

Conservation and evolution of the *rpsU*–*dnaG*–*rpoD* macromolecular synthesis operon in bacteria

James Versalovic,¹ Thearith Koeuth,¹ Robert Britton,¹ Kati Geszvain¹ and James R. Lupski^{1,2*}

¹Institute for Molecular Genetics and ²Department of Pediatrics, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA.

Summary

The macromolecular synthesis (MMS) operon contains three essential genes (*rpsU*, *dnaG*, *rpoD*) whose products (S21, primase, sigma-70) are necessary for the initiation of protein, DNA, and RNA synthesis respectively. PCR amplifications with primers complementary to conserved regions within these three genes, and subsequent DNA sequencing of *rpsU*–*dnaG* PCR products, demonstrate that the three genes appear to be contiguous in 11 different Gram-negative species. Within the Gram-negative enteric bacterial lineage, the S21 amino acid sequence is absolutely conserved in 10 species examined. The putative *nut_{eq}* antiterminator sequence in *rpsU* consists of two motifs, *boxA* and *boxB*, conserved in primary sequence and secondary structure. The terminator sequence, T₁, located between *rpsU* and *dnaG* is conserved at 31 positions in nine enterobacterial species, suggesting the importance of primary sequence in addition to secondary structure for transcription termination. The intergenic region between *rpsU* and *dnaG* varies in size owing to the presence or absence of the Enterobacterial Repetitive Intergenic Consensus (ERIC) DNA element. The *rpoD* gene contains rearrangements involving a divergent sequence, although two carboxy-terminal regions which encode functional domains are conserved in primary sequence and spacing. Our data suggest that primary sequence divergence and DNA rearrangements in both coding and non-coding sequences account for the interspecies variation in operon structure. However, MMS operon gene organization and *cis*-acting regulatory sequences appear to be conserved in diverse bacteria.

Introduction

The *Escherichia coli* macromolecular synthesis (MMS) operon contains three genes, *rpsU*, *dnaG*, and *rpoD*, whose gene products, S21, primase, and sigma-70 initiate protein, DNA, and RNA synthesis respectively (Lupski *et al.*, 1983; Burton *et al.*, 1983; Lupski and Godson, 1984; 1989). Ribosomal protein S21 facilitates translation initiation either by enabling the interaction of the mRNA ribosome-binding site (RBS) and the complementary 3' end of the 16S rRNA (Backendorf *et al.*, 1981; Kyriatsoulis *et al.*, 1986) or by promoting ribosomal subunit association (Melancon *et al.*, 1990). Primase synthesizes the primer RNA (pRNA) to begin leading (van der Ende *et al.*, 1985) and lagging (Lark, 1972) strand DNA synthesis for chromosomal replication. DnaG may also play a role in chromosome partitioning as evidenced by two alleles, *dnaG2903* and *parB*, which display abnormal nucleoid segregation phenotypes (Grompe *et al.*, 1991). Sigma-70 represents the principal sigma factor of *E. coli* and facilitates the recognition of specific promoter sequences by the RNA polymerase holoenzyme complex (Helmann and Chamberlin, 1988; Dombroski *et al.*, 1992). The three genes in the MMS operon encode products involved in initiation of the three principal informational macromolecules of the cell; DNA, RNA, and protein. The fact that these genes are contained in a single MMS operon suggests a need for co-ordinate regulation of macromolecular synthesis.

The presence of an RNA transcript that encompasses all three genes (Burton *et al.*, 1983) and polar effects of *dnaG* amber mutations on downstream *rpoD* gene expression (Nakamura, 1984) demonstrate that these genes are in one operon. The MMS operon exemplifies a complex operon structure whereby three genes which regulate distinct biosynthetic processes can be co-ordinately regulated in a single operon while individual genes within the operon can be expressed at different levels (Lupski and Godson, 1989). *Cis*-acting regulatory sequences, located within intergenic sequences as well as within structural genes, account for differences in expression between individual genes. Regulation by transcription termination (Smiley *et al.*, 1982; Lupski *et al.*, 1983; 1984; Burton *et al.*, 1983), antitermination (Lupski *et al.*, 1983; 1984; Peacock *et al.*, 1985; Almond *et al.*, 1989), codon usage differences (Smiley *et al.*, 1982; Konigsberg and Godson, 1983), retroregulation (Burton *et al.*, 1983), and multiple

promoters (Lupski *et al.*, 1983; 1984; Burton *et al.*, 1983; Nesin *et al.*, 1988) probably contribute to differences in gene expression.

The MMS operon was initially described in *E. coli*. Analysis of the MMS operon in the closely-related Gram-negative enterobacterium, *S. typhimurium* (Erickson *et al.*, 1985), revealed conservation of the operon gene structure in addition to significant similarities between coding and regulatory sequences. The distantly-related Gram-positive bacterium, *Bacillus subtilis* (Wang and Doi, 1986), contains the *dnaG* and *rpoD* genes juxtaposed in a single operon. Upstream of *dnaG* is a gene designated *P23* whose function is unknown. Recently DNA sequence information has established the presence of *dnaG* just upstream of *rpoD* in the obligate intracellular bacterium, *Rickettsia prowazekii* (Marks *et al.*, 1992; Marks and Wood, 1992). Conservation of MMS operon structure throughout extensive phylogenetic distances underscores the significance of co-ordinate regulation of genes initiating macromolecular synthesis.

We examined the conservation and evolution of MMS operon structure in diverse bacterial species. Degenerate oligonucleotide primers that matched conserved DNA sequences in MMS operon genes were used in the polymerase chain reaction (PCR) to determine if these coding sequences were contiguous. DNA sequencing of MMS operon PCR products from 11 different Gram-negative bacterial species enabled us to examine the level and pattern of primary sequence conservation within expressed and non-coding intergenic sequences and to study the evolution of the complex MMS operon. The MMS operon appears to be conserved in diverse bacterial

species. Conservation of the MMS operon structure throughout bacteria suggests functional importance for maintaining genes whose products are involved in initiating the synthesis of the major informational macromolecules of the cell within a single regulatory unit.

Results

PCR-based strategy to evaluate operon conservation

An approach based on the PCR was developed to evaluate the conservation of MMS operon structure in diverse bacterial species. Coding sequences in essential genes which are conserved by functional constraints represent potential primer binding sites for evolutionary studies. If two or more genes are contiguous, PCR with primers in opposing orientations will yield specific amplification products provided that the primer binding sites are conserved and the distances between these two primer binding sites do not exceed the limits of polymerase extension.

The feasibility of utilizing individual oligonucleotide primers to recognize single divergent DNA sequences was tested initially with *dnaG-rpoD* intergenic sequence primers. The *dnaG-rpoD* intergenic sequence differs in length and primary sequence between species (Smiley *et al.*, 1982; Burton *et al.*, 1983; Erickson *et al.*, 1985; Wang and Doi, 1986; Marks and Wood, 1992). PCR primers *Bs2*, *Ec1*, and *St2B*, exactly matched genus-specific *dnaG-rpoD* intergenic sequences from *B. subtilis*, *E. coli* and *S. typhimurium*, respectively (Table 1; Fig. 1A). With genomic template DNAs from these three bacteria

Table 1. PCR primer sequences and conditions.

PCR primers	Primer sequences	PCR (denature; anneal; extend)
ALL1I	(+) 5'- CAITGCTTTGGITGIGGIGCGIIIGGCAA-3'	90°C, 30 s;
ALL3I	(-) 5'- GCCTGICIGATCCACCAIGTIGCITAIGT-3'	47-60°C, 1 min;
		65°C, 8 min
ALL2I	See Versalovic <i>et al.</i> (1991b)	90°C, 30 s;
ALL4I	See Versalovic <i>et al.</i> (1991b)	55°C, 1 min;
		70°C, 5 min
<i>Bs2</i>	(+) 5'- GATCGCTTAACCTCATCATG-3'	90°C, 30 s;
ALL3I	see above	60°C, 1 min;
		70°C, 5 min
<i>Ec1</i>	(+) 5'- TTATCGTTGGCGGTAACAACCGTTGG-3'	90°C, 30 s;
ALL3I	see above	60°C, 1 min;
		70°C, 5 min
RPSU-5'ATG-BamHI	(+) 5'- TTGGATCCACATGCCGGTAATTAAAGTACGTG-3'	90°C, 1 min;
ALL1R-I-EcoRI	(-) 5'- TTGAATTCTTGCCIIICGICCCICAICCAAAGCAITG-3'	42°C, 1 min;
		70°C, 5 min
RPSU-5'ATG-EcoRI	(+) 5'- TTGAATTCACATGCCGGTAATTAAAGTACGTG-3'	90°C, 1 min;
ALL1R-I-BamHI	(-) 5'- TTGGATCCTTGCCIIICGICCCICAICCAAAGCAITG-3'	42°C, 1 min;
		70°C, 5 min
<i>St2B</i>	(+) 5'- CGACAGCTATACCGTCGACACC-3'	90°C, 30 s;
ALL3I	see above	60°C, 1 min;
		70°C, 5 min.

