J. Biochem. 132, 63–70 (2002)

Effective Structure of a Leader Open Reading Frame for Enhancing the Expression of GC-Rich Genes[1]

Masami Ishida*,[2] and Tairo Oshima†

Laboratory of Marine Biochemistry, Tokyo University of Fisheries, 4-5-7 Konan, Minato-ku, Tokyo 108-8477; and †Department of Molecular Biology, Tokyo University of Pharmacy and Life Science, Horinouchi, Hachioji, Tokyo 192-0392

Received January 22, 2002; accepted April 22, 2002

To overexpress broad kinds of GC-rich genes in *Escherichia coli*, we examined how the structures of leader open reading frames (leader ORFs) affect the expression of GC-rich genes, such as *polA*, *trpA*, and *trpB*, from *Thermus thermophilus*. When a leader ORF overlapped with the *polA*-initiation codon by 1 bp in the TGATG motif, gene expression increased by more than 3-fold compared to when a leader ORF was several-bp distant from the initiation codon. A 4-bp overlap with the ATGA motif was more effective than a 1-bp overlap with the TGATG motif. When a 4-bp overlapping leader ORF was placed in front of the successive *trpB* and *trpA* genes, the *trpA* gene was poorly expressed whereas the *trpB* gene was overexpressed. Mutation analysis revealed that the expression of the *trpA* gene was strongly enhanced by replacing G and C in the translation termination region of the leader ORF with A and T. In contrast, other mutations, such as alterations between synonymous codons in the *trpA*-coding region, produced diminished gene expression. Using the most effective leader ORF obtained from these results, new expression vectors were constructed.

Key words: expression vector, GC-rich gene, overexpression, thermostable protein, *Thermus thermophilus*.

Many kinds of important microorganisms in biotechnology have GC-rich genomic DNA. Genome projects are complete or in progress on several such organisms, including *Halobacterium* species, *Pseudomonas* species, *Streptomyces* species, and *Thermus* species. Also, for *T. thermophilus*, structural genomics is now under investigation. However, GC-rich genes from microorganisms are often expressed poorly in *Escherichia coli*. The overexpression of GC-rich genes in *E. coli* is, therefore, an important and urgent research subject.

In order to establish a general way to enhance the expression of GC-rich genes, we have studied the effects of the 5′ upstream region on the expression of *T. thermophilus* genes in *E. coli*. As a result, we found that the introduction of a short ORF, named leader ORF (30 to 40 bp long), enhanced the expression of the *leuB* gene (G + C mol% = 69.9) (*1–3*). Overexpression of the *leuB* gene and the *pfk1* gene (G + C mol% = 71.8) were observed when the translation termination region of a leader ORF overlapped with the initiation region of the gene by 4 bp, ATGA (*1, 2*). In this motif, ATG is the initiation codon for the downstream gene, and TGA is the termination codon for the leader ORF. From the structural similarity to overlapping motifs, such as ATGA and TGATG, found in certain gene pairs from enteric bacteria (*4, 5*), we supposed that the translation efficiency of GC-rich genes in *E. coli* would be enhanced by the efficient translation re-initiation with an overlapping leader ORF (*1, 2, 6*). The introduction of a short ORF was also applied to mammalian cDNA expression in *E. coli* (*7*). To our knowledge, however, no information is available concerning the effect of a short overlap with an artificial upstream ORF on the expression efficiency of genes and cDNAs. To establish a general way to overexpress GC-rich genes, the effect of the overlapping structure with a leader ORF on gene expression should be determined. Additionally, the *T. thermophilus trpBA* genes overlap one another with a similar 4-bp overlapping motif, GTGA (*8*). In our preliminary experiments, however, the expression level of the downstream *trpA* gene is extremely low compared with that of the upstream *trpB* gene. To find a way to improve the expression efficiency of such a poorly expressible gene will be important in establishing the general technique.

Based on this background, the present investigation was conducted to clarify in detail an effective structure for a leader ORF to overexpress GC-rich genes in *E. coli*. We selected three *T. thermophilus* genes, the *trpA* gene (G + C mol% = 70.2), the *trpB* gene (69.3), and the *polA* gene (67.9). The *T. thermophilus polA* gene encodes the thermostable DNA polymerase I, named *Tth* DNA polymerase (*9*).

[1] This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Culture, and Sports of Japan (09558081, 10044095, 11794038), and the Fund for Research Encouragement in Celebration of the Centennial Anniversary of the Founding of Tokyo University of Fisheries. The DNA sequences of new expression vectors, pLEAD4 and pLEAD5, are available from DDBJ under accession numbers AB049968 for pLEAD4 and AB049969 for pLEAD5.

[2] To whom correspondence should be addressed. Tel: +81-3-5463-0586, Fax: +81-3-5463-0589, E-mail: ishida@tokyo-u-fish.ac.jp

Abbreviations: CBB, Coomassie Brilliant Blue; ELISA, enzyme-linked immunosorbent assay; IPTG, isopropyl 1-thio-β-D-galactopyranoside; ORF, open reading frame; RT-PCR, reverse transcription and polymerase chain reaction.

Because of its reverse transcriptase activity, the enzyme can be applied to RT-PCR as a single enzyme (*10*), as well as DNA sequencing and PCR. In this paper, we report (i) the effect on gene expression of two overlapping structures between a leader ORF and the *polA* gene, (ii) the expression profile of the *trpBA* gene pair with a leader ORF, (iii) the effect of the nucleotide sequences around the overlapping structure on the expression of the *trpA* gene, and (iv) new expression vectors constructed on the basis of these results.

MATERIALS AND METHODS

Materials—The *T. thermophilus polA* gene was subcloned from plasmid pLED-NS (*9*), kindly provided by Toyobo (Osaka). The *T. thermophilus trpB* and *trpA* genes were subcloned from plasmid pTWAB2 (*1*). *Escherichia coli* strain JM109 {*recA*1 *supE*44 *endA*1 *hsdR*17 *gyrA*96 *relA*1 *thi* Δ(*lac-proAB*) F′(*traD*36 *proAB*$^+$ *lacI*^{q} *Z*′DM15)} (*11*) was used as the host.

DNA Manipulation—DNA manipulation of plasmids and synthetic oligonucleotides was carried out using standard techniques (*12*). Nucleotide sequences were determined by a DNA sequencer, ABI model 373A (Perkin Elmer). DNA sequences were analyzed with the software program GENETYX (Software Development, Tokyo).

