FEBS 24727 FEBS Letters 493 (2001) 139–143

Directed evolution of β-galactosidase from *Escherichia coli* by mutator strains defective in the 3′→5′ exonuclease activity of DNA polymerase III

Alessandra Stefan, Annalisa Radeghieri, Antonio Gonzalez Vara y Rodriguez, Alejandro Hochkoeppler*

Department of Industrial Chemistry, University of Bologna, Viale Risorgimento 4, 40136 Bologna, Italy

Received 5 February 2001; accepted 5 March 2001

First published online 15 March 2001

Edited by Gianni Cesareni

Abstract Directed evolution of *Escherichia coli* β-galactosidase into variants featuring β-glucosidase activity was challenged. To this end, mutagenesis of *lacZ* was performed by replication in *E. coli* CC954, a mutator strain containing a DNA polymerase III defective in 3′→5′ exonuclease activity. β-Galactosidase variants can be isolated upon mutagenesis of *lacZ* hosted into the self-transmissible episome F′128. Optimal evolution of *lacZ* can be achieved by propagation of *E. coli* CC954/F′128 cultures for 15 generations; further growth of mutator cultures for 37 or 55 generations imposes a high mutational load on *lacZ* and hinders the selection of efficiently evolved clones.

Key words: β-Galactosidase; Directed evolution; Mutator strain; Conjugative plasmid; DNA polymerase III

1. Introduction

It has recently been shown that rational protein engineering can be successfully applied to tailor enzyme variants stable at boiling temperatures [1] or featuring catalytic activities absent in their wild-type counterparts [2].

The engineering of proteins independently of structural information has recently been challenged by directed evolution of enzymes, i.e. the generation of random variants to be screened or selected for a desired phenotype. In particular, this approach has already been used for the generation of enzyme variants featuring improved thermal stability [3], activity in polar organic solvents [4], altered substrate specificity [5], and cofactor-independent activity [6]. Moreover, the directed evolution of an entire operon has also been reported, paving the way for metabolic engineering [7], and the rational engineering of a β-barrel protein scaffold has been shown to be efficiently improved by the concomitant engineering of this structural motif by directed evolution [8].

Three major factors affect the effectiveness of directed evolution of enzymes: (i) the strategy used for introducing random mutations into cloned genes encoding target enzymes; (ii) the design of efficient screening or selection procedures to identify the evolved clones; and (iii) the availability of procedures for the recombination of evolved mutants [3,9–13].

The method of choice for the generation of random mutants is usually error-prone PCR [14], which inserts mutations during DNA amplification reactions. The frequency of mutations introduced by mutagenic PCR cannot be finely modulated, and the yield of amplified DNA decreases under conditions yielding a higher mutation frequency [14]. Moreover, the use of error-prone PCR for the generation of mutant libraries demands back-insertion of the mutant clones into appropriate expression vectors, and transformation of bacterial recipients suitable for the screening or selection of the desired enzyme variants.

Mutator *Escherichia coli* strains have been isolated and characterized for quite a number of years (for a review see [15]). These strains are defective in replicative mispairing repair (e.g. the *mutD* genotype associated with defective 3′→5′ exonuclease activity of DNA polymerase III (DNA pol-III)) or in post-replicative mismatch repair systems [15]. The *mutD* genotype is dominant and confers the strongest mutator phenotype; interestingly, the use of *E. coli mutD* strains for the introduction of random mutations into a target gene is linked with the number and type of replication events occurring in the mutator background. In particular, the frequency of mutations introduced into a target gene could be theoretically tuned by controlling the number of generations occurring in growing cultures of *E. coli mutD5*. It is noteworthy that such a relationship stems from the assumption that the target gene, cloned into an expression vector, is replicated as chromosomal genes are, i.e. by DNA pol-III. It is known that pBR322 and ColE1-type plasmids depend on DNA polymerase I (DNA pol-I) for replication [16]: it has been shown that in *E. coli* strains defective in DNA pol-I the copy number of pBR322 is dramatically reduced when compared with minichromosomes replicated into the same hosts [16]. Whether or not DNA pol-III can afford the replication of few copies of pBR322 and other ColE1-type plasmids is still an open question.

