J. Microbiol. Biotechnol. (2004), **14**(6), 1280–1285

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Overexpression and Purification of Reverse Transcriptase of Retron EC83 by Changing the Downstream Sequence of the Initiation Codon

JEONG, DAE-WON[1] AND DONGBIN LIM[2]*

[1]*BK21 HLS, Seoul National University, Seoul 110-749, Korea*
[2]*Department of Bioinformatics and Life Science, College of Natural Sciences, Soongsil University, Seoul 156-743, Korea*

Received: May 14, 2004
Accepted: July 12, 2004

Abstract Retron is a prokaryotic genetic element, producing a short single-stranded DNA covalently linked to RNA (msDNA-RNA) by a reverse transcriptase (RT). In retron EC83, msDNA is further processed at between the 4th and the 5th nucleotides, leaving a 79 nucleotide-long single-stranded DNA as a final product. To investigate this site-specific cleavage in msDNA synthesis, we purified the RT protein of retron EC83. Initially, RT ORF was cloned under the tac promoter, but the expression was very poor largely because of poor translation. In order to facilitate translation, the nucleotide sequence for the first nine amino acids was randomized with synonymous codons. This change of downstream sequence of translational initiation codon greatly affected the efficiency of translation. We could isolate clones which greatly increased RT production, and their sequences were compared to those of the low producers. The overproduced protein was purified and was shown to have RT activity.

Key words: Retron, reverse transcriptase, translational control, overexpression

Genetic elements whose life cycles depend on reverse transcription are called retroelements. They include retroviruses, retrotransposons, some fungal plasmids, and group II introns. These elements are able to propagate by reverse transcription, and the enzyme reverse transcriptase (RT) is essential for their life cycle.

Retron is a generic name for a prokaryotic genetic element with a reverse transcriptase [21]. It produces a small single-stranded DNA (msDNA) covalently linked to RNA (msdRNA) [3]. Retrons are composed of genes for msDNA (*msd*), msdRNA (*msr*), and RT (*ret*). Genes are located on the chromosome in the order of *msr*, *msd*, and *ret*, and form a single operon [13]. Some retrons, in addition to the three genes required for msDNA synthesis, have extra open reading frames, but their functions are largely unknown [10].

MsDNA is synthesized through reverse transcription. A primary transcript covering *msr*, *msd*, and *ret* serves as both a primer and a template for reverse transcription [13]. Since the 2'-OH group of a special guanine residue located at the end of a pair of inverted repeats of the template RNA serves as a primer for RT, msDNA is covalently linked to the template RNA via a 2'-5' phosphodiester bond. DNA synthesis is terminated by an unknown mechanism approximately at the middle of the template, leaving a part of the template RNA as an msdRNA [14]. Recent boom of genome sequencing reveals that retron-type RT is present in a variety of prokaryotes, including an Archea *Methanococcoides burtonii*, but it is largely unknown if these RTs produce msDNA.

Several years ago, we found that retron EC83, a retron isolated from a clinical strain of *E. coli* [12], produced an msDNA which lacked an associated RNA. This is surprising not only because all previously known retrons produce msDNA covalently linked to RNA, but because, like any other retrons, the primary product of reverse transcription was msDNA linked to RNA [11]. We showed that the msDNA without an associated RNA is produced by site-specific cleavage of msDNA [5]. Since RT is the only protein encoded by retron EC83, it is suggested that either the RT is an endonuclease or that msDNA is cleaved by a host enzyme [11]. It is necessary to purify the RT of retron EC83 in order to investigate if the RT has such a site-specific endonuclease activity. In this report, we showed that the downstream sequence of RT initiation codon greatly influenced the translation of RT protein. The overexpressed RT protein was purified, and the RT activity was verified *in vitro*.