Oligonucleotide-Directed Mutagenesis of the polA Gene—The translation initiation region of the *T. thermophilus polA* gene on pLED-NS was truncated by PCR oligonucleotide-directed mutation. The first primer TPOL-C3 (5′-AAGAATTCCCATGGTTTCGAACCCAAAGGCCGGGTCC-3′) was designed to delete the translation initiation region (24 bp) of the *polA* gene and to introduce *Eco*RI, *Nco*I (for screening the mutant plasmid) and new *Csp*45I sites. The second primer TPOL-C5 (5′-GCCAAGACGGCGAGGTCC-TT-3′) was complementary to the template around the region 80-bp downstream from the unique *Eco*52I site, located 995 bp from the initiation codon of the *polA* gene. PCR was carried out for 30 cycles of 94°C-45s, 59°C-45s, and 72°C-90s in a Zymoreactor II-Model 1820 (Atto, Tokyo). The product was cleaved with *Eco*RI and *Eco*52I, and then inserted between the *Eco*RI and *Eco*52I sites of pLED-NS. The truncated *polA*-coding region was separated from the obtained plasmid by cleavage with *Eco*RI and *Xba*I, and recloned between the *Eco*RI and *Xba*I sites of pUC18. On the resulting plasmid, pTTP-UC, the *Eco*RI site was placed just behind the *lacZ*-initiation region (17 bp), and the *Csp*45I site was just in front of the truncated *polA*-coding region. The ds-cassettes used for mutation were as follows: 10D (+) 5′-AATTCTACTAGGAGGTTATGAGGAGGTTATTATGG-AGGCGATGCTTCCGCTCTT-3′/(–) 5′-CGAAGAGCGGAA-GCATCGCCTCCATAATAACCTCCTCATAACCTCCTAGT-AG-3′, 2D (+) 5′-AATTCTACTTTAAGGAGGTGATTATGG-AGGCGATGCTTCCGCTCTT-3′/(–) 5′-CGAAGAGCGGAA-GCATCGCCTCCATAATCACCTCCTTAAAGTAG-3′, 5M (+) 5′-AATTCTACTAGGAGGTTATGATGGAGGCGATGC-TTCCGCTCTT-3′/(–) 5′-CGAAGAGCGGAAGCATCGCCT-CCATCATAACCTCCTAGTAG-3′, 10DT (+) 5′-AATTCTA-CTAGGAGGTTATGAGGAGGTTATTATGACCGCGATGC-TTCCGCTCTT-3′/(–) 5′-CGAAGAGCGGAAGCATCGCGG-TCATAATAACCTCCTCATAACCTCCTAGTAG-3′, 2DT (+) 5′-AATTCTACTTTAAGGAGGTTATGATTATGACCGCGA-TGCTTCCGCTCTT-3′/(–) 5′-CGAAGAGCGGAAGCATCG-CGGTCATAATCACCTCCTTAAAGTAG-3′, 5MT (+) 5′-AA-TTCTACTAGGAGGTTATGATGACCGCGATGCTTCCGC-TCTT-3′/(–) 5′-CGAAGAGCGGAAGCATCGCGGTCATCA-TAACCTCCTAGTAG-3′, and 4MT (+) 5′-AATTCTAGGAG-GACTTTATGTCCGCGATGCTTCCGCTCTT-3′/(–) 5′-CGA-AGAGCGGAAGCGTCGCGGTCATAAAGTCCTCCTAG-3′. The cassettes were designed to sequence a cohesive end for *Eco*RI, a translation termination region (19 bp) for the upstream *lacZ*′-coding region, a connecting region, the translation initiation region (24 bp) of the *polA* gene (or a modified *polA* gene, in which the second codon GAG was replaced by ACC), and the cohesive end for *Csp*45I. Each 5′-phosphorylated ds-cassette was inserted between the *Eco*RI and *Csp*45I sites of pTTP-UC. The resulting plasmids, pTP-10D, pTP-2D, pTP-5M, pTP-10DT, pTP-2DT, pTP-5MT, and pTP-4MT, were confirmed by DNA sequencing.

Oligonucleotide-Directed Mutagenesis of the trpB and trpA Genes—The *Pst*I fragment, including the *T. thermophilus trpBA* genes on pTWAB2, was re-cloned into the *Pst*I site of pUC19, yielding pTWAB4. To introduce a 4-bp overlap between a leader ORF and the *trpB* gene on pTWAB4, the translation initiation region of the *trpB* gene was modified by PCR oligonucleotide-directed mutation. The first primer TWB1 (5′-AAAGCTTAAGGAGGTCTAGA-TGATTACCCTACCCGACTTTCCCCTG-3′) was designed for the following purposes: to introduce a *Hin*dIII site, a new ribosome binding sequence (AGGAGG), and a *Xba*I site (for screening of the mutant plasmids); to replace the second codon CTG of the *trpB* gene with ATT (Leu-2 to Ile); and to terminate the translation of the upstream *lacZ*′-coding region with the ATGA motif. The second primer TWB-BAL (5′-CACCGTGGCCACGCTCAC-3′) was complementary to the template around the unique *Bal*I site located 383 bp from the initiation codon of the *trpB* gene. PCR was carried out for 30 cycles of 94°C-45s, 58°C-30s, and 72°C-15s. The product was cleaved with *Hin*dIII and *Bal*I, and then inserted between the *Hin*dIII and *Bal*I sites of pTWAB4. The resulting plasmid, named pTWAB5, was confirmed by DNA sequencing. The *Sma*I–*Pst*I fragment containing the *trpA* gene on pTWAB5 was subcloned between the *Sma*I and *Pst*I sites of pUC18, and then the translation initiation region of the *trpA* gene was modified by a method similar to that used for the *trpB* gene. A first primer TWA1 (5′-AAGAATTCGGGAGGGGAGCTGTGAC-CAC-3′) was designed to introduce an *Eco*RI site, and to fuse the upstream *lacZ*′-coding region with the translation termination region of the *trpB* gene. Another first primer TWA2 (5′-AAGAATTCGAGGAGGGAGCTATGACCACCC-TCGAGGCCTTC-3′) was designed for the following purposes: to introduce an *Eco*RI site and a new ribosome binding sequence (AGGAGG), to replace the initiation codon GTG of the *trpA* gene with ATG, and to terminate a leader ORF with the ATGA motif. The second primer TWA-SAC (5′-CGGACGAGCTCCAAAGCGC-3′) was complementary to the template around the unique *Sac*I site located 243 bp from the initiation codon of the *trpA* gene. PCR was carried out for 30 cycles of 94°C-45s, 58°C-30s, and 72°C-10s. The products were cleaved with *Eco*RI and *Sac*I, and then inserted between the *Eco*RI and *Sac*I sites of the parent plasmid. After DNA sequencing, the resulting plasmids were named pTWSA1 and pTWSA2. The translation initiation region of the *trpA* gene on pTWSA2 was further modi-

