Method to protect a targeted amino acid residue during random mutagenesis

Daisuke Umeno, Kaori Hiraga and Frances H. Arnold*

Division of Chemistry and Chemical Engineering, California Institute of Technology, 210-41, 1200 East California Boulevard, Pasadena, CA 91125, USA

ABSTRACT

To generate a random mutant library that is free from mutation at a particular amino acid residue, we replace the codon of interest with a detachable, short DNA sequence containing a BsaXI recognition site. After PCR mutagenesis, this sequence is removed and intramolecular ligation of the sequences flanking the insert regenerates the gene. The three-base cohesive ends for ligation correspond to the codon for the targeted residue and any sequences with mutations at this site will fail to ligate. As a result, only the variants that are free from mutation at this site are in the proper reading frame. In a random library of C_{30} carotenoid synthase CrtM, this method was used to exclude readily accessible mutations at position F26, which confer C_{40} synthase function. This enabled us to identify two additional mutations, W38C and E180G, which confer the same phenotype but are present in the random library at much lower frequencies.

INTRODUCTION

PCR mutagenesis is widely used to generate libraries of mutant sequences which are used to search for improved or novel functions. The mutation rate is usually set relatively low, to discover single amino acid substitutions that confer the desired phenotype (1,2). The ability to discover those mutations can be substantially decreased by experimental factors such as mutation bias in the PCR. PCR-based mutagenesis generates certain types of mutations more frequently than others (e.g. using *Taq* polymerases, T/C or A/G transitions are favored) (3,4). Beneficial, but rarer mutations, are therefore less likely to be discovered, especially when a limited number of clones can be screened. When more clones are screened, variants with rare mutations are difficult to identify because they are overshadowed by other, more frequent mutations.

Previously, we discovered a single amino acid substitution that causes the C_{30} carotenoid synthase CrtM to acquire C_{40} synthase activity (5). DNA sequence analysis of 10 positive mutants (out of 116 identified from the random library) identified amino acid substitutions at position F26 that alone confer the desired phenotype. Nine of the 10 sequences carried the same F26S (T → A) or F26L (T → C) mutation. We believed that the T/C mutation was dominating the pool of positive mutants, hiding other mutations that would also allow CrtM to function as a C_{40} synthase. Using the current method to prepare new mutagenic libraries free from mutation at F26, we were able to identify two additional key residues, W38 and E180, that contribute to this phenotype.

MATERIALS AND METHODS

Materials

GGDP synthase (*crtE*), phytoene synthase (*crtB*) and phytoene desaturase (*crtI*) are from *Erwinia uredovora*, as described (5,6). Diapophytoene synthase (*crtM*) and diapophytoene desaturase (*crtN*) are from *Staphylococcus aureus*, as described (5,7). AmpliTaq polymerase (Perkin-Elmer, Boston, MA) was used for mutagenic PCR, while Vent polymerase (New England Biolabs, Beverly, MA) was used for cloning PCR. BsaXI was purchased from New England Biolabs. All chemicals and reagents were of the highest available grade.

Plasmid construction

crtN was subcloned into the EcoRI/NcoI site of pUCmodII (6), resulting in pUC-*crtN*. *crtB* was removed from previously constructed pUC-*crtE-crtB-crtI* (5), resulting in pUC-*crtE-crtI*. From these two plasmids, genes and promoters (lacP-*crtN* and lacP-*crtE-crtI*) were PCR amplified and subcloned into the SalI site of pACYC184, resulting in pAC-*crtN* and pAC-*crtI-crtE*, respectively. The carotene synthase gene (*crtM*) was cloned into the XbaI/XhoI site in pUC18Nmod (5), resulting in pUC-*crtM*. pUb-*crtM* was constructed by removing a BsaXI site (located upstream of the lac promoter) from pUC-*crtM*. To pUb-*crtM*, the following sequence was substituted for the sequence (TTT) encoding F26: 5′-*TTC***TGA**CTTAAG<u>ACT</u>-GACT<u>CTCC</u>GCTAGCC*TTC*-3′; underlined and italicized sequences are the recognition site and cut sites for BsaXI [*NNN* (N)$_9$AC(N)$_5$CTCC(N)$_7$*NNN*, N can be any base], and the bold bases correspond to an in-frame stop codon. To do this, two pieces of PCR fragments, Mu and Md, were first generated using pUb-*crtM* as a template (Fig. 1). The primers used were pucfwd (5′-GAACGTGTTTTTGTGGATAAG-AGG-3′) and Insrev (5′-<u>CTAGCGGAGAGTCAGTCTTA-AG</u>TCA*GAA*AGCGTAAGAAAAGCTTTTTGAATG-3′) for Mu, and Insfwd (5′-<u>CTTAAGACTGACTCTCCGCTAG</u>CC-

