RNA (1999), *5*:751–763. Cambridge University Press. Printed in the USA.

16S ribosomal RNA pseudouridine synthase RsuA of *Escherichia coli*: Deletion, mutation of the conserved Asp102 residue, and sequence comparison among all other pseudouridine synthases

JOEL CONRAD,[1] LINGHAO NIU,[1,3] KENNETH RUDD,[1] BYRON G. LANE,[2] and JAMES OFENGAND[1]

[1]Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, Florida 33101, USA
[2]Department of Biochemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

ABSTRACT

The gene for RsuA, the pseudouridine synthase that converts U516 to pseudouridine in 16S ribosomal RNA of *Escherichia coli*, has been deleted in strains MG1655 and BL21/DE3. Deletion of this gene resulted in the specific loss of pseudouridine516 in both cell lines, and replacement of the gene *in trans* on a plasmid restored the pseudouridine. Therefore, *rsuA* is the only gene in *E. coli* with the ability to produce a protein capable of forming pseudouridine516. There was no effect on the growth rate of *rsuA*$^-$ MG1655 either in rich or minimal medium at either 24, 37, or 42 °C. Plasmid rescue of the BL21/DE3 *rsuA*$^-$ strain using pET15b containing an *rsuA* gene with aspartate102 replaced by asparagine or threonine demonstrated that neither mutant was active in vivo. This result supports a role for this aspartate, located in a unique GRLD sequence in this gene, at the catalytic center of the synthase. Induction of wild-type and the two mutant synthases in strain BL21/DE3 from genes in pET15b yielded a strong overexpression of all three proteins in approximately equal amounts showing that the mutations did not affect production of the protein in vivo and thus that the lack of activity was not due to a failure to produce a gene product. Aspartate102 is found in a conserved motif present in many pseudouridine synthases. The conservation and distribution of this motif in nature was assessed.

Keywords: catalytic center; growth rate; overexpression; plasmid rescue; sequence alignment; superfamilies

INTRODUCTION

The pseudouridine (5-ribosyl uracil; Ψ) residues of ribosomal RNA are one of the many poorly understood features of the ribosome. In the large subunit of all those organisms examined, ranging from bacteria to human, they cluster in a nonrandom manner around the general region of the peptidyl transferase center (Ofengand & Bakin, 1997) which in turn is found at the interface region between the two subunits of the ribosome. This is the case even though their total number varies widely from species to species (Ofengand & Bakin, 1997). To obtain an understanding of the role of Ψ in the ribosome, it is first necessary to choose a well-characterized ribosome with a manageable number of Ψ residues. *Escherichia coli* meets these specifications, as it is the most well-studied ribosome known and has only 10 Ψ residues, one at position 516 in the SSU RNA (Bakin et al., 1994a) and nine at positions 746, 955, 1911, 1915, 1917, 2457, 2504, 2580, and 2605 in the LSU RNA (Bakin & Ofengand, 1993; Bakin et al., 1994b). The classical way to determine the function of something is to block its occurrence and ask what happens. In the case of Ψ, this is easily accomplished because the various Ψ in *E. coli* rRNA are made by a set of specific synthases (Ofengand & Fournier, 1998). Deletion of the genes for each of these enzymes in turn results in the absence of a specific Ψ (or group of Ψ) whose effect on the physiology and metabolism of the cell can then be determined.

Reprint requests to: James Ofengand, Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, P.O. Box 016129, Miami, Florida 33101, USA; e-mail: jofengan@molbio.med.miami.edu.

[3]Present address: Program in Molecular Pharmacology and Therapeutics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

Abbreviations: Ψ, pseudouridine; CMC, *N*-cyclohexyl-*N′*-β-(4-methylmorpholinium)ethylcarbodiimide; LSU, large subunit; PCR, polymerase chain reaction; SSU, small subunit.

So far, we have identified five synthases that are responsible for formation of nine of the ten Ψ in *E. coli* rRNA. RluA is responsible for formation of Ψ746 in LSU RNA (Wrzesinski et al., 1995a; Raychaudhuri et al., 1999) and RluC makes Ψ residues 955, 2504, and 2580 (Conrad et al., 1998). RluD forms Ψ residues 1911, 1915, and 1917 (Raychaudhuri et al., 1998). These reactions have been demonstrated both in vitro by reaction of affinity-purified recombinant enzyme with RNA transcripts, and in vivo by deletion or disruption of the relevant gene. Reaction of RluE to form Ψ2457 has so far only been shown in vitro (J. Conrad, C. Alabiad, & J. Ofengand, unpubl. results). Formation of Ψ516 in SSU RNA is catalyzed by RsuA, again in a reaction that was only demonstrated in vitro (Wrzesinski et al., 1995b).

In this work, we show that deletion of the *rsuA* gene causes the specific loss of Ψ516 in SSU RNA, and examine the effect of this deletion on the growth of the mutant *E. coli* cells. In addition, we have mutated an aspartate residue, D102, whose equivalent has been proposed by Huang et al. (1998) to be at the reaction center of all Ψ synthases and show that indeed mutation of D102 to threonine or asparagine inactivates the in vivo production of Ψ516.

RESULTS

An intact *rsuA* gene is necesary for formation of Ψ516 in SSU RNA

Previous work had shown that the protein product of the *rsuA* gene, overexpressed and affinity-purified to homogeneity on SDS gels, was able to make the only Ψ in *E. coli* SSU RNA, Ψ516. This only took place when the substrate was an in vitro transcript of part of the SSU RNA that had first been incorporated into an SSU subparticle (Wrzesinski et al., 1995b). The enzyme was highly specific for this site in this substrate, neither recognizing other U residues in the subparticle nor any U residues at all in other subparticles or in free SSU, LSU, or transfer RNA. The extreme specificity shown by the synthase in vitro did not, however, address the question of whether it was the only enzyme in the cell capable of SSU RNA Ψ516 formation, or what the effect of deletion of this Ψ residue would be. Therefore, the gene was deleted by insertion of the kanamycin resistance gene in strain MC1061 (Hamilton et al., 1989). The deletion was confirmed by Ψ sequencing of RNA extracted from the kanamycin-resistant strain that showed that Ψ516 was no longer present (data not shown).

To assess the effects of this gene deletion on growth rate uncomplicated by the other mutant genetic loci present in MC1061 (Hamilton et al., 1989), the deletion was transferred by bacteriophage P1 transduction into MG1655, whose sequenced genome (Blattner et al., 1997) provided a well-defined background. Transductants were selected by resistance to kanamycin. Ψ sequencing analysis of the ribosomal RNA from the mutant strain showed unequivocally that Ψ516 was absent (Fig. 1). To be sure that the loss of Ψ516 was a direct consequence of the deletion of *rsuA* and was not due to some downstream polarity or other indirect effect, the gene was replaced *in trans* by transformation of the deletion strain with a rescue plasmid that contained only the *rsuA* gene inserted into pTrc99A. Wild-type MG1655 and MG1655(*rsuA*$^-$) were transformed with both the rescue plasmid and the control vector pTrc99A, with selection on carbenicillin plates. Total RNA was isolated and sequenced for the presence of Ψ. Figure 1 shows the results. Comparing the wild-type (rsuA$^+$) lane with the rsuA$^-$ lane, it is clear that the stop corresponding to Ψ516 only occurred when CMC was used, and only in the rsuA$^+$ lane. There was no stop in the + CMC lane of the rsuA$^-$ sample. In this method of sequencing, stops due to the presence of Ψ are CMC-dependent, and reverse transcriptase halts one base 3′ to the CMC-Ψ residue (Bakin & Ofengand, 1993, 1998). However, when the rescue plasmid was introduced into the *rsuA*$^-$ strain, the CMC-dependent stop reappeared. We conclude that the loss of Ψ516 is a direct result of deletion of *rsuA* and that RsuA is the sole gene product capable of synthesizing Ψ516.

The strong stop seen at position 528 with or without CMC treatment is thought to be due to m^{7}G527. Whereas m^{7}G itself is not expected to halt reverse transcription, the alkaline treatment used in the sequencing method opens the imidazole ring (Hall, 1971) and is expected to result in reverse transcriptase arrest. In support of this conclusion, the A, C, U, G se-

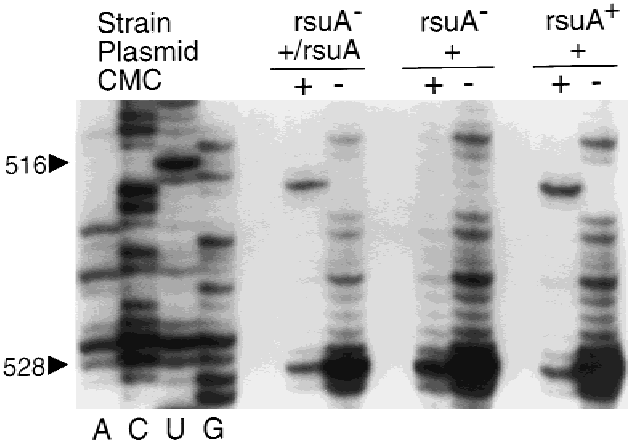

FIGURE 1. Pseudouridine sequencing analysis of the wild-type and *rsuA*-deleted strains of *E. coli* and plasmid-containing derivatives. Preparation of the wild-type (rsuA$^+$) and *rsuA*-disrupted (rsuA$^-$) MG1655 strains and the plasmids pTrc99A (+) and pTrc99A carrying the *rsuA* gene (+/rsuA), transformation of the strains with the plasmids, RNA preparation, and Ψ sequencing were done as described in Materials and Methods. The single naturally occurring Ψ516 in the SSU RNA is indicated by the arrow. RNA for A, C, U, G sequencing lanes was isolated from wild-type cells.

quence lanes, which used RNA extracted from cells but not treated with CMC or alkali, did not show this stop. Because the stops were also present in the RNA derived from the *rsuA*$^-$ strain, we conclude that the absence of Ψ516 in that strain did not block m^7G527 biosynthesis.