fied by a similar method with seven other kinds of first primers as follow: WA-N1 (5′-AGAATTCAAGGAGGTCTAGATGACCACCCTCGAGGGC-3′), WA-N2 (5′-AGAATTCAAGGAGGTCTAGATGATCACCCTAGAGGCCTTCGCCAAGGCC-3′), WA-L1 (5′-AGAATTCTAGGAGGGAATTATGACCACCCTCGAGGCC-3′), WA-L2 (5′-AGAATTCTAGGAGTGAATTATGACCACCCTCGAGGCC-3′), WA-C (5′-AGAATTCGAGGAGGGAGCTATGACTACTCTTGAGGCCTTCGCCAAGGCC-3′), WA-CN1 (5′-AGAATTCAAGGAGGTCTAGATGACTACTCTTGAGGCCTTCGCCAAGGCC-3′), and WA-CL2 (5′-AGAATTCTAGGAGTGAATTATGACTACTCTTGAGGCCTTCGCCAAGGCC-3′). Each PCR product was cleaved with *Eco*RI and *Sac*I, and then inserted between the *Eco*RI and *Sac*I sites of pTWSA2. The resulting plasmids, named pTWA-N1, pTWA-N2, pTWA-L1, pTWA-L2, pTWA-C, pTWA-CN1, and pTWA-CL2, were confirmed by DNA sequencing.

Oligonucleotide-Directed Mutagenesis to Construct pLEAD4 and pLEAD5—A leader ORF was introduced between the *tac* promoter and the *polA*-coding region on pLED-NS by oligonucleotide-directed mutation using PCR. A first primer LOR-PO5 (5′-AAGAATTCAGGAGGATTATCATATGACCATGATTACTGATATCAGGAGGATCACGTGATGGAGGCGATGCTTCCGC-3′) was designed to create new *Eco*RV and *Bbr*PI sites, and to introduce a 36-bp leader ORF that overlapped with the *polA*-coding region by 1 bp. The other first primer LOR-PO4 (5′-AAGAATTCAGGAGGATTGTCATATGACCATGATTGATATCAGGAGGATTTACGTATGACCGCGATGCTTCCGCTCTTTGA-3′) was designed to create new *Eco*RV and *Eco*105I sites, to introduce the leader ORF overlapping by 4 bp, and to replace the second codon GAG of the *polA* gene with ACC. TPOL-C5 was used as the second primer. PCR was carried out for 5 cycles of 94°C-45s, coming down to 57°C for 3 min, 57°C-30s, and 72°C-90s, followed by 25 cycles of 94°C-45s, 57°C-30s, and 72°C-90s. Each product was cleaved with *Eco*RI and *Eco*52I, and then inserted between the *Eco*RI and *Eco*52I sites of pLED-NS. The resulting plasmids, pPO536 and pPO45, were confirmed by DNA sequencing. Multiple cloning sites were introduced into pPO536 and pPO45 by replacing each *polA*-coding region with a ds-cassette. The ds-cassette LEAD4 [(+) 5′-GTATGCATTGAGCTCGGTACCTAGGCCTGGATCCA-3′/(–) 5′-AGCTTGGATCCAGGCCTAGGTACCGAGCTCAATGCATAC-3′] was designed to introduce seven cloning sites, *Eco*105I, *Eco*T52I, *Sac*I, *Kpn*I, *Stu*I, *Bam*HI, and *Hin*dIII into pPO45; and the other LEAD5 [(+) 5′-GTGATGCATGAGCTCGGTACCTAGGCCTGGATCCA-3′/(–) 5′-AGCTTGGATCCAGGCCTAGGTACCGAGCTCATGCATCAC-3′] was designed to introduce seven cloning sites, *Bbr*PI, *Eco*T52I, *Sac*I, *Kpn*I, *Stu*I, *Bam*HI, and *Hin*dIII, into pPO536. The 5′-phosphorylated ds-cassettes were inserted between the *Eco*105I and *Hin*dIII sites of pPO45 or between the *Bbr*PI and *Hin*dIII sites of pPO536. The resulting plasmids, pLEAD4 and pLEAD5, were confirmed by DNA sequencing.

Bacterial Growth and Enzyme Extraction—*E. coli* JM109 cells were grown in YT medium (1% Bacto tryptone, 0.5% Bacto yeast extract, 0.5% NaCl) containing 100 mg/liter of ampicillin and 1.0 to 1.25 mM of IPTG for 20 h at 37°C. Cells were collected by centrifugation, washed with saline, and disrupted by ultrasonication in 50 mM potassium phosphate buffer (pH 7.5) on ice. The cell-free extracts were heated at 70°C for 30 min, centrifuged at 12,000 ×*g* for 15 min, and the supernatants were obtained as heat-treated extracts.

Analysis of the Enzymes—DNA polymerase activity was measured in 100 μl aliquots at 37°C by incorporating digoxigenin- and biotin-labeled dUTP into DNA using an ELISA kit (Roche Diagnostics, Cat No. 1 669 885). One unit of activity was defined as 1 pmol of digoxigenin-labeled dUTP incorporated per 30 min. The activity of the tryptophan synthase β subunit, termed β activity, was measured as described by Miles *et al.* (*13*), except that the reaction was performed at 70°C. One unit of β activity was defined as 1 μmol of tryptophan produced per min. To compare the expression levels of the *T. thermophilus trpA* gene, the amount of α subunit was estimated by SDS-PAGE. SDS-PAGE to resolve proteins in the heat-treated extract was performed according to the method of Laemmli (*14*) using 12.5% polyacrylamide gels. Heat-treated extracts corresponding to 1.0 mg aliquots of *E. coli* were loaded onto gels. The gels were then stained with CBB-R 250. The N-terminal amino acid sequences of the proteins after SDS-PAGE were determined according to Matsudaira (*15*) using a gas-phase protein sequencer LF-3400D TriCart (Beckman Coulter, Fullerton, CA).