The present study reports on the directed evolution of β-galactosidase from *E. coli*. This enzyme has been evolved to feature β-glucosidase activity by random mutagenesis of the *lacZ* gene and identification of the best enzyme variants by selection of *E. coli* mutants growing on solid minimal medium with hydroquinone-β-D-glucopyranoside (arbutin) as the sole carbon source. Random mutagenesis was performed using *E. coli mutD5* for the mutagenic replication of the *lacZ*

*Corresponding author. Fax: (39)-51-2093673.
E-mail: hochko@ms.fci.unibo.it

0014-5793/01/$20.00
PII: S0014-5793(01)02293-1

gene, hosted into pBR322 or into the self-transmissible episome F′128. The evolution of the *lacZ* gene as a function of the number and types of replication events in the mutator background is reported. The mutations inserted into the evolved *lacZ* gene are also presented and discussed.

2. Materials and methods

2.1. Bacterial strains, plasmids, and growth media

E. coli CC954 [*ara*, Δ(*gpt-lac*), *rpsL*, *mutD5*, zae-502::Tn*10*/F′128 *lacI*^{q}, Δ*lacY4700*::*cat*], CSH109 [*ara*, Δ(*gpt-lac*), *rpsL*], and JM108 [*endA1*, *recA1*, *gyrA96*, *thi*, *hsdR17*, *relA1*, *supE44*, Δ(*lac-proAB*)] were grown in LB supplemented with the appropriate antibiotics. F$^-$ derivatives of *E. coli* CC954 were isolated by growth in non-selective medium (LB) and were identified by screening for Cam^s, LacZ$^-$, Pro$^-$ and Arg$^-$ phenotypes. Further screening for mutator phenotype revealed that one of these CC954 F$^-$ isolates has a mutation frequency comparable to the parent genotype. This strain was also grown in LB medium supplemented with the appropriate antibiotics.

Plasmid pUR278 [17] purification from *E. coli* K12 and preparation of *E. coli* CC954 F$^-$ competent cells were performed according to standard procedures [18].

When used, antibiotics were supplemented to growth media at the following concentrations: 50, 15, 25, 50, and 100 μg/ml for ampicillin, tetracycline, chloramphenicol, nalidixic acid, and rifampicin, respectively.

Conjugations were carried out by mating liquid donor and recipient cultures at logarithmic phase.

2.2. Mutagenesis, selections and screenings

Random mutagenesis of *lacZ* was performed as follows: (i) *E. coli* CC954 F′128 was grown in LB medium at 37°C and then conjugated with *E. coli* JM108; LacZ variants were selected by plating JM108 F′128 on solid M9 minimal medium [18] supplemented with 0.2% (w/v) hydroquinone-β-D-glucopyranoside (arbutin), 1 μg/ml thiamine, chloramphenicol, and previously spread with isopropyl-thiogalactopyranoside (IPTG) [18]; (ii) *E. coli* CC954 pUR278 was grown in LB medium at 37°C and pUR278 was isolated and purified from *E. coli* CC954 in order to transform *E. coli* JM108.

Selected (*E. coli* JM108 F′128) and transformed (*E. coli* JM108 pUR278) clones were screened on LB plates previously spread [18] with IPTG and with 5-bromo-4-chloro-3-indolyl-β-galactopyranoside (X-Gal) or 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-Gluc) [18]. Further screening was performed by assaying enzyme variants with 3.3 mM of *o*-nitrophenyl-β-D-galactopyranoside (ONPGal), *p*-nitrophenyl-β-D-galactopyranoside (PNPGal), *o*-nitrophenyl-β-D-glucopyranoside (ONPGluc), or *p*-nitrophenyl-β-D-glucopyranoside (PNPGluc). To this end selected clones were grown overnight and an aliquot of this culture was induced with IPTG for 3 h at 37°C. Induced and non-induced cells were permeabilized by addition of 2–3 drops of chloroform, and the hydrolysis of each substrate as a function of time was tested at pH 7.6 by measuring the increase in absorbance at 415 nm, and using 590 nm as the reference wavelength.

2.3. DNA sequencing

Sequencing of *lacZ* was performed on 500–700 gene fragments generated with the primers listed in Table 1. The amplified fragments were purified with Amicon Microcon PCR ultrafiltration tubes (Millipore), prepared for sequencing with the ABI Prism Big-Dye Terminator kit (PE Biosystems) and finally desalted with Sephadex G-50 spin columns (AutoSeq G-50, Amersham-Pharmacia).