Growth rate of the mutant strain

rsuA$^-$ cells lacking Ψ516 in their SSU RNA were viable and appeared to grow normally. To better uncover any subtle metabolic defects, growth rates were measured at different temperatures in both rich and minimal glucose media. The growth experiments were done in the MG1655 genetic background after transduction of the *rsuA*$^-$ gene from strain MC1061. Both wild-type MG1655 and MG1655(*rsuA*$^-$) were transformed with both the rescue plasmid and its control, and exponential growth rates were measured for all four strains (Table 1). The presence of SSU RNA Ψ516 in three of the constructs is shown in Figure 1. Although both rich and minimal media were tested over a temperature range from 24 to 42 °C, no significant difference in growth rate was observed. We conclude that at least under the described conditions, the loss of Ψ516 in the small subunit of the ribosome does not affect the growth rate of the cell.

Mutation of an essential aspartate residue blocks the activity of RsuA

Huang et al. (1998) showed that the replacement of Asp60 by Ala, Asn, Glu, Lys, or Ser in the pseudouridine synthase TruA resulted in the loss of catalytic activity while retaining the ability to bind to tRNA. There are equivalent Asp residues in a four amino acid motif, (G/H)(R/a)(L/t)(D), in all 10 of the known or suspected pseudouridine synthases of *E. coli* (lower case identifies a rare event). In particular, Asp102 occurs in a GRLD sequence in RsuA, and it is the only Asp residue

TABLE 1. Growth rate of *rsuA* deletion and rescue strains.

		Doubling time in minutes[a]		
Strain	Medium	24 °C	37 °C	42 °C
MG1655/pTrc99A	Rich[b]	86 (2)	23 (4)	21 (3)
MG1655/pTrc99A(rsuA)	Rich[b]	84 (2)	23 (4)	21 (3)
MG1655(*rsuA*$^-$)/pTrc99A	Rich[b]	80 (2)	23 (4)	21 (3)
MG1655(*rsuA*$^-$)/pTrc99A(rsuA)	Rich[b]	83 (2)	23 (4)	22 (3)
MG1655/pTrc99A	Minimal[c]	444 (1)	57 (4)	56 (4)
MG1655/pTrc99A(rsuA)	Minimal[c]	456 (1)	54 (4)	53 (4)
MG1655(*rsuA*$^-$)/pTrc99A	Minimal[c]	438 (1)	56 (4)	55 (4)
MG1655(*rsuA*$^-$)/pTrc99A(rsuA)	Minimal[c]	420 (1)	55 (4)	54 (4)

[a]Values in parentheses are the number of exponential phase doublings over which the doubling time was measured.
[b]LB broth (Zyskind & Bernstein, 1992).
[c]M9 (Zyskind & Bernstein, 1992) plus 0.4% glucose, 1 mM $MgSO_4$, and 0.1 mg/mL carbenicillin.

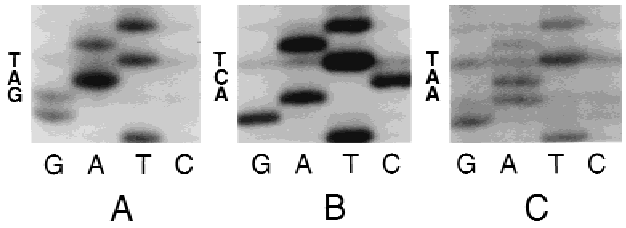

FIGURE 2. DNA sequencing analysis of pET15b plasmids carrying the *rsuA* gene with single amino acid substitutions at Asp102. **A**: pET15b carrying the wild-type gene; GAT codes for Asp102. **B**: The D102T mutant; ACT codes for Thr102. **C**: The D102N mutant; AAT codes for Asn102.

in such a sequence in the molecule. To test the possibility that Asp102 could also be an essential residue of this enzyme, we mutated it to Thr and Asn by megaprimer mutagenesis (Picard et al., 1994). To be sure that the mutations did not dramatically increase the sensitivity of RsuA to proteolysis, we constructed the mutants in pET15b, and used a BL21/DE3(*rsuA*$^-$) strain to obtain stable overexpression of the mutant proteins. DNA sequencing analysis of the pET15b constructs (Fig. 2) confirmed that the desired mutants had been produced. The *rsuA*$^-$ gene was transferred into BL21/DE3 by P1 transduction from MC1061(*rsuA*$^-$) with selection by kanamycin resistance. The BL21/DE3(*rsuA*$^-$) cells were then transformed with pET15b or the *rsuA* constructs in pET15b with selection on carbenicillin plates. The BL21/DE3 cells carrying either the vector or the various *rsuA* constructs were then induced with IPTG. After 3 h of growth, samples from each culture were taken both for protein analysis on SDS-PAGE gels and at the same time for ribosomal RNA isolation and Ψ sequencing analysis. Figure 3 shows that a

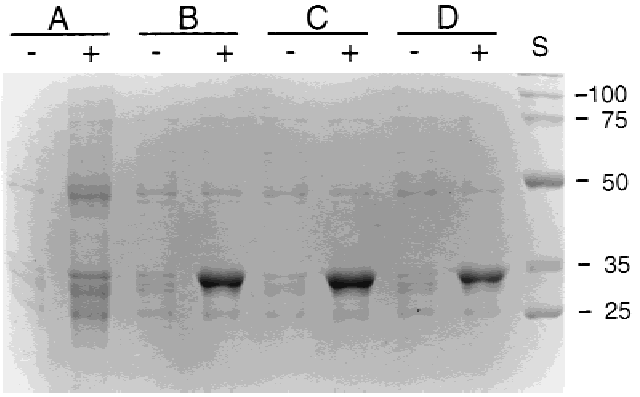

FIGURE 3. Overexpression of the wild-type and mutant *rsuA* gene products in BL21/DE3(*rsuA*$^-$) cells. Cells grown at 37 °C were transformed with the various pET15b/rsuA constructs, harvested either before (−) or after (+) induction with 1 mM IPTG, lysed by boiling in SDS, and analyzed on SDS-polyacrylamide gels. A: pET15b; B: pET15b/rsuA-D102N; C: pET15b/rsuA-D102T; D: pET15b/rsuA-D102D; S: molecular mass standards of the indicated sizes.

strongly overexpressed protein band at about 34 kDa, the expected size (Wrzesinski et al., 1995a), was found in the cells carrying both wild-type and mutant *rsuA* constructs whereas there was no such overexpressed protein band in the cells carrying the vector only. Furthermore, induction was required to produce the band. The intensity of the 34-kDa band appeared to be about the same in both wild-type and mutant constructs indicating that approximately the same amount of wild-type and mutant protein were present in the cell at the time RNA was extracted for sequencing. Ψ sequencing analysis of the rRNA showed that the mutant rescue plasmids were unable to form Ψ516 (Fig. 4). We conclude that the two mutant *rsuA* constructs produced stable proteins that were incapable of isomerizing U516 to Ψ as a result of the replacement of D102 by T102 or N102.

The Ψ synthase superfamily

The essential Asp102 of RsuA is an invariantly conserved residue within the second of two highly conserved motifs (Motif II) that unite the TruB, RsuA, and RluA families into one superfamily (Koonin, 1996). The TruA family does not show any statistically significant similarity with this superfamily (Koonin, 1996). Nonetheless, the TruA family contains the conserved motif GRTD, including the essential Asp60, a proposed catalytic center residue common to part of Motif II of the other Ψ synthase families (Huang et al., 1998). To evaluate sequence conservation within Motif II and to search for statistical support for inclusion of the TruA family into the Ψ synthase superfamily, homology relationships of the Ψ synthase families were reinvestigated.