(+), sense strand; (-), antisense strand; I, inosine.

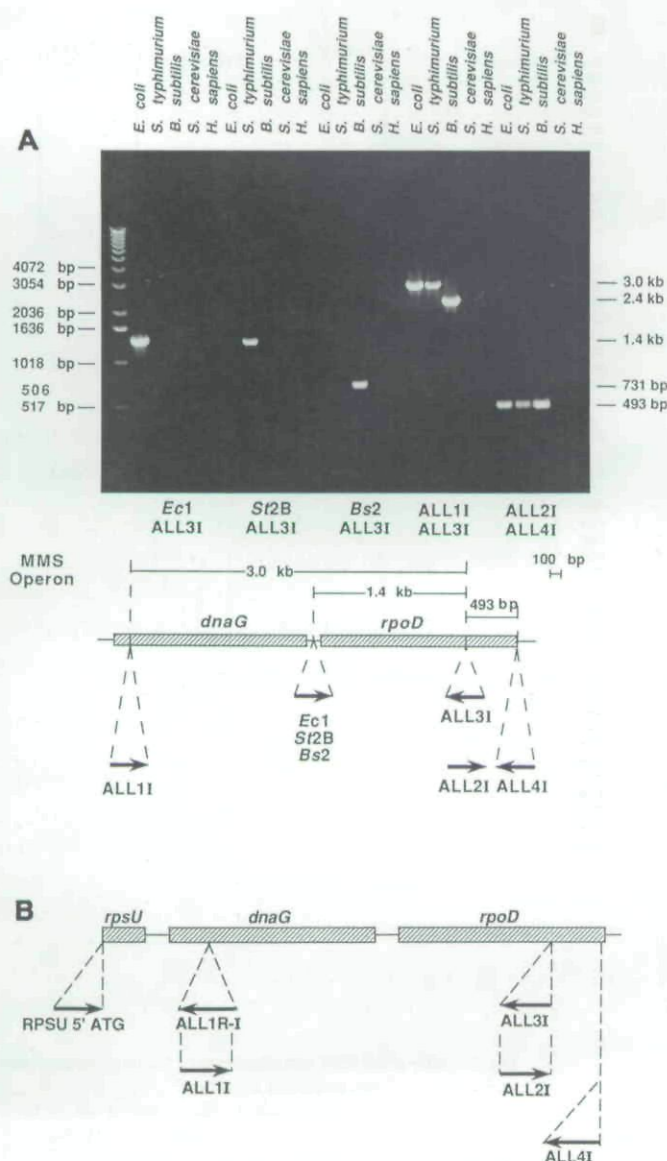


Fig. 1. MMS operon PCR amplifications with intergenic and intragenic primers.

A. Specific PCR products were obtained by amplification of MMS operon sequences in different bacterial species. Template DNAs include the eukaryotic species, *Saccharomyces cerevisiae* and *Homo sapiens*, as negative control samples. Each primer set listed below the gel was used with each of the five different template DNAs listed above the gel. The DNA molecular weight marker in lane 1 (far left) was a 1-kb ladder (BRL). Gel was 1% agarose-1 × Tris-acetate-EDTA and contained 0.5 µg of ethidium bromide per ml. B. Schematic diagram of oligonucleotide primers which match conserved intragenic sequences in *rpsU*-*dnaG*-*rpoD* MMS operon. Arrowheads represent the 3' ends of primers available for polymerase extension in the PCR. Arrow labels represent the names of primers whose sequences are listed in Table 1.

species, intergenic sequence primers amplified genus-specific PCR products of the expected size when used in combination with the conserved intragenic *rpoD* primer, ALL3I (Table 1; Fig. 1A).

The ability of oligonucleotide primers to recognize multiple conserved sequences from the MMS operon in different genera was then evaluated. Published coding sequences within *rpsU*, *dnaG*, and *rpoD* were analysed from different bacterial species (Wang and Doi, 1986; Burton *et al.*, 1983; Erickson *et al.*, 1985; Smiley *et al.*, 1982; Lupski *et al.*, 1983; Helmann and Chamberlin, 1988) and highly conserved domains within each gene were selected as primer binding sites. Conserved domains were selected as regions which contained a minimum number of nucleotide substitutions in sequences of 20–30

nucleotides with approximately 50% GC content. Since *rpsU* DNA sequences differed at only two positions between *E. coli* and *S. typhimurium*, primers which exactly matched the absolutely conserved amino terminus of *rpsU* (RPSU-5'ATG-BamHI, RPSU-5'ATG-EcoRI) were synthesized (Table 1). For *dnaG* and *rpoD*, degenerate primers were designed with inosine residues inserted at ambiguous positions (see the *Experimental procedures*; Table 1) to permit mismatches between less than perfectly conserved sequences in different species. PCR experiments with these primer pairs (Table 1), ALL1I, ALL3I and ALL2I, ALL4I, yielded amplification products of the expected sizes with *B. subtilis*, *E. coli*, and *S. typhimurium* genomic template DNAs (Fig. 1A). Oligonucleotide primers matching conserved intragenic sequences amplified

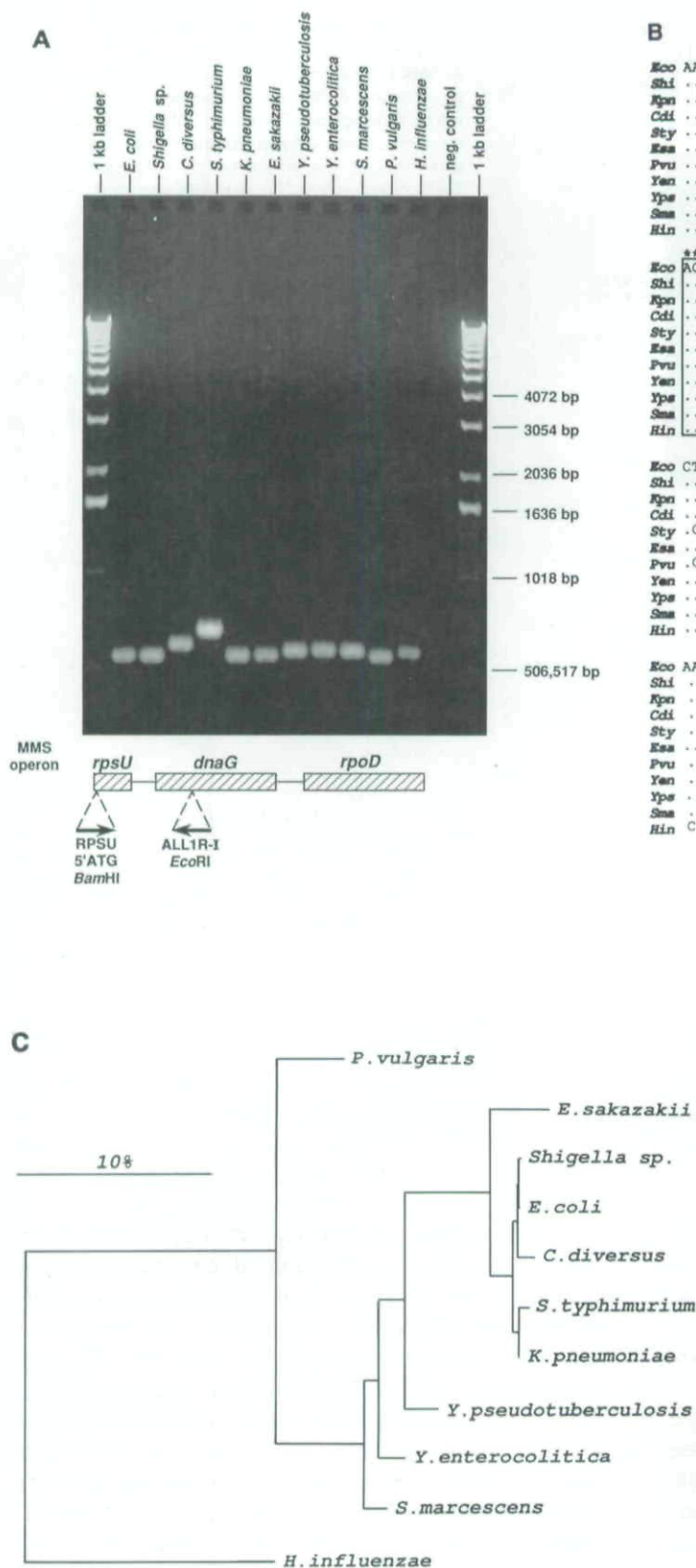


Fig. 2. *rpsU*-*dnaG* PCR amplifications of bacterial species and corresponding DNA sequence alignment.

