Vol. 199, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1994 Pages 410-417

MOLECULAR CLONING AND SEQUENCE ANALYSIS OF THE *proC* GENE ENCODING Δ^1-PYRROLINE-5-CARBOXYLATE REDUCTASE FROM AN EXTREMELY THERMOPHILIC EUBACTERIUM *THERMUS THERMOPHILUS*

Takayuki Hoshino*, Takehide Kosuge, Yuki Hidaka, Kazuyuki Tabata and Tadaatsu Nakahara

Institute of Applied Biochemistry, University of Tsukuba, Tsukuba, Ibaraki 305, Japan

Received January 18, 1994

A gene library of the extremely thermophilic bacterium, *Thermus thermophilus* HB27, was constructed in *Escherichia coli*, and recombinant plasmids able to complement *proC* mutants of HB27 were obtained. Using the plasmids, the complete nucleotide sequence of the *proC* gene encoding Δ^1-pyrroline-5-carboxylate reductase (P5CR) [EC 1.5.1.2] was determined. The deduced amino acid sequence showed a significant homology to those of P5CRs of *E. coli*, soybean and human. The *proC* gene of *T. thermophilus* was expressed in *E. coli* and the activity of heat-resistant P5CR was determined. The mutation sites of two HB27 *proC* mutants were also determined. © 1994 Academic Press, Inc.

It is well known that enzymes and proteins of thermophilic bacteria are stable at high temperatures. *Thermus thermophilus*, one of the extreme thermophiles for which extensive studies have been carried out, is a gram-negative, rod-shaped eubacterium (1) that can grow at temperatures between 50 and 82 °C. The strains belonging to the genus *Thermus* also produce heat-stable enzymes. Some of them are widely used in biotechnological applications. For example, *Taq* polymerase and *Tth* polymerase are routinely used in the polymerase chain reaction technology (2).

In this paper, we report the nucleotide sequence of the *proC* gene that encodes Δ^1-pyrroline-5-carboxylate reductase [EC 1.5.1.2] of *T. thermophilus*. The pathway of the proline biosynthesis defined in *Escherichia coli* is as follows (3): Glutamate is first activated by phosphorylation to form glutamyl phosphate and it is subsequently converted to glutamate semialdehyde. Glutamate semialdehyde spontaneously cyclizes to form Δ^1-pyrroline-5-carboxylate (P5C) which is then converted to proline. In the enteric bacteria, three enzymes are responsible for the reactions: γ-glutamyl kinase (ATP:L-glutamate 5-phosphotransferase [EC 2.7.2.11]), the product of *proB* gene, γ-glutamyl phosphate reductase (L-glutamate 5-semialdehyde:$NADP^+$ oxidoreductase [EC 1.2.1.41]), the product of *proA* gene, and P5C

*Corresponding author.

Abbreviations:
bp,basepair(s); kb, kilobasepair(s); Mb, megabasepair(s); P5C, Δ^1-pyrroline-5-carboxylate; P5CR, P5C reductase; ORF(s), open reading frame.

0006-291X/94 $5.00
Copyright © 1994 by Academic Press, Inc.
All rights of reproduction in any form reserved.

reductase (P5CR; L-proline:NAD(P)$^+$ 5-oxidoreductase [EC 1.5.1.2]), the product of *proC* gene. The final step in the proline biosynthetic pathway generates both proline and NADP$^+$, which is considered to stimulate the production of ribose-5-phosphate *via* the pentose phosphate shunt in human erythrocytes (4) and in soybean nodules (5). As a result, the amount of 5-phosphoribosyl 1-pyrophosphate increases, leading to an increase in *de novo* purine biosynthesis.

We are currently studying the proline biosynthetic pathway of *T. thermophilus* to determine whether the pathway is similar to that of mesophiles, and to analyze the thermostability of the proline biosynthetic enzymes by comparing their amino acid sequences with those of the mesophiles.