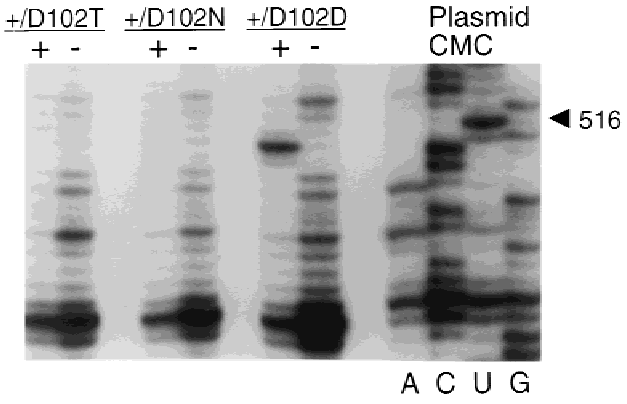

FIGURE 4. Pseudouridine sequencing analysis of strain BL21/DE3(*rsuA*$^-$) containing mutated rescue plasmids. pET15b/rsuA-D102T (+/D102T), pET15b/rsuA-D102N (+/D102N), and pET15b/rsuA (+/D102D). Plasmid constructs were prepared and transformed into the *rsuA*$^-$ strain, RNA was isolated, and Ψ sequencing was done as described in Materials and Methods. Ψ516 is indicated by the arrow. RNA for A, C, U, G sequencing lanes was isolated from wild-type cells.

We used the BLAST2 pairwise search algorithm and the reiterative homology search tool, PSI-BLAST (Altschul et al., 1997; Altschul & Koonin, 1998) to identify new putative Ψ synthases.

In the expanded database, we find 52 RluA members, 23 RsuA members, and 36 TruB members of the Ψ synthase superfamily compared to the 16, 10, and 5 reported previously (Koonin, 1996). The motif II region of these sequences are aligned in Figure 5. All three families are distinct. The RluA and RsuA family members were detected as statistically significant (*E* (Expect) value $< 10^{-5}$; only three members of the RluA family had *E* between 10^{-5} and 10^{-2}) by searching protein databases with the entire *E. coli* RluA and RsuA protein sequences, respectively, using the BLAST2 search algorithm. Only members of the specific family queried were found to have significant *E* values in both the RluA and the RsuA searches. RluA members were detected with an RsuA PSI-BLAST search (and vice-versa) as soon as the first iteration (first use of the position weight matrix) was performed. The TruB family is well separated from the RluA and RsuA families, but complete identification of this family required the use of PSI-BLAST to find the most distant members. The entire Ψ synthase superfamily was identified as statistically significant hits ($E < 10^{-3}$) using a PSI-BLAST search seeded with RluA, with the exception of a few TruB family members. This PSI-BLAST search was complicated somewhat by the fact that *E. coli* RluD, RluC, and RluA (and some other Ψ synthases) have a recently described amino-terminal nucleic acid binding domain that is shared with the (otherwise unrelated) proteins ribosomal protein S4, tyrosyl-tRNA synthase, hemolysin HlyA, and a new small heat shock protein Hsp15 (Korber et al., 1999).

An alignment of 111 Motif II sequences from members of the Ψ synthase superfamily identified the essential Asp (RsuA D102) and the Gly five residues away as invariant (Fig. 5). The single exception to this DXXXXG pattern, an open reading frame (ORF) from *Synecocystis* sp., has GXXXXG. This could be due to a nonfunctional variant, a pseudogene, or a single-base sequencing error, as it seems unlikely that a Gly could functionally substitute for Asp in this position. Specific residues at other positions are typical of each of the three families and of subgroups within the families. For example, the RsuA and TruB families can be completely distinguished from the RluA family because the former always has a Gly at position 2 of Motif II, whereas it is almost always His in the RluA family. The next two residues, RL, are invariant within the RsuA family and nearly so in the RluA family. Forty-six RluA members have RL, 4 have RI, and 2 have NL and QL, respectively. The TruB family never has an Arg in this position; it is usually Thr followed by an invariant Leu. Following the invariant G, there is almost always a group of four aliphatic residues. The Xs in the DXXXXG pat-

tern also follow family-specific patterns, with prolines frequently following the D in the TruB family and conservation of charged residues and Ser/Thr in the other positions distinguishing other families and subgroups. For example, the particular charged residues present seem to correspond with RluA family functional classes: MA for RluA, KD for RluD, RD for RluC, and R followed by an uncharged residue for YqcB.

Using the *E. coli* TruA protein sequence as a query, 35 TruA family members were identified in the sequence databases using BLAST2 and PSI-BLAST. The sequence region analogous to Figure 5 is shown in Figure 6. As with the TruB family, PSI-BLAST searches were necessary to identify all TruA family members, but no RluA, RsuA, or TruB family members were identified in any of the TruA searches, even when the Expect value for inclusion in a PSI-BLAST iteration was reduced to $E = 0.01$. Similarly, no TruA members were ever revealed in RluA, RsuA, or TruB PSI-BLAST or BLAST2 searches, even when distant members were used as queries. Additional searches using the MoST algorithm (Tatusov et al., 1994) failed to identify any TruA members when seeded with representative Motif II sequences. The search parameters of MoST were reduced in stringency until numerous obviously unrelated sequences were drawn in and still no TruA members could be detected at all, even among the unrelated sequences. Although exhaustive searches were performed, no evidence for a homologous relationship between TruA and the other three Ψ synthase families was obtained.

Alignment of the conserved sequence motif among 33 TruA family members is shown in Figure 6. A phylogenetic analysis of this region reveals several TruA subgroups as indicated by the spacing in Figure 6. The conserved GRT sequence that precedes the essential D60 of TruA is like the conserved GRL that precedes the essential RsuA D102 except that the T and L residues are invariant in each respective family. There is little similarity between TruA and RsuA beyond the essential D residue. In general, the TruA region is more highly conserved, the invariant G of the DXXXXG motif being largely replaced by Ala, and the aliphatic region is less hydrophobic and frequently interrupted by a charged residue.

Motif II has also been proposed to be a putative uridine-binding motif shared with dUTPases (Dut) and dCTP deaminases (Dcd) (Koonin, 1996). Dut and Dcd are two closely related families: a BLAST2 search with the *E. coli* Dcd sequence hits the Dut sequence of the Archael virus SIRV with an *E* value of 2×10^{-10} (Prangishvili et al., 1998). This high similarity has led to some probable misidentification of putative Dcd function in genome projects (Prangishvili et al., 1998). Other Dut/DcD superfamily members are much more distantly related, but can easily be detected by PSI-BLAST. Despite exhaustive searches, including the use of the MoST algorithm, we were unable to confirm a statistically significant similarity between the Dut/Dcd superfamily and the pseudouridine synthase superfamily. While Motif II of the pseudouridine synthase superfamily and Motif 3 of the Dcd/Dut superfamily (McGeoch, 1990; Prangishvili et al., 1998) may indeed both be uridine recognition motifs, whether they evolved from a common origin will likely remain unresolved until structure–structure comparisons can be made.

DISCUSSION

Specificity

Previous in vitro studies have demonstrated that RsuA was able to modify U516 of *E. coli* SSU RNA to Ψ516, but was unable to form any of the other known Ψ in either *E. coli* LSU RNA or tRNA (Wrzesinski et al., 1995a). The present work shows that the formation of Ψ516 in vivo is solely dependent upon the presence of *rsuA*. In its absence, Ψ516 formation does not occur, but when a plasmid-encoded *rsuA* gene was reintroduced into an *rsuA*-deficient strain, Ψ516 formation was restored. Thus RsuA is the only gene product in *E. coli* capable of Ψ516 formation. It is unlikely that RsuA makes Ψ at any other known Ψ site in RNA, that is, a dual specificity (Wrzesinski et al., 1995b), as Ψ516 is the only Ψ in *E. coli* SSU RNA (Bakin et al., 1994a) and RsuA did not react with either LSU RNA or tRNA in vitro (Wrzesinski et al., 1995a).

Mechanism of Ψ biosynthesis

Ψ formation requires cleavage of the uracil N_1–ribosyl $C_{1'}$ bond of uridine, rotation of the cleaved uracil to align the uracil C_5 with the ribosyl $C_{1'}$, and carbon–carbon bond formation between the two aligned moieties. Huang et al. (1998) have proposed a reaction mechanism for this isomerization that involves the β-carboxyl of a conserved aspartate residue within the Ψ synthase TruA. The aspartate is found within a conserved sequence motif H/G-R/a-L/t-D (where lower case identifies rare events) common to all 10 previously identified and putative Ψ synthases in *E. coli*. Within RsuA, a single such motif, GRLD, exists with the conserved aspartate located at residue 102. As in TruA, D102 appears to have an essential role in the conversion of U516 to Ψ516, since replacement of it with either threonine or asparagine blocked formation of Ψ516 formation in vivo. The essential nature of the aspartate in this conserved motif has now been demonstrated in four known Ψ synthases (Table 2). TruA (Huang et al., 1998) and RluA (Raychaudhuri et al., 1999) have been described and RsuA is shown in this work. Moreover, recent work (J. Conrad & J. Ofengand, unpubl. results) has shown that the aspartate residue in the same sequence motif in RluC is essential for the

