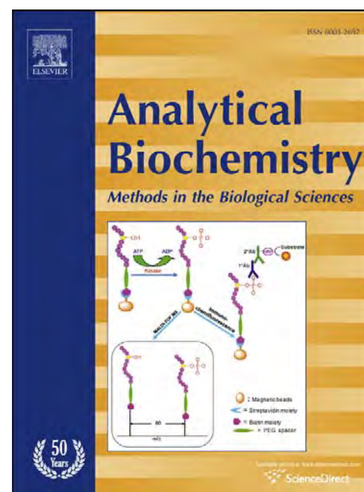
Accepted Manuscript



Easy preparation of a large-size random gene mutagenesis library in *Escherichia coli*

Chun You, Y.-H. Percival Zhang

PII: S0003-2697(12)00291-6
DOI: http://dx.doi.org/10.1016/j.ab.2012.05.022
Reference: YABIO 10976

To appear in: *Analytical Biochemistry*

Received Date: 27 April 2012
Revised Date: 21 May 2012
Accepted Date: 22 May 2012

Please cite this article as: C. You, Y.-H. Percival Zhang, Easy preparation of a large-size random gene mutagenesis library in *Escherichia coli*, *Analytical Biochemistry* (2012), doi: http://dx.doi.org/10.1016/j.ab.2012.05.022

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

1

2 # Easy preparation of a large-size random gene mutagenesis

3 # library in *Escherichia coli*

4

5 **Running title:** Constructing a large-size mutant library

6

7 **Chun You[1], Y.-H. Percival Zhang[1,2,3,4 *]**

8

9 [1] Biological Systems Engineering Department, Virginia Tech, 304 Seitz Hall, Blacksburg,

10 Virginia 24061, USA

11 [2] Institute for Critical Technology and Applied Science (ICTAS), Virginia Tech,

12 Blacksburg, Virginia 24061, USA

13 [3] DOE BioEnergy Science Center (BESC), Oak Ridge, Tennessee 37831, USA

14 [4] Gate Fuels Inc. 2200 Kraft Drive, Suite 1200B, Blacksburg, VA 24060, USA

15

16 [*]Corresponding authors: YPZ, Email: ypzhang@vt.edu

17 Tel: (540) 231-7414 [O]; Fax: (540) 231-3199

18

19

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

20 **Abstract**

21 A simple and fast protocol for the preparation of a large size mutant library for directed
22 evolution in *Escherichia coli* was developed based on the DNA multimers generated by
23 prolonged overlap extension PCR (POE-PCR). This protocol comprised (i) a linear DNA
24 mutant library was generated by error-prone PCR or shuffling, and a linear vector
25 backbone was prepared by regular PCR; (ii) the DNA multimers were generated based on
26 these two DNA templates by POE-PCR; and (iii) the one restriction enzyme-digested
27 DNA multimers were ligated to circular plasmids, following by transformation to *E. coli*.
28 Because the ligation efficiency of one DNA fragment was several orders of magnitude
29 higher than that of two DNA fragments for typical mutant library construction, it was
30 very easy to generate a mutant library with a size of more than10^7 protein mutants per 50-
31 L of the POE-PCR product. Via this method, four new fluorescent protein mutants were
32 obtained based on monomeric cherry fluorescent protein. This new protocol was simple
33 and fast because it did not require labor-intensive optimizations in restriction enzyme
34 digestion and ligation, did not involve special plasmid design, and enabled to construct a
35 large size mutant library for directed enzyme evolution within one day.

36

37 **Keywords**: directed evolution, cherry fluorescent protein, high transformation efficiency,
38 *E. coli*, prolonged overlap extension PCR, Simple Cloning

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

39
Introduction

40 Directed evolution is a powerful protein engineering tool to accelerate the evolution of
41 proteins with desired properties in laboratories [1-4]. Directed evolution of proteins is
42 usually conducted in *Escherichia coli* or *Saccharomyces cerevisiae* as a host [5]. *S.*
43 *cerevisiae* is a suitable platform for the evolution of proteins from eukaryote, and the
44 gene of interest can be easily and efficiently integrated into the chromosome of yeast by
45 homologous recombination for protein expression [6, 7]. However, *E. coli* is still the
46 preferable workhorse especially for bacterial proteins [1, 2, 8-11]. The preparation of a
47 protein mutant library in *E. coli* usually comprises three steps: (i) in vitro generation of
48 the gene mutant library by error-prone PCR or DNA shuffling, (ii) subcloning of the gene
49 mutant library into a protein expression plasmid, and (iii) transformation of the plasmid
50 library into competent cells. It is very easy to *in vitro* generate a very large DNA mutant
51 library (e.g., 1 μg of 1-kb dsDNA = ca.10^{12} DNA molecules) by error-prone PCR, DNA
52 shuffling, or their derived methods [2]. Also, high-efficiency competent *E. coli* cells
53 usually have transformation efficiencies of 10^{8-10} colony-forming units (cfu) per μg of
54 plasmid DNA. However, much smaller-size protein mutant libraries (e.g., 10^{3-6} mutants)
55 are usually generated during the subcloning step [10, 12]. This low efficiency is due to
56 low digestion efficiency of the vector and inserted DNA fragment, low efficiency of
57 ligation, and possible self-ligation of the digested plasmid [6, 13]. Because subcloning for
58 directed enzyme evolution usually requires the use of two restriction enzymes, each of
59 which has a sole recognition site in the resulting chimeric plasmid, resulting in limited
60 choices of restricting enzymes [11]. As a result, the construction of a reasonable size
61 protein mutant library requires careful design (e.g., restriction enzymes and sequences of

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

62 the targeted protein gene and vector) and a series of optimizations in digestion, ligation,
63 and transformation [6].
64
65 It is vital to construct plasmids containing a large DNA mutant library for directed
66 evolution in *E. coli* [10, 14, 15]. To address such problems as low digestion efficiencies
67 and limited choices of two different restriction enzymes, Maynard et al. [10] developed a
68 restriction enzyme-free construction of random DNA library by using numerous
69 megaprimers, T7 polymerase, T4 DNA ligase and a special host dut$^+$ ung$^+$ *E. coli*. But a
70 special *E. coli* host was required and the operation was pretty complicated and required
71 high experimental skills. Another method, megaprimer PCR of whole plasmid
72 (MEGAWHOP), which is derived from QuickChange site-directed mutagenesis method
73 [16, 17], have also been developed. However, this method required the addition of the
74 high-level parental plasmid (about 1-2 ng/μl) from dam$^+$ *E. coli* strain as the template for
75 PCR and of a large amount of *Dpn* I for digesting the parental plasmid completely.
76 Furthermore, the transformation efficiency of the MEGAWHOP product was about 10^4-
77 10^5 mutants/μg DNA [18], lower than that of the optimized transformation efficiency of
78 10^6 mutants/μg ligated DNA [6]. Zhang and Zhang (2011) developed a restriction
79 enzyme-free and ligase-free PCR-based method for the generation of a large secretory
80 protein mutant library in *Bacillus subtilis*, whereas DNA multimers generated by
81 prolonged overlap extension PCR (POE-PCR) are transformed into *B. subtilis* with very
82 high efficiencies [19]. However, this protocol works only for the purpose of subcloning
83 in *E. coli* but not for the propose of directed evolution due to low transformation
84 efficiencies, e.g., ca. 10^{1-4} per μg of DNA multimers [20].