RESULTS

Effect of a Short Overlap between the Leader ORF and polA Gene on Expression—The effect of an overlap between a leader ORF and the *T. thermophilus polA* gene on expression under the *lac* promoter in *E. coli* JM109 was determined (Fig. 1). The leader ORF was introduced in three ways: distant (10 bp) from the initiation codon (pTP-10D), a short distance from the initiation codon (2 bp, pTP-2D), and overlapping the initiation codon by 1 bp in the TGATG motif (pTP-5M). The DNA polymerase activity detected in heat-treated extracts of *E. coli* JM109 (pTP-5M) increased 3.1-fold compared with *E. coli* JM109 (pTP-10D) extracts. Shortening the distance of 10 bp (pTP-10D) to 2 bp (pTP-2D) resulted in a decrease in enzyme activity to about 40%. To compare the enhancing effect of a 4-bp overlap, *i.e.* the ATGA motif, on gene expression with that of the 1-bp overlap, the second codon GAG (coding for Glu) of the *polA* gene was replaced by ACC (Thr) (Fig. 2). Leader ORFs were introduced in front of the modified *polA* genes for the 10-bp distant (pTP-10DT), the 2-bp distant (pTP-2DT), the 1-bp overlapping of the TGATG motif (pTP-5MT), and the 4-bp overlapping of the ATGA motif (pTP-4MT). No significant difference in gene expression appeared between the wild *polA* gene (Fig. 1) and the modified *polA* genes (Fig. 2), although a slight increase in the activity of the latter was observed. Among the leader ORFs introduced in three fashions, *i.e.* 10-bp distant, 2-bp distant, and with the 1-bp overlap, the relative yields of the modified DNA polymerase (E2T) with pTP-10DT, pTP-2DT, and pTP-5MT (Fig. 2) were quite similar to those of the wild type enzyme with pTP-10D, pTP-2D, and pTP-5M (Fig. 1). Moreover, 10% higher activity was detected in the heat-treated extract of *E. coli* JM109 (pTP-4MT) than the heat-treated extract of *E. coli* JM109 (pTP-5MT) (Fig. 2). These results indicate that both 1-bp and 4-bp overlaps effectively enhance *polA*-gene expression in *E. coli*. Moreover, comparing the 1-bp and 4-bp overlaps, the latter was shown to be more effec-

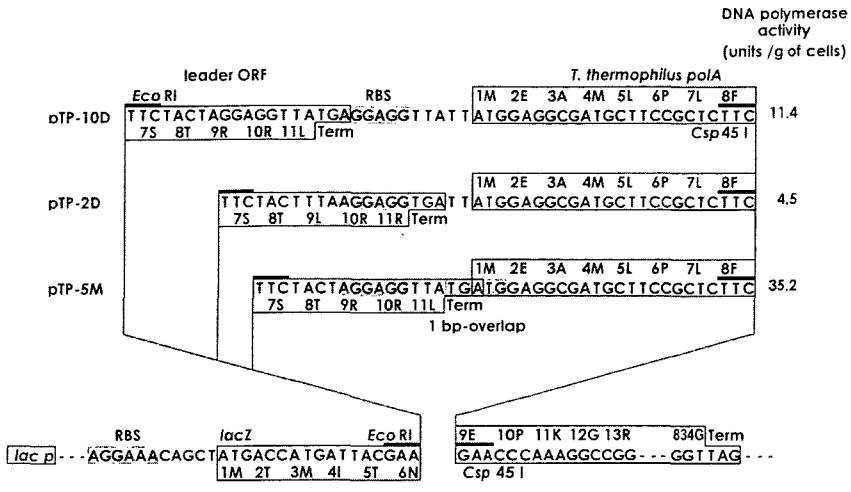

Fig. 1. **The effect of a 1-bp overlap between a leader ORF and the *polA* gene on gene expression under the *lac* promoter in *E. coli* JM109.** In all constructs, the ribosome binding sequence, AGGAGG, for the *polA* gene is placed in the same position (9-bp upstream from the initiation codon), and the leader ORF of 36 bp is designed to comprise a combination of the same codons. Boxed nucleotides, coding regions of the leader ORF and the *polA* gene; shaded nucleotides, a ribosome binding sequence ("RBS") and the TGATG motif; nucleotides with a thick upper line, restriction sites; "*lac p*" in a box, the *lac* promoter; "Term," a termination codon.

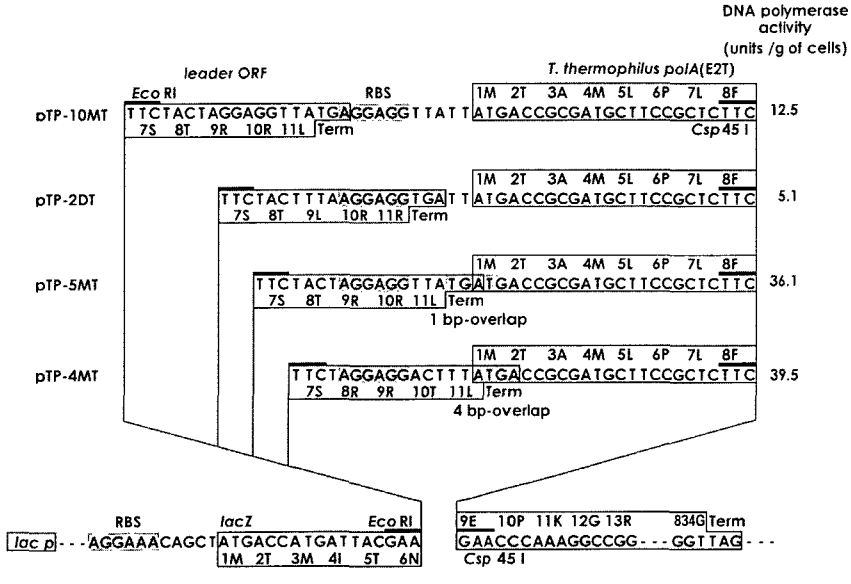

Fig. 2. **The effect of a 4-bp overlap on *polA* gene expression.** To create a 4-bp overlap, the second codon of the *polA* gene was replaced with ACC (E2T). Shaded 4 nucleotides, the ATGA motif. Other structures and representations are the same as in Fig. 1.

tive.