MATERIALS AND METHODS

Bacterial strains

Thermus thermophilus HB27 (6) and its proline auxotrophic mutants, pro3, pro4, pro5 and pro6 were used. The mutants were obtained by treating the wild type HB27 with N-methyl-N'-nitro-N-nitrosoguanidine. Four proline auxotrophs were tested for their growth on minimal medium (MM) (7) containing pyrroline-5-carboxylate (P5C). The mutants pro3 and pro5 were able to grow, unlike the pro4 and pro6 ones, suggesting that pro4 and pro6 were *proC* mutants. Thus, we designated them as *proC4* and *proC6*. *Escherichia coli* JM109 (8) was also used to construct the gene library of *T. thermophilus* and for the cloning experiments.

Construction of *T. thermophilus* gene library

Total DNA was isolated from *T. thermophilus* HB27 as described (9) and partially digested with *Msp*I. The digested fragments were run on a 0.8 % agarose gel and the fragments 2 to 4 kb in size were recovered by using GENECLEAN II (Funakoshi, Tokyo). They were ligated to *Acc*I-digested and dephosphorylated pUC19, and the ligation mixture was used to transform *E. coli* JM109. More than 20,000 transformants grown on H plates (10 g tryptone, 8 g NaCl, 12 g agar, per 1 l) containing 50 mg/l of ampicillin, 40 mg/l of isopropyl-1-thio-β-D-galactopyranoside (IPTG) and 40 mg/l of 5-bromo-4-chloro-3-indolyl-β-D-galactoside were all scraped, suspended in saline (0.85 % NaCl) containing 20 % (v/v) glycerol and stored at -30 °C.

Screening of the clones that complement *Thermus proC* mutants from the library

A part of the gene library stock was thawed on ice, diluted and plated onto H plates containing 50 mg/l of ampicillin, then incubated overnight at 37 °C. *E. coli* colonies thus obtained were replica-plated onto MM plates on which a fresh culture of either *proC4* or *proC6* mutant had been spread. The plates were incubated at 70 °C for 2 days. Assuming that an *E. coli* transformant contained a *Thermus* DNA fragment with a ProC$^+$ region, *proC* mutants could grow in the position where such an *E. coli* colony was detected, since transformation *via* homologous recombination between donor and host chromosomes takes place even on an agar plate. The recombinant plasmids were prepared by the alkaline lysis method (10) from the *E. coli* transformants thus screened.

Cloning of mutant *proC* genes from *proC4* and *proC6*

Total DNAs from *proC4* and *proC6* were digested with *Acc*I, run on an agarose gel and fragments about 1.0 kb in size were recovered. The fragments were ligated to pUC19 and introduced into *E. coli* JM109. The clones containing the entire mutant *proC* genes were screened by colony hybridization using the wild type 1.0 kb *Acc*I fragment (shaded box in Fig. 1) as a probe.

DNA sequencing

The restriction fragments to be sequenced were cloned into appropriate restriction sites of pUC19. DNA sequencing was performed using Taq Dye Primer Cycle Sequencing Kit (Applied Biosystems Japan, Tokyo) and an automatic sequencer model 373A (Applied Biosystems).

Expression of the *Thermus proC* gene in *E. coli*

The 1.0 kb *Acc*I fragment was recovered from pUC-pro4,6$^+$, blunt ended by T4 polymerase, then ligated to the *Sma*I site of an expression vector pPROK-C (Clontech). The *proC* expression plasmid pPROK-*proC* was thus obtained.