Pseudouridine synthase superfamily

Db|Access|Gene_Organism	Motif
RsuA	$GRLD +T G$$$$
	A S
sp|P33918|RSUA_ECOLI	AGRLDIDTTGLVLM
sp|P45124|RSUA_HAEIN	AGRLDVDTTGLVLL
sp|O66829|Y554_AQUAE	AGRLDVDAEGLLLI
sp|O32068|YTZF/BACSU	AGRLDKDTEGFLLL
sp|P75966|RLUE_ECOLI	AGRLDRDSEGLLVL
sp|P44827|RLUE_HAEIN	AGRLDRDSEGLLIL
sp|P32684|YJBC_ECOLI	IGRLDKDSQGLIFL
sp|P72581|Y612_SYNY3	VGRLDQDSEGLLLL
sp|O05668|YRSU_MYCLE	VGRLDADTEGLILL
sp|O33210|YRSU_MYCTU	VGRLDADTEGLMLL
sp|P35159|RLUB_BACSU	IGRLDYDTSGLLLL
tr|O84728|YJBC/CHLTR	VGRLDKETSGLILV
sp|P55986|YE59_HELPY	VGRLDFASEGVLLL
tr|O80967|RSUX/ARATH	VGRLDVATTGLIVV
tr|O83472|RSUX/TREPA	IGRLDVRSEGALLF
sp|O51155|Y129_BORBU	IGRLDFKSSGLLLF
em|e1342837|RSUX/RICPR	IGRLDLNSEGLLLL
sp|Q55578|Y361_SYNY3	VGRLDRNSTGALLL
sp|P37765|YCIL_ECOLI	VGRLDVNTCGLLLF
sp|P45104|YCIL_HAEIN	VGRLDINTSGLLLF
sp|P42395|YCIL_BUCAP	VGRLDINTKGLLLF
gi|3845303|RSUX/PLAFA	VGRLDRNTSGVLLL
sp|O67444|YE64_AQUAE	VGRLDYNTEGLLIL
TruB	GtLDP AtGLP
	V C $
sp|P09171|TRUB_ECOLI (2)	TGALDPLATGMLPI
pi|C57253|TRUB/ACICA	TGALDPLATGLLPI
sp|P72154|TRUB_PSEAE	TGSLDPLATGVLPL
sp|P74696|TRUB_SYNY3	GGTLDPLAEGVLPL
tr|O32785|TRUB/LACLA	GGTLDPQVTGVLPV
sp|Q57612|TRUB_METJA	GGTLDPKVTGVLPV
sp|O26140|TRUB_METTH	GGTLDPKVTGVLPL
sp|O59357|TRUB_PYRHO	GGTLDPKVSGVLPV
sp|O30001|TRUB_ARCFU	AGTLDPRVTGVLPI
dd|d1038085|TRUB/SYNP7	GGTLDPAVTGVLPI
sp|P32732|TRUB_BACSU	TGTLDPEVSGVLPI
sp|O66922|TRUB_AQUAE	TGTLDPIATGLLII
tr|O84096|TRUB/CHLTR	AGTLDPFATGVMVM
sp|O33335|TRUB_MYCTU	AGTLDPMATGVLVI
tr|O59721|TRUB/SCHPO (1)	GGTLDPLASGVLVV
sp|P48567|PUS4_YEAST	GGTLDPLASGVLVI
sp|O60832|DKC1_HUMAN (7)	SGTLDPKVTGCLIV
sp|O17919|NO50_CAEEL	SGTLDPKVSGCLIV
sp|O14007|CBF5_SCHPO	SGTLDPKVTGCLII
gb|AA532324|CBF5/CRYPV	SGTLDPKVTGCLLV
sp|O51743|TRUB_BORBU	AGTLDKFASGILVC
em|e1342797|TRUB/RICPR	AGTLDVEAEGILPL
sp|O83859|TRUB_TREPA	TGTLDRFADGLLLL
gb|AI120090|TRUB/MOUSE	GGTLDSAARGVLVV
em|Z45640|TRUB2/HUMAN	XXXLDAQASGVLVL
em|HS1186784|TRUB/HUMAN	GGTLDSARRGVLVV

Db|Access|Gene_Organism	Motif
RluA	$HRLD++TsG$$$$
	CN
sp|P39219|RLUA_ECOLI	VHRLDMATSGVIVV
sp|P44782|RLUA_HAEIN	VHRLDMATSGIIVF
sp|P33643|RLUD_ECOLI (3)	VHRLDKDTTGLMVV
sp|Q45480|YLYB_BACSU	VHRIDKDTSGLLMV
sp|P50513|RLUD_ZYMMO	VHRIDKDTSGLLVV
sp|P74346|YG29_SYNY3	VHRLDKDTTGAMVV
sp|O50310|YBC5_CHLVI	VHRLDKDTSGLIII
sp|O67638|YH58_AQUAE	VHRLDKETAGVMVI
tr|O84665|RLUD/CHLTR	VHRLDKDTSGLLIT
em|e1342564|RLUX/RICPR	VHRLDKETSGLLLI
tr|O83359|RLUX/TREPA	VHRLDKDTSGVLLT
sp|P70870|RLUD_BORBU	VHRLDKDTSGVLIC
sp|P54604|YHCT_BACSU	VHRLDQDTSGAIVF
sp|Q10786|Y04P_MYCTU	VHRLDVGTSGVMVV
sp|P23851|RLUC_ECOLI	VHRLDRDTSGVLLV
sp|P44433|RLUC_HAEIN	VHRLDRDTSGILLI
tr|O84108|RLUC/CHLTR	VHRLDRDTSGCILF
sp|O66114|YLP4_ZYMMO	VHRLDRDTSGCLLL
sp|P47451|Y209_MYCGE	VHRLDRDTSGAIVV
sp|P75485|Y209_MYCPN	VHRLDRDTSGVIML
sp|O31613|YJBO_BACSU	VTRLDRDTSGIMLV
sp|Q45826|YMDA_CHLAU	VHRLDRDTSGLLVI
tr|P70863|RLUX/BARBA	VHRLDRETSGLLVV
gb|3834304|RLUX/ARATH	VHRLDRETSGLLVM
sp|Q47417|YQCB_ERWCA	VHRLDRPTSGVLLL
sp|Q46918|YQCB_ECOLI	AHRLDRPTSGVLLM
sp|P44197|YQCB_HAEIN	IHRLDRPTSGVLLF
tr|O07166|YQCB/MYCTU	AHRLDRLTAGVLLF
tr|O83189|YQCB/TREPA	LHRLDRGTEGLIAF
sp|P47610|Y370_MYCGE	AHRIDRNTSGIVIG
sp|P75230|Y370_MYCPN	AHRIDRNTCGLVIG
tr|O51755|RLUX/BORBU	VHRLDRNTSGIIIF
tr|Q25257|RLUX/LEIDO	VHRLDAETSGCLLI
gb|AC006246|RLUX/DROME	IHRLDRLTSGLLLF
sp|O16686|YK27_CAEEL	LHRLDRATSGVLLF
sp|Q12069|YD36_YEAST	CNRLDRLTSGLMFL
sp|Q09709|YA32_SCHPO	CNRLDRLTSGLMFF
sp|P53294|YG3X_YEAST	CYRLDKITSGLLIL
sp|Q12362|RIB2_YEAST	CNRLDKPTSGLMFL
tr|Q06244|Y165C/YEAST	VHRLDHCVTGGMLI
sp|P43930|Y042_HAEIN	VHRLDKVTSGLLIL
sp|O25441|Y745_HELPY	VHRLDKDTSGGIVI
tr|O83259|RLUX/TREPA	LHRLDKDTAGVLLF
gb|4155455|RLUX/HELP2	LHRLDKETSGVILL
sp|O25610|Y956_HELPY	LHRLDKETSGVVLL
sp|O25114|Y347_HELPY	AHRLDYETSGLVLA
tr|Q25346|RLUX/LEIMA	CHNLDTETSGCVVL
sp|P72970|YF92_SYNY3	LHRLGTGTSGLLLL

FIGURE 5. (*Legend on facing page.*)

formation of Ψ955, Ψ2504, and Ψ2580 in the LSU RNA of *E. coli*.

Despite the above described conservation of this sequence motif, the role of the β-carboxyl of Asp at the active site still remains only a hypothesis, albeit an attractive one. The mutational results described above and listed in Table 2 only show that the conserved Asp is essential for function but do not prove a mechanism. For example, the conserved Asp could be required to maintain some specific conformation of the polypeptide chain that is needed for function. The Gly residue that is almost universally found five residues toward the C-terminus from the conserved Asp (see Figs. 5 & 6) could be an example of such a residue fulfilling a structural need as its lack of a side chain with the potential for a functional interaction makes it unlikely to be involved directly at the catalytic center.

Function of Ψ

No difference in exponential growth rates between wild-type and *rsuA*-deficient cells were found under laboratory conditions employing a range of temperatures in both rich and minimal media. This was also true for RluA, RluC, and TruA in *E. coli* (Table 2), and is also the case in eukaryotes where deletion of specific guide RNAs blocks formation of the specified Ψ, but has no effect on cell growth or metabolism (Ofengand & Fournier, 1998). So far, only inactivation of RluD has a strong growth inhibitory effect (Table 2). However, it is possible that more subtle growth differences in *rsuA*$^-$ cells may be uncovered when examining such conditions as survival in stationary phase as well as competition studies employing mixed cultures of mutant and wild-type strains such as those done recently on *rluA*$^-$ cells (Raychaudhuri et al., 1999). These have not yet been done with *rsuA*$^-$ cells. Moreover, it is possible that Ψ516 formation in the SSU of *E. coli* confers a growth advantage only under specific nonlaboratory conditions. Further experiments designed to better understand the relationship between Ψ516 formation and its effect on growth are clearly necessary.