*To whom correspondence should be addressed. Tel: +1 626 395 4162; Fax: +1 626 568 8743; Email: frances@cheme.caltech.edu
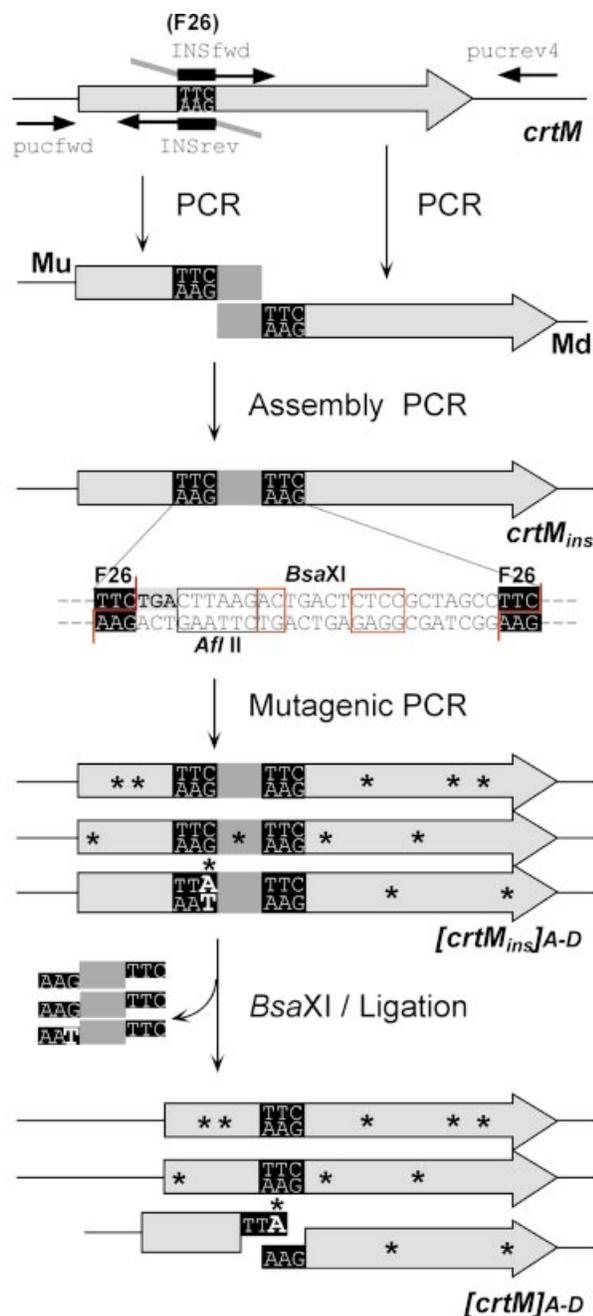
Figure 1. Schematic of method used to protect a specific amino acid residue during random mutagenesis. The residue of interest is replaced by a short, detachable DNA sequence containing a BsaXI recognition site. After random mutagenesis (here, error-prone PCR), this sequence is removed. Subsequent intramolecular ligation of the sequences flanking the insert regenerates the original codon. The three-base cohesive ends for ligation correspond to the codon for the target residue and any variants with mutations at these sites cannot be ligated. As a result, only the variants that are free from mutation at this site can restore their functional reading frame.

*TTC*GACTTGTTACCAGAAGATCAAAGAAAAGC-3′) and pucrev4 (5′-GATGAACGTGTTTTTTTGCGCAGACCG-3′) for Md. These fragments were assembled by a second PCR (underlined sequences in Mu and Md are complementary). The resultant PCR product was subcloned into pUb-*crtM* to yield pUb-*crtM$_{ins}$*.