E. coli JM109 cells carrying pPROK-*proC* were grown in 200 ml of Luria broth (11) containing 100 mg/l of ampicillin. At an A_{600} of 0.7, 1 mM of IPTG was added to induce the tac promoter of the vector. After 3 hr incubation, cells were harvested. Crude cellular extracts were prepared essentially as described (12) except that sonication was performed in 3 ml of buffer A (50 mM Tris-HCl, 1 mM dithiothreitol, pH 7.2) and that heat treatment at 72 °C for 1 hr was applied in the final step. P5CR activity was measured by a modified method as described (13). The reaction mixture (0.3 ml) contained 1 mM DL-1-pyrroline-5-carboxylate (Sigma), 0.15 mM NADPH, 50 mM Tris base (pH 7.0) and the cell free extract. The reaction was run at 70 °C. SDS polyacrylamide gel electrophoresis of the crude extracts was performed according to the method of Laemmli (14).

RESULTS

Cloning and sequencing of the HB27 *proC* gene

Two clones which were able to complement the *proC* mutations were obtained. One clone, pUC-pro4$^+$ rescued only the *proC4* mutation, whereas the other, pUC-pro4,6$^+$ rescued both the *proC4* and *proC6* mutations. The sizes of the inserted fragments were 1.4 kb (pUC-pro4$^+$) and 2.1 kb (pUC-pro4,6$^+$), respectively. Restriction analysis of the two fragments showed that they overlapped with each other (Fig. 1), and the total region covered by the two fragments reached 2.7 kb. Subcloning experiments revealed that both the *proC4* and *proC6* mutations were rescued by the 1.0 kb *Acc*I fragment (shaded box in Fig. 1). Since the length of the *proC* gene of *E. coli* is about 800 bases (15), and the shaded region is present almost at the center of the cloned area, it was likely that the entire *proC*-encoding frame was located in the 2.7 kb region. The complete nucleotide sequence of the region was determined by sequencing both strands.

Eight complete or truncated open reading frames (ORFs) with a length exceeding 300 bases were identified in the 2.7 kb region. Since the codon usage of *Thermus* is known to be highly biased (16), four ORFs with percentages of use of G or C in the third letter of the codons higher than 90 % are shown in Fig. 1. Only ORF-1 was included in the 1.0 kb *Acc*I fragment, where both the *proC4* and *proC6* mutations were located. Thus, it was most likely that ORF-1

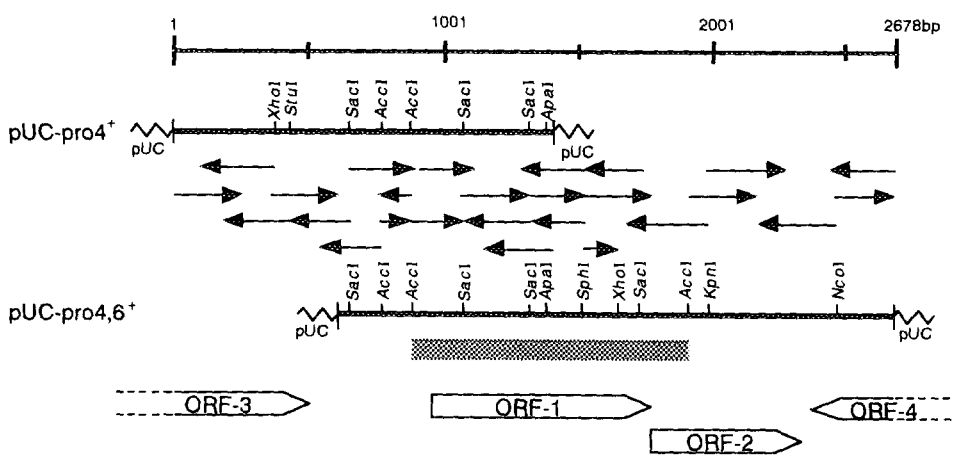

Fig. 1. Restriction maps of pUC-pro4$^+$ and pUC-pro4,6$^+$ and sequencing strategy. Only the restriction sites used for subcloning are shown. Each arrow represents the direction and distance of the sequencing. The shaded box represents a 1.0 kb *Acc*I fragment.