Although the location of Ψ516 at the base of the highly conserved "530" loop (Ofengand et al., 1993), which is known to be involved in the fidelity of codon recognition (summarized in Santer et al., 1993), is highly suggestive of a functional role in the small subunit of the ribosome, the location is not generally conserved in other organisms. A homologous Ψ exists in *Bacillus subtilis* (Wrzesinski et al., 1995a; Niu & Ofengand, 1999), and the existence of an ORF in *Haemophilus influenzae* with 73% homology to RsuA suggests the pres-

FIGURE 5. Alignment of 111 Motif II sequences of the Ψ synthase superfamily. The 14 amino acids of Motif II are aligned and grouped according to families and subgroups within the families. The subgroups are organized to highlight certain patterns of variations within the motif but do not indicate definitive functional or evolutionary groupings, although CLUSTALX, NJplot, BLAST2 results, and functional information were considered when making the subgroups. The motif consensus shown above the sequences summarizes the sequence patterns but does not necessarily indicate all pattern variations. Motif conventions are as follows. Single capital letters: invariant or nearly invariant; lower case letters: highly conserved; letter pairs indicate the two most common amino acids at that position; $: usually one of ILVM; +: charged. The existence of additional sequences in different organisms with exactly the same 14 amino acid sequence in Motif II are indicated by the numbers in parentheses. The additional sequences are: TRUB ECOLI orthologs: TRUB_YEREN (sp:O34273); TRUB_HAEIN (sp:P45142). TRUB_SCHPO ortholog: TRUB2/SCHPO (em:1319405). DKC1_HUMAN orthologs: CBF5_EMENI (sp:O43100); CBF5_YEAST (sp:P33322); CBF5_ASPFU (sp:O43102); CBF5_CANAL (sp:O43101); CBF5_KLULA (sp:O13473); DKC1_RAT (sp:P40615); NO60_DROME (sp:O44081). RLUD_ECOLI orthologs: RLUD_HAEIN (sp:P44445); RLUD_PSEAU (sp:P33640); RLUD/RICPR (em:1343115). The *B. subtilus* YTZF/BASCU protein sequence is a reconstruction of a probable frameshift mutation that fuses two adjacent protein sequences, YTZF_BACSU (O32068) and YTZG_BACSU (O32069). There is also a human ortholog to the YD36_YEAST subgroup of the RluA family (tr:Q92939) that is not listed in the figure because it is a C-terminal partial sequence with the Motif II region still unsequenced. The database record accession numbers are taken from a variety of databases abbreviated as sp: SWISS-PROT; pi: PIR; tr: TREMBL, em: EMBL; gb: Genbank; dd: DDBJ. The citations to the original sequence papers can be found within the database records. Locus names were taken from SWISS-PROT or are provisional designations devised as described in Materials and Methods. When no ortholog was obvious and no other name was available, a generic family name was given to the protein as a temporary identifier, that is, RLUX or RSUX. The SWISS-PROT organism codes used are ACICA: *Acinetobacter calcoaceticus*; AQUAE: *Aquifex aeolicus*; ARATH: *Arabidopsis thaliana*; ARCFU: *Archaeoglobus fulgidus*; ASPFU: *Aspergillus fumigatus*; BACSP: *Bacillus* sp. strain KSM-64; BACSU: *Bacillus subtilis*; BARBA: *Bartonella bacilliformis*; BUCAP: *Buchnera aphidicola*; BORBU: *Borrelia burgdorferi*; CHLTR: *Chlamydia trachomatis*; CAEEL: *Caenorhabditis elegans*; CANAL: *Candida albicans*; CHLAU: *Chloroflexus aurantiacus*; CHLVI: *Chlorobium vibrioforme*; CRYPV: *Cryptosporidium parvum*; DROME: *Drosophila melanogaster*; ECOLI: *Escherichia coli*; EMENI: *Emericella nidulans*; ERWCA: *Erwinia carotovora*; HAEIN: *Haemophilus influenzae*; HELPY: *Helicobacter pylori*; HUMAN: *Homo sapiens*; KLULA: *Kluyveromyces lactis*; LACLA: *Lactococcus lactis*; LEIDO: *Leishmania donovani*; LEIMA: *Leishmania major*; METJA: *Methannococcus jannaschii*; METTH: *Methanobacterium thermoautotrophicum*; MOUSE: *Mus musculus*; MYCGE: *Mycoplasma genitalium*; MYCLE: *Mycobacterium leprae*; MYCPN: *Mycoplasma pneumoniae*; MYCTU: *Mycobacterium tuberculosis*; PLAFA: *Plasmodium falciparum*; PSEAE: *Pseudomonas aeruginosa*; PYRHO: *Pyrococcus horikoshii*; RAT: *Rattus norvegicus*; RICPR: *Rickettsia prowazekii*; SCHPO: *Schizosaccharomyces pombe*; SYNP6: *Synechococcus* sp. PCC 6301; SYNP7: *Synechococcus* sp. PCC 7942; SYNY3: *Synechocystis* sp. PCC 6803; TREPA: *Treponema pallidum*; YEAST: *Saccharomyces cerevisiae*; YEREN: *Yersinia enterocolitica*; ZYMMO: *Zymomonas mobilis*. HELP2 is a second strain (J99) of *H. pylori* that has been recently sequenced. The Motif II sequence from this strain was included because it differed from the HELPY version.

```
TruA family

                                AGRTDAGVHA  Q$                                  A RTDKGVHA  N$
                                S    K                                          s    a S   Q
sp|P07649|TRUA_ECOLI            AGRTDAGVHGTGQV   sp|Q59069|TRUA_METJA          GGRTDKGVSALGNF
sp|P45291|TRUA_HAEIN            AGRTDSGVSGTGQV   sp|O58941|TRUA_PYRHO          ASRTDKGVSALGNV
tr|O87016|TRUA/PSEAE            AGRTDAAVHASGQV   sp|O26928|TRUA_METTH          AGRTDRGVHALGNF
sp|O06322|TRUA_MYCTU            AGRTDAGVHASGQV
sp|O28544|TRUA_ARCFU            AGRTDAGVHAYGQV   sp|Q12211|PUS1_YEAST          AARTDKGVHAGGNL
tr|O86776|TRUA/STRCO            AGRTDAGVHARGQV   em|e1359943|TRUA/SCHPO       AARTDKGVHAAGNV
sp|P73295|TRUA_SYNY3            AGRTDAGVHAAAQV   sp|P53167|PUS2_YEAST          AARTDKGVHAMLNL
sp|O24712|TRUA_SYNP6            AGRTDTGVHRAAQV   em|e1339974|TRUA2/SCHPO      AARTDKGVHTLRNL

sp|P70973|TRUA_BACSU            SGRTDSGVHAAGQV   sp|P31115|PUS3_YEAST          CGRTDKGVSAMNQV
sp|Q45557|TRUA_BACSP            SGRTDAGVHALGQV   tr|O65241|TRUA/ARATH          AGRTDKGVSALNQV
em|e1343126|TRUA/RICPR          SGRTDAGVHAIGQV   em|Z83321|TRUA3/ARATH         SARTDKGVSAVGQV
tr|O84469|TRUA/CHLTR            SGRTDAGVHAQGQI   em|e1349968|TRUA/CAEEL       AARTDRAVSAARQM
sp|O83802|TRUA_TREPA            SGRTDSGVHAVGQA
                                                 sp|O22928|PUSH_ARATH          SSRTDKGVHSLATS
sp|P70830|TRUA_BORBU            SGRTDKGVHAKKQI   tr|O04502|TRUA2/ARATH        GVLQDAGVHALSNV
sp|P56144|TRUA_HELPY            AGRTDKGVHANNQV   gb|AA696025|TRUA/DROME       SSRTDAGVHALHST
sp|Q50291|TRUA_MYCPN            SGRTDKGVHAINQT   gb|W48211|TRUA/MOUSE         SSRTDAGVHALSNA
sp|P47428|TRUA_MYCGE            SGRTDKGVHAINQT

sp|O66953|TRUA_AQUAE            CCRTDSGVHALDYI
```

FIGURE 6. Alignment of a highly conserved motif found in the TruA family of Ψ synthases. Thirty-five TruA family members were aligned and grouped using the methods and conventions described in Figure 5. Two sequences were omitted from the alignment. The conserved motif in the TruA homolog YQN3_CAEEL (sp: Q09524) appears to have two gaps and therefore was not aligned. A TruA homolog found in the EST database supposedly from *Arabidopsis thaliana* (gb: N37304) is very similar to the *E. coli* sequence and we suspect it represents a bacterial contaminant in that particular EST library. One additional SWISS-PROT organism code is used: STRCO,

ence of an equivalent Ψ in that organism. On the other hand, *Halobacter halobium* does not have a homologous Ψ and may not have any Ψ at all in its SSU RNA (Bakin & Ofengand, 1995). Moreover, yeast SSU RNA, despite having 14 Ψ does not have one at the position corresponding to Ψ516. The 32 mapped Ψ of mammalian SSU RNA (Ofengand & Fournier, 1998) likewise do not include a homolog of Ψ516 although since approximately four Ψ remain to be positioned, it cannot be said with certainty that no homolog of Ψ516 exists in mammals. Consequently, even if Ψ516 serves a functional role in *E. coli*, it cannot be a universal one. Nevertheless, the extreme substrate specificity of RsuA (see above) argues for an important role in the metabolism of *E. coli* since it seems unlikely that such a finely-tuned enzyme–substrate relationship would exist if Ψ516 formation was merely adventitious.