A. *rpsU*-*dnaG* PCR amplifications with the primer set, RPSU-5'ATG-BamHI and ALL1R-I-EcoRI (Table 1; Fig. 1B). No template DNA was added to the negative control lane. The DNA molecular weight marker was a 1-kb ladder (BRL). **B.** Multiple DNA sequence alignment of *rpsU* coding sequences from 11 different bacterial species. Primary DNA sequence of codons 9-72 are shown in the top line for *E. coli* (*Eco*) (Lupski *et al.*, 1983; Burton *et al.*, 1983). Dots represent bases that are identical to *E. coli*. *Shi*, *Shigella* (GenBank L01760); *Kpn*, *K. pneumoniae* (L01757); *Cdi*, *C. diversus* (L01754); *Sty*, *S. typhimurium* (Erickson *et al.*, 1985); *Esa*, *E. sakazakii* (L01755); *Pvu*, *P. vulgaris* (L01758); *Yen*, *Y. enterocolitica* (L01761); *Yps*, *Y. pseudotuberculosis* (L01762); *Sma*, *S. marcescens* (L01759); *Hin*, *H. influenzae* (L01756). Dots above boxed sequence indicate *nutA* boxA. Asterisks above boxed sequence indicate *nutA* boxB. The *E. coli* and *S. typhimurium* DNA sequences were determined in this study and matched published sequences (Lupski *et al.*, 1983; Burton *et al.*, 1983; Erickson *et al.*, 1985). **C.** Phylogenetic tree based on *rpsU* DNA sequence alignment and derived by DeSoete least-squares distance matrix method. *H. influenzae* represents tree root.

similarly-sized products from all bacterial species with established MMS operon structures. Eukaryotic template DNAs failed to yield any PCR products (Fig. 1A). All other pairwise combinations of primers: RPSU-5'ATG plus ALL1R-I, RPSU-5'ATG plus ALL3I, RPSU-5'ATG plus ALL4I, and ALL1I plus ALL4I (Fig. 1B) gave PCR amplification products of the expected sizes with *E. coli*, *S. typhimurium*, and *B. subtilis* (data not shown). These results demonstrate that primer pairs which recognize conserved sequences in adjacent genes may be used to evaluate operon structure in different species.

Conservation of MMS operon structure in bacteria

The PCR-based strategy to evaluate operon conservation was next applied to bacteria in which the MMS operon structure had not been previously examined. Since the size of the MMS operon (>4 kb) approximates the limits of polymerase extension in the PCR, our strategy divided this operon into two sections. First, oligonucleotides matching the conserved amino termini of *rpsU* (RPSU-5'ATG-*Bam*HI, RPSU-5'ATG-*Eco*RI; Table 1) and *dnaG* (ALL1R-I-*Bam*HI, ALL1R-I-*Eco*RI; Table 1) on opposite strands were used to examine the contiguity of *rpsU* and *dnaG*. Secondly oligonucleotide primers matching the amino terminus of *dnaG* (ALL1I; Table 1; Fig. 1B) and the carboxy terminus of *rpoD* (ALL3I; Table 1; Fig. 1B) on opposite strands were used to evaluate whether the contiguous arrangement of these two genes was maintained.

Nine different Gram-negative bacterial species yielded apparent *rpsU*-*dnaG* PCR products of similar size (Fig. 2A). Four exceptions, *Citrobacter amalonaticus*, *Citrobacter diversus*, *Citrobacter freundii*, and *S. typhimurium*, yielded larger *rpsU*-*dnaG* amplification products which were subsequently shown to be due to the presence of the ERIC repetitive DNA element in the intergenic sequence (Figs 2A and 5B; see below). DNA sequences obtained from these PCR products confirm that *rpsU* lies adjacent to *dnaG* in these organisms (Fig. 2B; data not shown).

PCR amplification between the 5' end of *dnaG* (ALL1I) and region 2 of *rpoD* (ALL3I) yielded similarly-sized products greater than 3 kb in length in 10 Gram-negative species, indicating that *dnaG* appears to lie adjacent to *rpoD* in these organisms (Fig. 3). One enterobacterial species, *Yersinia pseudotuberculosis*, failed to yield a distinct ALL1I-ALL3I or ALL1I-ALL4I amplification product. However, single *rpsU*-*rpoD* PCR amplification products of the expected size were obtained from *Y. pseudotuberculosis*, as well as other species examined in Fig. 3, with primer pairs RPSU-5'ATG, ALL3I and RPSU-5'ATG, ALL4I (data not shown). Together with the *rpsU*-*dnaG* sequencing data (Fig. 2B; data not shown), these results indicate that the MMS operon appears to be conserved in *Y. pseudotuberculosis* and the ALL1I primer

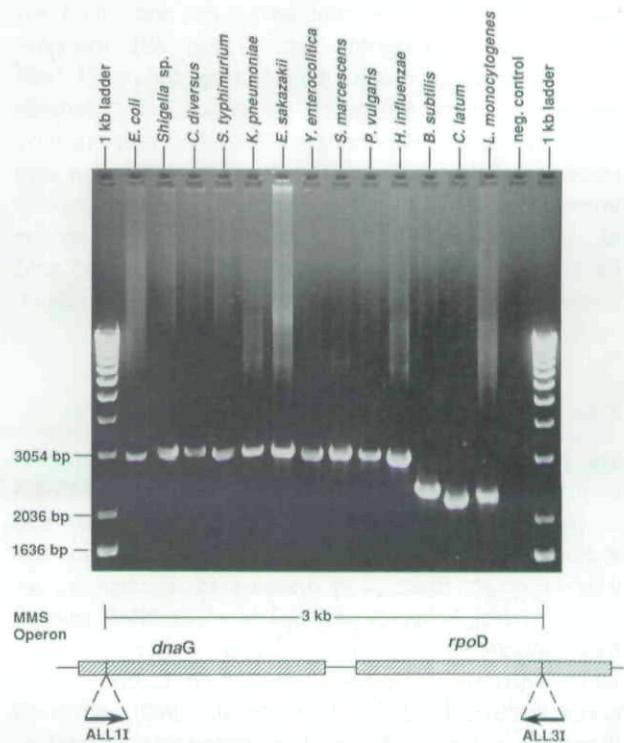


Fig. 3. *dnaG*-*rpoD* PCR amplifications of bacterial species. PCR amplifications were performed with the primer set, ALL1I and ALL3I (Table 1; Fig. 1B). No template DNA was added to the negative control lane.

binding site in *dnaG* may have diverged in this species. An apparent *dnaG*-*rpoD* PCR product which is smaller in size (~2.4 kb) appears in selected Gram-positive species (Fig. 3; see below). These *dnaG*-*rpoD* PCR products yielded prominent hybridization signals when probed with radio-labelled *E. coli* ALL1L-ALL3I PCR amplification product (data not shown), indicating the relatedness of these single PCR products (Fig. 3) with the *E. coli* *dnaG*-*rpoD* product. Therefore the contiguous operon structure, *rpsU*-*dnaG*-*rpoD*, appears conserved in enterobacteria and the related Gram-negative species, *Haemophilus influenzae*. The MMS operon from Gram-positive bacteria appears to lack *rpsU* (data not shown) but seems to maintain *dnaG* adjacent to *rpoD* (Fig. 3).

Conservation of *rpsU* coding sequence in Gram-negative bacteria

The *rpsU*-*dnaG* PCR products (Fig. 2A) from 11 different bacterial species were cloned and sequenced. Primer sequences were eliminated from analysis thereby enabling the inclusion of 192 bp of 216 total bp. No amino acid differences were detected in the 63 residues examined in the *rpsU* gene from 10 enterobacterial species (Table 2; Fig. 2B). One extreme example, *Proteus vulgaris*,

has 28 nucleotide differences with *E. coli* and yet retains 100% amino acid identity (Table 2; Fig. 2B). The phylogenetic tree generated (Fig. 2C) by the *rpsU* DNA sequence alignment (Fig. 2B) corresponds with enterobacterial species relationships established previously by ribosomal RNA sequence comparisons (Ochman and Wilson, 1987) and DNA hybridization studies (Sakazaki *et al.*, 1976). Only the most distantly-related, non-enteric bacterium, *H. influenzae*, reveals scattered amino acid substitutions, or ~80% identity, for the *rpsU* sequence under study (Table 2; Fig. 2B).

