

---

# Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*

---

MARK DEL CAMPO, JAMES OFENGAND, and ARUN MALHOTRA

Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, Florida 33101-6129, USA

## ABSTRACT

*Escherichia coli* pseudouridine synthase RluD makes pseudouridines 1911, 1915, and 1917 in the loop of helix 69 in 23S RNA. These are the most highly conserved ribosomal pseudouridines known. Of 11 pseudouridine synthases in *E. coli*, only cells lacking RluD have severe growth defects and abnormal ribosomes. We have determined the 2.0 Å structure of the catalytic domain of RluD (residues 77–326), the first structure of an RluA family member. The catalytic domain folds into a mainly antiparallel  $\beta$ -sheet flanked by several loops and helices. A positively charged cleft that presumably binds RNA leads to the conserved Asp 139. The RluD N-terminal S4 domain, connected by a flexible linker, is disordered in our structure. RluD is very similar in both catalytic domain structure and active site arrangement to the pseudouridine synthases RsuA, TruB, and TruA. We identify five sequence motifs, two of which are novel, in the RluA, RsuA, TruB, and TruA families, uniting them as one superfamily. These results strongly suggest that four of the five families of pseudouridine synthases arose by divergent evolution. The RluD structure also provides insight into its multisite specificity.

**Keywords:** 23S rRNA; RluA family; motifs;  $\Psi$ 1911;  $\Psi$ 1915;  $\Psi$ 1917

## INTRODUCTION

Cells devote time and metabolic energy to making posttranscriptional modifications to their RNAs. The most abundant single nucleoside modification is the isomerization of uridine (U) to pseudouridine (5-ribosyluracil,  $\Psi$ ). To date,  $\Psi$  has been detected in tRNA, rRNA, tmRNA, snRNA, and snoRNA, but not, so far, in mRNA. The precise role of  $\Psi$  in these RNAs has remained elusive, despite the many hints to its physiological importance that have been uncovered. For example, most ribosomal  $\Psi$  cluster in the vicinity of the peptidyl transferase center of the ribosome, the heart of ribosome function (Ofengand et al. 2001). Also, a single  $\Psi$  in U2 snRNA appears to stabilize the extrahelical branch point adenosine used in RNA splicing (Newby and Greenbaum 2002).

$\Psi$  synthases are the enzymes responsible for the U-to- $\Psi$  conversion in a reaction that requires no cofactor or energy

source. Amino acid sequence similarity has been used to group  $\Psi$  synthases into five families, each named for the first *Escherichia coli* family member to be characterized: TruA, TruB, RsuA, RluA, and TruD. Statistically significant sequence similarity is found between the RluA and RsuA families using the BLASTP algorithm (~20% identity over ~200 residues) and it can be extended to include only the TruB family using the PSI-BLAST algorithm (Ofengand and Rudd 2000). The sequence similarity between families is located primarily in short sequence motifs. The RsuA and RluA families share three motifs, the TruB family shares only motifs I and II, and only motif II is shared by the TruA family (Gustafsson et al. 1996; Koonin 1996; Huang et al. 1998b). Motif II contains an aspartate residue that is conserved among these four families and is essential for catalytic activity (Del Campo et al. 2001). Even the TruD family has an essential conserved aspartate, although it lacks a conserved motif II (Kaya and Ofengand 2003). In the past three years, the *E. coli* TruA (Foster et al. 2000), TruB (Hoang and Ferré-D'Amaré 2001), and RsuA (Sivaraman et al. 2002) crystal structures have shown that despite their sequence conservation being limited to three or less motifs, they share the same basic fold in their catalytic domain along with placement of residues in the active site—clearly

---

**Reprint requests to:** Arun Malhotra, Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Box 016129, Miami, FL 33101-6129, USA; e-mail: malhotra@miami.edu; fax: (305) 243-3955.

Article and publication are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.5187404>.

showing  $\Psi$  synthases arose by divergent evolution (Mueller 2002). In addition, the cocrystal structure of TruB with a 22mer T loop RNA elegantly showed that  $\Psi$  synthases gain access to their substrate U by a base flipping mechanism that places the U in close proximity to the catalytic aspartate (Hoang and Ferré-D'Amaré 2001).