ACCEPTED MANUSCRIPT

85

86 To create a large size library for directed evolution in *E. coli* and simplify plasmid design,
87 a new protocol was developed based on DNA multimers generated by POE-PCR (Figure
88 1). DNA multimers were digested by one restriction enzyme, whose cutting site can be
89 located any place of the vector backbone, yielding linear plasmid molecules. After
90 ligation by T4 DNA ligase, the circular plasmids were transformed to competent cells
91 with very high transformation efficiencies, resulting in a large size protein mutant library.
92 Labor intensive optimizations in two restriction enzyme digestion and ligation were not
93 needed so that a beginner can obtain a large size mutant library of more than10^7 cfu from
94 50 μL of the PCR product within one day.

95

96 **Materials and Methods**

97 **Reagents.** All chemicals were reagent grade or higher and were purchased from Sigma
98 (St. Louis, MO) or Fisher Scientific (Pittsburgh, PA) unless otherwise noted. All enzymes
99 in molecular biology experiments were purchased from New England Biolabs (NEB,
100 Ipswich, MA). The NEB regular Taq polymerase was used for error-prone PCR and NEB
101 high-fidelity Phusion DNA polymerase was used for regular cloning and POE-PCR. The
102 primers were synthesized by Integrated DNA Technologies (Coralville, IA). The PCR
103 thermocycler was Eppendorf temperature gradient Mastercycler (Hauppauge, NY).

104

105 **Strains and culture conditions.** The *E. coli* BL21(DE3) was used for recombinant
106 protein expression. All microorganisms were grown in the Luria-Bertani (LB) medium.
107 The ampicillin concentration in LB media was 100 μg/mL. The chemical competent cells

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

108 of *E. coli* BL21(DE3) were prepared as described elsewhere [13]. High transformation
109 efficiency *E. coli* BL21 (DE3) cells were also purchased from Invitrogen (Carlsbad, CA).
110
111 **Preparation of Plasmid.** The 706-bp DNA sequence encoding mCherry fluorescent
112 protein was amplified with a pair of primers (IF, TTAAC TTTAA GAAGG AGATA
113 TACAT ATGGT GAGCA AGGGC GAGGA GGATA; and IR, CAGTT CATTA
114 TCTGC CCACA GCTTA TCAGA ACCTG GCTTG) using the NEB Phusion
115 polymerase. The linear pET20b vector backbone was amplified with a pair of primers
116 (VF, CAAGC CAGGT TCTGA TAAGC TGTGG GCAGA TAATG AACTG; and VR,
117 TATCC TCCTC GCCCT TGCTC ACCAT ATGTA TATCT CCTTC TTAAA GTTAA)
118 using the NEB Phusion polymerase based on the plasmid pET20b. The insertion DNA
119 and vector backbone were assembled into DNA multimers by POE-PCR, as described
120 elsewhere [20]. The DNA multimers were transferred to *E. coli* Top 10, yielding plasmid
121 pET20b-mcherry-cbm.
122
123 **Preparation of DNA mutant library.** A DNA mutant library encoding mCherry
124 fluorescent protein was generated with a pair of primers (IF and IR) by error-prone PCR.
125 The reaction solution with a total volume of 50 μL contained 0.02 ng/μL plasmid
126 pET20b-mcherry-cbm, 0.2 mM dATP, 0.2 mM dGTP, 1 mM dCTP, 1 mM dTTP, 5 mM
127 $MgCl_2$, 0.2 mM $MnCl_2$, 0.05 U/μL the NEB regular Taq polymerase, 0.4 mM IF, and 0.4
128 mM IR [19]. The PCR reaction was conducted as the following: 94°C denaturation, 2
129 min; 13-16 cycles of 94°C denaturation for 30 s, 60°C annealing for 30 s, and 68°C

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

130 extension for 45 s; and 68°C extension for 5 min. The PCR product was cleaned with a
131 Zymo Research DNA Clean & Concentrator Kit (Irvine, CA).
132
133 The linear pET20b vector backbone was amplified with a pair of primers (VF and VR) by
134 using the NEB high-fidelity Phusion DNA polymerase. The PCR reaction system
135 contained dNTP, 0.2 mM for each; primers, 0.02 μM; pET20b template, 0.02 ng/μL; and
136 the Phusion polymerase, 0.04 U/μL. The PCR program was 98°C denaturation for 30 s;
137 30 cycles of 98°C denaturation for 10 s; 50 ^{o}C annealing for 10 s; and extension at 72 ^{o}C
138 at 3 kb per min for the targeted fragment. The PCR product was cleaned with a Zymo
139 Research DNA Clean & Concentrator Kit (Irvine, CA).
140
141 To generate DNA multimers by POE-PCR, 50-μL PCR reaction solution contained the
142 following: dNTP, 0.2 mM for each; the insertion DNA fragment, 5 ng/μL (i.e., 250 ng);
143 the vector backbone, equimolar with the insertion fragment; and the Phusion polymerase,
144 0.04 U/μL. POE-PCR was conducted at 98 ^{o}C denaturation for 30 s; 30 cycles of 98 ^{o}C
145 denaturation for 10 s, 60 ^{o}C annealing for 10 s, and extension at 72 ^{o}C at a rate of 2
146 kb/min for the length of the resulting plasmid. For some DNA fragments with
147 complicated structures, the longer extension time was recommended when the estimated
148 polymerase rate was 1.5 kb/min for ensuring the generation of DNA multimers. One or
149 two microliter of the PCR product was used to check the formation of th`e DNA
150 multimers with 0.8% agarose gel electrophoresis. Another one or two microliter of the
151 PCR product was digested by restriction enzymes for checking plasmid map with 0.8%

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

152 agarose gel electrophoresis. One μL of the PCR product (approximately 100 ng DNA)

153 was also transformed into *E. coli* as a control.

154

155 Approximately 45 μL of the POE-PCR product (i.e., 100 ng/μL DNA multimers) was

156 digested in 300 μL of the NEB XhoI digestion buffer containing 200 units of XhoI at 37

157 ^{o}C for 5 h. The mixture was. The digested DNA multimers were applied to 0.8% agarose

158 gel electrophoresis. The linear plasmid DNAs with the targeted size was purified with a

159 Zymo gel DNA recovery kit. The purified linear plasmid DNAs (approximately1.5 μg)

160 were ligated in 600 μL Quick ligation system containing 1,200 units of the NEB T4 DNA

161 ligase at 25°C for 5 min. A small fraction of the ligation product (e.g., 50 ng) was mixed

162 with 100 μL of the competent *E. coli* cells in a 1.5-mL regular centrifuge tube. The tube

163 was kept at 0 ^{o}C for 30 min, 42 ^{o}C for 90 s, and 0 ^{o}C for 5 min. One mL of the LB liquid

164 medium was added to the tube. The cell cultures in the tube were cultivated with well-

165 mixing at 37°C for 45 min. After centrifugation at 5,000 rpm for 5 min, one mL of the

166 supernatant was discarded. The cell pellets were resuspended in the remaining liquid, and

167 were spread on the petri plates containing the LB solid medium supplemented with 100

168 ng/mL ampicillin. The petri plates were incubated at 37 ^{o}C for 12-16 h or until the

169 colonies were easily examined by eyes.