Expression of a trpBA Gene Pair with an Overlapping Leader ORF—Figure 3 illustrates the introduction of a 4-bp overlapping leader ORF to the *trpB* and *trpA* genes of *T. thermophilus*. On pTWAB5, a *lacZ'*-leader ORF overlapped the translation initiation region of the *trpB* gene by 4 bp of the ATGA motif. To create the ATGA motif, the second codon CTG (Leu) of the gene was replaced with ATT (Ile). On pTWB11, most of the *trpA*-coding region of pTWAB5 was removed by deleting the *Sac*I fragment. Reading through the remaining *trpA*-coding region, 246 bp, on pTWB11 terminates with the termination codon of the downstream *'lacZ* gene. As shown in Fig. 4, SDS-PAGE analysis of the heat-treated extracts of *E. coli* JM109 (pTWAB5) and *E. coli* JM109 (pTWB11) indicates that the β subunit (L2I) was markedly overproduced. In contrast, the yield of the α subunit in the same extracts of pTWAB5-harboring cells was extremely low (less than 5%) compared with that of the β subunit. This may be the result of the enhancing effect of the leader ORF upstream of the *trpB* gene not reaching the downstream *trpA* gene. Therefore, a similar leader ORF was overlapped directly with the *trpA*-coding region by 4 bp (pTWSA1 and pTWSA2, Fig. 3). To leave the original overlapping structure between the *trpBA* genes, the *lacZ'*-initiation region was fused to the *trpB*-termination region on pTWSA1. The amount of α subunit produced in heat-treated extracts of *E. coli* JM109 (pTWSA1), however, dropped sharply compared with heat-treated extracts of *E. coli* JM109 (pTWAB5) (Fig. 4). On the other hand, on pTWSA2, a new ribosome binding sequence, AGGAGG, was introduced for the *trpA* gene, and the ATGA motif was created by replacing the initiation GTG codon with ATG. The production level in *E. coli* harboring pTWSA2 was considerably improved compared with *E. coli* harboring pTWSA1; however, the level was lower than that in *E. coli* harboring pTWAB5 (Fig. 4).

Effect of the DNA Sequence around the Overlapping Motif on trpA Gene Expression—From the difference in the expression level of the *trpA* gene between *E. coli* JM109 (pTWSA1) and *E. coli* JM109 (pTWSA2) (Figs. 3 and 4), we presumed that a nucleotide sequence around the overlap motif likely influences the expression efficiency. To confirm this, DNA sequences on both the 5′ and 3′ sides of the ATGA motif on pTWSA2 were varied as shown in Fig. 5.

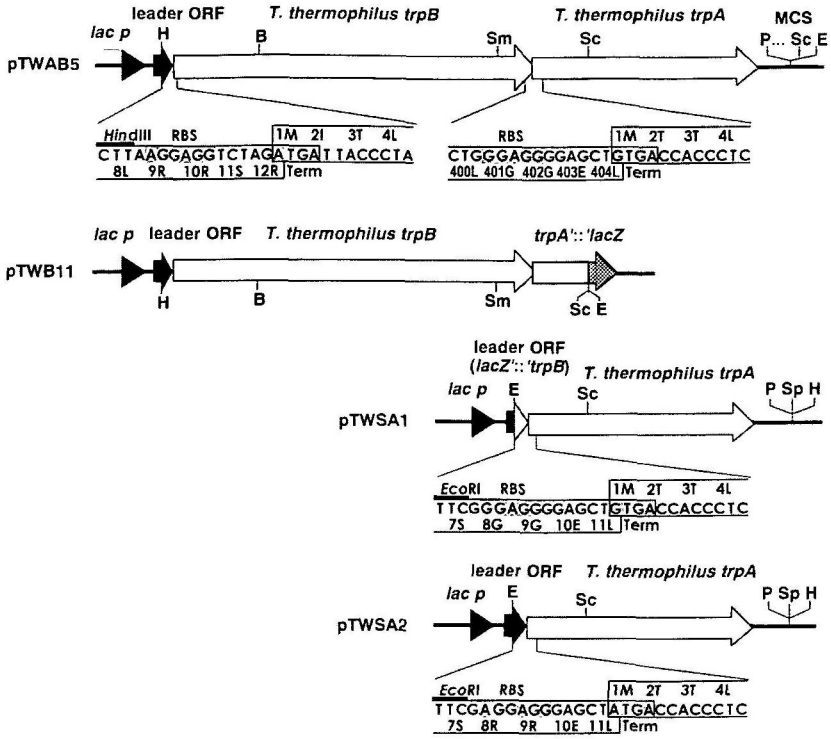

Fig. 3. **Introduction of a 4-bp overlapping leader ORF in front of the *trpB* and *trpA* genes.** Closed triangles, the *lac* promoter; closed arrows, 36 or 39-bp leader ORFs; open arrows, the *trpB*- and *trpA*-coding regions; open box, the translation initiation region of the *trpA* gene; shaded arrow, the translation termination region of the *'lacZ* gene (coding for a part of the α peptide of the β galactosidase); closed box, the translation initiation region of the *lacZ* gene; open triangle, the translation termination region of the *trpB* gene. One-letter abbreviations for recognition sites with restriction endonucleases: B, *Bal*I; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; Sc, *Sac*I; Sm, *Sma*I; Sp, *Sph*I; MCS, multiple cloning sites. Other representations are the same as in Fig. 1.

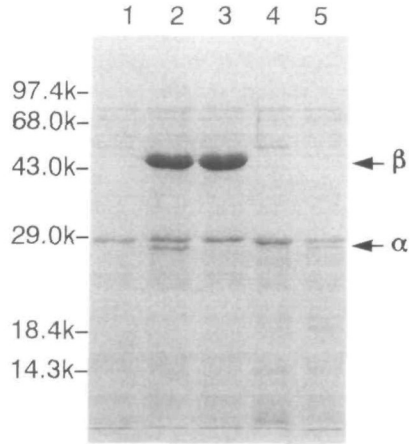

Fig. 4. **Production of the tryptophan synthase α and β subunits in a heat-treated extract of *E. coli* JM109 with a leader ORF-containing plasmid.** Plasmids used were: lane 1, pUC18 (control); lane 2, pTWAB5; lane 3, pTWB11; lane 4, pTWSA1; and lane 5, pTWSA2. Phosphorylase B (97,400), bovine serum albumin (68,000), ovalbumin (43,000), carbonic anhydrase (29,000), β-lactoglobulin (18,400), and lysozyme (14,300) were used as molecular mass standards. Arrows indicate the protein bands corresponding to the modified β subunit (L2I) ("β") and the α subunit ("α"). N-terminal amino acid sequences of both proteins were confirmed by a protein sequencer.