PCR mutagenesis of *crtM$_{ins}$* gene

A pair of primers, pucfwd and pucrev4, flanking *crtM$_{ins}$* were designed to amplify the 0.9 kb gene by PCR under mutagenic conditions: 5 U AmpliTaq (100 µl total volume); 15 ng of template (pUb-*crtM$_{ins}$*, 3.4 kb); 50 pmol of each primer; 0.2 mM of each dNTP; 5.5 mM MgCl$_2$. Three mutagenic libraries were made using four different MnCl$_2$ concentrations: 0.2, 0.1, 0.05 and 0.02 mM. The cycling scheme was 95°C for 2 min followed by 30 cycles of 95°C for 30 s, 52°C for 30 s and 72°C for 1 min, followed by 72°C for 10 min. This yielded 5 µg of the 0.9 kb fragment, corresponding to an amplification factor of ~1000, or 10 effective cycles. The PCR product from each library was purified using a Zymoclean gel-purification kit (Zymo Research, Orange, CA) followed by digestion with XhoI and XbaI. The fragments were ligated into the XhoI–XbaI site of pUb-*crtM$_{ins}$* resulting in pUC-*[crtM$_{ins}$]* libraries (square brackets indicate the randomly mutagenized gene). The ligation mixtures were transformed into XL10 Gold cells (Stratagene) to yield 11 000–29 000 colony forming units (c.f.u.). Super-coiled library plasmid DNA was obtained by overnight culture of the transformation and miniprep (~5 µg). These plasmid mixtures were digested with BsaXI (8 U) for 5 h at 37°C. After purification, this reaction mixture was subjected to intra-plasmid ligation. The ligation mixture was gel-purified and further treated with AflII to remove any plasmids that still contain the insert. The reaction mixtures (pUb-*[crtM]$_{A-D}$*) were directly transformed into *Escherichia coli* XL1 Blue cells harboring pUC-*crtE-crtI* for screening C$_{40}$ synthase function.

Screening C$_{30}$ and C$_{40}$ carotenoid synthase activity

To screen for C$_{40}$ synthase activity, mutant *crtM* sequences were placed in the XbaI–XhoI site of pUb-*crtM* and transformed into XL1 Blue cells harboring pAC-*crtE-crtI*. C$_{30}$ synthase activity was evaluated by transforming pUC-*[crtM]* into XL1 Blue cells harboring pAC-*crtN*. Colonies were lifted onto white nitrocellulose membranes (Pall, Port Washington, NY) and placed at room temperature for an additional 12–24 h. Selected colonies were picked and cultured overnight in test tubes containing 3.0 ml of liquid LB medium supplemented with carbenicillin and chloramphenicol (50 µg/ml each) and were shaken for 12 h at 37°C. A portion (20 µl) from each preculture was then inoculated into 2 ml of fresh TB medium for further evaluation of synthase activity. After shaking for 36 h at 30°C, cells were harvested and extracted with acetone (1 ml). The highest peak (475 nm) in each UV/vis spectrum was used to score C$_{40}$ activity, and 470 nm was used for C$_{30}$ activity.

RESULTS AND DISCUSSION

Our method effectively masks the targeted amino acid during mutagenesis by selectively removing variants with mutations at that site from the library. To accomplish this, we introduced a removable DNA fragment at the codon corresponding to position F26 of CrtM (Fig. 1). Here, the F26 codon is duplicated and placed at both ends of the inserted sequence. The resulting *crtM$_{ins}$* is inactivated, due to the inserted stop codon. This insert contains a recognition site for BsaXI (type IIb endouclease), which recognizes $(N)_{12}AC(N)_5CTCC(N)_{10}$

Table 1. Characterization of the CrtM random mutant libraries

Transformed	$MnCl_2$ (μM)	Number of variants screened	Inactive (%)[a]	Colonies screened[b]	Number of positives (%)[b]
pUb-*crtM$_{ins}$*[c]	–	–	100	–	–
pUb-*crtM$_{(wt)}$*[d]	–	–	1.22	–	–
pUb-*[crtM]$_A$*	20	28 000	16.9	1300	0
pUb-*[crtM]$_B$*	50	26 000	24.4	2700	0
pUb-*[crtM]$_C$*	100	16 600	29.7	7500	5 (0.067)
pUb-*[crtM]$_D$*	200	10 900	50.6	4600	2 (0.043)
Total				16 400	7 (0.043)