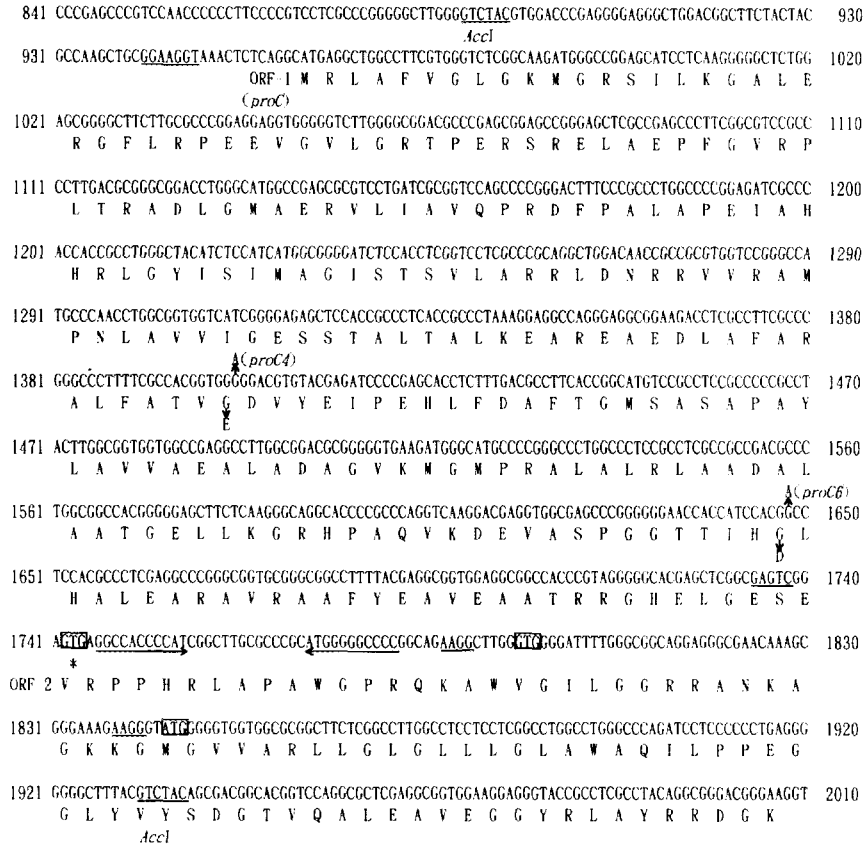

Fig. 2. Nucleotide sequence of the *proC* gene of *T. thermophilus*. Only the nucleotide sequence covering the *proC* gene is shown in this figure. As mentioned in the text, the nucleotide sequence of the 2,678 bp region was determined. The complete nucleotide sequence will appear in the GSDB, DDBJ, EMBL, NCBI nucleotide sequence databases under the accession number D25413. The nucleotide sequence is numbered from the first base of the 2,678 bp fragment. Amino acid sequences are shown under the DNA sequence as one-letter symbols. Arrows indicate inverted repeat. Possible translational initiation codons for ORF-2 are boxed. Putative ribosome binding sites are underlined. The mutation sites of *proC4* and *proC6* and corresponding amino acid substitutions are also shown.

encoded the *proC* gene of *T. thermophilus*. The nucleotide sequence of the *proC* coding region (nucleotide number (nt) 841 to 2010) is shown in Fig. 2 together with the deduced amino acid sequence. Preceding the first ATG codon at nt 960, there was a putative ribosome binding site (nt 941-947). There were other possible translational initiation codons, GTG at nt 975 and ATG at nt 990. Based on the results of homology among ProC proteins (see next section), we took the ATG at nt 960 as the initiation codon for the ORF-1. The molecular mass of the putative ORF-1 protein was estimated to be 27,819 Da, a value similar to the molecular mass of *E. coli* ProC protein (28,145 Da).