170

171 **Protein production and purification.** The strain *E. coli* BL21 Star (DE3) containing the

172 protein expression plasmid was cultivated in the LB medium supplemented with 1.2%

173 glycerol at 37 °C. Once the absorbance of the culture reached ~0.8, IPTG was added at a

ACCEPTED MANUSCRIPT

174 final concentration of 0.1 mM. After 16 h of cultivation at 20 ^{o}C, the *E. coli* cells were
175 harvested by centrifugation and re-suspended in a 50 mM HEPES buffer (pH 7.2). The
176 cells were lysed by sonication. After centrifugation, the His-tagged protein in the
177 supernatant was adsorbed to the Bio-Rad Profinity IMAC Ni-resin (Hercules, CA) and
178 then eluted with a 50 mM HEPES buffer (pH 7.2) containing 250mM imidazole after
179 washed by a 50 mM HEPES buffer (pH 7.2) containing 20 mM imidazole.
180
181 **Other assays.** For the quantification of the DNA multimers, two μL of the PCR product
182 was digested with the restriction enzyme XhoI and then was subjected to 0.8% agarose
183 gel electrophoresis and analyzed with Quantity One (Version 4.6.7). Protein mass
184 concentration was measured by the Bio-Rad Bradford protein dye reagent method with
185 bovine serum albumin as a reference. The plasmid mutants were sequenced by DNA
186 sequencing.
187

188 Results

189 Primer design

190 For the formation of DNA multimers by POE-PCR (Fig. 2), the insertion DNA fragment
191 and vector backbone must have the overlap regions at 3’ and 5’ termini. To generate the
192 insertion DNA mutant library by error-prone PCR, a primer pair of IF and IR were used
193 to amplify the linear DNA fragment, where mutation rates could be adjusted by the
194 concentrations of Mg^{2+}, Mn^{2+}, and dNTPs [6]. Primer IF consisted of two fragments – a
195 3’ terminus responsible for amplifying the inserted DNA fragment by error-prone PCR
196 and a 5’ terminus responsible for overlapping with the host vector DNA fragment by

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

197 POE-PCR. Primer IR consisted of two fragments – a 3' terminus responsible for
198 amplifying the inserted DNA fragment by error-prone PCR and a 5' terminus responsible
199 for overlapping with the host vector DNA fragment by POE-PCR. To generate the linear
200 vector backbone, a primer pair of VF/VR was used. IF had a reverse complementary
201 sequence with VR. IR had a reverse complementary sequence with VF, too. It was
202 possible to add a unique restriction enzyme cutting site in the primer IF/VR or IR/VF
203 (Fig. 3).
204
205 **Transformation of DNA multimers and circular plasmids to *E. coli***
206 When the insertion DNA mutant library and the backbone vector was mixed together
207 without primers added, DNA multimers were generated by POE-PCR [20], where 1.5-2
208 fold of longer extension time was used than that in regular overlap extension PCR. The
209 DNA multimers were digested by one restricting enzyme, whose sole cutting site could
210 be located in the backbone vector, yielding linear plasmid DNAs. Circular plasmids were
211 obtained from linear plasmid DNAs with a T4 ligase, and then were transferred to
212 competent *E. coli* (Fig. 2).
213
214 Wild-type mCherry fluorescent protein was chosen for the proof-of-the-concept
215 experiment for facile preparation of a large size mutant library for directed evolution.
216 This recombinant protein was expressed under the control of T7 promoter (Fig. 4a). In
217 the LB solid plate containing a trace amount of lactose as an inducer, the color of the
218 expressed fluorescent protein mutants could be easily identified by eye. The 4.2-kb linear
219 vector backbone and the 706-bp *mcherry* gene were generated by regular PCR, as shown

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

220 in lane 1 and lane 2, respectively (Fig. 4b). Large molecular weight DNA multimers
221 generated by POE-PCR did not migrate from the well (Fig. 4b, Lane 3). The NdeI/XhoI-
222 digested DNA multimers exhibited two bands (Fig. 4b, lane 4), in good agreement with
223 the desired chimeric plasmid pET20b-mcherry-cbm digested with NdeI/XhoI (Fig. 4b,
224 lane 7). The XhoI-digested multimers exhibited a single band of 4.9 kb (Fig. 4b, lane 5),
225 indicating linear plasmids. After ligation, circular plasmids (Fig. 4b, lane 6) had a slower
226 migration rate than linear plasmids (lane 5). Circular plasmids can be transformed to
227 competent *E. coli* cells. By using commercial competent cells with a transformation
228 efficiency of 1.8×10^9 cfu per μg of intact plasmid DNA, the circular plasmids resulted
229 in an efficiency of 1.3×10^7 cfu per μg of DNA (Table 1). DNA multimers without
230 digestion and ligation can be also transformed to *E. coli* but nearly three order of
231 magnitude of lower transformation efficiency was obtained (Table 1).
232
233 One hundred fifty ng of DNA multimers from POE-PCR and fifty ng of circular DNAs
234 isolated from Lane 6 (Fig. 4b) were transformed into low-transformation efficiency self-
235 made BL21(DE3) competent cells, respectively. DNA multimers resulted in a
236 transformation efficiency, 4.2×10^2 per μg of DNA (Fig. 4c and Table 1), which was
237 good for regular subcloning [20] but was not enough for directed evolution. The circular
238 plasmids had an efficiency of 8.6×10^4 cfu per μg of DNA (Fig. 4d and Table 1). Clearly,
239 circular plasmids obtained from DNA multimers through one restriction enzyme
240 digestion and ligation was vital to construct a large size mutant library. Almost all the
241 colonies on the plate of Fig. 4c and d were red, suggesting that a high possibility of

ACCEPTED MANUSCRIPT

242 obtaining desired positive clones through direct transformation DNA multimers and
243 transformation of circular plasmids.
244
245 **mCherry protein variants**
246 Via this method, a large DNA mutant library encoding mCherry fluorescent protein was
247 generated by error-prone PCR. When 5 mM $MgCl_2$ and 0.2 mM $MnCl_2$ was added in the
248 PCR solution, approximately 70% colonies were colorless (Fig. 5a), suggesting that most
249 of mutants were negative. The size of this mutant library was estimated to be 1.95×10^7
250 cfu when 1.5 μg of circular plasmids was transformed to commercial high-efficiency
251 competent cells. Among them, several fluorescent protein mutants exhibiting different
252 colors were easily identified in petri dish plates. Four purified protein mutants exhibited
253 different colors under natural visible lights and UV excitation (Fig. 5 b and c). Plasmids
254 encoding fluorescent protein mutants were sequenced. Four of them had three to five base
255 changes, resulting in two to three amino acid changes (Table 2). Among them, one
256 mutant was yellow under natural visible light and had a very weak fluorescent signal. The
257 other three mutants also had much weaker colors and fluorescent signals than wild-type
258 mCherry fluorescent protein. For example, for the mutant protein 1, a mutation site
259 occurred at the amino acid site of 71, the amino acid at this site was essential. Upon
260 deleting this amino acid, the protein lost its color totally (data not shown). Despite all the
261 fluorescent intensities of the mutant protein were decreased, these mutants could provide
262 some useful information about influences of amino acid sites on fluorescent emissions.
263
264 **Discussion**