*Corresponding author
Phone: 82-2-820-0452; Fax: 82-2-828-7190;
E-mail: dblim@ssu.ac.kr

MATERIALS AND METHODS

Plasmid Construction

RT ORF of retron EC83 was amplified by PCR, using primers with *Nco*I or *Bam*HI site at the ends, and the amplified fragment was cloned into the *Nco*I and *Bam*HI sites of pTrc99AHind, which was made by removal of a *Hin*dIII site of pTrc99A (Pharmacia) through *Hin*dIII digestion, filled-in by Klenow and self-ligation. The resulting plasmid was called pTrc99A-161RT. To replace the DNA fragment encoding the first 11 amino acids of RT, pTrc99A-161RT was digested with *Nco*I and *Hin*dIII, and the large fragment was gel-purified. This fragment was ligated with two synthetic oligonucleotides with sequences of 5'-ATGTC-NATHGAYATHGARCANCANCTNCAAAAA-3' and 5'-AGCTTTTTTGNAGNGTNGTYTCDATRTCDATNGA-3' (N=A, G, C, T; Y=C, T; R=G, A; H=C, T). Ligated mixture was transformed into *E. coli* CC107 [15], and each transformant was tested for the production of RT by SDS-PAGE after induction of tac promoter.

Antibody Preparation

Anti-RT antibody was raised in two rabbits against fusion protein of maltose binding protein (MBP) and intact RT. Purification of MBP-RT fusion protein was described previously [4]. Two female rabbits were injected subcutaneously at multiple sites with a total of 0.5 mg of MBP-RT which had been emulsified 1:1 (vol/vol) in Freund's complete adjuvant. At 4 and 8 weeks, the rabbits were boosted with 0.5 mg of the same fusion protein preparation mixed at 1:1 in Freund's incomplete adjuvant. Antiserum was collected and applied to a protein A-Sepharose column (Pharmacia-LKB) to purify the immunoglobulin G fraction.

SDS-PAGE and Western Blot

Proteins were separated by SDS-PAGE (3% stacking gel, pH 6.8, 10% separating gel, pH 8.8) in a minigel apparatus and stained with Coomassie Brilliant Blue R250. For the immunological detection, proteins were blotted onto nitrocellulose membrane. Then, the membrane was incubated with blocking solution containing 1% (w/v) nonfat milk for 1 h, washed and probed with anti-MBP-RT rabbit polyclonal antibody diluted to 1:5,000. The band was detected with an anti-rabbit IgG conjugated with alkaline phosphatase and developed by using the Bio-Rad AP color reagent kit.

Purification of RT

E. coli CC107 (pTrc99A-HD-RT), one of the RT overproducer constructed by oligo nucleotide replacement (see Plasmid construction section), was cultivated in LB medium containing ampicillin (50 μg/ml), at 28°C until OD_{600} reached 0.6 [16]. IPTG (a final concentration of 0.3 mM) was added and further incubated for 2 h. Cells were harvested, washed with wash buffer (20 mM Tris, pH 7.5, 0.1 M NaCl) and resuspended in extraction buffer (20 mM Tris·HCl, pH 7.5, 0.7 M NaCl, 10 mM $MgCl_2$, 7 mM β-mercaptoethanol). Cells were disrupted by sonication and centrifuged at 15,000 rpm for 20 min at 4°C. Supernatant was collected and applied to a Phenyl-Toyopearl column (Tosoh, Japan). Proteins were eluted by extraction buffer, containing a linear gradient of NaCl from 1 M to 0 M. Fractions (from 0.2 M to 0 M NaCl), containing RT protein, were dialyzed against ssDNA column buffer (20 mM Tris·HCl, pH 7.5, 50 mM NaCl, 10 mM $MgCl_2$, 7 mM β-mercaptoethanol) and applied to a single-stranded DNA affinity column (Sigma-Aldrich). After washing, proteins were eluted with ssDNA column buffer, containing a linear gradient of NaCl from 0.1 to 1 M. RT was eluted at around 0.8 M NaCl. Fractions, containing RT protein, were dialyzed against ssDNA column buffer, mixed with an equal volume of glycerol, and kept at - 20°C until use.

