Biochemical and Biophysical Research Communications **264,** 24–28 (1999)
Article ID bbrc.1999.1480, available online at http://www.idealibrary.com on IDEAL®

# Cloning, Sequencing, and Characterization of Ribosomal Protein and RNA Polymerase Genes from the Region Analogous to the α-Operon of *Escherichia coli* in Halophilic Archaea, *Halobacterium halobium*

Kazuyoshi Sano, Atsuo Taguchi, Hiromi Furumoto, Taizo Uda, and Takuzi Itoh[1]
*School of Bioresources, Hiroshima Prefectural University, Shobara City, Hiroshima 727-0023, Japan*

Received September 1, 1999

**A determination was made of the nucleotide sequence of the 3215-bp region of a ribosomal protein gene cluster (HS13, HS4, HS11, and HeL18), RNA polymerase (RNA poly D), and tRNA genes (tRNAser and tRNAarg) of halophilic Archaea *Halobacterium halobium,* which is analogous to the α-operon of *Escherichia coli* (tRNAser-HS13-HS4-HS11-RNA poly D-tRNAarg-HeL18). The seven-gene string was preceded by a pseudoknot-like structure similar to the proposed S4 ribosomal protein binding site of the α-operon mRNA leader in *E. coli.* Using an inducible expression system *H. halobium* HS4 was produced in large amounts in *E. coli,* and immunoblot analysis showed the S4 to constitute a 21-kDa polypeptide component of the ribosome. Analysis of the deduced amino acids sequence revealed that the HS13, HS4, and HS11 sequences including the RNA polymerase subunit are more similar to their eukaryotic than to their bacterial counterparts. HeL18, located downstream of the gene cluster analogous to the *E. coli* α-operon (S13-S11-S4-RNA poly D-L17), was similar to both the eukaryotic (eL18) and eubacterial ribosomal protein L15 located in the spc-operon, but not to L17 positioned as the terminal gene of the bacterial α-operon.** © 1999 Academic Press

Ribosomes are essential cell components that are present in all organisms. The constituents of ribosomes, diverse rRNAs and proteins (rp), should be quite useful for studying molecular evolution with respect to the conservation of functionally important structures. Until the latter half of 1970s, the biological kingdom was believed to consist of eukaryotes and prokaryotes. However, based on studies of the structure of ribosomal RNA, Woese *et al.,* in 1977 observed a group that were markedly different in evolution and phylogenic features from other prokaryotes, namely Archaea, which led to other prokaryotes being designated as Bacteria (1). All living cells belong to one of three lines of descent or phylogenetic domains; Bacteria, Archaea, and Eucarya (eukaryotes) (2). According to structural studies on rp, protein elongation factors and RNA polymerase, the Archaea form a specific group that are apparently different in evolution, and most of the proteins of this group as gene products are much more analogous to those of Eucarya (3–8). However, the gene order and composition of these units are conserved among Archaea, and are very similar to those in Bacteria. These results substantiate the consideration that Archaea represents an intermediate evolutionary stage between Bacteria and Eucarya (8). In addition, complete or partial sequences and gene order in a chromosome for many rps are available from Archaea, permitting a more extensive evolutionary comparison. Such comparisons can provide important structural and functional information on molecules or gene transitional information during the course of evolution, by identifying both core regions of the proteins that have been highly conserved and regions subject to fewer constraints on gene order in a chromosome.

In the present study, the authors cloned and sequenced the *H. halobium* α-operon of rp in order to examine how *H. halobium* rp genes are organized. The phylogenetic relationships and rp gene evolution in the three kingdoms were also investigated.

## MATERIALS AND METHODS

*Bacterial strains, plasmids, and media.* *Halobacterium halobium* S9 strain was grown at 40°C as already described (3). Cloning of *H. halobium* DNA was carried out by using the *E. coli* host/vector systems (JM109/pUC18 or pUC19, XL1-Blue/Super Cos1) and the expression system BL21(DE3)/pET15b (Novagen). *E. coli* strains

Abbreviations used: bp, base pair(s); kb, 1000 bp; ORF, open reading frame; rp, ribosomal protein; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis.