3. Results and discussion

Directed evolution of β-galactosidase from *E. coli* into a β-glucosidase has been challenged by random mutagenesis of *lacZ* and by demanding growth of *E. coli* on solid M9 medium containing hydroquinone-β-D-glucopyranoside (arbutin) as the only carbon source. Random mutagenesis was performed by replication of *lacZ* in a mutator background, i.e. *E. coli* CC954 (*mutD5*). To this end, *lacZ* was hosted in two different expression vectors, F′128 and pUR278; F′128 is a F′ factor containing the *lac* operon [19], and pUR278 is a derivative of pBR322 containing *lacZ* under the control of the *lac* promoter [17].

The mutator phenotype of *E. coli* CC954 was characterized by determining the frequency of mutations yielding strains resistant to rifampicin or nalidixic acid. These phenotypes are conferred by mutations mapping to the genes *rpoB* (4026 bp, encoding the β-subunit of RNA polymerase) and *gyrA* (2625 bp, encoding the A subunit of DNA gyrase), respectively. To test the frequency of *rpoB* and *gyrA* mutations, cultures of *E. coli* CC954 (Nal^s, Rif^s) were grown for 10 generations in LB medium at 37°C, and were then plated on LB supplemented with rifampicin or nalidixic acid. *E. coli* CSH109 was used as a reference strain and its mutation frequency also determined. Resistance to rifampicin was acquired at a frequency equal to $(2 \pm 0.95) \times 10^{-7}$/(cell gene generation) and $(4.3 \pm 2.2) \times 10^{-10}$/(cell gene generation) by *E. coli* CC954 and CSH109, respectively. The corresponding mutation frequencies yielding resistance to nalidixic acid were $(0.9 \pm 0.6) \times 10^{-7}$/(cell gene generation) and $(4.1 \pm 1.6) \times 10^{-10}$/(cell gene generation) for *E. coli* CC954 and CSH109, respectively. The mutation frequencies determined by observing the appearance of Nal^r and Rif^r strains are thus consistent and indicate a difference of three orders of magnitude in the mutation frequency by *E. coli* CSH109 and CC954. Moreover, considering the similar molecular masses of genes *rpoB*, *gyrA*, and *lacZ* (3075 bp), it is reasonable to suppose that *lacZ*, hosted in *E. coli* CC954, could be randomly mutated at a frequency of 1.5×10^{-7}/(cell generation). To test this hypothesis *lacZ*, hosted into F′128, was randomly mutagenized by replication in *E. coli* CC954 (mutator phenotype), and by subsequent transfer of F′128 into a non-mutator background (*E. coli* JM108) by conjugation.

The frequency of *E. coli* JM108 F′128 Arb$^+$ clones as a function of the number of generations of *E. coli* CC954 independent cultures is shown in Fig. 1. Arb$^+$ mutants can be selected at a frequency of ca. 2×10^{-5}/(cell generation) upon growth of *E. coli* CC954 for 15 or 37 generations, while the

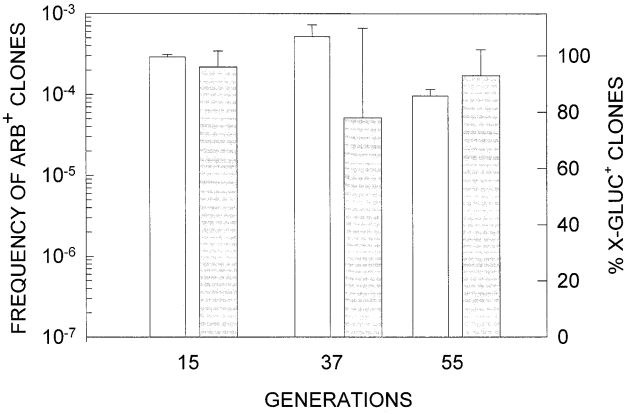

Fig. 1. *E. coli* CC954 F′128 cultures were grown for 15, 37, or 55 generations before conjugation with *E. coli* JM108. The corresponding frequencies of *E. coli* JM108 F′128 clones able to grow on solid M9 medium supplemented with arbutin are shown (empty bars) along with the percentage of selected clones featuring β-glucosidase activity as determined with X-Gluc. Error bars represent S.D. ($n = 3$).