RT Assay

In order to use as a template and primer, msDNA-RNA compound of *E. coli* B was isolated by modified alkaline lysis method [13] and diluted to OD_{260}=10 [13]. Reverse transcription was started by addition of 200 ng of RT protein to 40 μl of a solution, containing 50 mM Tris·HCl, pH 7.8, 10 mM $MgCl_2$, 60 mM NaCl, 10 mM DTT, 2 μl of template-primer, 120 μM each of dGTP, dTTP, dCTP, and 250 μCi of [α-^{35}S]dATP. After incubation at 37°C for 1 h, cold dATP (120 μM) was added and incubated for an additional 1 h. The reaction mixture was divided into two fractions, and RNAase A (final 50 μg/ml) was added to one of them. After RNA digestion for 20 min, both fractions were extracted with phenol/chloroform and precipitated with ethanol. Nucleic acid was dissolved in 50 μl of stop solution (95% formamide containing XC and BPB) and analyzed on 6% sequencing gel [18].

RESULTS AND DISCUSSION

Inefficient Translation of RT

Initial attempts to purify RT of retron EC83 by employing popular T7 promoter system were unsuccessful because of low expression level. This prompted us to consider other expression vectors. Plasmids with *tac* promoter are more flexible than T7 promoter system in the following aspects. First, expression vectors, based on *tac* promoter, usually contain *lac*I, the lac repressor gene, so that most *E. coli* strains can be used as a host. In contrast, for plasmids with T7 promoter, only *E. coli* strains, carrying T7 RNA polymerase gene such as BL21, should be used as a host [9]. Secondly, controlling the level of expression is simpler with tac promoter than with T7 promoter system. With these considerations, we cloned the entire RT ORF into the

expression plasmid pTrc99A in such a way that the transcription and translation are controlled by tac promoter and Shine-Dalgarno sequence of lacZ (Materials and Methods). The resulting plasmid pTrc99A-161RT, like previous T7 promoter based constructs, did not show the RT protein band on SDS-PAGE of total protein. Western blot analysis revealed that, even after induction, it produced a very little amount of RT protein (see below). Taken together, it is very likely that the low production of RT is due to the inefficient translation of RT mRNA [17].

It is generally accepted that translation of mRNA in *E. coli* is affected by the composition of codons. Genes with rare codons, such as AGG (Arg), AGA (Arg), AUA (Ile), CUA (Leu), GGA (Gly), and CCC (Pro), are known to be less efficiently translated [6, 7, 19]. We examined the codons of RT ORF, but it was not particularly rich in such codons, suggesting that there might be different reasons behind the inefficient translation observed.

Efficient Translation by Altering Sequences Downstream of the Initiation Codon

Several studies showed that the nucleotide sequence downstream of the initiation codon, sometimes called downstream box, strongly influences translation in *E. coli* [1, 2, 6, 8, 20]. We tested this possibility by altering the nucleotide sequence, corresponding to the N-terminal 9 amino acids, into various sequences which encode the same amino acids. This was done by taking advantage of the *Hin*dIII site at the nucleotide sequence of 10^{th} and 11^{th} codons. We replaced the DNA fragment from the initiation codon to the *Hin*dIII site with synthetic oligonucleotides, representing 4096 possibly different sequences (Fig. 1). These pools were transformed into *E. coli* CC107, and the

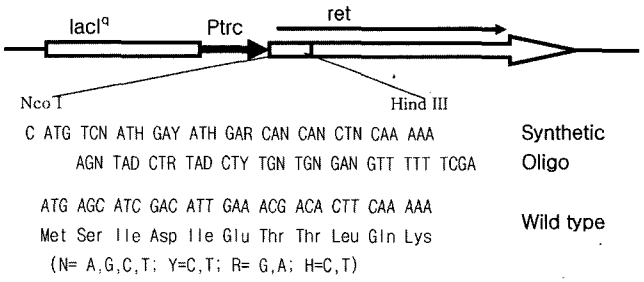

Fig. 1. Randomization of nucleotide sequence, corresponding to the first 10 amino acids.
The *Nco*I-*Hin*dIII fragment of plasmid pTrc99A-161RT was replaced with degenerate synthetic oligonucleotides, having synonymous codons. The drawing shows the structure of control region of pTrc99A. Ptrc and ret indicate tac promoter of plasmid pTrc99A and the gene for RT of retron EC83, respectively.

production of RT protein by each colony was examined by SDS-PAGE. Individual clones produced various amounts of RT protein ranging from a level barely detectable by Western blot to very dark band after Coomassie staining. After testing 400 different colonies, we collected clones which produced very high or very low levels of RT. Figure 2 shows the result of SDS-PAGE of fourteen such clones.