Comparison of the deduced amino acid sequence of *Thermus proC* gene with those of other *proC* genes

We compared the putative amino acid sequence of the ORF-1 with those of other *proC* genes in SWISS-PROT data base. As shown in Fig. 3, the ORF-1-encoded protein showed a strong homology to the ProC proteins of *E. coli* (15), soybean (17) and human (18). As the sequence

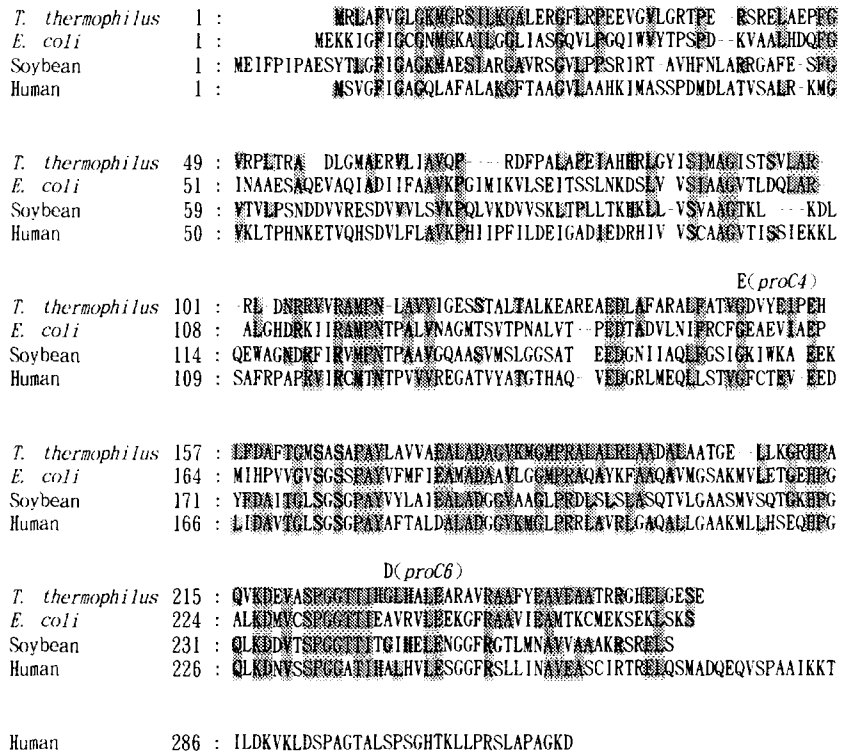

Fig. 3. Amino acid sequence alignment of P5CRs. Identical amino acids are shaded. Amino acid substitutions of the mutants are also shown.

identity with *E. coli* protein was 32.8 %, it was suggested that ORF-1 encoded the *Thermus proC* gene.

Attempts were made to detect homologies between other ORF proteins and the proteins in the data base. Neither homology or similarity was found.

Determination of the mutation sites of *proC4* and *proC6* mutants

The 1.0 kb *Acc*I fragments were cloned from the *proC4* and *proC6* mutants, respectively, and their nucleotide sequences were determined. The results are shown in Fig. 2. As both mutation sites were located in the ORF-1, it was assumed that the ORF-1 encoded the *Thermus proC* gene. Mutations consisted of single base (G to A) conversions at nt 1,402 for *proC4* and at nt 1,648 for *proC6*, respectively, resulting in the amino acid substitution of Glu for Gly_{148} and Asp for Gly_{230}. The mutation sites are also shown in Fig. 3.

Identification of the putative *Thermus proC* gene product produced in *E. coli*

The 1.0 kb *Acc*I fragment in pUC-pro4,6^{+} was transferred into an expression vector pPROK-C. Heat treated cell free extracts were prepared from *E. coli* JM109, JM109 carrying pPROK-C and JM109 carrying pPROK-*proC*. Fig. 4 shows the SDS polyacrylamide gel electrophoresis analysis of the extracts. A unique band whose molecular mass was about 29,000 Da, which is similar to the molecular mass calculated from the deduced amino acid sequence, was observed in the extract of JM109 carrying pPROK-*proC*. The P5CR activities of the extracts were also examined at 70 °C. Although the activity was low, it was clearly detected in the extract of the cells carrying pPROK-*proC*.