Table 1
Sequences of the primers used to amplify 500–700 bp fragments of *lacZ* and to perform cycle sequencing reactions

Forward	Reverse
GCTATGACCATGATTACGGATTCAC	CGCATCGTAACCGTGCATCTGCC
GCGTTAACTCGGCGTTTCATCTGTG	GGAAATCGCTGATTTGTGTAGTCGG
GCCGATCGCGTCACACTACGTCTG	GATGGTTCGGATAATGCGAACAGC
GCAGCGCGATCGTAATCACCCGAG	GCGTCTCTCCAGGTAGCGAAAGCC
CGGCTTACGGCGGTGATTTTGGCG	GCAGGAGCTCGTTATCGCTATGACGG
GCATCTGACCACCAGCGAAATGG	CCTGATGCTGCCACGCGTGAGCG
CCAGCTGGCGCAGGTAGCAGAGCGG	GACACCAGACCAACTGGTAATGG

corresponding frequency determined for 55 generations is significantly decreased (Fig. 1, empty bars). While this decrease is most likely due to the absence of selection pressure during the growth of *E. coli* CC954 on LB medium, the frequencies observed upon the growth of the mutator strain for 15 or 37 generations are in reasonable agreement with the hypothesis previously formulated, yielding ca. 10^{-5} Arb^+/(cell generation) vs. 10^{-7} Rif^r or Nal^r/(cell generation). Moreover, it should be considered that the frequency of Arb^+ mutants depends on a decrease of specificity of *E. coli* β-galactosidase, while the selection of Rif^r or Nal^r mutants demands stricter conditions [20,21].

Fig. 1 also shows the relative frequency of Arb^+ clones featuring β-glucosidase activity as determined with X-Gluc. According to this analysis, 80–90% of the Arb^+ clones also feature the X-$Gluc^+$ phenotype independently of the number of generations of *E. coli* CC954 culture. This behavior was studied in more detail by quantitatively assaying the β-glucosidase activity of Arb^+ clones (see below).

In order to test whether or not ColE1-type vectors can be efficiently mutagenized by *E. coli* strains defective in DNA pol-III $3' \rightarrow 5'$ exonuclease activity, the purification of pUR278 from cultures of *E. coli* CC954 grown for 15, 37, or 55 generations was performed and the corresponding preparations used to transform *E. coli* JM108. LB solid medium supplemented with X-Gluc was then used to screen the transformants thus obtained (ca. 20 000 transformants were screened); independently of the type of plasmid preparation, pUR278 failed to yield transformants featuring β-glucosidase activity (blue colonies), indicating that DNA pol-III does not efficiently support the replication of this vector. This finding is in agreement with previous observations indicating that the copy number of ColE1-type plasmids is dramatically decreased in *E. coli* strains defective in DNA pol-I but competent in DNA pol-III activity [16]. Accordingly, a role of DNA pol-III appears unlikely either in initiation of replication or in complementary strand elongation of ColE1-type plasmids.

The rate of hydrolysis of nitrophenyl-glucosides by Arb^+, X-$Gluc^+$ *E. coli* JM108 F′128 clones was determined and compared with the parent activity, i.e. the rate of hydrolysis of nitrophenyl-galactosides. In this case four Arb^+, X-$Gluc^+$ clones were randomly chosen from each round of mutagenesis and their activity determined. β-Galactosidase (Fig. 2A) and β-glucosidase (Fig. 2B) activities decrease when *E. coli* CC954 F′128 cultures are grown for 37 or 55 generations before conjugation with *E. coli* JM108, as compared with the activities featured by *E. coli* JM108 F′128 clones obtained by conjugation with *E. coli* CC954 cultures grown for 15 generations; accordingly, to optimize directed evolution of enzymes it appears useful to apply selection pressure after mutator strain cultures (featuring mutation frequencies of ca. 10^{-7}/(cell gene generation)) have been grown for ca. 15 generations. The extent of directed evolution achievable under these optimal conditions was studied in detail by screening β-glucosidase and β-galactosidase activities of 36 *E. coli* JM108 F′128 clones obtained by conjugation with *E. coli* CC954 F′128 previously grown for 15 generations, and by calculating the β-glucosidase/β-galactosidase activity ratio of each clone. Two independent experiments were performed: the two highest ONPGluc/ONPGal activity ratios obtained were 0.0395 and 0.0402, while the PNPGluc/PNPGal activity ratio was found to increase among the clones up to limiting