The RT band from each high producer became very dark after induction. From the low producers, however, the RT was detected only by Western blot (Fig. 2). In all higher producers, RT production was greatly increased by the induction of tac promoter. However, in most low producers, the level of RT was not so much increased by induction, as revealed by Western blot. This showed that, in low producers, the increase of mRNA was not followed by corresponding increase of translated protein, suggesting that inefficient translation might be the major hurdle of RT

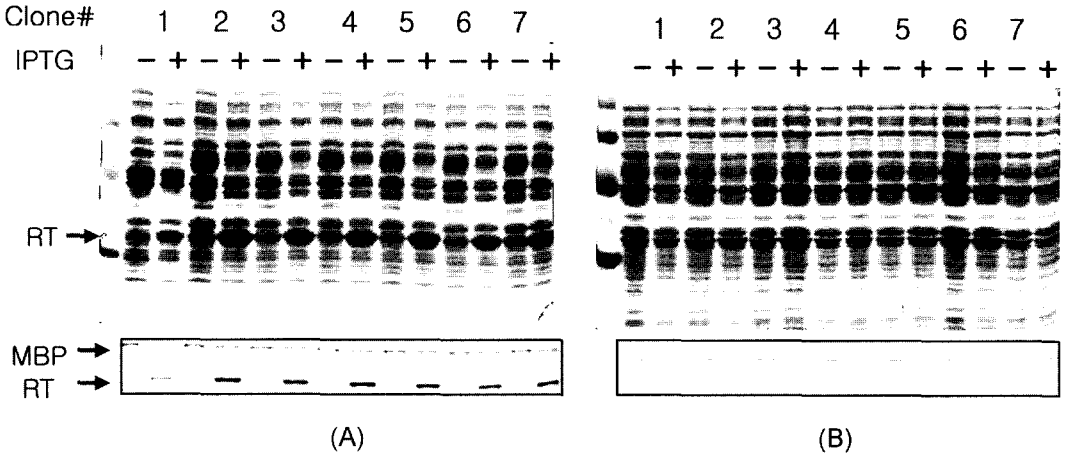

Fig. 2. SDS-PAGE and Western blot of high and low producers of RT.
After screening of 400 clones, 12 clones producing high (A) and low (B) levels of RT were chosen. Proteins produced by these clones were analyzed before (-) and after (+) the addition of IPTG. The same clones were used for Western blot, and the results are shown under the gel. The positions of RT and MBP are indicated on the left. Both RT and MBP were recognized, because the antibody was raised against MBP-RT fusion protein. Note that RT in high producers was greatly increased by the addition of IPTG but, in low producers, the same induction increased RT production by less than two times.

```
5' -ATG AGC ATC GAC ATT GAA ACG ACA CTT-3'   - WT
5' -ATG TCN ATH GAY ATH GAR CAN CAN CTN-3    - Oligo

5' -ATG TCC ATT GAT ATT GAA ACA ACG CTT-3'
5' -ATG TCT ATC GAT ATC GAA ACG ACC CTC-3'
5' -ATG TCA ATC GAT ATT GAG ACT ACA CTC-3'   - High
5' -ATG TCG ATT GAC ATC GAG ACC ACA CTG-3'
5' -ATG TCT ATT GAT ATC GAA ACT ACT CTC-3'

5' -ATG TCA ATA GAC ATA GAA ACG ACA CTC-3'
5' -ATG TCA ATA GAT ATA GAA ACC ACG CTA-3'
5' -ATG TCG ATA GAC ATA GAG ACG ACT CTA-3'   - Low
5' -ATG TCG ATT GAT ATA GAA ACC ACG CTC-3'
5' -ATG TCC ATC GAC ATA GAG ACA ACC CTG-3'
```

Fig. 3. Nucleotide sequence of high and low RT producers.
Nucleotide sequences of 5 high producers (plasmids corresponding to the lanes 2, 3, 4, 5, and 6 of Fig. 2A) and 5 low producers (plasmids to the lanes 2, 3, 4, 5, and 6 of Fig. 2B) are shown. For comparison, the nucleotide sequences of wild-type and oligos are also shown.

