than the WT DCase protein (Figure 3A). Of the three single mutations (A18T, Y30N and K34E), the A18T and Y30N mutants were solubly expressed in approx. 35% of the total DCase proteins, whereas the mutant K34E produced up to approx. 50% soluble expression. To check whether cumulative effects are present among the three point mutations on protein solubility, three double mutants (A18T/Y30N, A18T/K34E and Y30N/K34E) and one triple mutant (A18T/Y30N/K34E) were constructed using site-directed mutagenesis. Solubility testing indicated that more soluble proteins could be obtained by any combination of more than one mutation (Figure 3B). The triple mutant DCase-M3 (A18T/Y30N/K34E) showed the highest solubility in all mutants, approx. 3-fold higher than that of the WT DCase. In comparison with the WT DCase, solubility of the evolved DCase mutants had all been improved from 10% to 60%, with the highest for DCase-M3 where more than 80% of the target proteins were soluble.

### Comparative characterization of the WT DCase and DCase-M3

In order to evaluate the effect of three mutations (A18T, Y30N and K34E) on the enzyme properties of DCase-M3 and WT DCase, biochemical characterisation was performed. Both the WT DCase and DCase-M3 enzymes were produced at a purity of above 95% on Ni²⁺-affinity columns, as determined by SDS/PAGE (Figure 4A). We comparatively analysed two categories of steady-state characteristics of the enzymes. The first was the kinetic properties of enzymes, such as $K_m$, $k_{cat}$, and $k_{cat}/K_m$. Analytical results showed that DCase-M3 had similar $k_{cat}$ and $K_m$ values compared with WT DCase (Table 2). The second was the thermodynamic stability of the enzymes. The equilibrium unfolding transitions of the proteins by urea were carried out by monitoring fluorescence emission at 340 nm. The curves of the fluorescence change versus the increase of urea concentration are shown in Figure 4B. Profiles of the WT DCase and the DCase-M3 showed almost identical and typical sigmoidal curves, indicative of a two-state unfolding model. According to the unfolding model, the Gibbs free energy changes ($\Delta G_{H2O}$) were calculated (Table 3). The $\Delta G_{H2O}$ of the WT DCase (5.72 ± 0.41 kcal/mol; 1 cal = 4.184 J) was similar to that for Dcase-M3 within the experimental error (6.13 ± 0.60 kcal/mol). Furthermore, values of $C_m$ and $m$ were only slightly changed (Table 3). Taken together, the results suggested that the kinetic properties and thermodynamic parameters of DCase-M3 were essentially identical with those of the WT DCase.

### Table 1 Randomly evolved mutations leading to improved solubility of DCase

<table>
<thead>
<tr>
<th>Amino acid substitution</th>
<th>Base change (codon)</th>
<th>Secondary structure at position</th>
<th>Location</th>
<th>Conserved residue change</th>
<th>Hydrophilicity change*</th>
</tr>
</thead>
<tbody>
<tr>
<td>A18T</td>
<td>GCG → ACG</td>
<td>Turn</td>
<td>Surface</td>
<td>No</td>
<td>+2.5</td>
</tr>
<tr>
<td>Y30N</td>
<td>TAC → AAC</td>
<td>α-Helix</td>
<td>Surface</td>
<td>No</td>
<td>+2.2</td>
</tr>
<tr>
<td>K34E</td>
<td>AAA → GAA</td>
<td>Surface</td>
<td>Yes</td>
<td>−0.4</td>
<td></td>
</tr>
</tbody>
</table>

* Increase or decrease of hydrophilicity of the mutated sites are indicated by ‘+’, ‘−’ respectively. Hydropathy indices are obtained by the method of Kyte and Doolittle [34].

#### Figure 3

Solubility analysis of target proteins that were expressed using the pET expression system at 22°C

The mutations are listed on the right-hand side of the panels. (A) Solubility analysis of WT DCase and evolved recombinant DCases. (B) Expression test of a single mutant (Y30N), three double mutants (A18T/Y30N, A18T/K34E, and Y30N/K34E) and a triple mutant (A18T/Y30N/K34E). tot, total cell; ppt, precipitant fraction; sup, supernatant fraction.
Additional site-directed mutation of the three residues: Ala^{18}, Tyr^{30} and Lys^{34}

To explore further the molecular mechanism for increased soluble expression, three amino acid residues of DCase were mutated using overlap extension PCR. Replacements for the amino acids Ala^{18}, Tyr^{30} and Lys^{34} were primarily selected on the basis of the hydrophobicity/hydrophilicity of the amino acids [34], and secondarily on their classification [35]. All mutant proteins were expressed in E. coli. Both soluble and insoluble fractions were fractionized and analysed by SDS/PAGE. Solubility tests of various point mutants of DCase are listed in Table 4. For the Ala^{18} residue, five mutant proteins, A18Y, A18T, A18E, A18D and A18N, showed a marked increase in their solubility, whereas the other mutants did not increase in solubility compared with the WT DCase. For the Tyr^{30} residue, three mutant proteins, Y30E, Y30D and Y30N, were generated and they were all significantly more soluble than the WT DCase. For the Lys^{34} residue, two mutated DCases, K34E and K34D, had higher solubility than the WT DCase (Table 4).