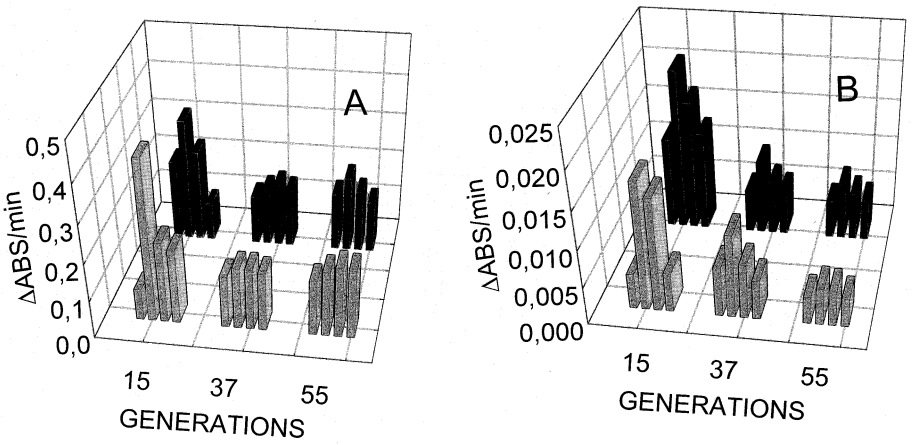

Fig. 2. β-Galactosidase (A) and β-glucosidase (B) activity of *E. coli* JM108 F′128 clones obtained by conjugation with *E. coli* CC954 F′128 previously grown for 15, 37, or 55 generations. Activities as determined with ONPGal and ONPGluc are reported (gray bars) along with activities determined with PNPGal and PNPGluc (black bars).

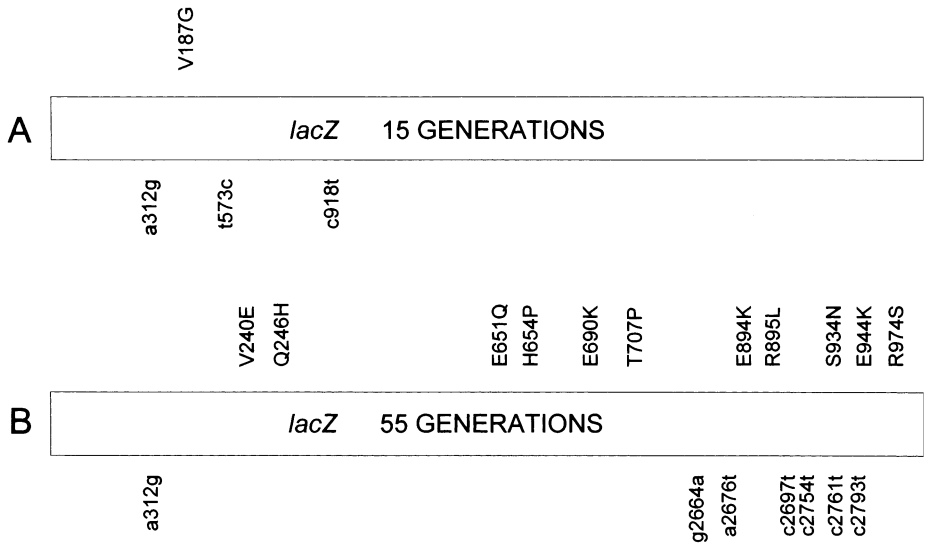

Fig. 3. Sequences of *lacZ* genes amplified after propagation of *E. coli* CC954 F′128 for 15 (clone A) or 55 (clone B) generations. Translated and silent mutations are indicated above and below each bar, respectively.

values of 0.087 and 0.065. The use of these clones as primers for further cycles of mutagenesis, selection and screening is currently under investigation.

To analyze in detail the mutations yielding β-glucosidase-competent clones two evolved *lacZ* genes were sequenced. In particular, *lacZ* was sequenced from the *E. coli* clone featuring the highest β-glucosidase/β-galactosidase activity ratio (clone A, isolated from cultures of *E. coli* CC954 propagated for 15 generations) and from the clone featuring the lowest β-glucosidase activity (clone B, isolated from cultures of *E. coli* CC954 propagated for 55 generations). Comparing these sequences with the sequence of wild-type *lacZ*, clone A features one translated (V187G) and three silent mutations, while clone B features 11 translated and seven silent mutations (Fig. 3). This observation is therefore consistent with the corresponding extent of random mutagenesis imposed on the two clones and indicates that the introduction of one to four mutations per mutagenesis round is optimal for the directed evolution of a target enzyme. Moreover, it is interesting to note that V187 resides in β-strand 10 of the first domain (spanning residues 51–217) of β-galactosidase, and is located quite far away from domain 3 spanning residues 334–627 and containing the active site pocket [22]. V187 is highly conserved among homologous β-galactosidases [22], and the altered specificity of β-galactosidase observed upon the introduction of the V187G mutation could be related to a distorted conformation of β-strand 10 and to a concomitant altered coordination of magnesium by D193. To test this hypothesis further, experiments are in progress.