[1] To whom correspondence should be addressed. Fax: 81-8247-4-1776. E-mail: takuzi@hiroshima-pu.ac.jp.

0006-291X/99 $30.00

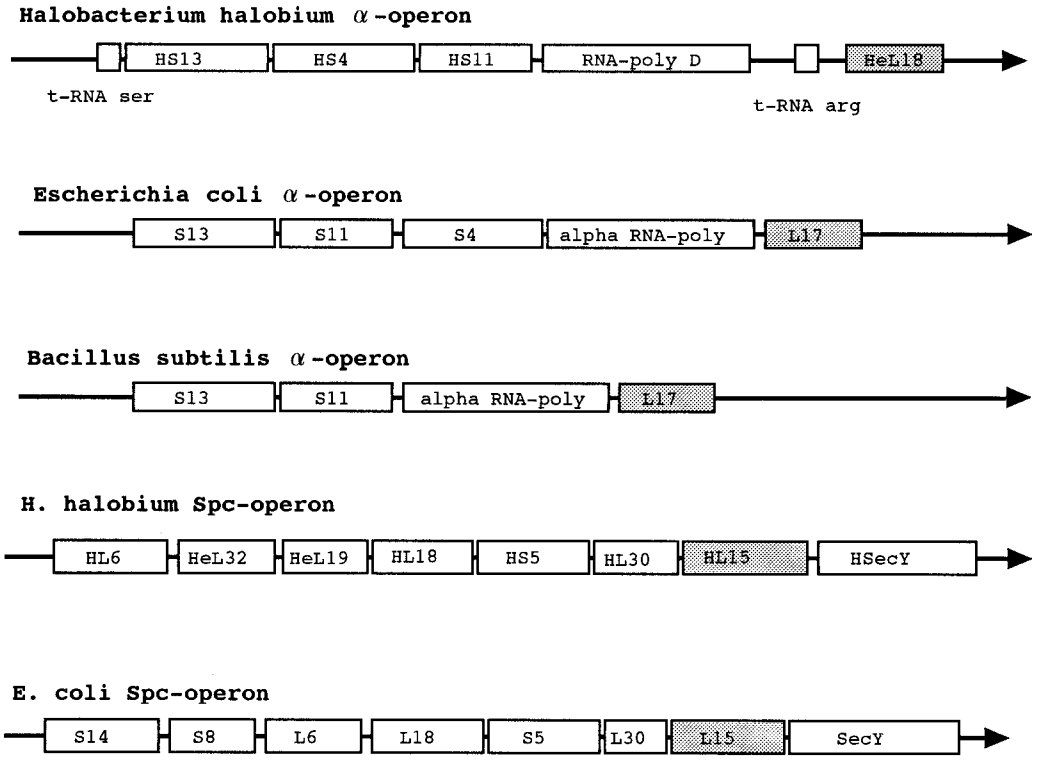


**FIG. 1.** Comparison of the gene arrangement of the *H. halobium* ribosomal protein gene cluster with those of *E. coli* and *B. subtilis.*

were grown in yeast/tryptone medium at 37°C with 100 μg/ml ampicillin.