[a]Fraction of variants without C_{30} carotenoid synthase activity. Estimated from the fraction of white colonies, when transformed in XL1 Blue cells harboring pAC-*crtN*.
[b]Screening was conducted for C_{40} carotenoid synthase activity. Libraries were transformed into XL1 Blue cells harboring pAC-*crtE-crtI*.
[c]BsaXI digestion and religation were not conducted for this sample.
[d]In place of PCR product, fragment prepared by gel extract from XbaI–XhoI digestion of pUb-*crtM$_{ins}$* was used.

and excises the complete insert exactly at both ends, leaving 3′ overhangs. BsaXI treatment to remove the insert followed by intramolecular ligation recovers the functional reading frame. Frame restoration proceeds only when the two ends are complementary. Mutation in the cut sites of BsaXI generates non-complementary cohesive ends and the reading frame cannot be recovered. If mutation occurs in the recognition site of BsaXI, the insert is no longer removable, and the mutant also remains out of frame and inactive. Thus, the functional clones in a mutagenic library prepared by this method should be essentially free from mutation at F26.

Construction of mutant libraries

We conducted mutagenic PCR under four different conditions using pUb-*crtM$_{ins}$* as the template, resulting in four libraries, pUb-*[crtM]$_{insA-D}$*. Each supercoiled plasmid library was obtained by subcloning into the pUb vector, followed by transformation into XL10 Gold cells. A fraction of the supercoiled plasmid was subjected to the BsaXI treatment to remove the insert, resulting in three-base 3′ sticky ends (TTC) on both sides, followed by ligation.

Before screening for C_{40} synthase activity, we transformed the resultant mixture into XL1 Blue cells harboring pAC-*crtN* to check what fraction was functional. Colonies harboring wild-type CrtM and its active variants exhibit a yellow color on agar plates; non-active ones should remain pale (5). Before removal of the insert (pUb-*crtM$_{ins}$*), no colored colonies were observed upon transformation into XL1 Blue cells harboring pAC-*crtN* (Table 1), due to inactivation of the CrtM gene by the insert. When the insert was removed from pUb-*crtM$_{ins}$* by BsaXI, followed by ligation (Fig. 1), 99% of the population

[see pUb-*crtM$_{(wt)}$* in Table 1] showed yellow pigmentation when transformed in XL1 Blue cells carrying pAC-*crtN*. Thus, we estimate that 1% of the sequences are undigested and retain the insert, while 99% have CrtM variants in the right frame (free from insert). For mutagenic libraries (after BsaXI treatment and ligation), the inactive ratio was 17, 24, 30 and 51% for libraries A, B, C and D, respectively. As is the general trend, the active fraction decreased with increased concentration of $MnCl_2$ in the PCR, reflecting the increased mutation frequency.

Screening for C_{40} synthase function

Each library (A–D) was transformed into XL1 Blue cells harboring pAC-*crtE-crtI*. Here, CrtM variants with C_{40} synthase activity should be pink, due to the production of lycopene and 3,4,3′,4′-tetradehydrolycopene (5). Colonies with significant pigmentation were picked and verified for C_{40} carotenoid production. Out of 16 400 colonies, we isolated seven and sequenced them (Table 2). A total of 15 base mutations were found. Notably, the frequency (~0.05%) of positive clones in the libraries generated by this method was 10 times lower than that of libraries produced by random mutations of the whole ORF [0.5%, see Umeno *et al.* (5)]. Because the previous PCR conditions were almost identical to the ones in the present work, together with the fact that ~99% of the population is in-frame and functional, we conclude that preparing the libraries by this method removed ~90% of the positive variants.