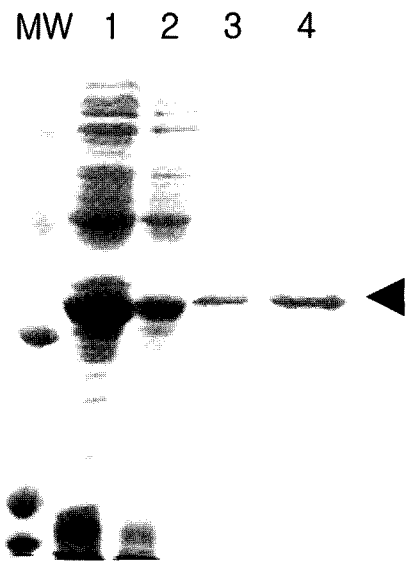

Fig. 4. Purification of RT protein.
Extracts were prepared from *E. coli* CC107, harboring RT overproducing plasmid, by sonication and centrifugation, and RT protein was purified by column chromatography. Protein in each purification step was analyzed by 10% SDS-PAGE. Lane 1, total protein; lane 2, cytosolic extract after centrifugation; lane 3, RT fraction after Phenyl Toyopearl; lane 4, RT protein after single-stranded DNA affinity column chromatography. Arrowhead indicates the position of RT protein. Molecular weight markers indicate 200, 97, 68, 43, 29, and 18 kDa.

production in these cases. This supported our assumption that RT mRNA from wild-type was poorly translated.

Sequence Analysis of High and Low Producers

Each of five high and low producers was sequenced, and the results are shown in Fig. 3. It is well established that proteins with rare codons were not efficiently translated [7, 19]. In particular, six codons [AGG (Arg), AGA (Arg), AUA (Ile), CUA (Leu), GGA (Gly), and CCC (Pro)] are known to affect translation in *E. coli*. Two isoleucines and one leucine were found in the first 9 amino acids where the codons were altered. Interestingly, the fifth codon for Ile changed from AUU to AUA in all five low producers (Fig. 3). Moreover, the third AUC codon for Ile altered to AUA in three out of five clones. These observations suggest that the AUA exacerbated the already low production of wild-type RT. In five high producers, the AUA codon was not found. However, it is not clear how such a big difference of translation could be achieved simply by changing a few downstream codons. We speculate that downstream of the start codon, as a part of translational control element, might be operating here [1, 2, 8, 20]. Whatever the exact mechanism, our purpose to obtain RT-overproducing clones was fulfilled by randomizing the downstream sequence and examining proteins produced by each clone.

Purification and Activity of RT Protein

When *E. coli* CC107 (pTrc99A-HD-RT) was cultivated at 28°C, the majority of the overproduced RT was soluble and easily purified by column chromatography. Among several column materials, hydrophobic resin was found to be very efficient. RT protein bound tightly to Phenyl-Toyopearl and was eluted by less than 0.2 M NaCl concentration. Since retron RT is known to bind msDNA [4], single-stranded DNA column was used. The RT purified by using these two columns showed a single band in SDS-PAGE. The protein profiles from purification steps are shown in Fig. 4.