The Ψ synthase superfamily

Although previous reports have implied that the RluA/RsuA/TruB Ψ synthase superfamily may be homologous to the TruA Ψ synthase family (Huang et al., 1998) as well as to the Dcd/Dut dUTPase/dCTP deaminase superfamily (Koonin, 1996), we were unable to support these relationships with statistically significant sequence alignments. The limited similarity detected could be the result of convergent evolution, extreme sequence divergence, or chance. On the other hand, it is clear that the identical Asp in the conserved motif of TruA, RsuA, RluA, and RluC are all essential for function (Table 2). Whether this residue acts as a nucleophilic catalyst (Huang et al., 1998) or as a uridine-binding motif (Koonin, 1996), or both remains to be established. We note that the TruA family does not conserve the invariant Gly of Motif II, which is instead mostly an Ala, although it is Gly in *E. coli* TruA. The similarities between the GRTD motif in TruA and the analogous ones in the RsuA/RluA/TruB superfamily coupled with the functional results in Table 2 suggest that the TruA family may in fact be related to the superfamily despite the lack of sequence similarity. If future three dimensional structural comparisons should fail to confirm a homologous relationship, it would be a remarkable example of either convergent evolution or divergent evolution that is so complete as to retain only the very essence of a catalytic site.

Ψ synthase distribution among organisms

The distribution of the known and putative Ψ synthases in all currently completely sequenced genomes is listed

TABLE 2. Effects of deletion and "active site" aspartate mutations of *E. coli* pseudouridine synthases on cell growth and synthase activity.

Synthase	Substrate (sites)	Growth effect of deletion	Aspartate mutation	Synthase activity (%)[a]	
				In vivo	In vitro
RsuA	16S rRNA (516)	no effect[b]	D102T	—[b]	n.d.[c]
			D102N	—[b]	
RluA	23S rRNA (746), tRNA (32)	no effect[d]	D64T	—[d]	<2[d]
			D64N	—[d]	<2[d]
RluC	23S rRNA (955, 2504, 2580)	no effect[e]	D144T	—[f]	n.d.
			D144N	—[f]	
RluD	23S rRNA (1911, 1915[g], 1917)	severe growth inhibition[h]	n.d.	n.d.	n.d.
RluE	23S RNA (2457)	n.d.	n.d.	n.d.	n.d.
TruA	tRNA (38–40)	no effect[i]	D60X[j]	n.d.	<0.01[k]
TruB	tRNA (55)	n.d.	n.d.	n.d.	n.d.

Identification of synthase genes and sites of action were as follows. RsuA (Wrzesinski et al., 1995a), RluA (Wrzesinski et al., 1995b), RluC (Conrad et al., 1998), RluD (Raychaudhuri et al., 1998), RluE (J. Conrad, C. Alabiad, & J. Ofengand, unpubl. results), TruA (Marvel et al., 1985), TruB (Nurse et al., 1995).

[a] Assayed in vivo as the presence (+) or absence (−) of Ψ. Assayed in vitro using purified overexpressed protein with values expressed as percent of wild-type activity.

[b] This work.

[c] n.d.: not determined.

[d] Inhibited only in competition with wild-type (Raychaudhuri et al., 1999).

[e] Conrad et al. (1998).

[f] J. Conrad & J. Ofengand, in prep.

[g] N^3-methyl Ψ (Kowalak et al., 1996).

[h] Gene was disrupted by miniTn10(cam) insertion such that the N-terminal 60% of the protein could be made (Raychaudhuri et al., 1998).

[i] No effect on rich medium or when uracil was added to minimal medium (Tsui et al., 1991). The 3–4-fold decrease in growth rate observed on unsupplemented minimal medium is thought to be due to the *rph* mutation in the test strain (Jensen, 1993).

[j] X is Ala, Asn, Glu, Lys, or Ser.

[k] Huang et al. (1998).

in Table 3, which contains, in addition, the current but incomplete data for humans. The results were grouped according to the three major categories, Prokarya, Archaea, and Eukarya because, according to our current knowledge, the Prokarya and Eukarya use two distinct methods for specifying Ψ residues in their ribosomal RNA molecules (Ofengand & Fournier, 1998). The situation in Archaea is not known. Whereas prokaryotes have a set of synthases each of which is specific for one or a few particular Ψ residues in their rRNAs, eukaryotes appear to specify the sites for the many more Ψ in their rRNAs by means of small nucleolar guide RNAs, found in ribonucleoprotein particles, which also contain, in addition to other proteins, the single Cbf5-like putative Ψ synthase (Lafontaine et al., 1998; Watkins et al., 1998). Consequently, eukaryotes should contain an analog of the yeast Cbf5 protein plus tRNA and sn(o)RNA Ψ synthases, whereas prokaryotes should have homologs of both rRNA and tRNA enzymes. So far as is known, the TruB family contains only those synthases that are specific for Ψ55 in tRNA and the TruA family contains the ones which make Ψ at all the other sites in tRNA. The RsuA and RluA families are so far rRNA-specific with the exception that *E. coli* RluA makes Ψ32 in tRNA in addition to Ψ746 in LSU RNA, an example of dual specificity (Wrzesinski et al., 1995b). There is no rRNA specificity corresponding to the RsuA and RluA families. For example, while the arch-type *E. coli* RsuA is specific for Ψ516 in SSU RNA (Wrzesinski et al., 1995a), the RsuA family member RluB (*B. subtilis*) specifies Ψ2633 in LSU RNA (Niu & Ofengand, 1999) and another RsuA family member, RluE (*E. coli*), specifies Ψ2457 in its LSU RNA (Table 2).

As shown in Table 3, *Mycoplasma genitalium* and *Mycoplasma pneumoniae* have no TruB synthase but two that are RluA-like and one that is TruA-like. Either these organisms have no Ψ55 in any of their tRNAs, a most unusual case, or one of the RluA synthases carries out this reaction. If the latter, there would be only one synthase left to make the Ψ in rRNA. Unfortunately, there is no Ψ sequence information on either the tRNAs or rRNAs of either of these organisms. It is noteworthy that *M. genitalium*, with only 467 protein coding genes, has nevertheless devoted three, or 0.6%, to Ψ synthases. A similar percentage is also found for five other organisms (Table 3). The Archaeal organisms appear to have no synthase for rRNA. *Methanobacterium thermoautotrophicum* is known to have Ψ54Ψ55 in at least two of its tRNAs (Gu et al., 1984), which would account for the TruB synthase. Either these

TABLE 3. Family distribution of Ψ synthases in sequenced genomes.

Organism/Family	RsuA	RluA	TruB[a]	TruA	Sum	No. of ORFs	% Ψ
Prokarya							
E. coli	4	4	1	1	10	4,289	0.23
B. subtilis	2	3	1	1	7	4,100	0.17
H. influenzae	3	5	1	1	10	1,709	0.59
M. genitalium	0	2	0	1	3	467	0.64
M. pneumoniae	0	2	0	1	3	677	0.44
M. tuberculosis	1	2	1	1	5	3,918	0.13
R. prowazekii	1	2	1	1	5	834	0.60
H. pylori	1	3	0	1	5	1,566	0.32
C. trachomatis	1	2	1	1	5	894	0.56
B. burgdorferi	1	2	1	1	5	850	0.59
T. pallidum	1	3	1	1	6	1,031	0.58
Synechocystis sp.	2	2	1	1	6	3,169	0.19
A. aeolicus	2	1	1	1	5	1,522	0.33
Archaea							
M. jannaschii	0	0	1	1	2	1,715	0.12
M. thermoautotrophicum	0	0	1	1	2	1,869	0.11
A. fulgidus	0	0	1	1	2	2,407	0.08
P. horikoshii	0	0	1	1	2	2,064	0.10
Eukaryotes							
S. cerevisiae	0	4	2 (1)	3	9	6,537	0.14
C. elegans	0	1	1 (1)	2	4	19,099	0.02
H. sapiens[b]	0	1	3 (1)	0	—	—	—

Data taken from Figures 5 and 6.
[a]Values in parentheses are those ORFs with homology to Cbf5.
[b]Incomplete genome.

archaeal organisms do without Ψ in their rRNA or they possibly possess a Cbf5-like synthase with guide RNAs, and the synthase is too divergent in sequence to be detected by the procedures used here. There is no information on the presence of Ψ in the rRNA of these organisms, although *Halobacter halobium* has four Ψ in its LSU RNA alone (Ofengand & Bakin, 1997).