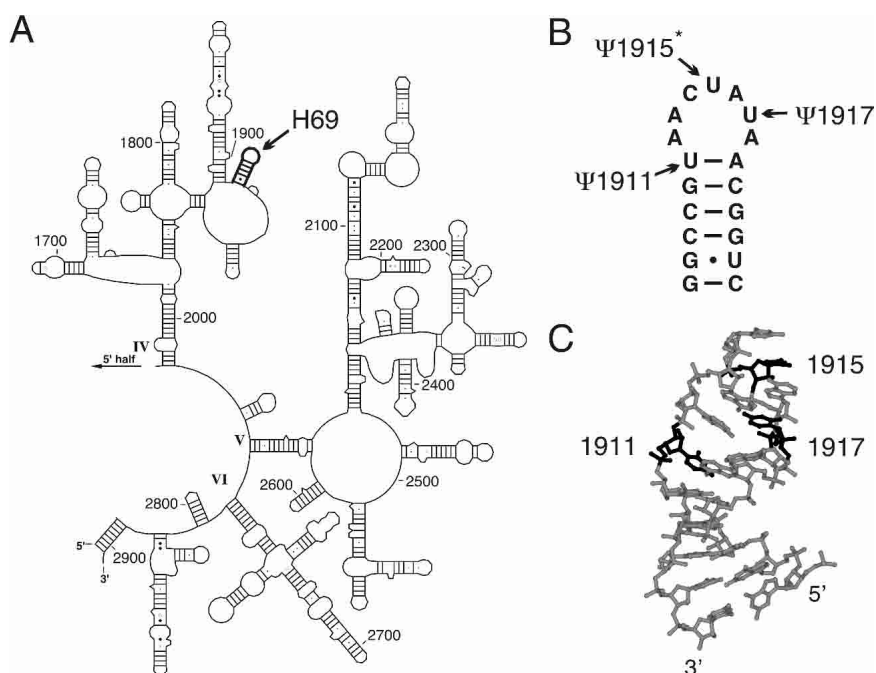
In *E. coli*, the entire collection of  $\Psi$  in both rRNA and tRNA and the 11 responsible  $\Psi$  synthases have been determined (Kaya and Ofengand 2003). Deletion of each of the synthases in turn has shown that RluD is by far the most physiologically important synthase. Cells lacking functional RluD have a growth rate 20% of wild type (Gutgsell et al. 2001) and their ribosome particles are abnormal (Ofengand et al. 2001). Hence, cells require RluD and/or its  $\Psi$  products for normal ribosome function. RluD is an RluA family member responsible for making  $\Psi$  1911, 1915, and 1917 in 23S RNA (Fig. 1A,B,C). These three  $\Psi$  are in the loop of helix 69 (H69) and are the most conserved ribosomal  $\Psi$  known (Ofengand 2002). In the 70S ribosome structure, H69 is the main feature of the B2a intersubunit bridge and makes contacts with both P- and A-site tRNAs (Yusupov et al. 2001; Bashan et al. 2003). Here we describe the structure of the catalytic domain of RluD, the first structure of any RluA family member. The structure provides clues toward understanding the multisite specificity of RluD and

reveals new similarities among the first four  $\Psi$  synthase families.