4. Conclusions

The present report has shown that the use of *E. coli mutD5* strains for the directed evolution of enzymes is strongly affected by: (i) the type of plasmid that hosts the target gene to be evolved; ColE1-type plasmids are most likely replicated by DNA pol-I and their random mutagenesis has to be performed with other in vitro or in vivo methods; (ii) when using strains mutating at a frequency of ca. 10^{-7}/(cell gene generation) selection pressure has to be applied (or screening has to be performed) every ~15 generations of growth of the mutator strain harboring the gene to be evolved; the mutational load can in this way be kept to a minimum and the directed evolution of a gene more conveniently challenged.

Acknowledgements: We are indebted to Prof. Colin Manoil and to Dr. Jeannie Bailey (University of Washington, Seattle, WA, USA), for the generous gift of *E. coli* CC954 and for useful discussions. We are also indebted to Prof. Angela Algeri and to Prof. Bruno Andrea Melandri (University of Bologna, Italy) for encouragement and for critical reading of the manuscript. A.S. was supported by a grant from Consorzio per lo Sviluppo dei Sistemi a Grande Interfase (C.S.G.I., Italy).

References

[1] Van den Burg, B., Vriend, G., Veltman, O.R., Venema, G. and Eijsink, G.H. (1998) Proc. Natl. Acad. Sci. USA 95, 2056–2060.
[2] Quéméneur, E., Moutiez, M., Charbonnier, J. and Ménez, A. (1998) Nature 391, 301–304.
[3] Giver, L., Gershenson, A., Freskgard, P. and Arnold, F.H. (1998) Proc. Natl. Acad. Sci. USA 95, 12809–12813.
[4] Chen, K. and Arnold, F.H. (1993) Proc. Natl. Acad. Sci. USA 90, 5618–5622.
[5] Zhang, J. and Dawes, G. (1997) Proc. Natl. Acad. Sci. USA 94, 4504–4509.
[6] Joo, H., Lin, Z. and Arnold, F.H. (1999) Nature 399, 670–673.
[7] Crameri, A., Dawes, G., Rodriguez Jr., E., Silver, S. and Stemmer, W.P.C. (1997) Nature Biotechnol. 15, 436–438.
[8] Altamirano, M.M., Blackburne, J.M., Aguayo, C. and Fersht, A.R. (2000) Nature 403, 617–622.
[9] Stemmer, W.P.C. (1994) Nature 370, 389–391.
[10] Shao, Z., Zhao, H., Giver, L. and Arnold, F.H. (1998) Nucleic Acids Res. 26, 681–683.
[11] Stemmer, W.P.C. (1994) Proc. Natl. Acad. Sci. USA 91, 10747–10751.
[12] Zhao, H., Giver, L., Shao, Z., Affholter, J.H. and Arnold, F.H. (1998) Nature Biotechnol. 16, 258–261.
[13] Volkov, A.A., Shao, Z. and Arnold, F.H. (1999) Nucleic Acids Res. 27, 18.
[14] Leung, D.W., Chen, E. and Goeddel, D.V. (1989) Technique 1, 11–15.
[15] Miller, J.H. (1998) Mutat. Res. 409, 99–106.
[16] Kogoma, T. and Maldonado, R.R. (1997) J. Bacteriol. 179, 2109–2115.
[17] Rüther, U. and Müller-Hill, B. (1983) EMBO J. 2, 1791–1794.
[18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular

Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[19] Miller, J.H. (1992) A Short Course in Bacterial Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[20] Yonezawa, M., Takahata, M., Banzawa, N., Matsubara, N., Watanabe, Y. and Narita, H. (1995) Microbiol. Immunol. 39, 243–247.
[21] Jin, D.J. and Gross, C.A. (1988) J. Mol. Biol. 202, 45–58.
[22] Jacobson, R.H., Zhang, X.-J., DuBose, R.F. and Matthews, B.W. (1994) Nature 369, 761–766.