Saccharomyces cerevisiae, mouse, and human mitochondrial ribosomes have one Ψ in their LSU RNA (Ofengand & Bakin, 1997). The corresponding synthase should be encoded in a nuclear gene. This may explain why *Caenorhabditis elegans* has both an RluA-like synthase, presumable for its mitochondrial rRNA, and a Cbf5-like enzyme for its cytoplasmic ribosomes. The remaining two synthases are of the TruA type and therefore one should specify Ψ55 in tRNA, and the other one all other tRNA Ψ. By contrast, *S. cerevisiae* has too many synthases. The tRNA-specific properties of two of the three TruA-like enzymes have been described (Simos et al., 1996; Grosjean et al., 1997; Lecointe et al., 1998), and one of them even has a dual specificity for an snRNA (Massenet et al., 1999). The TruB-like synthases have been accounted for as the tRNA Ψ55 enzyme (Becker et al., 1997), and the putative rRNA Ψ synthase working in conjunction with guide RNAs (Lafontaine et al., 1998). One of the four RluA-like synthases, YD36 (Q12069), is probably the mitochondrial rRNA Ψ synthase (J. Ofengand, unpubl. results). There are three more enzymes with no known function. Perhaps *S. cerevisiae* has a dual system for making Ψ in rRNA, one that is analogous to the bacterial system and is used for certain critical Ψ residues, and another using Cbf5 and guide RNAs for the remainder of the 44 Ψ in *S. cerevisiae* SSU and LSU RNAS (Bakin et al., 1994b; Bakin & Ofengand, 1995).

MATERIALS AND METHODS

Gene deletion

The *rsuA* gene was deleted by the method of Hamilton et al. (1989). The insert, cloned into the *Xba*I and *Kpn*I sites of pMAK705, was prepared by PCR as described in Figure 2 of Supekova et al. (1995). It contained 845 bases 5′ to the AUG start and 1,015 bases 3′ to the UAA termination codon. Fifty-three nucleosides of the N-terminal portion of the gene and 53 residues of the C-terminus were retained, with the remainder being replaced by the kanamycin resistance gene, obtained by PCR amplification from pUC4K (Pharmacia, cat. #27-4958-01). The host strain for pMAK705 was the leucine auxotroph MC1061, as described by Hamilton et al. (1989). The deleted *rsuA* gene was moved into strains MG1655 (Blattner et al., 1997) and BL21/DE3 (Novagen, Inc.) by bacteriophage P1 transduction (Miller, 1992). Selection was done on rich (LB; Zyskind & Bernstein, 1992) medium containing 0.05 mg/mL kanamycin.

Wild-type rescue plasmids

The preparation of wild-type rescue plasmid pET15b/rsuA has been described previously (Wrzesinski et al., 1995a).

Wild-type rescue plasmid pTrc99A/rsuA was constructed by insertion into the *Nco*I and *Bam*HI sites of pTrc99A (Pharmacia, cat. #275007-01) of a segment of DNA PCR-amplified from pET15b/rsuA. The N-terminal primer used in the PCR reaction extended from −24 to +6, where the A of the initiating AUG is +1, and carried base changes at positions −13, −11, and −10 to create an *Nco*I site upstream of the AUG start site. The C-terminal primer, in reverse orientation, extended from +673 to +705, where the last sense nucleotide is +693. Mismatches at residues 697, 698, 699, and 701 were introduced to create a *Bam*HI site. The PCR product was purified by agarose gel electrophoresis and digested with *Nco*I and *Bam*HI (New England Biolabs, Beverly, Massachusetts). The digested and purified product was ligated with a similarly digested and purified pTrc99a vector for 16 h at 16 °C.

Mutant rescue plasmids

Mutant rescue plasmids were prepared by the megaprimer PCR mutagenesis procedure (Picard et al., 1994). PCR reactions were performed using the pTrc/rsuA rescue plasmid as template and three oligonucleotide primers—two outer primers that were upstream and downstream of the mutation site, and one mutagenic primer. The upstream and downstream primers contained the restriction sites *Nde*I and *Bam*HI, respectively, so that the product could be ligated directly into pET15b. Mutagenesis was carried out in three steps. The initial PCR reaction was performed with either mutagenic primer 5′-GGGCGGTTG**ACT**ATTGATACCACCGGTCT-3′ for the D102T mutation or 5′-GGGCGGTTG**AAT**ATTGATACCACCGGTCT-3′ for the D102N mutation (mutation sites shown in bold) and the downstream primer 5′-GAATTCGGATCCGGTTTATATGCTTG-3′ having a *Bam*HI site (underlined). A 100-μL reaction contained 50 ng of template plasmid DNA, 30 pmol each of the mutagenic primer and downstream primer, 3 U of *Pfu* DNA polymerase (Promega), 0.2 mM dNTPs, 20 mM Tris, pH 8.75, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 0.1% Triton X-100 and 0.1 mg/mL bovine serum albumin. The mixture was denatured at 95 °C for 240 s, then ten cycles of amplification (95 °C, 30 s; 47 °C, 60 s; 72 °C, 70 s) were performed, followed by 5 min extension at 72 °C. Fifty picomoles of the upstream primer were added (5′-GGCAGCCATATGCGACTTGATAAA-3′, *Nde*I site underlined) and the reaction mixture subjected to the same amplification program. Finally, 50 pmol of downstream primer were added and the sample was subjected to the same amplification program again. The amplified product was purified by gel electrophoresis, digested with *Nde*I and *Bam*HI, and ligated with similarly digested and purified pET15b for 16 h at 16 °C. The ligation mixture was transformed into Novablue cells (Novagen) by standard methods yielding three positive clones out of three tested for D102T and three positive out of three tested for D64N. DNA sequencing of the isolated plasmids verified that the expected mutation had been produced at the desired site (see Fig. 2).

Strain constructs

For Figure 1 and the growth rate experiment, both wild-type MG1655 and mutant MG1655(*rsuA*$^-$) cells were transformed with pTrc99A or pTrc99A/rsuA with selection on LB plus 0.1 mg/mL carbenicillin generating the four experimental strains: MG1655/pTrc99A, MG1655/pTrc99A(rsuA), MG1655 (*rsuA*$^-$)/pTrc99A, and MG1655(*rsuA*$^-$)/pTrc99A(rsuA). For the overexpression experiments of Figures 3 and 4, BL21/DE3(*rsuA*$^-$) cells were transformed with plasmids pET15b/rsuA, pET15b/rsuA-D102T, and pET15b/rsuA-D102N with selection as above.

RNA isolation

Total RNA used for Ψ sequencing was isolated from the experimental cell lines as described previously (King & Schlessinger, 1983) with omission of the LiCl step and with the following modification. To induce overexpression of the *rsuA* gene, 1 mM IPTG was added to the growing cell culture once an OD_{600} of approximately 0.5 had been reached, and the culture was harvested after incubation for an additional 60 min.

Growth rate measurements

Exponential growth rates were determined in both rich and minimal medium at 24 °C, 37 °C, and 42 °C. Cultures were started with a single colony isolate and shaken at the prescribed temperature overnight. Overnight cultures were diluted 1:100 (minimal medium) or 1:50 (rich medium) and grown to OD_{600} of 0.1–0.2. These exponentially growing cultures were again diluted 1:100 or 1:50 respectively, and cellular growth was monitored by A_{600} measurements as a function of time. Each culture contained 0.1 mg/mL carbenicillin and 1 mM IPTG.

Sequence analysis

Database searches of the nonredundant (NR) protein and DNA databases, the dbEST DNA database of EST sequences, and the HTGS DNA database of high throughput genomic sequences at the National Center for Biotechnology Information (Benson et al., 1999) were performed using BLASTP2 and TBLASTN2 (Altschul et al., 1997; Altschul & Koonin, 1998). The BLOSUM62 matrix and other default BLAST parameters were used, except during some PSI-BLAST searches in which the minimum *E* value for inclusion was reduced to 0.01. Iterative motif searches were performed using the MoST algorithm, (Tatusov et al., 1994). MoST searches of the NR database were performed using progressively lower *r* values (down to $r = 0.1$) to search for distant motif family members. The clustering of the 14 amino acid Motif II and TruA sequences was done using CLUSTALX (Thompson et al., 1997) and a neighbor-joining phylogenetic tree drawing program, NJplot (Perrière & Gouy, 1996). Organism codes are taken from a listing maintained by the SWISS-PROT database (expasy.hcuge.ch/cgi-bin/speclist). LOCUS names for protein sequences not yet in SWISS-PROT are presented in a modified SWISS-PROT format, for example, RLUD/CHLTR instead of RLUD-CHLTR, to indicate the choice of protein name is ours, not necessarily the one that will be chosen by SWISS-PROT (Bairoch & Apweiler, 1999).

Other methods and materials

Ψ sequencing was performed as described previously (Bakin & Ofengand, 1993, 1998). *Pfu* DNA polymerase was from Promega. All other enzymes and primers were obtained, and polyacrylamide gel electrophoresis performed, as described previously (Raychaudhuri et al., 1998).

ACKNOWLEDGMENTS

The initial part of the experimental work described here was performed at the now-defunct Roche Institute of Molecular Biology. Supported by funds from Hoffman-La Roche, Inc., The Medical Research Council of Canada (B.G.L.), and by a Markey Foundation grant to the Department of Biochemistry and Molecular Biology, University of Miami School of Medicine.