Conservation of the *nut_{eq}* antitermination sequence

The MMS operon putative transcription antitermination sequence *nut_{eq}* is located within the *rpsU* gene upstream of terminator *T₁* (Lupski *et al.*, 1983). The *E. coli nut_{eq}* site shows high sequence similarity to the lambda *nut* site (Lupski *et al.*, 1983) and appears to function as an antiterminator when cloned upstream of MMS operon rho-independent terminators *T₁* and *T₂* (Almond *et al.*, 1989). Known *nut* sequences including *nut_{eq}* contain two conserved motifs, *boxA* (Olson *et al.*, 1982) and *boxB* (Rosenberg *et al.*, 1978), which are important for antiterminator function (Das, 1992). The putative *nut_{eq}* sequences among 11 different Gram-negative bacterial species were analysed (Fig. 2B). The *E. coli nut_{eq}* *boxA* primary sequence contains the core GCTCT sequence found in *boxA* sequences in bacterial *rrn* operons and phages lambda, 21, and P22 (Friedman *et al.*, 1990). This sequence is completely conserved in nine of 11 bacteria analysed, with single base changes observed in *Enterobacter sakazakii* and *Haemophilus influenzae* (Fig. 2B). *boxB* secondary structure is conserved in all bacteria studied, with a GAAAAA loop in all cases except one (Fig. 2B). All changes in the stem preserve *boxB* secondary structure if G:U base pairing is allowed. The single nucleotide change in the loop occurs in *E. sakazakii*, with the third base position in the loop representing an A to G

transition (Fig. 2B). An analogous loop substitution previously failed to affect antitermination at the lambda *nut* site (Doelling and Franklin, 1989). Cautious interpretation of sequence similarity is necessary here because the *nut_{eq}* sequence is embedded within the highly conserved *rpsU* coding sequence.

Conservation of the transcription terminator *T₁* in Gram-negative bacteria

The *rpsU*–*dnaG* intergenic sequences were obtained from 11 bacterial species. A structure resembling a transcription terminator was identified in the *rpsU*–*dnaG* intergenic region of all MMS operons examined. Not only was the terminator stem-loop secondary structure conserved in all of these bacteria, but the primary sequence was conserved in 10 enterobacterial species (Fig. 4). Five bacteria had 100% sequence identity with the *E. coli T₁* stem loop at 31 positions, while three others had a single base insertion within this conserved stem, increasing the thermodynamic stability of the terminator (Fig. 4). The *T₁* stem loop of *P. vulgaris* has 83.3% identity with that of *E. coli* at 30 positions, with three of five substitutions occurring within the loop (Fig. 4). *H. influenzae*, which is distantly related to *E. coli*, had less apparent primary sequence similarity to the *E. coli T₁* sequence but did possess a more stable terminator stem-loop structure in its *rpsU*–*dnaG* intergenic sequence (Fig. 4). The DNA sequence flanking *T₁* in the *rpsU*–*dnaG* intergenic region was less conserved (data not shown).

Evolution of the *rpsU*–*dnaG* intergenic ERIC repetitive DNA element

A 126bp ERIC repetitive DNA sequence was identified in the *rpsU*–*dnaG* intergenic sequence of *S. typhimurium* (Sharples and Lloyd, 1990; Hulton *et al.*, 1991). Interestingly this ERIC element comprised the difference in length between the *rpsU*–*dnaG* intergenic sequences of *S.*

Table 2. *rpsU* primary sequence conservation.

Bacterial species	No. DNA changes	% DNA similarity	No. amino acid changes	% Amino acid similarity
<i>Escherichia coli</i>	—	—	—	—
<i>Shigella</i> sp.	0	100.0	0	100.0
<i>Klebsiella pneumoniae</i>	1	99.4	0	100.0
<i>Citrobacter diversus</i>	2	99.0	0	100.0
<i>Salmonella typhimurium</i>	2	99.0	0	100.0
<i>Enterobacter sakazakii</i>	9	95.3	0	100.0
<i>Yersinia pseudotuberculosis</i>	13	93.2	0	100.0
<i>Yersinia enterocolitica</i>	15	92.2	0	100.0
<i>Serratia marcescens</i>	15	92.2	0	100.0
<i>Proteus vulgaris</i>	28	85.4	0	100.0
<i>Haemophilus influenzae</i>	59	69.3	13	79.7

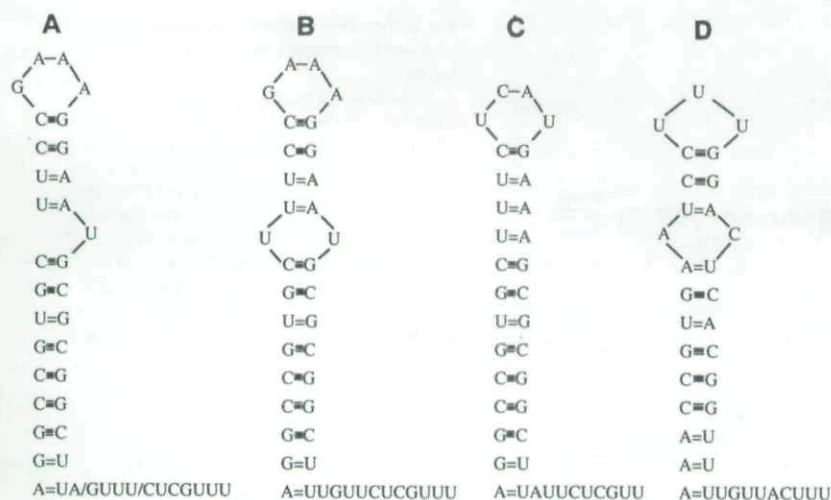


Fig. 4. T₁ terminator structures and standard free energy values for Gram-negative bacteria. T₁ primary sequences, stem loop structures, and free-energy values are depicted for 10 different enterobacterial species and *Haemophilus influenzae*. Standard free energies were calculated using the FOLD program of GCG (Zuker and Stiegler, 1981) for determination of thermodynamic stability of RNA molecules.

A. *E. coli*, *Shigella* sp., *C. diversus*, *S. typhimurium*, *K. pneumoniae*, and *E. sakazakii*. $\Delta G = -66.2 \text{ kJ mol}^{-1}$.

B. *Y. pseudotuberculosis*, *Y. enterocolitica*, and *S. marcescens*. $\Delta G = -69.3 \text{ kJ mol}^{-1}$.

C. *P. vulgaris*. $\Delta G = -71.4 \text{ kJ mol}^{-1}$.

D. *H. influenzae*. $\Delta G = -81.5 \text{ kJ mol}^{-1}$.

typhimurium (235bp) and *E. coli* (110bp), which completely lacks an ERIC sequence (Fig. 5A). Newly obtained *rpsU-dnaG* intergenic sequences from eight of 10 enterobacterial genera (Fig. 2B) lacked ERIC sequences (data not shown). The size of the *rpsU-dnaG* intergenic sequence of *C. diversus* is intermediate in length by PCR amplification (Figs 2A and 5B) and sequence analysis (data not shown). A partial ERIC sequence was identified in *C. diversus* (Fig. 5A) and this element represented the difference in size of the *rpsU-dnaG* intergenic region (data not shown). This ERIC remnant (ERIC') consists of the ends of ERIC, 35 and 21 bp, minus the 70bp conserved central core of the sequence (Fig. 5A). ERIC' from *C. diversus* shares 84% DNA sequence identity with the corresponding positions of the *S. typhimurium rpsU-dnaG* ERIC sequence. Nine clinical isolates of *C. diversus* yielded *rpsU-dnaG* PCR products of identical sizes, suggesting that the same partial ERIC element is present in all strains of this species (Fig. 5B). Interestingly, three *C. freundii* and four *C. amalonaticus* isolates yielded *rpsU-dnaG* PCR products which are similar in size to those obtained with six different *Salmonella* isolates (Fig. 5B). Subsequent DNA sequencing of these products indicated that an intact ERIC element is present between *rpsU* and *dnaG* in *C. freundii* and *C. amalonaticus* (data not shown). However, the *rpsU-dnaG* intergenic ERIC element in these two species lies in the reverse orientation relative to the ERIC element present in *S. typhimurium* (Fig. 5A).

Evolution of the *rpoD* gene in bacteria

Comparisons of amino acid sequences from *E. coli rpoD* (Burton *et al.*, 1981) and sigma factors from other species established the existence of four conserved domains in these genes (Gribskov and Burgess, 1986; Helmann and

Chamberlin, 1988; Lonetto *et al.*, 1992). Genetic (Gardella *et al.*, 1989; Siegle *et al.*, 1989) and biochemical (Dombroski *et al.*, 1992) evidence suggests that the most highly conserved domains, regions 2 and 4, interact with the -10 and -35 promoter boxes respectively.

A substantial size difference was observed initially between the *rpoD* genes of *E. coli* (Fig. 6A; (Burton *et al.*, 1981)) and *B. subtilis* (Fig. 6B; (Wang and Doi, 1986)). The larger *E. coli rpoD* gene contains a divergent sequence between conserved *rpoD* regions 1 and 2 (Fig. 6A) that is absent in *B. subtilis*, other Gram-positive bacterial species, and the cyanobacterium, *Anabaena* sp. (Fig. 6B). Between *E. coli* and *Pseudomonas aeruginosa*, this divergent *RpoD* sequence is identical at 36.4% of positions and contrasts with 92.4% identity in the carboxy terminus which includes conserved regions 2, 3, and 4 (Tanaka and Takahashi, 1991). PCR amplification between *dnaG* and *rpoD* yields smaller products from Gram-positive bacteria (~2.4 kb) (Fig. 3) which may correspond to the lack of this divergent *rpoD* sequence. *P. aeruginosa* and *Myxococcus xanthus* belong to the same phyla (Woese, 1987) as *E. coli* and contain larger *rpoD* genes with 700–800bp separating conserved regions 1 and 2 (Fig. 6A). Other enterobacterial species and *H. influenzae* yielded *dnaG-rpoD* PCR products of similar size as *E. coli* (Fig. 3) suggesting that these species also contain the larger *rpoD* gene.