## RESULTS

### Overall structure of RluD

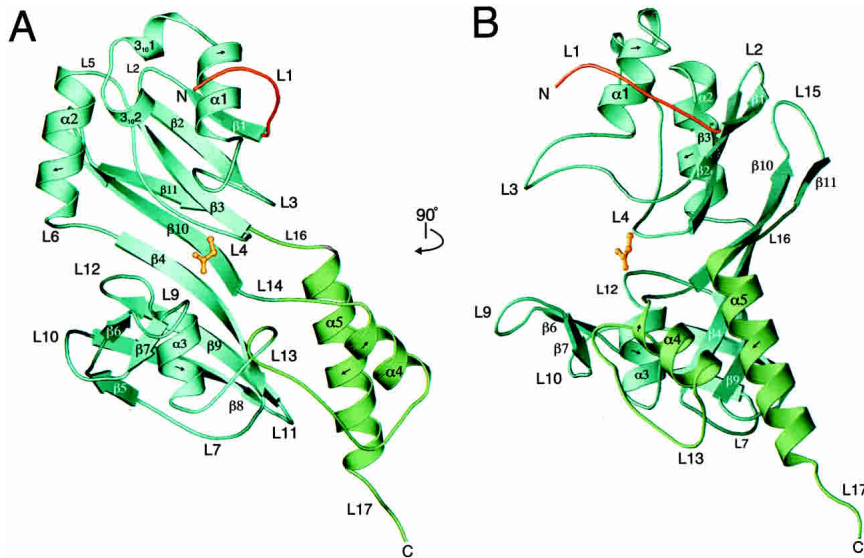
The 2.0 Å structure of the catalytic domain of RluD ( $R = 0.219$ ,  $R_{\text{free}} = 0.232$ ) consists of 250 amino acid residues (Phe 77 to Leu 326 at the C terminus) and 261 water molecules. The N-terminal domain consisting of a 23 residue affinity tag and residues M1 to R76 is disordered in our electron density maps and could not be modeled. The catalytic domain (Fig. 2A,B) folds into an extended nine-strand  $\beta$ -sheet consisting of two antiparallel sheets ( $\beta$ 1- $\beta$ 2- $\beta$ 3 and  $\beta$ 11- $\beta$ 10- $\beta$ 4- $\beta$ 9- $\beta$ 8- $\beta$ 5) joined by a parallel interaction between one strand in each sheet ( $\beta$ 3 and  $\beta$ 10). The front face of the sheet contains secondary structure elements clustered into two groups: four helices ( $\alpha$ 1- $3_{10}$ 1- $3_{10}$ 2- $\alpha$ 2) and several large loops in one group and one helix ( $\alpha$ 3), two short strands ( $\beta$ 6- $\beta$ 7), and more loops in the other. The juxtaposition of these groups forms a deep, central cleft in RluD with average dimensions 25 Å long by 10 Å wide by 14 Å deep. The essential, catalytic aspartate residue (Asp 139) is located at the base of this cleft. Off to the far side of the main sheet is a subdomain, which we term the tail region, consisting of two helices ( $\alpha$ 4 and  $\alpha$ 5) and several loops.



**FIGURE 1.** The substrate of RluD is helix 69 in 23S RNA. (A) A schematic of the secondary structure of the 3' half of *E. coli* 23S RNA from the Comparative RNA Web site (<http://www.rna.icmb.utexas.edu>). H69 is indicated by an arrow. (B) Closer view of H69 showing the  $\Psi$  made by RluD. The asterisk indicates N3-methyl- $\Psi$ . (C) H69 from the structure of the 50S ribosomal subunit of *Deinococcus radiodurans*, which was modeled with unmodified uridines (PDB entry 1NKW; Harms et al. 2001). The sequence of H69 in *D. radiodurans* is identical to that of *E. coli*. In this panel, uridines 1911, 1915, and 1917 are black and the remainder of H69 is gray. Panel C was generated with PyMOL (DeLano Scientific).

### RNA binding cleft and the active site

One or more residues from helices  $3_{10}2$ ,  $\alpha$ 2, and  $\alpha$ 4, strands  $\beta$ 3,  $\beta$ 4,  $\beta$ 6,  $\beta$ 7,  $\beta$ 9, and  $\beta$ 10, and loops L3, L4, L6, L9, L10, L12, L13, and L14 come together to make up the walls and floor of the central cleft in RluD. The walls of the cleft have a highly positive character contributed by residues Arg 131, Arg 137, Arg 165, Arg 169, Arg 191, Lys 206, and Arg 232 (Fig. 3A,C). This positive cleft could be used to bind and position a largely negative substrate RNA, consistent with other  $\Psi$  synthase structures (Foster et al. 2000; Hoang and Ferré-D'Amaré 2001; Sivaraman et al. 2003). The cleft walls start out wide ( $\sim 20$  Å at the widest point) and taper inwards ( $\sim 2.7$  Å at the narrowest point) over many side chain and backbone protrusions, eventually leading to a roughly spherical, central cavity ( $\sim 6.5$  Å long,  $6.5$  Å wide,  $5.5$  Å deep). The cavity is formed largely by the side chains of Asp 139, Thr 142, Leu