ACCEPTED MANUSCRIPT

265 We developed a facile preparation for constructing a large size mutant library (e.g., more
266 than10^7 protein mutants per 50 μL of the PCR product) for directed enzyme evolution in
267 *E. coli*. Frist, DNA multimers was generated by POE-PCR so that neither restriction
268 enzymes nor ligase were required. Second, DNA multimers were digested by one
269 restriction enzyme, whose site can be located in any place of plasmid, resulting in the
270 linear plasmids. Third, the linear plasmids were circularized by ligase and then
271 transformed to *E. coli* cells. Compared to the classical library construction method
272 through the digestion of two restriction enzymes followed by ligation, this new methods
273 had several advantages. (i) In the classic two-fragment ligation protocol, the optimization
274 of the ratio between two DNA fragments was vital to achieving high transformation
275 efficiency [6]. This labor-intensive optimization was not needed in this method because
276 of one-fragment ligation. (ii) Larger size libraries were generated through one-fragment
277 circularization than through two-fragment ligation. It was very easy for a beginner to
278 construct a large mutant library size with the transformation efficiency of 10^6-10^7 cfu/μg
279 using commercial competent cells. In contrast, experts with wise selection of restriction
280 enzymes, careful preparation of DNA fragments, and intensive optimization of the
281 ligation reaction between the inserted fragment and vector and transformation conditions
282 could obtain transformation efficiencies of 10^6 -10^7 cfu/μg of vector DNA [6]. (iii) No
283 optimization was required in the digestion through one restriction enzyme by using a very
284 high restriction enzyme loading, where the high-affinity restriction enzyme featuring
285 decreased star activity, such as BamHI-HF^{TM}, EcoRI-HF^{TM}, HindIII-HF^{TM}, and so on,
286 was highly recommended. (iv) The cutting site of this restriction enzyme can be located
287 in any place of the plasmid. (Note: it was recommended that the cutting site was located

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

288 in the vector backbone because the restriction site in the inserted mutant library may be
289 not available after mutagenesis. If there was not appropriate cutting site in vector
290 backbone, the primers can be designed containing a unique restriction enzyme cutting
291 site). (v) No extra amino acids were introduced in the ends of the target protein, while
292 several amino acids were often added in the classical protocol by designing two
293 restriction enzymes' cutting sites. Sometime, such extra amino acids may influence the
294 properties of target proteins [11]. Compared to other methods for mutant library
295 construction, like MEGAWHOP [16], MegAnneal [10], our method had its unique
296 advantages, such as large size of library, very low wild-type background contamination
297 (Fig. 4d), independence of *E. coli* strains.
298
299 The central of this method was the formation of DNA multimers by using POE-PCR
300 without restriction enzymes and ligase. The formation efficiency of DNA multimers were
301 influenced by several factors: (i) overlap length between the insertion fragment and
302 vector backbone, (ii) annealing temperature in POE-PCR, (iii) DNA template
303 concentration in POE-PCR, and (iv) extension time. Generally speaking, the longer
304 overlap region, the higher efficiency of the DNA multimer formation [19, 20]. According
305 to our experience, the DNA multimers could be obtained even when the overlap length
306 between two fragments was as short as 20 bp, like typical overlap extension PCR.
307 However, we recommended the overlap lengths of 40~50 bp so to ensure the formation
308 of DNA multimers. Like typical overlap extension PCR, melting temperatures of the
309 primers between 55° and 65°C generally yielded the best results. To construct a large size
310 mutant library in POE-PCR, the concentration of the inserted DNA mutant library was

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

311 recommended to be approximately 5 ng per µL (i.e., 3×10^{9} DNA molecules per
312 microliter, the total library size was 1.5×10^{10}) and the molar ratio of the DNA mutant
313 fragment to the linear vector backbone was 1:1. To ensure the formation of DNA
314 multimers, the extension time of POE-PCR was estimated to be based on the length of
315 resulting plasmid divided by a rate of 1.5 to 2 kb per min for the Phusion polymerase [20].
316

317 Although a 700-bp DNA fragment generated by ep-PCR and a 4.2-kb plasmid backbone
318 were assembled for demonstration purpose here, this technology would be very useful to
319 engineer large-size proteins. For big-size proteins, it may be inefficient to generate a very
320 large size DNA library with a constant mutagenesis rate for the whole gene sequence [9].
321 Via this method, it was feasible to generate two or more small-size DNA fragments with
322 different mutation rates for hot and cold DNA sequences and assemble them together. In
323 our lab, DNA multimers can be assembled based on up to three DNA fragments with
324 variable lengths of insertion (0.3 to7.0 kb) and a vector backbone (3.5 to 8.5 kb) by using
325 POE-PCR. The largest length of chimeric plasmid through DNA multimers was
326 approximately 12 kb [20].
327

328 In conclusion, this POE-PCR-based protocol was simple and fast, and had great
329 flexibility for the constructing a large size mutant library for directed enzyme evolution in
330 *E. coli*.
331

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

332 **ACKNOWLEDGMENT.** This work was partially supported by the DOE BioEnergy
333 Science Center, the CALS Biodesign and Bioprocessing Research Center, Shell Game
334 Changer Program, and DOE ARPA-E Petro (DE-AR0000203).

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

335 **References**

336 [1] P.A. Romero, F.H. Arnold, Exploring protein fitness landscapes by directed evolution. Nat
337 Rev Mol Cell Biol 10 (2009) 866-876.

338 [2] H. Lin, V.W. Cornish, Screening and selection methods for large-scale analysis of protein
339 function. Angew Chem Int Ed 41 (2002) 4402-4425.

340 [3] H. Zhao, L. Giver, Z. Shao, J.A. Affholter, F.H. Arnold, Molecular evolution by staggered
341 extension process (StEP) *in vitro* recombination. Nat Biotechnol 16 (1998) 258-261.

342 [4] M.B. Quin, C. Schmidt-Dannert, Engineering of Biocatalysts: from Evolution to Creation.
343 ACS Catal 1 (2011) 1017-1021.

344 [5] C. Roodveldt, A. Aharoni, D.S. Tawfik, Directed evolution of proteins for heterologous
345 expression and stability. Current opinion in structural biology 15 (2005) 50-56.

346 [6] F.H. Arnold, G. Georgiou, Directed Evolution Library Creation: Methods and Protocols,
347 Humana Press, Totowa, 2003.

348 [7] F.H. Arnold, A.A. Volkov, Directed evolution of biocatalysts. Curr Opin Chem Biol 3 (1999)
349 54-59.

350 [8] W. Liu, J. Hong, D.R. Bevan, Y.-H.P. Zhang, Fast identification of thermostable beta-
351 glucosidase mutants on cellobiose by a novel combinatorial selection/screening approach.
352 Biotechnol Bioeng 103 (2009) 1087-1094.

353 [9] X. Ye, C. Zhang, Y.H.P. Zhang, Engineering a large protein by combined rational and
354 random approaches: stabilizing the *Clostridium thermocellum* cellobiose phosphorylase. Mol
355 BioSyst 8 (2012) 1815-1823.

356 [10] J.C. Pai, K.C. Entzminger, J.A. Maynard, Restriction enzyme-free construction of random
357 gene mutagenesis libraries in *Escherichia coli*. Anal Biochem 421 (2012) 640-648.