In contrast to the polymorphic length of the 5' half of *rpoD*, the 3' half of *rpoD* consists of highly conserved domains 2 and 4 which are separated by conserved distances in known sigma factors (Helmann and Chamberlin, 1988; Lonetto *et al.*, 1992). PCR amplifications between regions 2 and 4 yielded specific products approximately 500bp in size in all species examined (Fig. 7), indicating conservation of distance between regions 2 and 4.

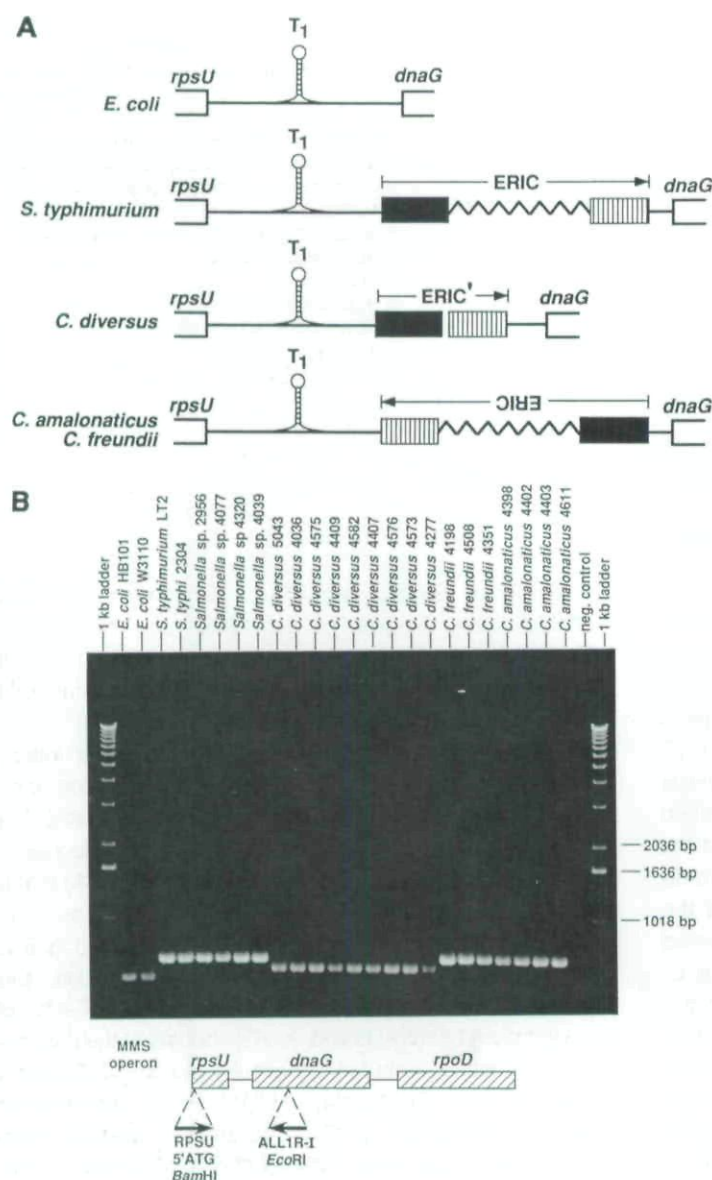


Fig. 5. Evolution of the ERIC element in the *rpsU*–*dnaG* intergenic sequence of enterobacteria.

A. Schematic diagram illustrates the evolution of the ERIC repetitive DNA sequences in the *rpsU*–*dnaG* intergenic sequence of *S. typhimurium*, *C. freundii*, *C. amalonaticus*, and *C. diversus*. The shaded boxes represent the ends of ERIC; the wavy line represents the central core sequence. The arrowheads indicate the orientation of ERIC. B. PCR amplifications between *rpsU* and *dnaG* with the primers RPSU-5'ATG-BamHI and ALL1R-I-EcoI in different enterobacterial strains. No template DNA was added to the negative control lane.

Discussion

The MMS operon represents a unique operon because it contains genes whose products initiate the synthesis of the three principal informational macromolecules of the cell, DNA, RNA, and protein. In this paper we have demonstrated that these genes are maintained in a contiguous operon structure in diverse bacterial species. Specific domains within coding regions and regulatory sequences display high levels of primary sequence conservation. In contrast, divergent coding sequences and extragenic repetitive DNA elements tolerate DNA rearrangements during evolution while maintaining the MMS operon structure.

The *rpsU*, *dnaG*, and *rpoD* genes represent conserved coding sequences for essential biochemical functions. *rpsU* illustrates dramatic differences in amino acid and nucleotide substitution rates which likely reflect effects of selective pressures on protein structure and function. Presumably few amino acid differences in S21 can be tolerated without disturbing its role in translation initiation. We have demonstrated that the 3' end of *rpoD* contains highly conserved regions 2 and 4 which are separated by a fixed distance in different species. This fixed distance may reflect its role in promoter recognition of conserved sequence motifs (–10 and –35 boxes) which are separated by fixed distances in the DNA (Hawley and

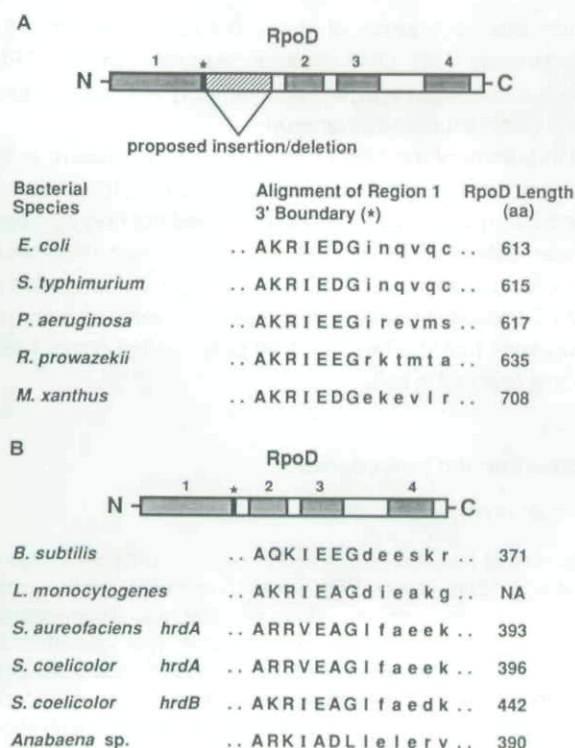


Fig. 6. Evolution of RpoD divergent sequences in bacteria.

A. Schematic diagram illustrates the conserved regions 1, 2, 3, and 4 in RpoD from proteobacteria and the corresponding amino acid sequence alignment. *E. coli* (Burton *et al.*, 1981), *S. typhimurium* (Erickson *et al.*, 1985), *P. aeruginosa* (Tanaka and Takahashi, 1991), *M. xanthus* (Inouye, 1990).

B. Schematic diagram and the corresponding amino acid sequence alignment of RpoD from Gram-positive bacteria and cyanobacteria. *B. subtilis* (Wang and Doi, 1986), *L. monocytogenes* (L. Katz, personal communication), *S. aureofaciens* (*hrdA*, GenBank accession No. M90410), *S. coelicolor* (*hrdA*, GenBank accession No. X52980; *hrdB*, GenBank accession No. X52983), *Anabaena sp.* (Brahams and Haselkorn, 1991). Multiple amino acid alignment of the region 1 boundary junction of RpoD is shown. Asterisk indicates the location of alignment. Left side of alignment (upper case letters) represents carboxy-terminal end of conserved region 1. Right side of alignment (lower case letters) signifies amino-terminal end of divergent region between regions 1 and 2. Dots indicate continuation of RpoD amino acid sequence on either side of this alignment. NA, indicates that the complete amino acid sequence is not available.

McClure, 1983). Region 2 of RpoD can be aligned by primary amino acid sequence with the human transcription initiation factor RAP30 (Sopta *et al.*, 1989), indicating extreme functional conservation of this domain. Therefore both primary sequence and gene structure are preserved in these essential genes.