**FIGURE 2.** Model of the catalytic domain of RluD. (A) Ribbon cartoon of secondary structure elements shown in cyan, except for the linker to the N-terminal domain (residues 77–84 [L1]) in red and the tail region (residues 249–276 [L13– $\alpha$ 4-L14] and 300–326 [L16– $\alpha$ 5-L17]) in green. The side chain of catalytic Asp 139 (yellow) is shown in ball and stick representation. Note: The  $C_{\alpha}$  of Asp 139 appears to be separated from L4 because the ribbon cartoon is smoothed. (B) A 90° turn of the view shown in A. Generated with PyMOL (DeLano Scientific).

228, Ile 236, Arg 237, and Leu 279 (Fig. 3C). The presence of the catalytically essential Asp 139, as well as other conserved residues described in  $\Psi$  synthase active sites (see below), defines this cavity as the active site of RluD. The side chain of Asp 139 hydrogen bonds to two of eight water molecules that occupy the active site.

### Crystal contacts in the putative RNA binding cleft

Despite solving the structure of RluD in the absence of any RNA, the presumed RNA-binding cleft is not empty. In our crystals, symmetry-related RluD molecules pack such that the negatively charged tail region (Fig. 3B) of one symmetry mate (in green) is bound in the positively charged cleft of another (Fig. 3D, in red). The tail region of a symmetry mate occludes most of the cleft and completely occludes the catalytic pocket of RluD. This interaction involves 28 cleft residues and 17 tail region residues through numerous bridging water molecules, hydrogen bonds, and van der Waals contacts (Fig. 3E). Approximately 14% of the total solvent-accessible surface area (SASA) of a single RluD molecule is buried in this crystal packing interaction: 7% of the cleft and 7% of the tail region. This is on the order of a protein–RNA interaction, considering that ~9% of the SASA of TruB is buried in contacts with a 22mer RNA substrate (Hoang and Ferré-D’Amaré 2001). However, this interaction is obviously a crystal packing artifact, because, in nature, the cleft is needed for binding RNA. It also renders this crystal form useless for binding small substrate nucleosides or RNAs.

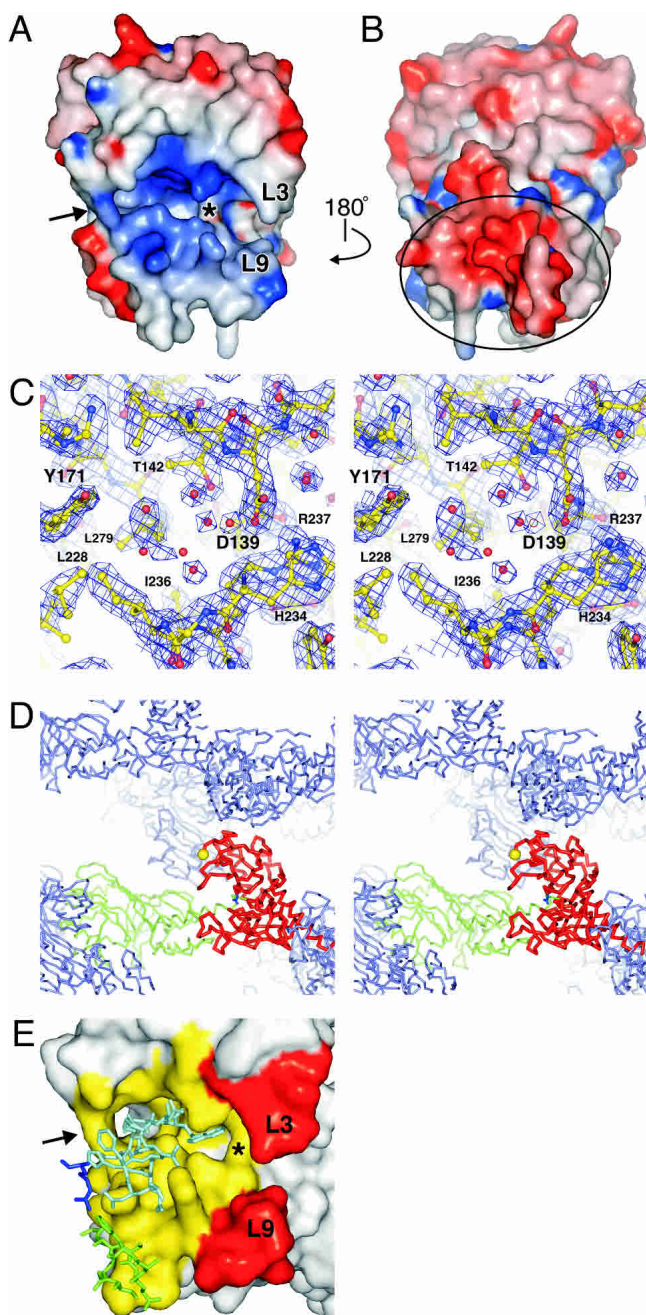
Part of the cleft is actually closed off by this crystal packing interaction. The side chain of Arg 165 from loop L6 closes one side of the cleft (see arrow, Fig. 3A,E) because it is making a salt bridge to Glu 322 in the tail region of a symmetry mate. In this position, Arg 165 also makes a single hydrogen bond to the carbonyl of Glu 229 in loop L12. In the absence of the salt bridge, the side chain of Arg 165 could be in another conformation, reopening one end of the cleft, and effectively making the entire cleft larger.