Received January 29, 1999; returned for revision March 4, 1999; revised manuscript received March 13, 1999

REFERENCES

Altschul SF, Koonin EV. 1998. Iterated profile searches with PSI-BLAST—a tool for discovery in protein databases. *Trends Biochem Sci 23*:444–447.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res 25*:3389–3402.

Bairoch A, Apweiler R. 1999. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res 27*:49–54.

Bakin A, Kowalak JA, McCloskey JA, Ofengand J. 1994a. The single pseudouridine residue in *Escherichia coli* 16S RNA is located at position 516. *Nucleic Acids Res 22*:3681–3684.

Bakin A, Lane BG, Ofengand J. 1994b. Clustering of pseudouridine residues around the peptidyl transferase center of yeast cytoplasmic and mitochondrial ribosomes. *Biochemistry 33*:13475–13483.

Bakin A, Ofengand J. 1993. Four newly located pseudouridylate residues in *Escherichia coli* 23S ribosomal RNA are all at the peptidyl transferase center: Analysis by the application of a new sequencing technique. *Biochemistry 32*:9754–9762.

Bakin A, Ofengand J. 1995. Mapping of the thirteen pseudouridine residues in *Saccharomyces cerevisiae* small subunit ribosomal RNA to nucleotide resolution. *Nucleic Acids Res 23*:3290–3294.

Bakin A, Ofengand J. 1998. Mapping of pseudouridine residues in RNA to nucleotide resolution. In: Martin R, ed. *Methods in molecular biology: Protein synthesis: Methods and protocols*. Totowa, New Jersey: Humana Press, Inc. pp 297–309.

Becker HF, Motorin Y, Planta RJ, Grosjean H. 1997. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Ψ55 in both mitochondrial and cytoplasmic tRNAs. *Nucleic Acids Res 25*:4493–4499.

Benson DA, Boguski MS, Lipman DJ, Ostell J, Ouellette BFF, Rapp BA, Wheeler DL. 1999. GenBank. *Nucleic Acids Res 27*:12–17.

Blattner FR, Plunkett G3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science 277*:1453–1474.

Conrad J, Sun D, Englund N, Ofengand J. 1998. The *rluC* gene of *Escherichia coli* codes for a pseudouridine synthase which is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23S ribosomal RNA. *J Biol Chem 273*:18562–18566.

Grosjean H, Szweykowska-Kulinska Z, Motorin Y, Fasiolo F, Simos G. 1997. Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs: A review. *Biochimie 79*:293–302.

Gu X-R, Nicoghosian K, Cedergren, RJ. 1984. Glycine and asparagine tRNA sequences from the archaebacterium, *Methanobacterium thermoautotrophicum*. *FEBS Lett 176*:462–466.

Hall RH. 1971. *The modified nucleosides in nucleic acids*. New York, New York: Columbia University Press. pp 146–147.

Hamilton CM, Aldea H, Washburn BK, Babitzke P, Kushner SR. 1989. New method for generating deletions and gene replacements in *Escherichia coli*. *J Bacteriol 171*:4617–4622.

Huang L, Pookanjanatavip M, Gu X, Santi DV. 1998. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry 37*:344–351.

Jensen KJ. 1993. The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J Bacteriol 175*:3401–3407.

King TC, Schlessinger D. 1983. S1 nuclease mapping analysis of ribosomal RNA processing in wild-type and processing deficient *Escherichia coli*. *J Biol Chem 258*:12034–12042.

Koonin EV. 1996. Pseudouridine synthases: Four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res 24*:2411–2415.

Korber P, Zander T, Herschlag D, Bardwell JCA. 1999. A new heat shock protein that binds nucleic acids. *J Biol Chem 274*:249–256.

Kowalak JA, Bruenger E, Hashizume T, Peltier JM, Ofengand J, McCloskey JA. 1996. Structural characterization of 3-methylpseudouridine in Domain IV from *E. coli* 23S ribosomal RNA. *Nucleic Acids Res 24*:688–693.

Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D. 1998. The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. *Genes & Dev 12*:527–537.

Lecointe F, Simos G, Sauer A, Hurt EC, Motorin Y, Grosjean H. 1998. Characterization of yeast protein Deg1 as pseudouridine synthase (Pus3) catalyzing the formation of Ψ_{38} and Ψ_{39} in tRNA anticodon loop. *J Biol Chem 273*:1316–1323.

Marvel CC, Arps PJ, Rubin BC, Kammen HO, Penhoet EE, Winkler ME. 1985. hisT is part of a multigene operon in *Escherichia coli* K-12. *J Bacteriol 161*:60–71.

Massenet S, Motorin Y, Lafontaine DLJ, Hurt EC, Grosjean H, Branlant C. 1999. Pseudouridine mapping in the *Saccharomyces cerevisae* spliceosomal UsnRNAs reveals that pseudouridine synthase Pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol Cell Biol 19*:2142–2154.

McGeoch DJ. 1990. Protein sequence comparisons show that the ‘pseudoproteases’ encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucleic Acids Res 18*:4105–4110.

Miller JH. 1992. *A short course in bacterial genetics.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 357–364.

Niu L, Ofengand J. 1999. Cloning and characterization of the 23S RNA pseudouridine 2633 synthase from *Bacillus subtilis*. *Biochemistry 38*:629–635.

Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J. 1995. Purification, cloning, and properties of the tRNA Ψ55 synthase from *Escherichia coli*. *RNA 1*:102–112.

Ofengand J, Bakin A, Nurse, K. 1993. The functional role of conserved sequences of 16S ribosomal RNA in protein synthesis. In: Nierhaus KH, Subramanian AR, Erdmann VA, Franceschi F, Wittman-Liebold B, eds. *The translational apparatus.* New York, New York: Plenum Press. pp 489–500.

Ofengand J, Bakin A. 1997. Mapping to nucleotide resolution of pseudouridine residues in large subunit ribosomal RNAs from representative eukaryotes, prokaryotes, archaebacteria, mitochondria, and chloroplasts. *J Mol Biol 266*:246–268.

Ofengand J, Fournier M. 1998. The pseudouridine residues of rRNA: Number, location, biosynthesis, and function. In: Grosjean H, Benne R, eds. *Modification and editing of RNA.* Washington, DC: ASM Press. pp 229–253.

Perrière G, Gouy M. 1996. WWW-query: An on-line retrieval system for biological sequence banks. *Biochimie 78*:364–369.

Picard V, Ersdal-Badju E, Lu A, Bock SC. 1994. A rapid and efficient one-tube PCR-based mutagenesis technique using Pfu DNA polymerase. *Nucleic Acids Res 22*:2587–2591.

Prangishvili D, Klenk HP, Jakobs G, Schmiechen A, Hanselmann C, Holz I, Zillig W. 1998. Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. *J Biol Chem 273*:6024–6029.

Raychaudhuri S, Conrad J, Hall BG, Ofengand J. 1998. A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of *Escherichia coli*. *RNA 4*:1407–1417.

Raychaudhuri S, Niu L, Conrad J, Ofengand J. 1999. Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA. *J Biol Chem.* Submitted.

Santer M, Santer U, Nurse K, Bakin A, Cunningham P, Zain M, O'Connell D, Ofengand J. 1993. Functional effects of a G to U base change at position 530 in a highly conserved loop of *Escherichia coli* 16S RNA. *Biochemistry 32*:5539–5547.

Simos G, Tekotte H, Grosjean H, Segref A, Sharma K, Tollervey D, Hurt EC. 1996. Nuclear pore proteins are involved in the biogenesis of functional tRNA. *EMBO J 15*:2270–2284.

Supekova L, Supek F, Nelson N. 1995. The *Saccharomyces cerevisiae* VMA10 is an intron-containing gene encoding a novel 13-kDa subunit of vacuolar H(+)-ATPase. *J Biol Chem 270*:13726–13732.

Tatusov RL, Altschul SF, Koonin EV. 1994. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc Natl Acad Sci USA 91*:12091–12095.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res 25*:4876–4882.

Tsui H-CT, Arps PJ, Connolly DM, Winkler ME. 1991. Absence of *hisT*-mediated tRNA pseudouridylation results in a uracil requirement that interferes with *Escherichia coli* K-12 cell division. *J Bacteriol 173*:7395–7400.

Watkins NJ, Gottschalk A, Neubauer G, Kastner B, Fabrizio P, Mann M, Luhrmann R. 1998. Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure. *RNA 4*:1549–1568.

Wrzesinski J, Bakin A, Nurse K, Lane B, Ofengand J. 1995a. Purification, cloning, and properties of the 16S RNA Ψ516 synthase from *Escherichia coli. Biochemistry 34*:8904–8913.

Wrzesinski J, Nurse K, Bakin A, Lane BG, Ofengand J. 1995b. A dual-specificity pseudouridine synthase: Purification and cloning of a synthase from *Escherichia coli* which is specific for both Ψ746 in 23S RNA and for Ψ32 in $tRNA^{Phe}$. *RNA 1*:437–448.

Zyskind JW, Bernstein SI. 1992. *Recombinant DNA laboratory manual.* San Diego, California: Academic Press. p 187.