*Isolation of clones and sequencing.* For the cosmid library, insert DNA consisted of partially digested DNA with Sau3A. Insert DNA was ligated with Super Cos1 *Xba*I/*Bam*HI-digested cosmid vector. The Super Cos1 cosmid library was packaged using Gigapack III packaging extracts as described in the manufacturer's instructions (Stratagen). Using a mixture of synthetic oligonucleotides corresponding to the conserved amino acid sequence of the rp S4 from the Archaea (T-K-F-Y-E-T-P), colony hybridization experiments were performed at 37°C as described in (5). Synthetic oligonucleotides were labeled using [$\gamma$-$^{32}$P]ATP (Amersham; sp act 3000 Ci/mmol) and T4 polynucleotide kinase. Restriction endonuclease enzymes and T4 DNA ligase were obtained from Takara Shuzo Co. and used according to the manufacturer's instructions. Recombinant cosmids were extracted and subsequent colony hybridization was carried out essentially according to the methods of Maniatis *et al.* (9). DNA sequencing was carried out on both strands using an AutoRead Sequencing Kit (Pharmacia Biotech.), with universal and *de novo* synthesized primers. Reactions were run on a Pharmacia sequencing instrument, sequencing data were analyzed using the DNASIS program (Hitachi), and open reading frames were matched with protein sequences in the Swiss Prot protein databank using BLAST X and BLAST P with BLOSUM62 substitution matrix and default parameters (10). The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with Accession No. AB030282.

*Expression of the ribosomal protein S4 gene product.* An inducible T7 promoter overexpression system was used for the production of the *H. halobium* rp S4 in *E. coli* (11). The S4 coding region was amplified by PCR using a 5′ oligonucleotide primer extended by a *Nde*I restriction site (5′-CCCCATATGGCGCTGCCCGGCGAGAACACCAAGTT-3′) and a 3′ reverse complement oligonucleotide primer containing a *Bam*HI site (5′-TTGGGATCCTCACTCCTGAGCACCGGCGCGAGCCG-3′). PCR was performed in the reaction mixture (100 μl) containing 100 pmol of each primer, four dNTPs at 0.2 mM each, 20 mM Tris–HCl (pH 8.2), 2 mM $MgCl_2$, 10 mM KCl, 6 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 1% nuclease-free BSA, 100 ng *H. halobium* genomic DNA and 2.5 U Pfu DNA polymerase (Stratagene). The reaction mixture was subjected to 30 cycles of polymerization at 98°C for 1 min and 70°C for 5 min. The amplification product was digested with NdeI and BamHI; the resulting fragment was ligated into vector pET 15b and transformed into *E. coli* strain BL 21(DE3). Expression of the S4 gene was induced by addition of 1 mM isopropyl-1-thio-β-D-galacto-pyranoside (IPTG) for 3 h. Samples were taken from induced and uninduced cell suspensions and centrifuged. For analysis of the cell protein the cell pellet was suspended in a sample buffer, heated, and run on an SDS–PAGE according to the methods of Laemmli (12).

*Preparation of the anti-S4 antibodies and Western blotting.* The overproduced S4 polypeptides were purified from inclusion bodies using His · Tag/His · Bind metal chelation affinity chromatography under denaturing conditions according to the protocol provided by the manufacturer (Novagene). Denatured S4 protein was dialyzed stepwise against 50 mM Tris–HCl (pH 8.0) and 500 mM NaCl. Finally, the protein solution was concentrated about tenfold (Ultrafree-15 centifugal filter, Millipore) and aliquots were stored frozen at −80°C or refrigerated at 4°C. A total of 200 μg protein was used to immunize a mouse for antibody production. Proteins from transformed *E. coli* whole cell extracts and rps extracted from *H. halobium* 70S ribosome with 67% acetic acid were transferred to a PVDF-membrane (3). Transfer of proteins from slab gels into a PVDF-membrane was performed by horizontal semi-dry blotting in a buffer system consisting of 20% (v/v) methanol with 50 mM Tris, 20 mM glycine (pH 8.3), 0.1% SDS with a constant current of 2 mA/cm$^2$ for 30 min. The Western blot analysis was performed as described in (13). Blotted filters were incubated overnight with polyclonal mouse antibody against S4 (1:1000 dilution) in a TBS buffer containing 20 mM Tris–HCl (pH 7.5), 500 mM NaCl, 0.5% BSA and 0.05% Tween

20. Bound antibodies were detected by anti-mouse IgG goat antibodies conjugated with alkaline phosphatase.

## RESULTS AND DISCUSSION

*Nucleotide sequence analysis.* To identify DNA fragments carrying the S4 gene of *H. halobium*, a genomic library of cosmids of *H. halobium* was screened with synthetic oligonucleotides corresponding to the conserved amino acid sequence of rp S4. By colony and Southern hybridization experiments using a DNA probe, six positive cosmid clones were isolated and the whole sequence of the approximate 3-kb DNA fragment, hybridized with the probe, was determined by subcloning appropriate restriction fragments that were sequenced using the dideoxy-chain termination method.