### Four families of $\Psi$ synthases are structural neighbors

A Dali (Holm and Sander 1993) search identified the  $\Psi$  synthases RsuA, TruB, and TruA as the closest structural neighbors to the catalytic domain of RluD. The resulting superposition of their  $C_{\alpha}$  backbones and alignment of their sequences are shown in Figures 4A and 5, respectively. RsuA is the closest neighbor ( $Z$  score = 11.4; RMSD = 2.8 Å over 145  $C_{\alpha}$  atoms), followed by TruB ( $Z$  score = 8.5; RMSD = 3.0 Å over 142  $C_{\alpha}$ ), and then TruA ( $Z$  score = 8.0; RMSD = 3.7 Å over 159  $C_{\alpha}$ ). The superimposed structures clearly illustrate the core fold of the four  $\Psi$  synthase families, which includes the central beta-sheet arrangement, two helices, and two loops ( $\beta$ 2,  $\alpha$ 1, L4,  $\beta$ 3,  $\beta$ 4,  $\beta$ 8,  $\beta$ 9, L12,  $\alpha$ 3,  $\beta$ 10; see Fig. 5), and the conserved placement of the active site (Fig. 4A). However, the superimposed structures also reveal that each synthase has its own additional features. A pseudouridine synthase and archaeosine tRNA guanine transglycosylase (PUA) RNA-binding domain has been appended to TruB at its C terminus (Aravind and Koonin 1999; Ferré-D’Amaré 2003), and an S4 RNA-binding domain has been added to RsuA at its N terminus (Aravind and Koonin 1999; Sivaraman et al. 2002). Interestingly, the tail region of RluD is a unique feature. Only TruA has a loop in this area, but it is used for dimerization contacts.

Five residues in the active sites of RluD, RsuA, TruB, and TruA are conserved. The catalytic aspartate, a basic arginine or lysine to make a salt bridge to the aspartate, and a tyrosine that provides a stacking interaction for the uracil base have all been previously noted (Hoang and Ferré-D’Amaré 2001). There are also two hydrophobic residues: a leucine and a second residue that can be an isoleucine or valine. These five residues appear to be completely conserved among these four families of  $\Psi$  synthases (M. Del Campo, unpubl. observations). In fact, the extended main chain atoms (N,  $C_{\alpha}$ ,  $C_{\beta}$ , C, and O) of these five residues from