Structural analysis of the engineering designed DCase-M3

Crystal structures of DCase from several sources have been solved [30,36]. The protein from Agrobacterium sp. KNK712, solved at 1.7 Å (1 Å = 0.1 nm) resolution, was employed to model the WT DCase and the DCase-M3. Sequence identity between the DCase from B. pickettii and the DCase from Agrobacterium sp. KNK712 was approx. 99% in the optimal alignment. The DCase-M3 model was constructed by replacing the 18–39 residue fragment containing the three mutated sites: A18T, Y30N and K34E using the SCWRL program [31]. In the structural models of the

Table 2  Kinetic parameters of WT DCase and DCase-M3

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$k_{cat}/K_m$ (mM$^{-1}$·s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>0.82 ± 0.1</td>
<td>5.4 ± 0.3</td>
<td>6.6</td>
</tr>
<tr>
<td>DCase-M3</td>
<td>0.80 ± 0.1</td>
<td>5.9 ± 0.2</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Table 3  Thermodynamic properties of WT and DCase-M3 by urea denaturation (25°C)

<table>
<thead>
<tr>
<th>Protein</th>
<th>$\Delta G_{H_2O}$ (kcal/mol)</th>
<th>$C_m$ (M)</th>
<th>$m$ (kcal/mol per M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>5.72 ± 0.41</td>
<td>3.56 ± 0.03</td>
<td>1.61 ± 0.10</td>
</tr>
<tr>
<td>DCase-M3</td>
<td>6.13 ± 0.60</td>
<td>3.21 ± 0.10</td>
<td>1.91 ± 0.13</td>
</tr>
</tbody>
</table>

Table 4  Solubility test of point mutants of DCase

<table>
<thead>
<tr>
<th>Mutation</th>
<th>$\Delta S_f$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A18L</td>
<td>–</td>
</tr>
<tr>
<td>A18T</td>
<td>~34</td>
</tr>
<tr>
<td>A18Y</td>
<td>~40</td>
</tr>
<tr>
<td>A18E</td>
<td>~47</td>
</tr>
<tr>
<td>A18D</td>
<td>~44</td>
</tr>
<tr>
<td>A18N</td>
<td>~42</td>
</tr>
<tr>
<td>A18R</td>
<td>~Y30N</td>
</tr>
<tr>
<td>A18R</td>
<td>~Y30R</td>
</tr>
<tr>
<td>Y30L</td>
<td>–</td>
</tr>
<tr>
<td>Y30A</td>
<td>–</td>
</tr>
<tr>
<td>Y30G</td>
<td>–</td>
</tr>
<tr>
<td>Y30E</td>
<td>~81</td>
</tr>
<tr>
<td>Y30D</td>
<td>~122</td>
</tr>
<tr>
<td>Y30N</td>
<td>~55</td>
</tr>
<tr>
<td>Y30K</td>
<td>~K34E</td>
</tr>
<tr>
<td>Y30R</td>
<td>~K34R</td>
</tr>
<tr>
<td>K34L</td>
<td>–</td>
</tr>
<tr>
<td>K34A</td>
<td>–</td>
</tr>
<tr>
<td>K34G</td>
<td>–</td>
</tr>
<tr>
<td>K34H</td>
<td>–</td>
</tr>
<tr>
<td>K34E</td>
<td>~81</td>
</tr>
<tr>
<td>K34D</td>
<td>~37</td>
</tr>
<tr>
<td>K34R</td>
<td>–</td>
</tr>
</tbody>
</table>
Directed evolution and structural analysis of DCase

Figure 5  Model of DCase based on its similarity with DCase of Agrobacterium sp. KNK712

(A) The three-dimensional structure of the DCase-M3 homotetramer and monomer. In the DCase-M3 homotetramer (left-hand panel), the three replacements are coloured: A18T (cyan), Y30N (red) and K34E (blue). In the DCase-M3 monomer (right-hand panel), the three mutations in DCase-M3 are indicated as space-filling models including A18T, Y30N and K34E, which are located on the surface of the protein. The catalytically important residues (Glu47, Lys127 and Cys172) are displayed as ball and stick model, and the N- and C-termini are also labelled in the structure. The Figure was generated using RasMol. (B) Charge distribution of the region near the amino acid at position 34 on the molecular surface of modelled WT and evolved DCases (K34E and K34D). The electrostatic potential is coloured: uncharged amino acids (white), negative amino acids (red) and positive amino acids (blue). All Figures were generated using GRASP.