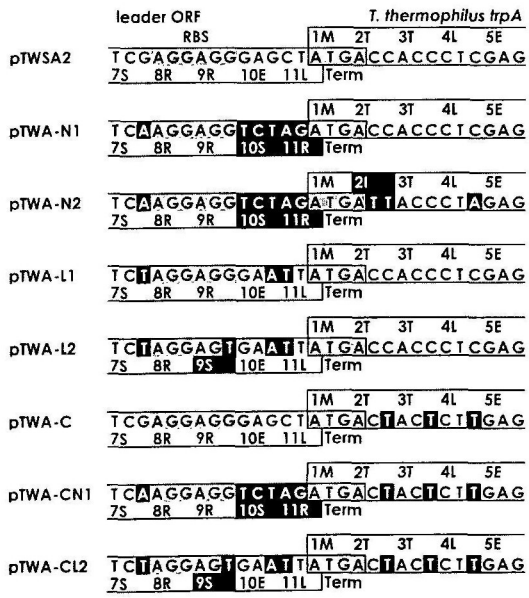

Fig. 5. **Base replacements around the ATGA motif between a leader ORF and the *trpA* gene on pTWSA2.** Nucleotides and amino acids replaced are written in reverse letters. Other representations are the same as in Fig. 1.

For pTWA-N1, 12 bp in the leader ORF upstream of the *trpA* gene of pTWSA2 were replaced with those in the leader ORF upstream of the highly expressive *trpB* gene of pTWAB5. In addition to the 12 bp in the leader ORF on pTWA-N2, 8 bp in the *trpA*-coding region were replaced with those of the *trpB*-coding region, resulting in an amino acid replacement, Thr-2 → Ile, in the gene product. Also, three G and C nucleotides in the leader ORF of pTWSA2 were replaced with A and T in pTWA-L1 by alterations within synonymous codons. The AGG codon (coding for Arg-9) in the leader ORF of pTWA-L1 was, moreover, changed to AGT (Ser) in pTWA-L2. For pTWA-C, three NNC codons adjacent to the initiation codon of the *trpA* gene of pTWSA2 were replaced with the respective synonymous NNT codons. For pTWA-CN1 and pTWA-CL2, 12 bp in the leader

ORF of pTWA-C were further replaced with those from pTWAB5 and pTWA-L2, respectively. Figure 6 displays the α subunit produced in the heat-treated extracts of *E. coli* JM109 cells harboring each plasmid. The greatest yield of α subunit was produced with pTWA-L1 (lane 4); the production level was about 10 times higher than with pTWSA2. Also, a nearly 2-fold higher yield of the α subunit was obtained with pTWA-N1 (lane 2). Comparing pTWA-L1 and pTWA-L2 (lanes 4 and 5), production levels of the α subunit decreased notably by replacing G (at position –6 from the *trpA*-initiation codon) with T, suggesting that G at this position is essential for ribosomal binding. On the other hand, production levels of the α subunit with pTWA-C, pTWA-CN1, and pTWA-CL2 were somewhat decreased (lanes 6 to 8) compared with pTWSA2. The modified α subunit (T2I) was scarcely detected in heat-treated extracts of *E. coli* JM109 harboring pTWA-N2 (lane 3).

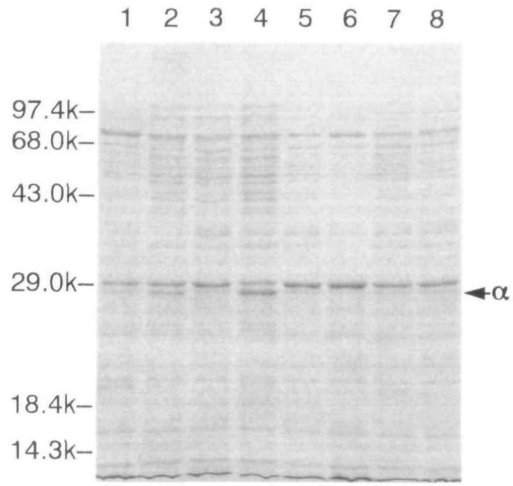

Fig. 6. **Production of the tryptophan synthase α subunit in a heat-treated extract of *E. coli* JM109 with a mutant plasmid.** Plasmids used were: lane 1, pTWSA2 (parental plasmid); lane 2, pTWA-N1; lane 3, pTWA-N2; lane 4, pTWA-L1; lane 5, pTWA-L2; lane 6, pTWA-C; lane 7, pTWA-CN1; lane 8, pTWA-CL2. Molecular mass standards are the same as those in Fig. 4. The arrow indicates the protein band corresponding to the α subunit. The N-terminal amino acid sequence of the proteins produced in the heat-treated extracts of *E. coli* JM109 (pTWA-N1) and *E. coli* JM109 (pTWA-L1) were confirmed by protein sequencing.

Construction of Expression Vectors Containing Effective Leader ORF Structures—The above results provide a design for the most effective leader ORF (details below). In addition to designing a leader ORF to enhance translation, we examined whether a strong promoter for further enhancing transcription would enhance gene expression. We used pLED-NS as the parental plasmid appropriate for this experiment. The *polA* gene on pLED-NS is transcribed under the strong *tac* promoter and the strong *rrnB* terminator. Two leader-ORF–containing plasmids, pPO45 and pPO536, were constructed by incorporating a well-designed leader ORF into pLED-NS (Fig. 7). For pPO45, the leader ORF overlaps with the *polA* (E2T)-coding region by 4 bp of the ATGA motif; for pPO536, the leader ORF overlaps with the *polA*-coding region by 1 bp in the TGATG motif. In both leader ORFs, the initiation regions were designed to be similar to that of the *lacZ* gene, and the termination regions were composed of A or T, except for the *Eco*105I and *Bbr*PI restriction sites. As a result, higher DNA polymerase activities (115 and 87 units/g cells, respectively) were detected in heat-treated extracts of *E. coli* JM109 (pPO45) and *E. coli* JM109 (pPO536) compared with 71 units/g cells for *E. coli* JM109 (pLED-NS). Similar increases were also confirmed by SDS-PAGE analysis (data not shown). These results show that, in addition to introducing a well-designed leader ORF, the use of a strong promoter further enhances *polA* gene expression in *E. coli*. Moreover, under the *tac* promoter as well as under the *lac* promoter, a 4-bp overlapping leader ORF is more effective than a 1-bp overlapping one.