The sequence was analyzed with the codon preference program deduced from the haloarchaeal codon usage table showing a very high GC content in the third codon position (8). Inspection of the sequence showed a 5′-noncoding region (332 bp) followed by two genes for tRNA and five ORFs spaced from one another at very short intervals (0–39 bp) on the same strand except for the interval (124 bp) between the second tRNA gene and its upstream ORF (Fig. 1). The presence of promoter-like AT rich sequences located upstream of the first tRNAser and second tRNAarg gene suggest that both constituents of tRNA and ORFs are organized as transcriptional units (tRNAser-HS13-HS4-HS11-RNA poly D and tRNAarg-HeL18).

*Expression of the ribosomal protein S4 gene product.* To confirm the determined nucleotide sequence and identify the proteins constituting the ribosome, the obtained ORF of the S4 gene (positions 952–1462) was expressed in *E. coli* under the control of the phage T7 promoter. Induction of the *E. coli* strain BL21 (DE3), transformed with the vector pET15b harboring the *H. halobium* S4 coding region, resulted in the production of a protein with a molecular mass of 22 kDa as seen from *E. coli* crude extracts analyzed on SDS–PAGE stained with Coomassie blue (Fig. 2A). A significant time-dependent increase of the respective band was evident. Western blot analysis of proteins extracted from the 70S ribosomes of the *H. halobium* showed the respective 21 kDa band reacting strongly with the anti-recombinant S4 (*E. coli* His-Tag/*H. halobium* S4) antiserum (Fig. 2B).

*Gene cluster coding for proteins homologous to ribosomal proteins and RNA polymerase subunit D protein.* By comparison of the derived amino acid sequences with ORFs in a data bank, the order of the six genes (tRNAser-HS13-HS4-HS11-RNA poly D-tRNAarg-HeL18) determined was found to resemble that found in the *E. coli* $\alpha$-operon (S13-S11-S4-RNA poly$\alpha$-L17). The HS13 proteins, deduced from the DNA sequence, showed 68, 30, and 19% sequence homologies with the deduced *Haloarcula marismortui* HmaS13 (14), *Drosophila melanogaster* S13 (15), and *E. coli* S13 ribosomal proteins (16), respectively (all protein sequences are deduced from the DNA sequences hereafter). The proteins predicted for HS11 and HS4 showed 43 and 15% sequence homologies with *E. coli* S11 and S4 rps, respectively. Additionally, all these proteins were more similar to eukaryotic S11 and to S4 rps than to eubacterial *E. coli*, as has been found for HS13 and many other genes for transcription and translation (5–8, 17).

High homology with the RNA polymerase D subunit from *H. halobium* was observed in the eukaryotic RNA polymerase II D subunit, but the similarity to the RNA polymerase alpha subunit of bacterial origin was weak and mainly restricted to the short regions.