358 [11] W. Liu, X.-Z. Zhang, Z.-M. Zhang, Y.-H.P. Zhang, Engineering of *Clostridium*
359 *phytofermentans* endoglucanase Cel5A for improved thermostability Appl Environ Microbiol 76
360 (2010) 4914-4917.

361 [12] M. Abou-Nader, M.J. Benedik, Rapid generation of random mutant libraries.
362 Bioengineered Bugs 1 (2010) 337-340.

363 [13] J. Sambrook, D.W. Russel, Molecular Cloning: A Laboratory Manual. New York, USA.
364 (2001) Cold Spring Harbor Laboratory.

365 [14] E. Muñoz, M.W. Deem, Amino acid alphabet size in protein evolution experiments:
366 better to search a small library thoroughly or a large library sparsely? Protein Eng Des Sel 21
367 (2008) 311-317.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

368 [15] B.J. Hackel, K.D. Wittrup, The full amino acid repertoire is superior to serine/tyrosine for
369 selection of high affinity immunoglobulin G binders from the fibronectin scaffold. Protein Eng
370 Des Sel 23 (2010) 211-219.

371 [16] K. Miyazaki, M. Takenouchi, Creating random mutagenesis libraries using megaprimer
372 PCR of whole plasmid. Biotechniques 33 (2002) 1033-1034, 1036-1038.

373 [17] U. Krauss, K.E. Jaeger, T. Eggert, Rapid sequence scanning mutagenesis using in silico
374 oligo design and the Megaprimer PCR of whole plasmid method (MegaWHOP). Methods Mol
375 Biol 634 (2010) 127-135.

376 [18] H.H. Hogrefe, J. Cline, G.L. Youngblood, R.M. Allen, Creating randomized amino acid
377 libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. Biotechniques 33 (2002)
378 1158-1160, 1162, 1164-1155.

379 [19] X.Z. Zhang, Y.-H.P. Zhang, Simple, fast and high-efficiency transformation system for
380 directed evolution of cellulase in *Bacillus subtilis*. Microb Biotechnol 4 (2011) 98-105.

381 [20] C. You, X.-Z. Zhang, Y.-H.P. Zhang, Simple Cloning via direct transformation of PCR
382 product (DNA multimer) to *Escherichia coli* and *Bacillus subtilis*. Appl Environ Microbiol 78 (2012)
383 1593-1595.

384

385

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

386 **Figure legends**

387 **Figure 1.** The scheme of the mutant library construction. First, the linear insertion DNA
388 mutants were generated by error-prone PCR or DNA shuffling; the linear vector
389 backbone was generated by high-fidelity PCR. Second, DNA multimers were formed by
390 prolonged overlap extension PCR (POE-PCR) [20]. Third, linear plasmid DNAs were
391 obtained after DNA multimers were digested by a restriction enzyme. Forth, the circular
392 plasmid mutant library was ligated by T4 DNA ligase. Fifth, the circular plasmids were
393 transformed to *E. coli* with high efficiencies.

394

395 **Figure 2.** The mechanism of the formation of DNA multimers and of circular plasmids.
396 In POE-PCR, DNA multimers were generated from both 3' and 5' overlapped inserted
397 DNA mutants and vector backbone, both of which functioned as both templates and
398 primers. DNA multimers can be digested to linear DNAs by one restriction enzyme,
399 whose cutting site was located in the vector backbone.

400

401 **Figure 3**. The schematic representation of primer design. A primer pair of IF/IR was used
402 to amplify the linear DNA fragment. A primer pair of VF/VR was used to amplify the
403 vector backbone. Primer IF and IR both consisted of two fragments, a 3' terminus of 20-
404 25 bp of gene-specific sequence and a 5' terminus of 20-25 bp of vector-specific
405 sequence. IF had a reverse complementary sequence with VR, as well as IR and VF.

406

407 **Figure 4.** a). The map of the vector pET20b-mcherry-cbm. The arrows show the
408 transcription directions for these genes. *mcherry,* wild-type mCherry-encoding sequence;

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

409 *cbm,* family 17 carbohydrate-binding module-encoding sequence. IF, IR, VF, and VR
410 denote the position of the primers for PCR. b). DNA analysis by agarose gel
411 electrophoresis. Lane M, the NEB 1-kb DNA ladder; Lane 1, a linear vector of pET20b
412 containing *cbm* generated by regular PCR; Lane 2, *mcherry* DNA generated by regular
413 PCR ; Lane 3, DNA multimers generated by POE-RCR; Lane 4, digested DNA
414 multimers digested with NdeI and XhoI; Lane 5, XhoI-digested DNA multimers with;
415 Lane 6, the circular plasmids after XhoI-digestion followed by T4 ligation; and Lane 7,
416 the digested circular plasmids with NdeI and XhoI. c). Transformation of 150 ng DNA
417 multimers to self-made *E. coli* BL21(DE3) competent cells. d). Transformation of 50 ng
418 of the ligated circular plasmid to self-made *E. coli* BL21(DE3) competent cells.
419
420 **Figure 5.** The screening of mCherry fluorescence protein mutants expressed in *E. coli*
421 BL 21 (DE3) in a Petri dish plate (a), where colonies containing wild-type or neutral
422 mutants were red, a yellow colony represented a mutant with a new fluorescent color, and
423 most white colonies reflected negative mutagenesis. Purified fluorescent proteins with the
424 same concentrations under natural visible light (b) and under UV excitation (c). B, blank
425 buffer; WT, wild-type mCherry fluorescent protein; 1-4, mCherry mutant 1, 2, 3, and 4.