It should be noted that our method does not guarantee that mutations at the protected site are 100% removed. T4 DNA ligase, for instance, can ligate mismatched cohesive ends (8),

Table 2. Mutations in CrtM variants isolated in this work

Name of clone	C_{40} synthase activity[a]	Base mutation (amino acid substitution) Non-synonymous	Synonymous/outside ORF
C_1	+	T446C (V156A), A539G (G180E)	
C_2, C_5	+++	G114T (W38C)	
C_3, C_4	+++	G114T (W38C)	A-14G (in RBS)
D_1	++++	T36A (H12Q), G114T (W38C)	T582C, A717G
D_2	+++	G114T (W38C)	A564G, A768G

[a]Determined based on the hue of colonies on LB agar (solid) plates when the CrtM variant was expressed in XL1 Blue cells harboring pAC-*crtE-crtI*.

although at a very low rate. Thus, a small number of variants with a mutation at the targeted site could still undergo ligation, with some fraction of those appearing in the library, depending on the *E.coli* DNA repair system.

Random mutagenesis methods are frequently biased. In PCR mutagenesis, mutations of the type T/A-to-x occur far more often than others (3,4). Thus, the frequency of mutation at F26 (TTT) is expected to be relatively high. Indeed, nine out of 10 sequenced mutants with C_{40} activity had mutations at this site (5). That none of the positive variants from the BsaXI libraries had a mutation at position F26 allowed us to discover other mutations that confer the same phenotype (Table 2). Clones C_2 and C_5 have exactly the same genotype and are believed to be from the same PCR product. C_3 and C_4 also share genotype, but we do not know if they are from the same PCR product or not, because they contain only one mutation. D_1 and D_2 are unique and have more mutations. Notably, all of the above mutants share the mutation at W38. C_1, the only clone free from this mutation, has two mutations, leading to V156G and E180G. The latter was also found previously in mutant M_{10} (M4V, H12R, F59S, Q81R, E180G) (5). Our assumption that E180G was responsible for the C_{40} function was confirmed by creating a CrtM mutant with E180G alone. This variant showed similar C_{40} synthase activity to C_1.

CONCLUSION

Because BsaXI excises DNA irrespective of the sequences at cut sites, this method can be used to protect any single amino acid simply by substituting TTC with nucleotides coding the amino acid of interest. Multiple residues can be protected simultaneously by inserting other detachable sequences. By applying this method, we can rapidly find multiple functional solutions to specific problems (different positive mutants), even when the frequencies of the solutions are very different. Using this method, we have identified amino acids that control the size specificity of carotenoid synthase CrtM. Further engineering at these key residues has allowed us to create novel synthase activities (D. Umeno and F. H. Arnold, manuscript submitted for publication).

REFERENCES

1. Cirino,P.C., Mayer,K.M. and Umeno,D. (2003) Generating mutant libraries using error-prone PCR. In Arnold,F.H. and Georgiou,G. (eds), *Directed Evolution Library Creation: Methods and Protocols. Methods in Molecular Biology.* Humana Press Inc., Totowa, NJ, Vol. 231, pp. 3–10.
2. Matsumura,I. and Ellington,A.D. (2001) Mutagenic polymerase chain reaction of protein-coding genes for *in vitro* evolution. In Braman,J. (ed.), *In Vitro Mutagenesis Protocols. Methods in Molecular Biology.* Humana Press Inc., Totowa, NJ, Vol. 182, pp. 259–267.
3. Shafikhani,S., Siegel,R.A., Ferrari,E. and Schellenberger,V. (1997) Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. *Biotechniques*, **23**, 304–310.
4. Lin-Goerke,J.L., Robbins,D.J. and Burczak,J.D. (1997) PCR-based random mutagenesis using manganese and reduced dNTP concentration. *Biotechniques*, **23**, 409–412.
5. Umeno,D., Tobias,A.V. and Arnold,F.H. (2002) Evolution of C_{30} carotenoid synthase CrtM for function in a C_{40} pathway. *J. Bacteriol.*, **184**, 6690–6699.
6. Schmidt-Dannert,C., Umeno,D. and Arnold,F.H. (2000) Molecular breeding of carotenoid biosynthetic pathways. *Nat. Biotechnol.*, **18**, 750–753.
7. Umeno,D. and Arnold,F.H. (2003) A C_{35} carotenoid biosynthetic pathway. *Appl. Environ. Microbiol.*, **69**, 3573–3579.
8. Wiaderkiewicz,R. and Ruiz-Carrillo,A. (1987) Mismatch and blunt to protruding-end joining by DNA ligases. *Nucleic Acids Res.*, **15**, 7831–7848.