DISCUSSION

Due to low solubility and improper folding, proteins often aggregate when over-expressed in heterologous expression systems. Directed evolution has been demonstrated as a powerful and efficient tool to overcome the difficulties of expressing target proteins in a soluble form [10,13,37,38]. In the present study, we have applied a directed evolution approach to improve the solubility of DCase in *E. coli*. The success of this approach was attributed in a large part to an efficient and high-throughput colorimetric screening system for monitoring protein solubility/folding in vivo using structural complementation of β-gal. This system greatly simplified the screening procedure and allowed visual inspection of colonies. Consistent with a previous report [13], the detailed analysis of β-gal activity and native PAGE showed that there was clear correlation between β-gal activity and solubility/folding of the evolved target proteins, and it was...
completely feasible to apply this method to screen clones with more soluble DCases (Figure 1).

The causes of IB formation have been widely investigated. The major factors that have been have been suggested include, the protein’s amino acid composition in *E. coli*, its hydrophobicity and its overall net charge [39]. Interestingly, when comparing the amino acid sequences of the WT DCase with the evolved DCases-M3, two prominent differences were the hydrophobicity and charge properties of the mutated residues. The replacement by a hydrophilic residue may be able to counteract the hydrophobic site, resulting in improved solubility [39,40]. By mutation of Ala18 and Tyr30 with selected amino acids (Table 4), solubility tests revealed that increases in hydrophilicity of the residues at positions 18 and 30 improved its solubility. However, when proteins were substituted by positively charged residues (lysine or arginine), which had the strongest hydrophilic effect, no increase of solubility was observed. It was speculated that the positive effect of hydrophilicity (increase in solubility) and the negative effect of positive charge (decrease in solubility) might counteract with each other. Moreover, there is an obvious charge change in mutation K34E. Therefore, in this case the major force to reduce aggregation seems to be the increased negative charge. Additional replacement at the Lys34 position showed that mutated proteins replaced with negative charges (glutamate or aspartate) were more soluble than the WT DCase (Figure 4). Although a reduced pI has been reported in several soluble muteins [7,9], no change of pI was observed for the A18T or Y30N single muteins and double mutein (A18T/Y30N). However, the single mutant K34E and triple mutant DCases-M3 (A18T/Y30N/K34E) had lower pI (6.57 and 6.57) values than WT DCase (pI 6.84). Although more proof is still needed, the results suggested that the pI change may be not the major factor affecting solubility in DCase, and it could still be partially involved in the increase in solubility of some of the muteins.

The effects of the mutations on soluble expression of DCases can be partially explained by homology structural modelling. There were three to four hydrophilic residues at intervals in the N-terminus of the WT DCase from the primary structure of protein, and molecular modelling predicted that both α1 and α2 had comparatively typical amphiphilic helices, that is, one surface of each helix is mainly hydrophobic side-chains that interact favourably with the solvent, while the opposite surface is mainly hydrophobic side-chains to satisfy the core [41]. Notably, one of the most hydrophobic cores was formed in this region, which is composed of five secondary structure elements (α1, α2, β1, β2 and β3). According to the ‘hydrophobic collapse model’, the native polypeptide conformation forms by rearrangement of a compact collapsed structure during protein folding [42], and formation of α1 and α2 may be a very significant initiating process. Hence, the three mutations, A18T, Y30N and K34E, located on this domain could cause profound effects on folding or formation of the native polypeptides. In addition, the other mechanism may contribute to the delay in folding of passenger proteins by increasing negative charge. Recent evidence shows that protein aggregation can also be decreased by peptide extensions with large net negative charge; the net negatively charged peptide extensions could prolong the time that nascent polypeptides remain in productive folding pathways by increasing electrostatic repulsion between nascent polypeptides [7]. In the present study, when the residue at position 34 was replaced by amino acids with a negative charge, the solubility of muteins was increased (Table 4). This conclusion was further supported by analysing the charge distribution of region near the amino acid at position 34 on the molecular surface of modelled WT DCase and evolved DCases (K34E and K34D; Figure 5B). Significantly, the monomer interface of the evolved DCases showed a more favourable charge distribution relative to the WT protein. In the evolved K34D or K34E proteins, Asp34 or Glu34 balanced the positive potential of Arg27 and Arg38 from the symmetry-related subunit. In contrast, in the model of WT DCase, Arg27, Lys34 and Arg38 displayed an asymmetric negative and positive charge distribution. The favourable charge distribution will help proper folding of proteins by suppressing formation of non-specific off-pathway aggregates [37].

In summary, by combining error-prone PCR and DNA shuffling with a high-throughput colorimetric screening assay, we have developed a powerful directed evolution approach to obtain a DCase with better solubility when expressed in *E. coli*. There are two key parameters that may correlate with DCase solubility/folding in *E. coli* and which are charge average and hydrophilicity. The present study provides new insights into protein solubility/folding in *E. coli* and is also important for various rational design studies in improving protein solubility of similar types of proteins.

This work was supported by a grant from the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-YW-G108, 007) and a grant from the Ministry of Science and Technology of China (National Basic Research Program of China, 2007CB707803). We thank Professor Haiyan Liu (Key Laboratory of Structural Biology, University of Science and Technology of China, Chinese Academy of Science, Hefei, China) and Professor Hongyu Hu (Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China) for their kind help in molecular modelling and determination of thermodynamic parameters respectively.

REFERENCES


© 2007 Biochemical Society