New expression plasmid vectors, pLEAD4 and pLEAD5 (Fig. 8), were constructed on the basis of pPO45 and pPO536. For both vectors, the *polA*-coding regions of the parental plasmids were deleted and multiple cloning sites were inserted. Thus, pLEAD4 is designed to overlap a foreign gene with a leader ORF by 4 bp of the ATGA motif, and pLEAD5 by 1 bp in the TGATG motif. As with other common structures, both plasmids consist of the *tac* promoter, the *rrnB* terminator, and multiple cloning sites, *i.e.* *Eco*105I (or *Bbr*PI), *Eco*T22I, *Sac*I, *Kpn*I, *Stu*I, *Bam*HI, and *Hin*dIII sites.

DISCUSSION

The present results clearly demonstrate that gene ex-

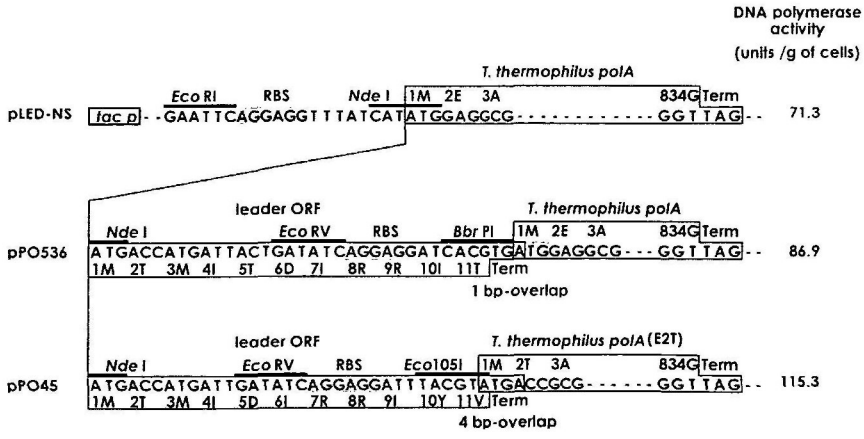

Fig. 7. **Effect of the introduction of an overlapping-leader ORF on *polA* gene expression under the *tac* promoter in *E. coli* JM109.** Leader ORFs overlap with the *polA*-coding regions of pPO536 and pPO45 by 1 bp and 4 bp, respectively. Details are in the text. "*tac p*" in a box, the *tac* promoter. Other representations are the same as in Fig. 1.

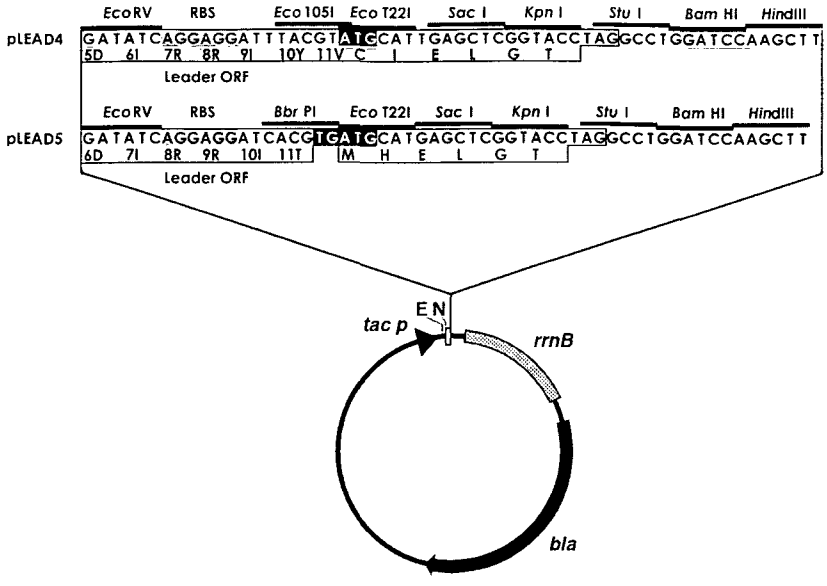

Fig. 8. **Structures of the expression vector plasmids pLEAD4 (2,657 bp) and pLEAD5 (2,659 bp).** In order to clone a foreign gene into pLEAD4 or pLEAD5, it is the best to add an *Eco*105I site (for pLEAD4) or a *Bbr*PI site (for pLEAD5) in front of the initiation codon of the gene. Also, one of the restriction sites within the multiple cloning sites on pLEAD4 or pLEAD5 is needed in the 3′ flanking region of the gene. The addition of such restriction sites to both 5′ and 3′ ends of the coding region is easily achieved by PCR. Five nucleotides of the TGATG motif on pLEAD5 and three nucleotides in the ATGA motif on pLEAD4 are written in reverse letters. Closed triangle ("*tac p*"), the *tac* promoter; open box, the translation initiation region of a leader ORF; shaded box ("*rrnB*"), the transcription terminator for the rRNA operon; closed arrow ("*bla*"), the coding region of the β-lactamase. One-letter abbreviations for restriction sites: E, *Eco*RI; N, *Nde*I. Other representations are the same as in Fig. 1.

pression in *E. coli* is more efficient when a leader ORF overlaps with the gene at the ATGA motif or TGATG motif rather than when the ORF is several-bp distant from the gene (Figs. 1 and 2). Since both overlapping motifs are exactly the same as those found in known gene pairs, they most likely function to enhance the efficiency of translation re-initiation between a leader ORF and a GC-rich gene. In addition, changing a promoter to a stronger one increased the expression of the *polA* gene coupled with a leader ORF (Figs. 1, 2, and 7), as well as the *leuB* gene coupled with a leader ORF. This also supports the view that a leader ORF activates the translation stage of GC-rich genes in *E. coli.*

The *T. thermophilus trpB* gene, as well as the *polA*, *leuB*, *pfk1* genes, was overexpressed by the introduction of a 4-bp overlapping leader ORF. In contrast, the expression of the *trpA* gene was hardly improved by a similar technique (Figs. 3 and 4). Under normal physiological conditions for *T. thermophilus*, the 4-bp overlap of GTGA likely functions in the equimolar expression between paired *trpB* and *trpA* genes. In *E. coli* cells, however, the expression of the *trpA* gene is not as high as that of the *trpB* gene. DNA sequence analysis showed no inhibitory secondary structures in the translation initiation regions of either the *trpB* or *trpA* gene. Because of the high G + C contents (69.3% in *trpB* and 70.2% in *trpA*), the codons frequently used in the *trpBA* genes are quite different from those optimally used in *E. coli* genes (*16*). However, no marked difference in codon usage was found between the *trpB* and *trpA* genes. Moreover, no stretches of codons that are rarely used in *E. coli* were found in the translation initiation regions of either of these genes.