*An eukaryotic L18 gene homolog of Archea and a bacterial specific L17 gene of Bacteria.* Adjacent to the RNA polymerase D subunit, the gene for tRNAarg was found, followed by the another ORF (HeL18) in *H. halobium*. In the case of *E. coli* and *B. subtilis*, the gene for the L17 ribosomal protein is found downstream of the RNA polymerase alpha gene (Fig. 1) (16, 18). A computer search for homologous proteins to the HeL18 sequence revealed a similarity to eukaryotic rp eL18 as well as eubacterial L15 located in the other ribosomal

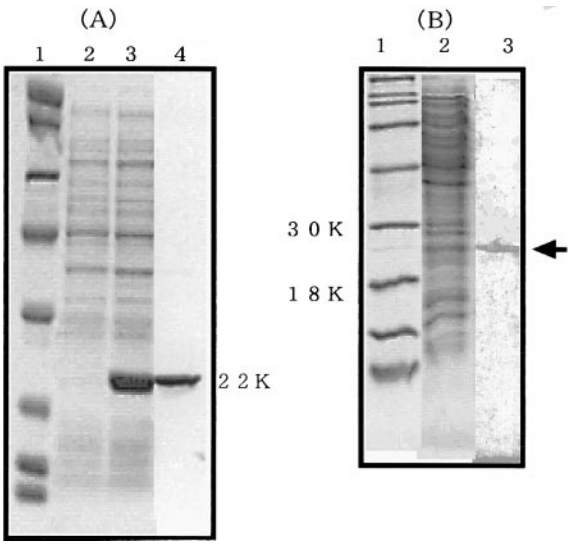


**FIG. 2.** SDS–PAGE analysis of *H. halobium* S4. (A) SDS–PAGE analysis of *H. halobium* S4 expressed in *E. coli*. Cell extracts from uninduced (lane 2) and induced (lane 3, 3 h after induction with IPTG) *E. coli* strain BL21 (DE3) harboring the *H. halobium* S4 coding region and the recombinant purified His-tag S4 (1 $\mu$g) (lane 4) were loaded on an 12% polyacrylamide gel, electrophoresed, and stained with Coomassie blue; lane 1, molecular mass marker proteins. (B) Western blot. Purified 70S ribosomal proteins from *H. halobium* (lanes 2 and 3) were loaded on gel identical to that described in A, transferred to a nitrocellulose membrane and probed with anti-S4 ribosomal protein antibodies. Ribosomal protein (lane 2) was stained with Coomassie blue, and lane 3 was visualized by immunostaining as described under Materials and Methods.