Transcription terminator T_1 partially accounts for the extreme difference in expression between *rpsU* and *dnaG* (estimated ratio = 1000:1) (Lupski and Godson, 1989). T_1 of the *E. coli* MMS operon (Smiley *et al.*, 1982) is a rho-independent terminator in the *rpsU*–*dnaG* intergenic sequence that has previously been demonstrated to have transcription termination function (Lupski *et al.*, 1983;

1984; Burton *et al.*, 1983; Katayama *et al.*, 1989). Recent studies indicate that both primary sequence and secondary structure comprise important elements of transcription terminator function. In lambda, specific nucleotide substitutions which theoretically enhance the thermodynamic stability of the stem-loop structure can diminish the ability of this sequence to function as an effective transcription terminator (Cheng *et al.*, 1991). The *hly* operon in the distantly related enterobacterial species *E. coli* and *P. vulgaris* contains an intergenic rho-independent terminator which is identical at 37 of 46 positions (Koronakis *et al.*, 1988). The results of the present study demonstrate that despite considerable divergence in the surrounding *rpsU*–*dnaG* intergenic sequence, transcription terminator T_1 displays a high degree of primary sequence conservation in enteric bacteria. These data support a role for primary sequence, in addition to secondary structure, in terminator function.

In contrast to conserved coding and regulatory sequences, DNA rearrangements within non-essential coding sequences may occur if essential gene function is unaffected. The divergent sequence in *rpoD* between

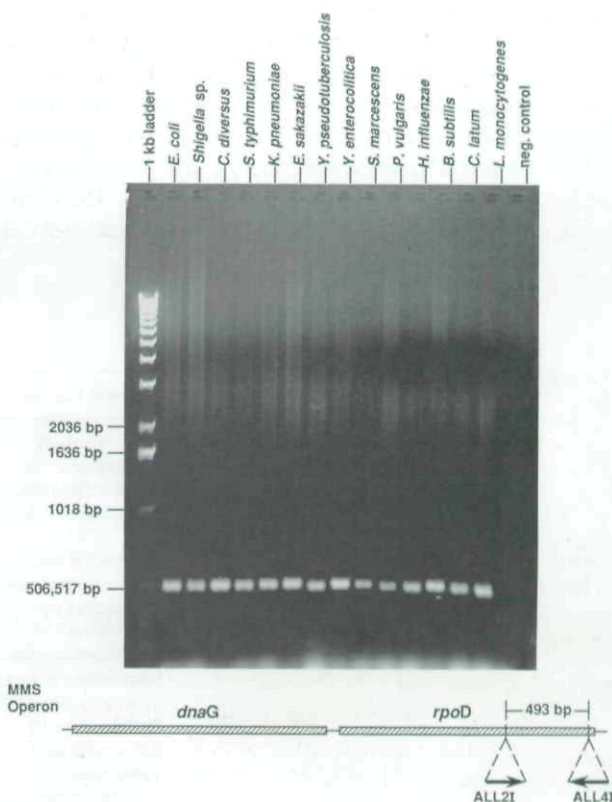


Fig. 7. PCR amplifications within *rpoD* genes between conserved regions 2 and 4. Intra-*rpoD* PCR amplification between conserved carboxy-terminal domains, regions 2 and 4. PCR primer set included ALL2I and ALL4I (Table 1; Fig. 1B). No template DNA was added to the negative control lane.

conserved regions 1 and 2 appears dispensable by DNA sequence analysis. The lack of this divergent sequence in the Gram-positive lineage suggests that it is non-essential for sigma factor function. It is likely that an insertion or deletion event occurred in this region of *rpoD* during bacterial evolution. An enrichment of amino acid substitutions was described previously in this same RpoD region from *S. typhimurium* (Erickson *et al.*, 1985). Interestingly, the *E. coli rpoD800* allele has a 42 bp deletion within region 1 adjacent to the site of this evolutionary rearrangement (Hu and Gross, 1983; Helmann and Chamberlin, 1988).

ERIC sequences (Hulton *et al.*, 1991), otherwise known as intergenic repeat units (IRUs) (Sharples and Lloyd, 1990), show high levels of primary sequence conservation among different loci in different species (Sharples and Lloyd, 1990; Hulton *et al.*, 1991). ERIC-like elements appear to be conserved in diverse bacterial species, such as *Klebsiella pneumoniae*, *Rhizobium meliloti*, *Deinococcus radiophilus* and *Xenorhabdus luminescens* (Versalovic *et al.*, 1991a; Lupski and Weinstock, 1992; Meighen and Szittner, 1992). We describe the first example of a truncated ERIC element at the same location as an intact ERIC element in another species. An insertion/deletion event has apparently occurred which results in an ERIC remnant (ERIC') lacking the central 70 bp core in *C. diversus* (Fig. 5A). *C. freundii* and *C. amalonaticus* are more closely related to each other than to *C. diversus* by biochemical profiles (Farmer, 1981). Interestingly, these two species share an inverted, intact ERIC sequence (Fig. 5A) which probably arose secondary to the inverted repeat structure of ERIC (Sharples and Lloyd, 1990). The complete absence of ERIC in the MMS operon *rpsU-dnaG*

intergenic sequence of most enterobacteria and the occurrence of DNA rearrangements within ERIC sequences argue against an essential role in gene regulation or structure in this operon.

In summary the MMS operon structure appears to be conserved in diverse bacterial species. Both intergenic and intragenic sequences have evolved not only by single base substitutions but also by DNA rearrangements involving blocks of contiguous sequence. Conservation of MMS operon structure during millions of years of evolution suggests that the MMS operon plays an important functional role in the cell.

Experimental procedures

Isolation of genomic DNA

Sources of bacterial strains and/or genomic DNA are shown in Table 3. Chromosomal DNA was prepared by the method published previously for Gram-negative and Gram-positive bacteria (Versalovic *et al.*, 1991a). DNA was quantified by spectrofluorometry (Versalovic *et al.*, 1991a). Genomic DNA samples from all bacterial species examined were checked for *E. coli* genomic DNA contamination by PCR screening with oligonucleotide primers *Ec1* and *ALL3I* (Table 1) and 36 cycles of amplification.

Oligonucleotide synthesis and design

Oligonucleotides were synthesized using a Model 380B DNA synthesizer (Applied Biosystems Inc.) located in the Nucleic Acids Core Facility in the Institute for Molecular Genetics, Baylor College of Medicine. DNA sequence information from published sequence data (Wang and Doi, 1986; Burton *et al.*, 1983; Erickson

Table 3. Bacterial strains used as DNA sources.

Bacterial species	Strain	Source
<i>Bacillus subtilis</i>	DB2	R. Doi
<i>Caryophanon latum</i>		New England Biolabs, Inc.
<i>Citrobacter amalonaticus</i>	4398, 4402, 4403, 4611	C. Woods
<i>Citrobacter diversus</i>	4036, 4277, 4407, 4409, 4575, 4576, 4582, 5043*	C. Woods
<i>Citrobacter freundii</i>	4198, 4351, 4508	C. Woods
<i>Enterobacter sakazakii</i>	4585*	E. Mason
<i>Escherichia coli</i>	HB101*	M. Winkler
<i>Escherichia coli</i>	W3110	Y. Kohara
<i>Haemophilus influenzae</i>	1775*	E. Mason
<i>Klebsiella pneumoniae</i>	*	New England Biolabs, Inc.
<i>Listeria monocytogenes</i>	5032	E. Mason
<i>Proteus vulgaris</i>	ATCC 13315*	New England Biolabs, Inc.
<i>Salmonella</i> sp.	2956, 4039, 4077, 4320	E. Mason
<i>Salmonella typhi</i>	2304	E. Mason
<i>Salmonella typhimurium</i>	LT2*	B. D. Erickson
<i>Serratia marcescens</i>	*	New England Biolabs, Inc.
<i>Shigella</i> sp.	170*	E. Mason
<i>Yersinia enterocolitica</i>	8081v.08*	V. Miller
<i>Yersinia pseudotuberculosis</i>	P-;inv::Tet*	R. Isberg

* Indicates sources of *rpsU-dnaG* sequence information shown in Figs 2B and 4.

et al., 1985; Smiley *et al.*, 1982; Lupski *et al.*, 1983; Helmann and Chamberlin, 1988) was used for primer design. Inosines (I) were incorporated into ambiguous positions of oligonucleotide primers matching conserved coding sequences. Recognition sequences of specific restriction enzymes (e.g. *Bam*HI, *Eco*RI) were added to the 5' ends of RPSU 5'ATG and ALL1R-I. Dried oligonucleotide pellets were resuspended in water (Fisher) and quantified by UV-VIS spectrophotometry with absorption measured at 260 nm.

PCR conditions

PCR amplification reactions (Versalovic *et al.*, 1991a) were performed in an automated thermal cycler (Perkin-Elmer/Cetus DNA Thermal Cycler 9600) with different oligonucleotide primer sets (Table 1) by initial denaturation (94°C, 9 min) followed by 30 cycles of denaturation, annealing, and extension with a single final extension. The temperature and time of the last two steps varied according to the primers utilized (Table 1). Five microlitres of each PCR reaction were electrophoresed in 1% agarose gels containing 1× TAE (Tris acetate-EDTA; (Maniatis *et al.*, 1982)), 0.5 µg ml⁻¹ ethidium bromide (EtBr). Gels were photographed with 60 s exposure to Polaroid Type 55 film.