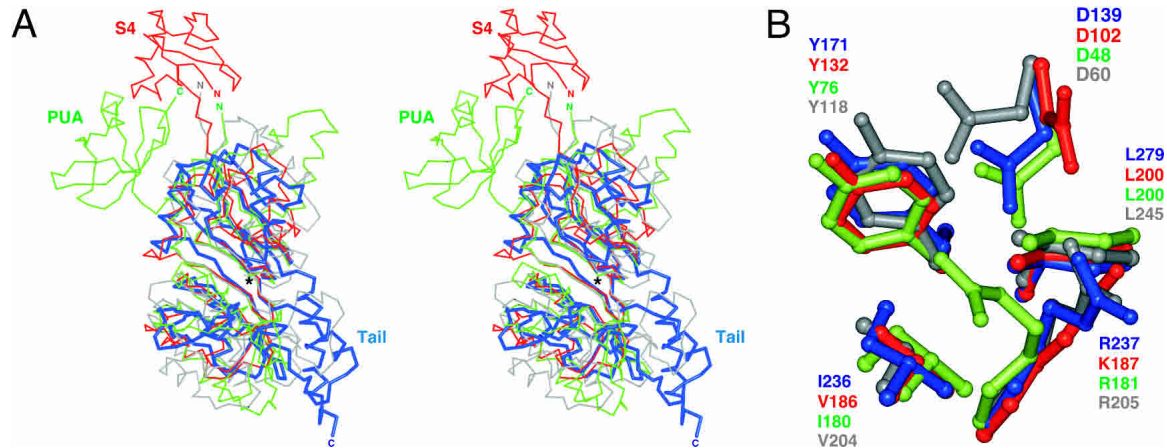
RsuA, TruB, and TruA superimpose onto RluD with RMSD's of 0.51 Å, 0.67 Å, and 0.75 Å, respectively. The side chains are aligned as shown in Figure 4B. Overall the side chains are similar with some conformational differences. Asp 102 of RsuA points away from the active site (Mueller 2002). Arg 181 of TruB points back into a position that cannot be adopted by the basic residues of RluD, RsuA, or TruA because of the presence of a histidine (His 234, His 184, and His 202, respectively) in that location (data not shown). Despite these differences, the close similarity of the active sites in size and composition implies that the base-flipping mechanism observed for TruB (Hoang and Ferré-D'Amaré 2001) is likely also used by RluD, RsuA, and TruA.



### Motifs: something old, something new

It was suggested (Ofengand and Rudd 2000), but never shown, that motif III exists in the TruB family and motifs I and III exist in the TruA family. From the structural alignment in Figure 5, it is clear that TruB contains motif III and TruA contains both motifs I and III. Motif I in the TruA family is clearly different from the other three families, consistently having a YD in place of the KP. The five conserved amino acids in the active site (Fig. 4B) are distributed among motifs II, IIa, III, and IIIa (Fig. 5). The conserved aspartate is in motif II, the arginine/lysine residue is in motif III adjacent to the isoleucine/valine residue, the tyrosine is found in a new motif we designate as IIa, and the newly recognized leucine occurs in a new motif IIIa. Motif IIa occurs after motif II in a  $\beta$ -strand ( $\beta$ 4 in RluD) and generally follows the form “+xYx###” where + is a positive side chain, x is any residue, and # is a hydrophobic residue. Motif IIIa is located adjacent to motif III in a  $\beta$ -strand ( $\beta$ 10 in RluD) and would technically be part of motif III if it were not for the insertion of a long loop following motif III in some  $\Psi$  synthase sequences, like RluD and TruA (Fig. 5). Motif IIIa differs between families, but it generally follows the form “Lxxxx#”. To determine whether all five motifs were conserved in other members of the RluA, RsuA, TruA, and TruB families, we examined sequences from each family obtained by BLAST searches (<http://www.ncbi.nlm.nih.gov/BLAST>) against databases of finished genomes at the National Center for Biotechnology Information (NCBI). All five motifs were found in 50 RluA members from 15 bacteria and 3 eukaryotes, 25 RsuA members from 15 bacteria and 1 eukaryote, 34 TruB members from 13 bacteria, 16 archaea (Yeast Cbf5-like), and 3 eukaryotes, and 37 TruA members from 16 bacteria, 12 archaea, and 3 eukaryotes (data not shown). There are no known RluA or RsuA family members in the archaea (Ofengand 2002). This preliminary