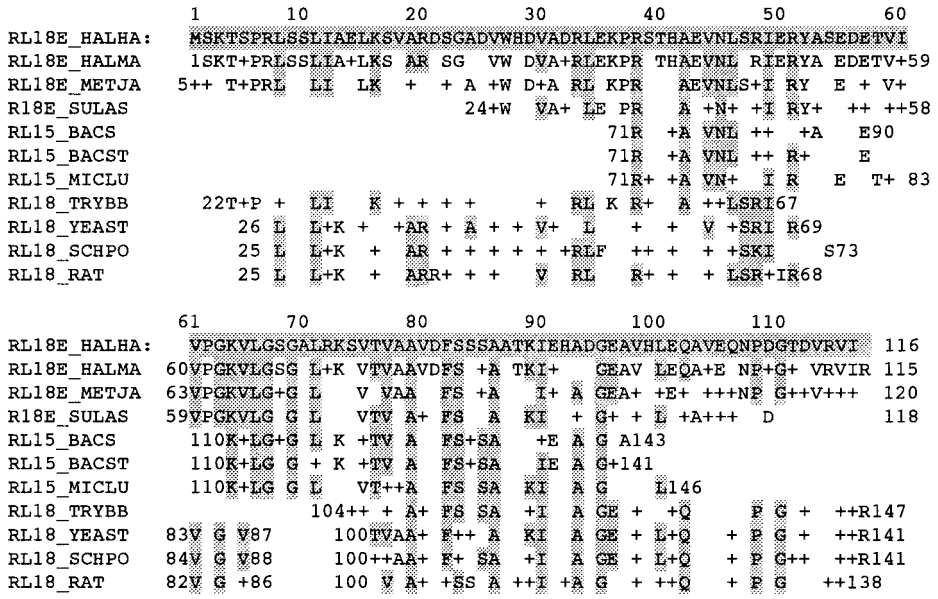


**FIG. 3.** Sequence alignment of ribosomal protein eL18 and L15. Only the amino acids that are same as the *H. halobium* sequence are indicated. Positions in which sequence conservation is >40% identity are highlighted in black. + indicates conservative substitutions. Sequences were extracted from the EMBL or SWISSPROT databases. Protein names are as follows: RL18E_HALHA, *H. halobium* HeL18; RL18E_HALMA, *H. marismortui* eL18; RL18E_METJA, *Methanococcus jannaschii* eL18; R18E_SULAS, *Sulfolobus acidocaldarius* eL18; RL15_BACS, *B. subtilis* L15; RL15_BACST, *B. stearothermophilus* L15; RL15_MICLU, *Micrococcus luteus* L15; RL18_TRYB, *Trypanosoma brucei* eL18; RL18_YEAST, *Saccharomyces cerevisiae* eL18; RL18_SCHPO, *Schizosaccharomyces pombe* eL18; RL18_RAT, *Rattus norvegicuse* eL18.

gene cluster in *E. coli* (Spc-operon, Fig. 1), but no similarity to eubacterial L17 linked to the RNA polymerase alpha gene. Identical or conserved amino acids were distributed among the whole sequence of HeL18 and eukaryotic eL18, but were limited to the central region (amino acids 38–100) of HeL18 and the bacterial L15 protein of the spc-operon (Fig. 3). This suggested that the N-terminal or C-terminal nonconserved regions of the amino acid sequence of HeL18 might have evolved to a bacterial specific sequence, on

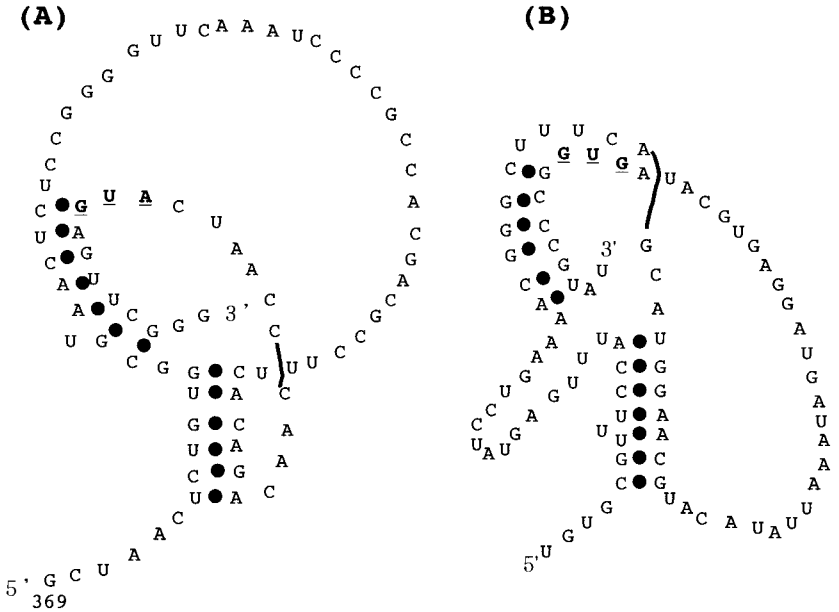


**FIG. 4.** Diagram of the α-operon leader pseudoknot secondary structure in *H. halobium* (A) and *E. coli* (B) (21). The initiation codon and hydrogen bond are indicated by an underline and closed circle, respectively.

the assumption that the amino acid sequences of haloarchaeal HeL18 kept the universal ancestor state. The haloarcheal L15 homolog (HL15) to bacterial L15 of the spc-operon has been found already in the analogous gene cluster, although in Eukarya no homolog has been found (unpublished data). These results indicate that the *H. halobium* genome contains both homologs of bacterial rp L15 and eukaryotic eL18 on the separated gene cluster. The existence of such bacterial and eukaryotic genes duplicated in Archaea have also been found for the initiation factor gene (hIF-2 and hIF-2$\gamma$) (19, 20).