Cloning of *rpsU*-*dnaG* PCR products prior to DNA sequencing

Oligonucleotide primers matching *rpsU* (RPSU 5'ATG) and *dnaG* (ALL1R-I) coding sequences were modified by the addition of *Eco*RI (RPSU-5'ATG-*Eco*RI, ALL1R-I-*Eco*RI) or *Bam*HI (RPSU-5'ATG-*Bam*HI, ALL1R-I-*Bam*HI) sites at the 5' ends (Table 1). PCR products obtained after amplification of template DNAs (RPSU-5'ATG-*Eco*RI, ALL1R-I-*Bam*HI or RPSU-5'ATG-*Bam*HI, ALL1R-I-*Eco*RI) were immediately extracted once with 100 µl of phenol and once with 100 µl of chloroform (Fisher). Aqueous phase was precipitated at -20°C for 1 h in 0.33 M sodium acetate and 2.5 volumes of ethanol (Maniatis *et al.*, 1982). Precipitated DNA was resuspended in 15 µl water (Millipore). PCR-amplified DNA and pTZ19R (Mead *et al.*, 1986) vector DNA were digested by *Eco*RI (25 U; Boehringer Mannheim) and by *Bam*HI (20 U; Boehringer Mannheim) and gel-purified. DNA was precipitated and resuspended in water as described above. Resuspended DNA (1–3 µl) was ligated into *Eco*RI-, *Bam*HI-digested pTZ19R with T4 DNA ligase (1 U; Boehringer Mannheim) (Maniatis *et al.*, 1982) at 16°C, 12 h.

Ligated DNA (1 µl) was added to transformation-competent (Chung *et al.*, 1989) XL1-Blue *E. coli* cells (Stratagene). Cells were plated on MacConkey agar (Difco) containing 1% supplemented lactose (Sigma), 10 µg ml⁻¹ tetracycline (Sigma), and 80 µg ml⁻¹ timentin (Beecham Laboratories). White colonies indicated the presence of insert (Jennings and Beacham, 1990) in the multiple cloning site of pTZ19R and were selected for plasmid miniprep (Zhou *et al.*, 1990) analysis.

DNA sequencing

DNA sequencing was performed by the Nucleic Acids Core Facility in the Institute for Molecular Genetics at Baylor College of Medicine with automated fluorescent DNA sequencing technology (Smith *et al.*, 1986). The *rpsU*, *rpsU*-*dnaG* intergenic, and *dnaG* DNA sequences were directly stored by the Model 373A

sequencing station (Applied Biosystems, Inc.). Sequences were first inspected by visual examination of hard-copy chromatograms containing the actual DNA sequence output from the sequencing station. Sequences were imported into the Eugene molecular biology analysis system (Lawrence *et al.*, 1989). Three independent clones from each species contained the *rpsU*-*dnaG* insert in each of two orientations (sixfold coverage for each species examined). Final DNA sequences contained *rpsU* codons 9–72 (including termination codon), the entire *rpsU*-*dnaG* intergenic region, and the first five codons of *dnaG*.

DNA sequence comparisons

DNA sequences were initially aligned and compared with the JHEIN program (Hein, 1990) within Eugene (Lawrence *et al.*, 1989). Sequences were exported to a Macintosh II (Apple Computer Inc.) microcomputer and analysed with GENEWORKS version 2.0 (Intelligence Inc.). Multiple DNA sequence alignments were generated within GENEWORKS by the CLUSTAL method (Higgins and Sharp, 1988). Sequence alignments were subsequently edited manually. Alignments were next exported to the Genetic Data Environment (GDE) version 2.0. Phylogenetic trees were generated using the DeSoete method within GDE. Terminator T₁ and the antitermination site (*nut*_{eq}) RNA sequences were first analysed with HAIRPINS in Eugene. These sequences were next exported to Genetics Computer Group (GCG Inc.) version 7.0 (GCG Inc.) (Devereux *et al.*, 1984) in a VAX/VMS minicomputer (Digital Equipment Corp.). RNA secondary structures were analysed with the FOLD program (Zuker and Stiegler, 1981) within GCG.

Acknowledgements

We are grateful to the individuals who provided strains and/or genomic DNA. Thanks are due to the New England Biolabs (NEB) and, particularly Marta Meda and Joe Bonventre, for their co-operation in obtaining samples. We are grateful to Dr Richard Gibbs of the Nucleic Acids Core Facility in the Institute for Molecular Genetics, Baylor College of Medicine. Thanks are also extended for the use of the computing facilities at the Molecular Biology Computational Resource (MBCR), Baylor College of Medicine, and to Dr David Wheeler for his advice on DNA sequence analysis. We appreciate the critical reviews of Drs Zachary Burton, Leonard Katz, and Andrew Pease. This research was supported by a research grant from Abbott Laboratories to J.R.L. J.V. is supported by the National Institutes of Health Medical Scientist Training Program at Baylor College of Medicine and the National Institutes of Health Minority Predoctoral Fellowship Program (MPFP) (1F31GM14601-01 SRC-7). J.R.L. acknowledges support from the PEW Scholars Program in Biomedical Sciences.

References

- Almond, N., Yajnik, V., Svec, P., and Godson, G.N. (1989) An *Escherichia coli* *cis*-acting antiterminator sequence: the *dnaG* *nut* site. *Mol Gen Genet* **216**: 195–203.
- Backendorf, C., Ravensbergen, C.J.C., Van der Plas, J., van Boom, J.H., Veenman, G., and Van Duin, J. (1981) Basepairing potential of the 3' terminus of 16S RNA: dependence on the