ACCEPTED MANUSCRIPT

Figure 1

ACCEPTED MANUSCRIPT

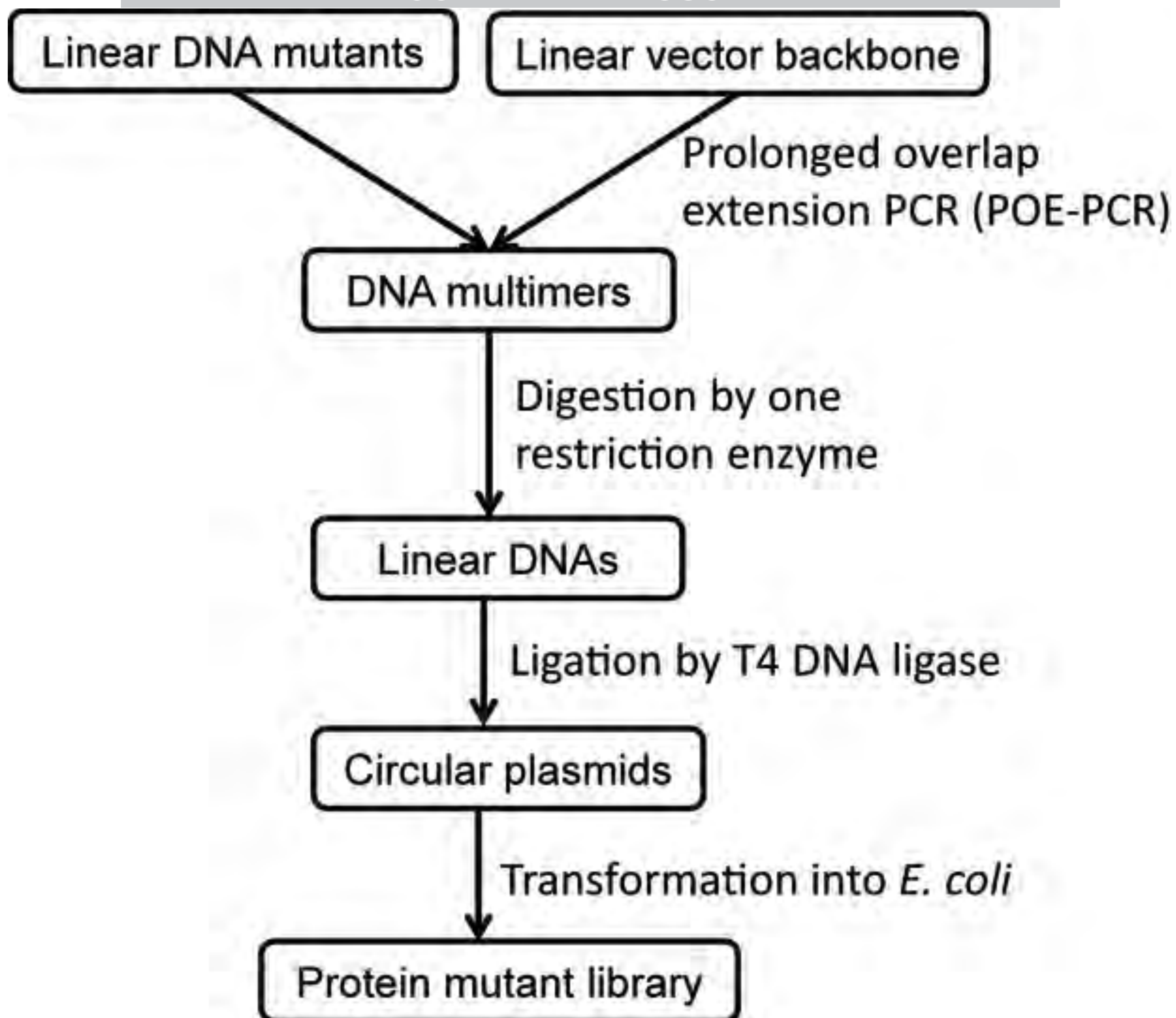
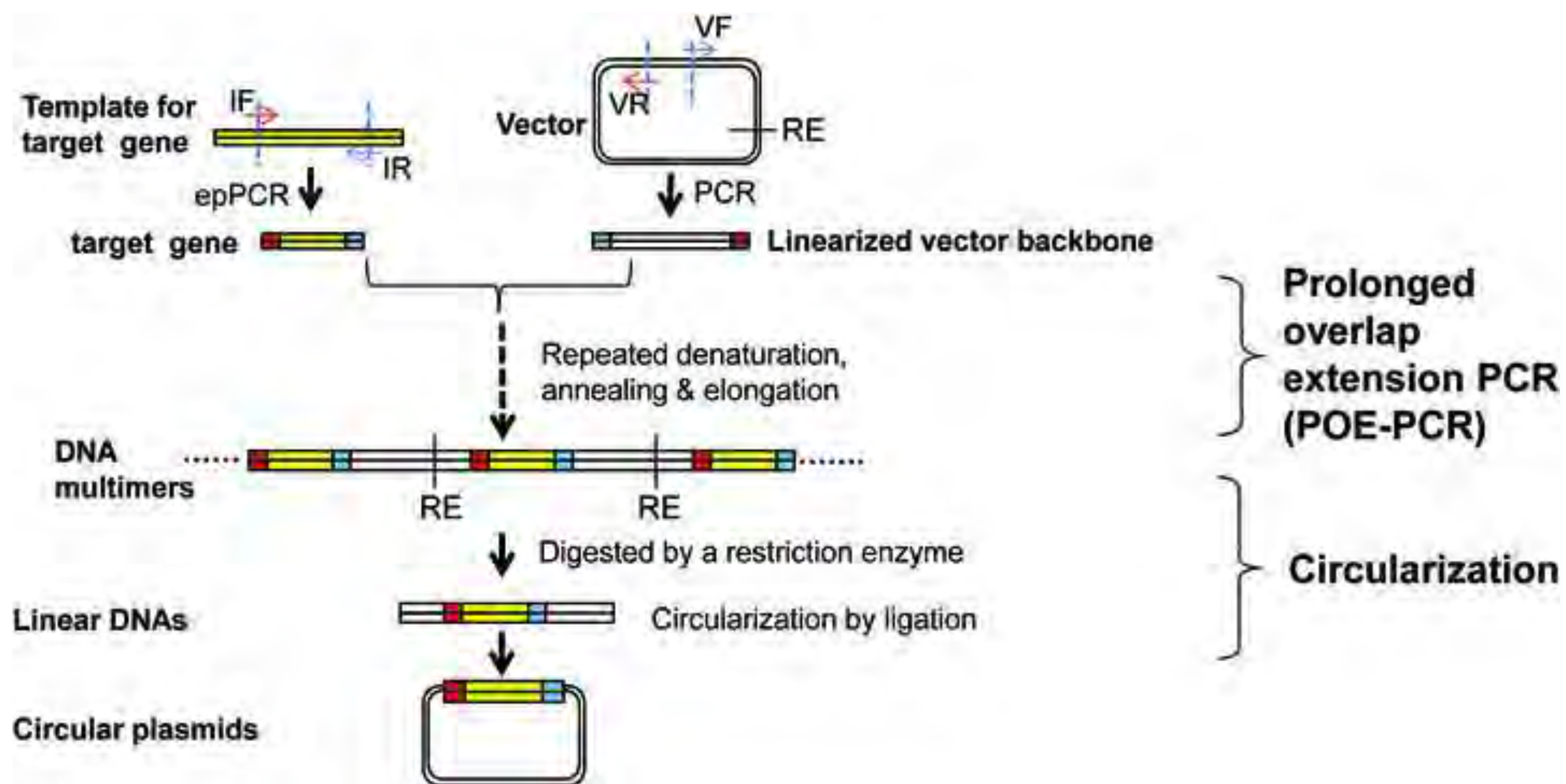