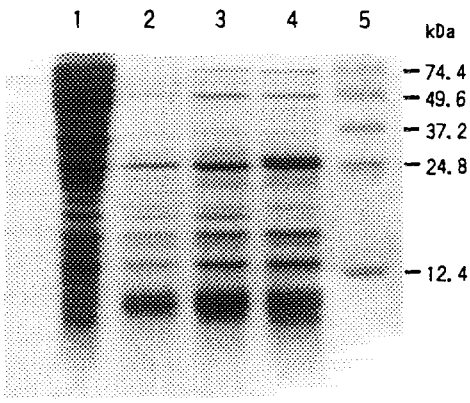

Fig. 4. SDS-polyacrylamide gel electrophoresis of the crude extracts. Crude extracts were heat-treated (except for lane1) and soluble materials were analyzed by SDS/15% polyacrylamide gel electrophoresis with 4.5 % stacking gel at 150 V for 3.5 hr. Lane 1, JM109 (pPROK-*proC* ,not heat-treated); lane2,JM109;lane3,JM109(pPROK-C);lane4, JM109(pPROK-*proC*);lane5, molecular weight markers.

DISCUSSION

We cloned the 2.7 kb chromosomal region of *T. thermophilus* HB27 and determined its nucleotide sequence. There were at least four ORFs with a *Thermus*-type biased codon usage. One of them, ORF-1, was found to encode the *proC* gene. We have recently constructed a physical map of the *T. thermophilus* HB27 chromosome, and the location of the *proC* gene has already been determined (19). The genome size was 1.82 Mb. Since the value was smaller than that of the chromosomes of other eubacteria, we speculated that genes may be located close to one another on the *Thermus* genome. In this case, ORF-2, ORF-3 and ORF-4 could code for other proteins, although we could not find any homologous proteins in the data base.

On the other hand, the ORF-1 protein showed a significant homology to the P5CRs of *E. coli*, soybean and human. Subcloning experiments and results of cloning and sequencing of the mutant chromosomal DNAs suggested that ORF-1 encoded the *proC* gene of *T. thermophilus*. The *proC* structural gene consists of 783 bases (261 amino acids) with a molecular mass of 27,819 Da. As shown in Fig. 3, there are several conserved regions among the ProC proteins of the four organisms. Those regions are likely to be essential for P5CR activities. It is noteworthy that both the *proC4* and *proC6* mutations occurred in conserved regions.

The codon usage of *Thermus proC* is summarized in Table 1 in comparison with that of *E. coli proC*. As already reported for other *Thermus* genes (16), the codon usage of *Thermus proC* is highly biased. The percentage of use of G or C in the third letter is 95.4 %. The length of the ProC proteins is 261 amino acids for *T. thermophilus* and 269 amino acids for *E. coli*. As the lengths are almost the same, it would be important to compare the amino acid composition between the two proteins. There is no Cys residue in *Thermus* P5CR, whereas *E. coli* protein has four Cys residues. Another major difference in the amino acid composition between the two organisms was that *Thermus* preferred Arg (26 Arg is used whereas *E. coli* uses only 7 Arg) to Lys (*Thermus:E. coli*, 6:16) for positively charged amino acids (Table 1).