- functional state of the 30S subunit and the presence of protein S21. *Nucl Acids Res* **9**: 1425–1444.
- Brahamsha, B., and Haselkorn, R. (1991) Isolation and characterization of the gene encoding the principal sigma factor of the vegetative cell RNA polymerase from the cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **173**: 2442–2450.
- Burton, Z., Burgess, R.R., Lin, J., Moore, D., Holder, S., and Gross, C.A. (1981) The nucleotide sequence of the cloned *rpoD* gene for the RNA polymerase sigma subunit from *E. coli* K12. *Nucl Acids Res* **9**: 2889–2903.
- Burton, Z.F., Gross, C.A., Watanabe, K.F., and Burgess, R.R. (1983) The operon that encodes the sigma subunit of RNA polymerase also encodes ribosomal protein S21 and DNA primase in *E. coli* K12. *Cell* **32**: 335–349.
- Cheng, S.-W.C., Lynch, E.C., Leason, K.R., Court, D.L., Shapiro, B.A., and Friedman, D.I. (1991) Functional importance of sequence in the stem-loop of a transcription terminator. *Science* **254**: 1205–1207.
- Chung, C.T., Niemela, S.L., and Miller, R.H. (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc Natl Acad Sci USA* **86**: 2172–2175.
- Das, A. (1992) How the phage lambda *N* gene product suppresses transcription termination: Communication of RNA polymerase with regulatory proteins mediated by signals in nascent RNA. *J Bacteriol* **174**: 6711–6716.
- Devereux, J., Haeberli, P., and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Research* **12**: 387–395.
- Doelling, J.H., and Franklin, N.C. (1989) Effects of all single base substitutions in the loop of *boxB* on antitermination of transcription by bacteriophage λ 's N protein. *Nucl Acids Res* **17**: 5565–5577.
- Dombroski, A.J., Walter, W.A., Record, Jr, M.T., Siegele, D.A., and Gross, C.A. (1992) Polypeptides containing highly conserved regions of transcription initiation factor σ^{70} exhibit specificity of binding to promoter DNA. *Cell* **70**: 501–512.
- Erickson, B.D., Burton, Z.F., Watanabe, K.K., and Burgess, R.R. (1985) Nucleotide sequence of the *rpsU-dnaG-rpoD* operon from *Salmonella typhimurium* and a comparison of this sequence with the homologous operon of *Escherichia coli*. *Gene* **40**: 67–78.
- Farmer III, J.J. (1981) The genus *Citrobacter*. In *The Prokaryotes*. Starr, M.P., Stolp, H., Truper, H.G., Balows, A., and Schlegel, H.G. (eds). Berlin: Springer-Verlag, pp. 1140–1159.
- Friedman, D.I., Olson, E.R., Johnson, L.L., Alessi, D., and Craven, M.G. (1990) Transcription-dependent competition for a host factor: the function and optimal sequence of the phage λ *boxA* transcription antitermination signal. *Genes Dev* **4**: 2210–2222.
- Gardella, T., Moyle, H., and Susskind, M.M. (1989) A mutant *Escherichia coli* σ^{70} subunit of RNA polymerase with altered promoter specificity. *J Mol Biol* **206**: 579–590.
- Gribskov, M., and Burgess, R.R. (1986) Sigma factors from *E. coli*, *B. subtilis*, phage SPO1, and phage T4 are homologous proteins. *Nucl Acids Res* **14**: 6745–6763.
- Grompe, M., Versalovic, J., Koeuth, T., and Lupski, J.R. (1991) Mutations in the *Escherichia coli* *dnaG* gene suggest coupling between DNA replication and chromosome partitioning. *J Bacteriol* **173**: 1268–1278.
- Hawley, D.K., and McClure, W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl Acids Res* **11**: 2237–2255.
- Hein, J. (1990) Unified approach to alignment and phylogenies. *Meth Enzymol* **183**: 626–645.
- Helmann, J.D., and Chamberlin, M.J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839–872.
- Higgins, D.G., and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237–244.
- Hu, J., and Gross, C.A. (1983) Marker rescue with plasmids bearing deletions in *rpoD* identifies a dispensable part of *E. coli* sigma factor. *Mol Gen Genet* **191**: 492–498.
- Hulton, C.S.J., Higgins, C.F., and Sharp, P.M. (1991) ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. *Mol Microbiol* **5**: 825–834.
- Inouye, S. (1990) Cloning and DNA sequence of the gene coding for the major sigma factor from *Myxococcus xanthus*. *J Bacteriol* **172**: 80–85.
- Jennings, M.P., and Beacham, I.R. (1989) MacConkey agar as alternative to Xgal in the detection of recombinant plasmids. *Biotechniques* **7**: 1082.
- Katayama, T., Murakami, Y., Wada, C., Ohmori, H., Yura, T., and Nagata, T. (1989) Genetic suppression of a *dnaG* mutation in *Escherichia coli*. *J Bacteriol* **171**: 1465–1491.
- Konigsberg, W., and Godson, G.N. (1983) Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. *Proc Natl Acad Sci USA* **80**: 687–691.
- Koronakis, V., Cross, M., and Hughes, C. (1988) Expression of the *E. coli* hemolysin secretion gene *hlyB* involves transcript anti-termination within the *hly* operon. *Nucl Acids Res* **16**: 4789–4800.
- Kyriatsoulis, A., Maly, P., Greuer, B., Brimacombe, R., Stoffler, G., Frank, R., and Blocker, H. (1986) RNA-protein cross-linking in *Escherichia coli* ribosomal subunits: localization of sites on 16S RNA which are cross-linked to proteins S17 and S21 by treatment with 2-iminothiolane. *Nucl Acids Res* **14**: 1171–1186.
- Lark, K. (1972) Genetic control over the initiation of the synthesis of the short deoxynucleotide chains in *E. coli*. *Nature New Biol* **240**: 237–240.
- Lawrence, C.B., Shalom, T.Y., and Honda, S. (1989) EuGene: a software package for nucleotide and protein sequence analysis for UNIX systems. *Mol Biology Information Resource*.
- Lonetto, M., Gribskov, M., and Gross, C.A. (1992) The σ^{70} family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**: 3843–3849.
- Lupski, J.R., and Godson, G.N. (1984) The *rpsU-dnaG-rpoD* macromolecular synthesis operon of *E. coli*. *Cell* **39**: 251–252.
- Lupski, J.R., and Godson, G.N. (1989) DNA-DNA, and DNA-RNA-protein: Orchestration by a single complex operon. *BioEssays* **10**: 152–157.
- Lupski, J.R., Smiley, B.L., and Godson, G.N. (1983) Regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon and the initiation of DNA replication in *Escherichia coli* K-12. *Mol Gen Genet* **189**: 48–57.
- Lupski, J.R., Ruiz, A.A., and Godson, G.N. (1984) Promotion, termination, and anti-termination in the *rpsU-dnaG-rpoD* macromolecular synthesis operon in *E. coli* K-12. *Mol Gen Genet* **195**: 391–401.
- Lupski, J.R., and Weinstock, G.M. (1992) Short, interspersed repetitive DNA sequences in prokaryotic genomes. *J Bacteriol* **174**: 4525–4529.
- Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Marks, G.L., Winkler, H.H., and Wood, D.O. (1992) Isolation and characterization of the gene coding for the major sigma factor

- of *Rickettsia prowazekii* DNA-dependent RNA polymerase. *Gene* **121**: 155–160.
- Marks, G.L., and Wood, D.O. (1993) Characterization of the gene coding for the *Rickettsia prowazekii* DNA primase analogue. *Gene* **123**: 121–125.
- Mead, D.A., Szczesna-Skorupa, E., and Kemper, B. (1986) Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Prot Eng* **1**: 67–76.
- Meighen, E.A., and Szittner, R.B. (1992) Multiple repetitive elements and organization of the *lux* operons of luminescent terrestrial bacteria. *J Bacteriol* **174**: 5371–5381.
- Melancon, P., LeClerc, D., Destroismaisons, N., and Brakier-Gingras, L. (1990) The anti-Shine-Dalgarno region in *Escherichia coli* 16S ribosomal RNA is not essential for the correct selection of translational starts. *Biochem* **29**: 3402–3407.
- Nakamura, Y. (1984) Amber *dnaG* mutation exerting a polar effect on the synthesis of RNA polymerase sigma factor in *Escherichia coli*. *Mol Gen Genet* **196**: 179–182.
- Nesin, M., Lupski, J.R., and Godson, G.N. (1988) Role of the 5' upstream sequence and tandem promoters in regulation of the *rpsU-dnaG-rpoD* macromolecular synthesis operon. *J Bacteriol* **170**: 5759–5764.
- Ochman, H., and Wilson, A.C. (1987) Evolutionary history of enteric bacteria. In *Escherichia coli and Salmonella typhimurium*. Ingraham, J.L., Brooks Low, K., Magasanik, B., Schaechter, M., and Umberger, H.E. (eds). Washington, D.C.: American Society for Microbiology, pp. 1649–1654.
- Olson, E.R., Flamm, E.L., and Friedman, D.I. (1982) Analysis of *nutR*: a region of phage lambda required for antitermination of transcription. *Cell* **31**: 61–70.
- Peacock, S., Lupski, J.R., Godson, G.N., and Weissbach, H. (1985) *In vitro* stimulation of *Escherichia coli* RNA polymerase sigma subunit synthesis by NusA protein. *Gene* **33**: 227–234.
- Rosenberg, M., Court, D., Shimatake, H., and Brady, C. (1978) The relationship between function and DNA sequence in an intercistronic regulatory region of λ . *Nature* **272**: 414–422.
- Sakazaki, R., Tamura, K., Johnson, R., and Colwell, R.R. (1976) Taxonomy of some recently described species in the family *Enterobacteriaceae*. *Int J Syst Bacteriol* **26**: 158–179.
- Sharples, G.J., and Lloyd, R.G. (1990) A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. *Nucl Acids Res* **18**: 6503–6508.
- Siegele, D.A., Hu, J.C., Walter, W.A., and Gross, C.A. (1989) Altered promoter recognition by mutant forms of the σ^{70} subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* **206**: 591–603.
- Smiley, B.L., Lupski, J.R., Svec, P.S., McMacken, R., and Godson, G.N. (1982) Sequences of the *Escherichia coli dnaG* primase gene and regulation of its expression. *Proc Natl Acad Sci USA* **79**: 4550–4554.
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* **321**: 674–679.
- Sopta, M., Burton, Z.F., and Greenblatt, J. (1989) Structure and associated DNA-helicase activity of a general transcription initiation factor that binds to RNA polymerase II. *Nature* **341**: 410–414.
- Tanaka, K., and Takahashi, H. (1991) Cloning and analysis of the gene (*rpoDA*) for the principal σ factor of *Pseudomonas aeruginosa*. *Biochim Biophys Acta* **1089**: 113–119.
- van der Ende, A., Baker, T.A., Ogawa, T., and Kornberg, A. (1985) Initiation of enzymatic replication at the origin of the *Escherichia coli* chromosome: primase as the sole priming enzyme. *Proc Natl Acad Sci USA* **82**: 3954–3958.
- Versalovic, J., Koeuth, T., and Lupski, J.R. (1991a) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucl Acids Res* **19**: 6823–6831.
- Versalovic, J., Koeuth, T., McCabe, E.R.B., and Lupski, J.R. (1991b) Use of the polymerase chain reaction for physical mapping of *Escherichia coli* genes. *J Bacteriol* **173**: 5253–5255.
- Wang, L.-F., and Doi, R.H. (1986) Nucleotide sequence and organization of *Bacillus subtilis* RNA polymerase major sigma operon. *Nucl Acids Res* **14**: 4293–4307.
- Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev* **51**: 221–271.
- Zhou, C., Yang, Y., and Jong, A.Y. (1990) Mini-prep in ten minutes. *Biotechniques* **8**: 172–173.
- Zuker, M., and Stiegler, P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucl Acids Res* **9**: 133–148.

Copyright of Molecular Microbiology is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.