The purified protein was tested for its reverse transcriptase activity. It is shown that msDNA-RNA compound is a very good template and primer for RT [13]. Therefore, we used msDNA-RNA compound, produced by retron EC86 of *E. coli* B, as a substrate for the RT activity test (Fig. 5). In this assay, msDNA and the associated msdRNA serve simultaneously as a primer and a template, respectively, and RT extends msDNA up to the branched point [13]. As shown in lane 1 of Fig. 5, reaction products were larger than msDNA itself (indicated by a in Fig. 5), and RNAase A treatment reduced the size, showing that reaction products were covalently linked to RNA. Most of reaction products after removal of RNA were shorter than fully extended products (144 nucleotide-long DNA indicated by b in Fig. 5). This suggests that the extension of msDNA by retron83 RT was not complete. Also, the covalent association of extended product with msdRNA showed that retron83 RT did not cleave msDNA EC86. It is highly likely that retron EC83 RT cleaved the reverse transcription product in its own msDNA synthesis.

Unlike other msDNAs, msDNA produced by retron EC83 is cleaved off from the associated RNA. This endonucleolytic cleavage occurs between the 4th and the 5th

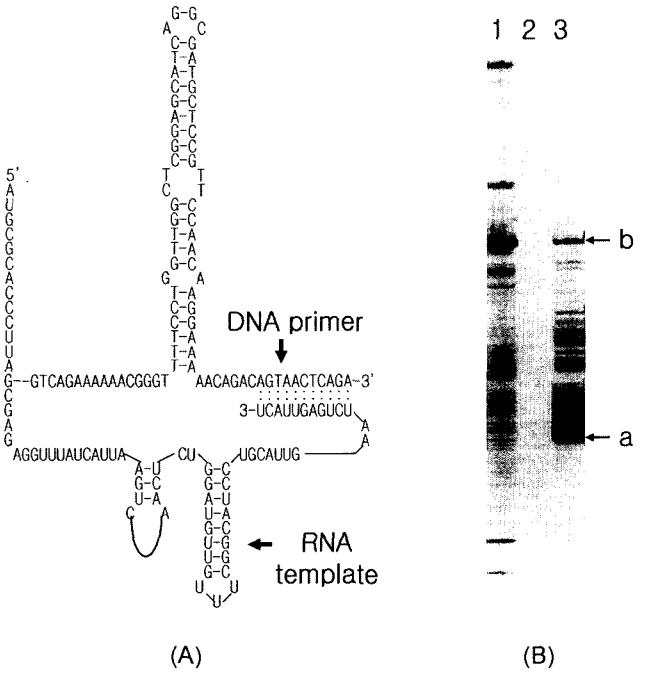

Fig. 5. Reverse transcriptase activity of the purified protein.
(A) Structure of the *E. coli* B msDNA-RNA compound used as a substrate to test the RT activity of the purified protein. It was shown that RT extends msDNA (indicated as DNA primer) up to the branched point of msdRNA (indicated as RNA template) using the 3' end of msDNA [13]. (B) Reaction products analyzed on a 6% sequencing polyacrylamide gel. Lane 1, reaction products before the digestion of template RNA by RNAase A. Lane 2, reaction products without the addition of RT. Lane 3, reaction products after the digestion of template RNA by RNAase A. Arrow with a indicates the position corresponding to the intact msDNA (84 nucleotides long). Arrow with b indicates the position corresponding to the msDNA fully extended to the branched guanine (144 nucleotides long).

nucleotide of msDNA [5, 11]. Since no other proteins of retron were required for this site-specific cleavage in msDNA synthesis, RT is most likely responsible for this cleavage. Since RT protein is purified, we are now able to test the suggested endonucleolytic activity of retron EC83 RT in future.

Acknowledgment

This work was supported by the Soongsil University Research Fund.