In *E. coli*, the *proC* gene is mapped in the position of 9 min. Other genes responsible for proline biosynthesis, *proA* and *proB* occur as an operon in the map position of 6 min (20). We

Table 1. Codon usage in the *proC* gene of *T. thermophilus* HB27

First letter	Second letter				Third letter
	U	C	A	G	
U	Phe 3(6)	Ser 0(3)	Tyr 0(1)	Cys 0(1)	U
	Phe 6(3)	Ser 5(5)	Tyr 4(2)	Cys 0(3)	C
	Leu 0(0)	Ser 0(2)	Term0(0)	Term1(1)	A
	Leu 5(0)	Ser 2(2)	Term0(0)	Trp 0(1)	G
C	Leu 2(3)	Pro 0(1)	His 0(3)	Arg 1(1)	U
	Leu13(3)	Pro10(2)	His 7(1)	Arg 8(5)	C
	Leu 1(0)	Pro 0(5)	Gln 0(3)	Arg 0(0)	A
	Leu10(12)	Pro 3(6)	Gln 2(5)	Arg12(1)	G
A	Ile 0(10)	Thr 0(1)	Asn 0(4)	Ser 0(0)	U
	Ile 9(11)	Thr 7(7)	Asn 2(3)	Ser 4(5)	C
	Ile 0(0)	Thr 0(0)	Lys 0(14)	Arg 0(0)	A
	Met 8(14)	Thr 4(4)	Lys 6(2)	Arg 5(0)	G
G	Val 0(3)	Ala 1(6)	Asp 0(4)	Gly 1(7)	U
	Val 8(5)	Ala32(16)	Asp 9(6)	Gly11(10)	C
	Val 0(8)	Ala 0(6)	Glu 1(15)	Gly 1(3)	A
	Val12(9)	Ala14(10)	Glu22(2)	Gly10(4)	G

The numbers in parentheses represent codon usage in the *proC* gene of *E. coli* (15).

have already cloned a *Thermus* DNA fragment that complemented both the pro3 and pro5 mutants. The fragment was mapped on the opposite site of the *proC* on the genome map (19). Sequencing of the cloned fragment is now in progress.

ACKNOWLEDGMENT

This work was partly supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

REFERENCES

1. Oshima, T. and Imahori, K. (1974) Int. J. Syst. Bacteriol. 24, 102-112.
2. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Science 239, 487-491.
3. Leisinger, T. (1987) Biosynthesis of proline. In Neidhardt, F.C. (ed.) *Escherichia coli* and*Salmonella typhimurium*: Cellular and molecular biology. American Society for Microbiology,Washington , D.C. pp. 345-351.
4. Yeh, G.C. and Phang, J.M. (1988) J. Biol. Chem. 263, 13083-13089.
5. Kohl, D.H., Schubert, K.R., Carter, M.B., Hagedorn, C.H. and Sheaver, G. (1988) Proc. Natl. Acad. Sci. USA 85, 2036-2040.
6. Sakaki, Y. and Oshima, T. (1975) J. Virol. 15, 1449-1453.

7. Koyama, Y., Hoshino, T., Tomizuka, N. and Furukawa, K. (1986) J. Bacteriol. 166, 338-340.
8. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119.
9. Saito, H. and Miura, K. (1963) Biochim. Biophys. Acta 72, 619-629.
10. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
11. Miller, J. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
12. Rossi, J.J., Vender, J., Berg, C.M. and Coleman, W.H. (1977) J. Bacteriol. 129, 108-114.
13. Hayzer, D.J. and Leisinger, T.H. (1980) J. Gen. Microbiol. 118, 287-293.
14. Laemmli, U.K. (1970) Nature 227, 680-685.
15. Deutch, A.H., Smith, C.J., Rushlow, K.E. and Kretschmer, P.J. (1982) Nucleic Acids Res. 10, 7701-7714.
16. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. and Oshima, T. (1984) J. Biol. Chem. 259, 2956-2960.
17. Delauney, A.J. and Verma, D.P. (1990) Mol. Gen. Genet. 221, 299-305.
18. Dougherty, K.M., Brandriss, M.C. and Valle, D. (1992) J. Biol. Chem. 267, 871-875.
19. Tabata, K., Kosuge, T., Nakahara, T. and Hoshino, T. (1993) FEBS Lett. 331, 81-85.
20. Bachmann, B.J. (1990) Microbiol. Rev. 54, 130-197.