Mutation analysis revealed that the DNA sequence around the ATGA motif significantly influences the expression efficiency of the *T. thermophilus trpA* gene (Figs. 5 and 6). The most effective mutation for enhancing expression is replacing G and C → A and T in the translation termination region of the leader ORF (pTWA-L1 and pTWA-N1). In the region between the ribosome binding sequence for the *trpA* gene and the ATGA motif, A + T contents increased to 3/5 on pTWA-N1 and 4/5 on pTWA-L1 compared with that (2/5) on pTWSA2. Expression levels of the *trpA* gene were enhanced with increasing A + T content. High A + T contents are also present in a similar region in front of the *polA*-initiation codon of pLED-NS, on which the expression level is relatively high, despite the absence of a leader ORF (Fig. 7). These findings indicate that one factor needed to enhance the expression of GC-rich genes in *E. coli* is increased A + T content in the translation termination region of a leader ORF. In contrast, replacing C → T in the *trpA*-coding region had a negative influence on gene expression (Figs. 5 and 6). This may be due to differences in translation efficiency dependent on the specific codon in the translation initiation region. Looman *et al.* (*17*) reported that the expression efficiency of the *E. coli lacZ* gene is somewhat lower when the second codon is ACT rather than ACC. It is also worth mentioning that AGR codons—those least used in *E. coli* genes—are used with great frequence (3 times) in the leader ORFs on expressible plasmids pTWA-N1 and pTWB11. In both leader ORFs and the *trpA* gene, therefore, codon usage different from that in *E. coli* genes is not critical for gene expression in *E. coli.*

The present results lead to the conclusion that an effective leader ORF for overexpressing GC-rich genes in *E. coli* should be designed as follows: (i) the initiation region of the leader ORF should be similar to that of a gene, such as the *lacZ* gene, which can be translated efficiently in *E. coli*; (ii) a ribosome binding sequence, such as AGGAGG, for the downstream gene should be placed within the leader ORF; (iii) the termination region of the leader ORF should be rich in A and T nucleotides; and (iv) the end of the leader ORF should overlap with a downstream gene by the ATGA motif or the TGATG motif. In order to induce the overexpression of various GC-rich genes using a leader ORF following the above design, two new expression vectors, *i.e.* pLEAD4 and pLEAD5, were constructed (Fig. 8). The vectors have been used on a trial basis in several laboratories other than ours. As a result, some genes were overexpressed by use of the vector, whereas the others were overexpressed by conventional methods. This can be due to the difference in the activating-point between the introduction

of an overlapping leader ORF and other methods. For overexpressing many kinds of GC-rich genes from important microorganisms in biotechnology, pLEAD4 and pLEAD5, together with known expression systems, may prove to be very useful tools.

We thank Messrs. Masahiro Yazaki and Toshio Horii for their generous assistance.

REFERENCES

1. Ishida, M. and Oshima, T. (1994) Overexpression of genes of an extreme thermophile, *Thermus thermophilus*, in *Escherichia coli* cells. *J. Bacteriol.* **176**, 2767–2770
2. Ishida, M., Yoshida, M., and Oshima, T. (1997) Highly efficient production of *Thermus thermophilus* enzymes: A practical method to overexpress in *Escherichia coli* of GC-rich genes from an extreme thermophile. *Extremophiles* **1**, 157–162
3. Suzuki, T., Tanaka, Y., Ishida, M., Ishizuka, M., Yamagishi, A., and Oshima, T. (1997) Screening of a mutant plasmid with high expression efficiency of GC-rich *leuB* gene of an extreme thermophile, *Thermus thermophilus*, in *Escherichia coli*. *J. Biochem.* **121**, 1031–1034
4. Yanofsky, C. and Crawford, I.P. (1987) The tryptophan operon in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology* (Neidhardt, F.C., Ingraham, J.L., Low, K.B., Magasanik, B., and Umbarger, H.E., eds.) pp. 1453–1472, American Society for Microbiology, Washington, DC
5. Lijestroem, P., Laamanen, I., and Palva, E.T. (1988) Structure and expression of the *ompB* operon, the regulatory locus for the outer membrane porin regulon in *Salmonella typhimurium*. *J. Mol. Biol.* **201**, 663–673
6. Draper, D.E. (1996) Translation initiation in *Escherichia coli and Salmonella Cellular and Molecular Biology* (Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M., and Umbarger, H.E., eds.) Vol. 1, pp. 902–908, ASM Press, Washington, DC
7. Schoner, B.E., Belagaje, R.M., and Schoner, R.G. (1991) Enhanced translational efficiency with two-cistron expression system. *Methods Enzymol.* **185**, 94–103
8. Koyama, Y. and Furukawa, K. (1990) Cloning and sequence analysis of tryptophan synthase genes of an extreme thermophile, *Thermus thermophilus* HB27: Plasmid transfer from replica-plated *Escherichia coli* recombinant colonies to competent *T. thermophilus* cells. *J. Bacteriol.* **172**, 3490–3495
9. Asakura, K., Komatsubara, H., Soga, S., Yomo, T., Oka, M., Emi, S., and Urabe, I. (1993) Cloning, nucleotide sequence, and expression in *Escherichia coli* of DNA polymerase gene (*polA*) from *Thermus thermophilus* HB8. *J. Ferment. Bioeng.* **76**, 265–269
10. Myers, T.W. and Gelfand, D.H. (1991) Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry* **30**, 7661–7666
11. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119
12. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
13. Miles, E.W., Bauerle, R., and Ahmed, S.A. (1987) Tryptophan synthase from *Escherichia coli* and *Salmonella typhimurium*. *Methods Enzymol.* **142**, 398–414
14. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
15. Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**, 10035–10038
16. Nakamura, Y., Gojobori, T., and Ikemura, T. (1997) Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Res.* **25**, 244–245
17. Looman, A.C., Bodlaender, J., Comstock, L.J., Eaton, D., Jhurani, P., De Boer, H.A., and Van Knippenberg, P.H. (1987) Influence of the codon following the AUG initiation codon on the expression of a modified *lacZ* gene in *Escherichia coli*. *EMBO J.* **6**, 2489–2492