References

1. Etchegaray, J. P. and M. Inouye. 1999. Translational enhancement by an element downstream of the initiation codon in *Escherichia coli*. *J. Biol. Chem.* **274:** 10079-10085.
2. Etchegaray, J. P. and M. Inouye. 1999. A sequence downstream of the initiation codon is essential for cold shock induction of cspB of *Escherichia coli*. *J. Bacteriol.* **181:** 5852-5854.
3. Furuichi, T., S. Inouye, and M. Inouye. 1987. Biosynthesis and structure of stable branched RNA covalently linked to the 5' end of multicopy single-stranded DNA of *Stigmatella aurantiaca*. *Cell* **48:** 55-62.
4. Jeong, D. W., K. Kim, and D. Lim. 1997. Evidence for the complex formation between reverse transcriptase and multicopy single-stranded DNA in retron EC83. *Mol. Cells* **7:** 347-351.
5. Kim, K., D. W. Jeong, and D. Lim. 1997. A mutational study of the site-specific cleavage of EC83, a multicopy single-stranded DNA (msDNA): Nucleotides at the msDNA stem are important for its cleavage. *J. Bacteriol.* **179:** 6518-6521.
6. Kim, M. D., K.-N. Kim, and Y. J. Choi. 2003. Enhanced expression of β-xylosidase of *Bacillus stearothermophilus* No. 236 by change of translational inhibition codon in *Escherichia coli* and *Bacillus subtilis*. *J. Microbiol. Biotechnol.* **13:** 584-590.
7. Kim, J. S., H. Y. Seong, M. W. Kim, J. S. Ku, and S. Y. Choi. 2003. Effects of minor arginyl tRNA and isoleucyl tRNA on the expression of *Clostridium botulinum* neurotoxin light chain in *Escherichia coli*. *J. Microbiol. Biotechnol.* **13:** 287-291.
8. Kuroda, H. and P. Maliga. 2001. Sequences downstream of the translation initiation codon are important determinants of translation efficiency in chloroplasts. *Plant Physiol.* **125:** 430-436.
9. Lee, J.-H., Y.-D. Cho, J. J. Choi, Y.-J. Lee, H.-S. Hoe, H.-K. Kim, and S.-T. Kwon. 2003. High-level expression in *Escherichia coli* of alkaline phosphatase from *Thermus caldophilus* GK24 and purification of the recombinant enzyme. *J. Microbiol. Biotechnol.* **13:** 660-665.
10. Lim, D. 1991. Structure of two retrons of *Escherichia coli* and their common chromosomal insertion site. *Mol. Microbiol.* **5:** 1863-1872.
11. Lim, D. 1992. Structure and biosynthesis of unbranched multicopy single-stranded DNA by reverse transcriptase in a clinical *Escherichia coli* isolate. *Mol. Microbiol.* **6:** 3531-3542.
12. Lim, D., T. A. Gomes, and W. K. Maas. 1990. Distribution of msDNAs among serotypes of enteropathogenic *Escherichia coli* strains. *Mol. Microbiol.* **4:** 1711-1714.
13. Lim, D. and W. K. Maas. 1989. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* **56:** 891-904.
14. Lim, D. and W. K. Maas. 1989. Reverse transcriptase in bacteria. *Mol. Microbiol.* **3:** 1141-1144.
15. Maas, W. K., C. Wang, T. Lima, G. Zubay, and D. Lim. 1994. Multicopy single-stranded DNAs with mismatched base pairs are mutagenic in *Escherichia coli*. *Mol. Microbiol.* **14:** 437-441.
16. Oh, Y. P., S. T. Jeong, D.-W. Kim, E.-C. Kim, and K.-H. Yoon. 2002. Simple purification of shiga toxin B chain from

recombinant *Escherichia coli*. *J. Microbiol. Biotechnol.* **12:** 986-988.
17. Oh, K.-S., Y.-S. Park, and H.-C. Sung. 2002. Expression and purification of an ACE-inhibitory peptide multimer from synthetic DNA in *Escherichia coli*. *J. Microbiol. Biotechnol.* **12:** 59-64.
18. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. 2nd Ed. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY. U.S.A.
19. Sorensen, M. A., C. G. Kurland, and S. Pedersen. 1989. Codon usage determines translation rate in *Escherichia coli*. *J. Mol. Biol.* **207:** 365-377.
20. Sprengart, M. L., E. Fuchs, and A. G. Porter. 1996. The downstream box: An efficient and independent translation initiation signal in *Escherichia coli*. *EMBO J.* **15:** 665-674.
21. Temin, H. M. 1989. Retrons in bacteria. *Nature* **339:** 254-255.