It is interesting to note that the homolog of the eubacterial L17 gene, located downstream of the RNA polymerase alpha-subunit gene in bacteria, is lacking in *H. halobium* and no homolog of it could be found in other Archaea or Eukarya. Similar observation were made for the eubacterial L16 gene, which is located between S3 and L29 gene in the S10 operon (8). These results suggested that L17 and L16 may be a specific protein found in bacterial ribosomes alone and participate in the somewhat bacterial specific structure and function of protein synthesis.

The order of the haloarchaeal gene HS13-HS4-HS11-RNA poly D is similar to the $\alpha$-operon in eubacterial *E. coli* when the genes for S4 and S11 are reversed (Fig. 1). In the *E. coli* $\alpha$-operon, the ribosomal protein S4 binds a large domain of the 16S ribosomal RNA and also a pseudoknot structure in the $\alpha$-operon mRNA leader region, where it represses its own synthesis (21). Computer analysis of the haloarchaeal $\alpha$-operon leader sequence (positions 368–447) indicated that this sequence has the potential to fold into a pseudoknot structure, similar to the proposed S4 binding site in *E. coli* (Fig. 4). That the gene cluster and pseudoknot structure are similar to those of *E. coli* suggests that there are no obvious fundamental differences in the operation of rp synthesis in the two organisms. However, the presence of two homologs each of bacterial L15 and the eukaryotic ribosomal protein eL18 in *H. halobium* suggest that Archaea seem to partially maintain the chimeric features of Bacteria and Eukarya in their translation system.

## ACKNOWLEDGMENT

We thank C. Davenport for critical reading of the manuscript.

## REFERENCES

1. Woese, C. R., and Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5088–5090.
2. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 4576–4579.
3. Itoh, T., Kumazaki, T., Sugiyama, M., and Otaka, E. (1988) *Biochim. Biophys. Acta* **949,** 110–118.
4. Itoh, T. (1988) *Eur. J. Biochem.* **176,** 297–303.
5. Itoh, T. (1989) *Eur. J. Biochem.* **186,** 213–219.
6. Fujita, T., and Itoh, T. (1995) *Biochem. Mol. Biol. Int.* **37,** 107–115.
7. Leffers, H., Gropp, R., Lottspeich, F., Zillig, W., and Garrett, R. A. (1989) *J. Mol. Biol.* **206,** 1–17.
8. Miyokawa, T., Urayama, T., Shimooka, K., and Itoh, T. (1996) *Biochem. Mol. Biol. Int.* **39,** 1209–1220.
9. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
11. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorf, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
12. Laemmli, U. K. (1970) *Nature* **227,** 680–685.
13. Blake, M. S., Johnston, K. H., Russel-Jones, G. J., and Gotschlich E. C. (1989) *Anal. Biochem.* **136,** 175–179.
14. Scholzen, T., and Arndt, E. (1992) *J. Biol. Chem.* **267,** 12123–12130.
15. Garwood, J., and Lepesant, J. A. (1994) *Gene* **141,** 231–235.
16. Bedwell, D., Davis, G., Gosink, M., Post, L., Nomura, M., Kestler, H., Zengel, J. M., and Lindahl, L. (1985) *Nucleic Acids Res.* **13,** 3891–3903.
17. Yuki, Y., Kanechika, R., and Itoh, T. (1993) *Biochim. Biophys. Acta* **1216,** 335–338.
18. Boylan, S. A., Suh, J. W., Thomas, S. M., and Price, C. W. (1989) *J. Bacteriol.* **11,** 2553–2562.
19. Kyrpides, N. C., and Woese, C. R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3726–3730.
20. Hasegawa, Y., Sawaoka, N., Kado, N., Ochi, M., and Itoh, T. (1998) *Biochem. Mol. Biol. Int.* **46,** 495–507.
21. Baker, A., and Draper, D. E. (1995) *J. Biol. Chem.* **270,** 22939–22945.