Figure 2

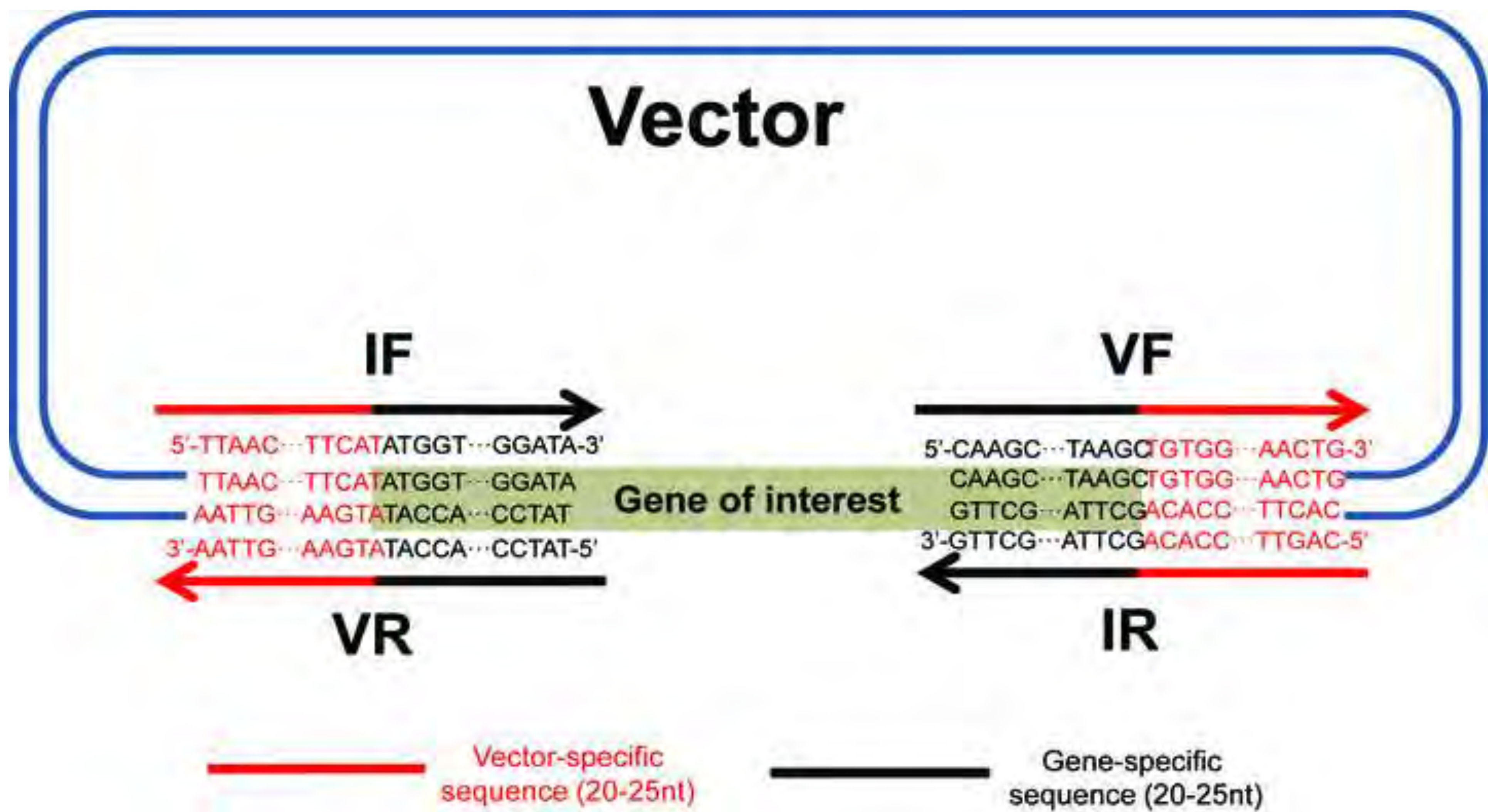
ACCEPTED MANUSCRIPT



ACCEP

Figure 3

ACCEPTED MANUSCRIPT



ACCE

Figure 4

ACCEPTED MANUSCRIPT

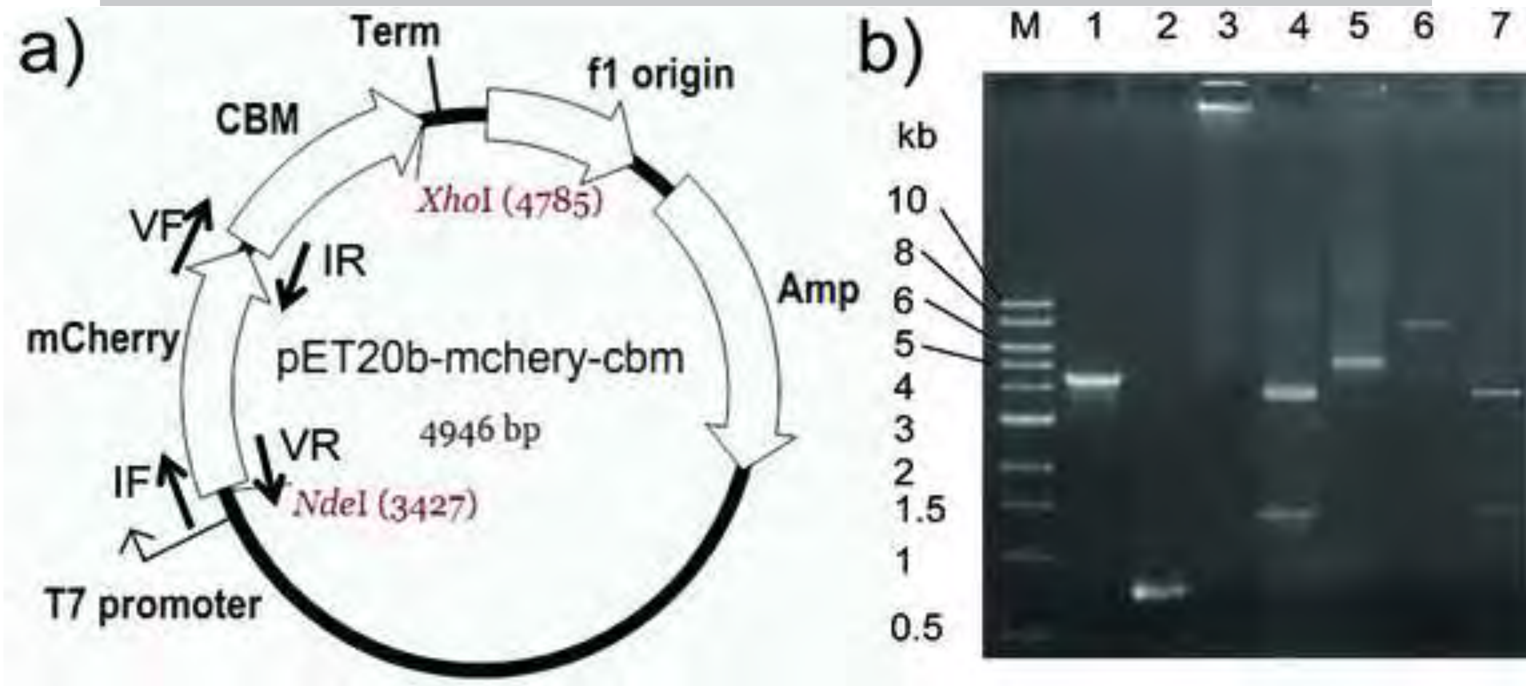
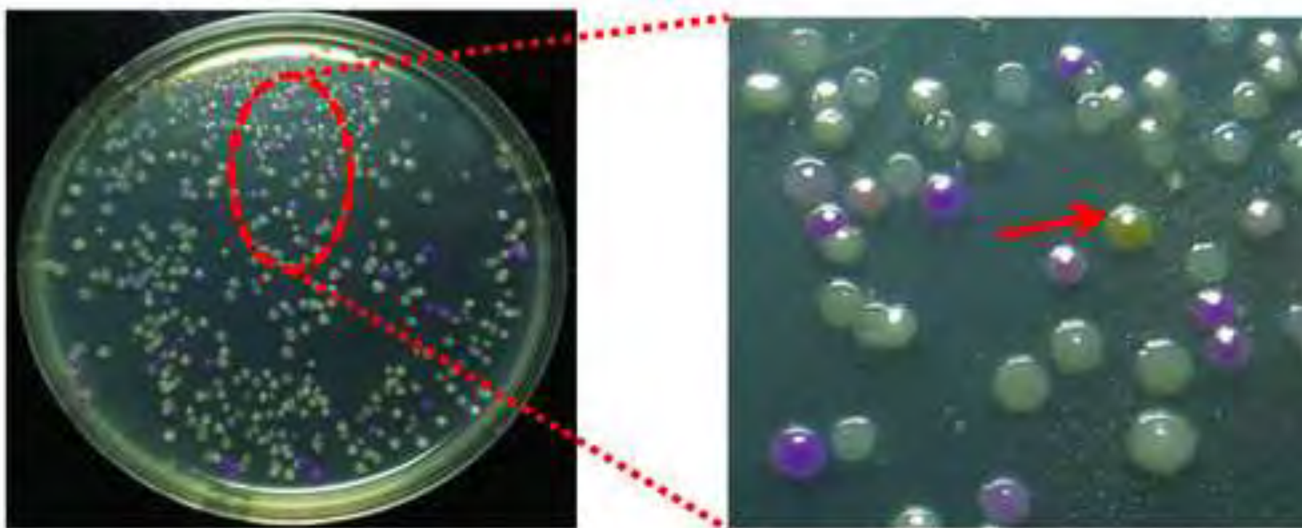


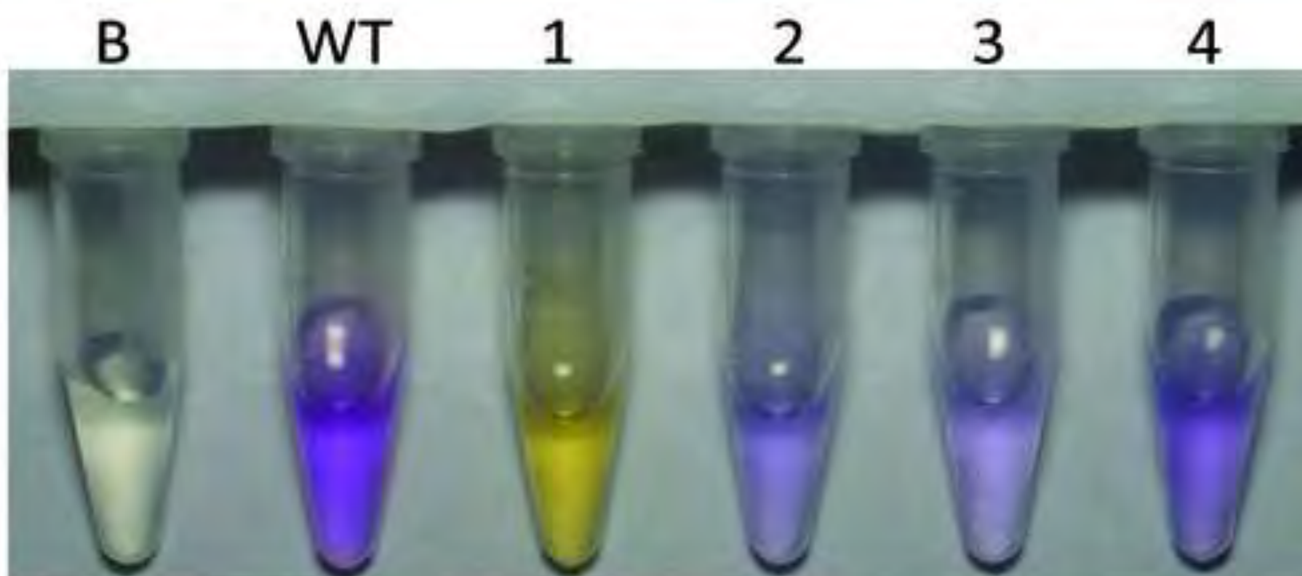
Figure 5

ACCEPTED MANUSCRIPT

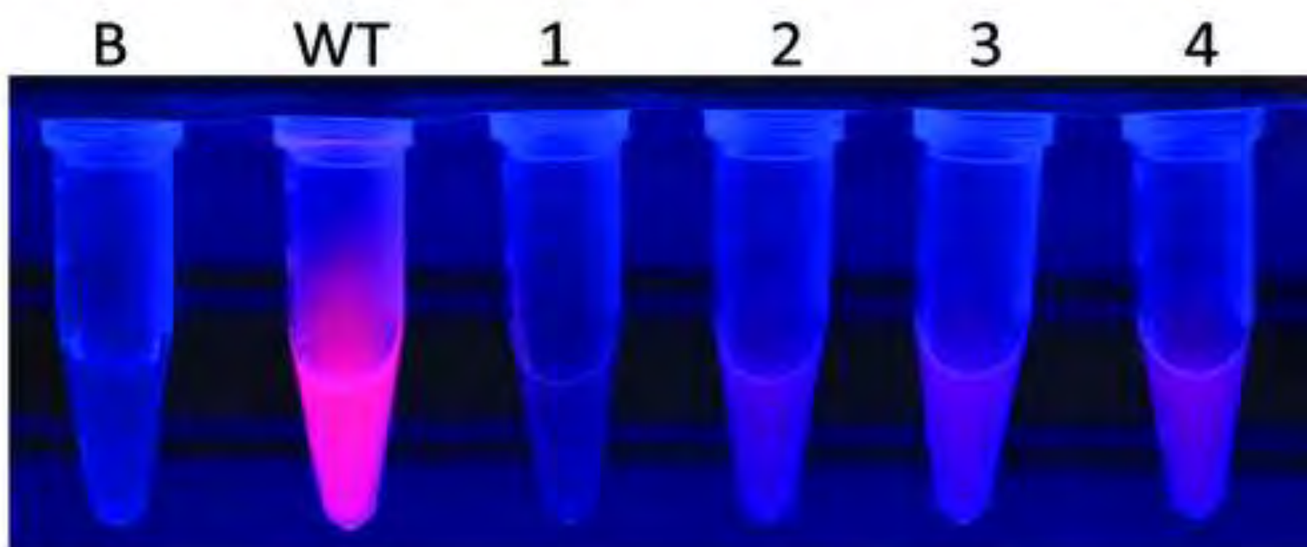
a)



b)



c)



ACCEPTED MANUSCRIPT

426 **Table 1.** Transformation efficiency of treated and untreated DNA multimers on self-
427 made and commercial competent *E.coli* BL21(DE3) cells

Source	Transformation efficiency (cfu/µg)	
	Commercial*	Self-made
DNA multimers	4.5×10^4	4.2×10^2
Circular plasmids	1.3×10^7	8.6×10^4
Plasmid	1.8×10^9	6.7×10^6

428 *The commercial competent cells were purchased from Invitrogen.
429

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

430

431 **Table 2**. Amino acid substitutions in the selected mCherry mutants and the
432 corresponding base mutations.

Mutant	Mutation DNA	Amino acid
1	A92T, T204C, A211T	Q31V, silent, M71L
2	A41T, T204A, T269A, T605C, G686C	K14M, silent, L90Q, I202T, G229A
3	T168A, T467G, A560G, T605C	silent, W156C, K187R, I202T
4	G325T, A425G, A503T	V109L, Q142R, Q168L

433

434

ACCEPTED MANUSCRIPT