**FIGURE 3.** Active site and crystal packing. (A) Molecular surface of RluD colored by local electrostatic potential (using GRASP; Nicholls et al. 1991): positive is blue and negative is red. This view is from above the positively charged putative RNA-binding cleft. Note the central, small, uncharged active site pocket flanked by the catalytic Asp 139 (asterisk). Arg 165 is marked by an arrow and loops L3 and L9 are indicated. (B) Same surface representation as in A turned 180° to view the negatively charged tail region (circled). (C) Stereo view of the electron density in the active site pocket from a  $2F_o - F_c$  map contoured at 2.0  $\sigma$ . Carbons are yellow, nitrogens are blue, and oxygens are red; lone oxygens represent water molecules. (D) Stereo view of the crystal packing of RluD molecules shown as  $C_\alpha$  backbones. The cleft of the central molecule (red) contacts the tail region of a symmetry related molecule (green). The yellow sphere marks the position of the N-terminal residue Phe 77. (E) The interaction between the cleft of one RluD molecule and the tail region of a symmetry-related molecule. Molecular surface of RluD is shown in gray, residues involved in the contact are yellow, and residues from two loops (L3 and L9) with highest main chain  $B$ -factors are red. Residues 219 (blue), 262–269 (green), and 314–326 (cyan) of a symmetry mate are shown in stick representation. Note the position of Asp 139 (asterisk) and Arg 165 (arrow). All panels generated with PyMOL (DeLano Scientific).



**FIGURE 4.** The conserved core fold of four families of  $\Psi$  synthases. (A) Stereo view of the  $C_{\alpha}$  backbones of RsuA (PDB entry 1KSK; red), TruB (PDB entry 1K8W; green), and a monomer of TruA (PDB entry 1DJO; gray) superimposed onto RluD (blue) by the Dali server (Holm and Sander 1993). Nonoverlapping domains are labeled. An asterisk marks the location of the catalytic site in all four enzymes. (B) Superimposed side chains (starting at the  $C_{\alpha}$ ) from five conserved active site residues in RluD, RsuA, TruB, and TruA. The N,  $C_{\alpha}$ ,  $C_{\beta}$ , C, and O atoms of all five residues from RsuA, TruB, and TruA were superimposed onto the equivalent five residues of RluD using LSQMAN (Kleywegt 1999). Colors are the same as in A. All panels generated with PyMOL (DeLano Scientific).

analysis shows that motifs I, II, IIa, III, and IIIa are conserved in the RluA, RsuA, TruB, and TruA families. Thus, it appears the first four families of  $\Psi$  synthases to be characterized are more similar in sequence than originally described (Koonin 1996) and comprise one  $\Psi$  synthase superfamily.

### Clues toward understanding the multisite specificity of RluD

RluD has been shown to make all three substrate  $\Psi$  on H69 of an *in vitro* transcript of 23S RNA. Thus, ribosomal proteins are not essential (Raychaudhuri et al. 1998; Wrzesinski et al. 2000). There are three ribosome structures available that include H69, so it would be reasonable to try to dock these structures into the putative RNA binding cleft of RluD. However, H69 is disordered in the 50S structure from *Haloarcula marismortui* (Ban et al. 2000) and in the 70S ribosome from *Thermus thermophilus*, only the backbone of this helix is discernable because of limited resolution (Yusupov et al. 2001). The only structure of H69 that is ordered and to sufficient resolution is from the 50S of another eubacterium *Deinococcus radiodurans* (Fig. 1C; Harms et al. 2001). Attempts at docking this RNA into the cleft of RluD did not yield any orientations without steric clashes (data not shown) because the cleft, although big enough to accommodate this RNA in several ways, does not provide a complementary surface to the H69 structure. However, it did define some restraints. It is clear that uridines 1911, 1915, or 1917 cannot get into the active site cavity without base flipping from their positions in the stem-loop, and the active site can only accommodate one uridine at a time. Thus, more than one conformation of the RNA and/or

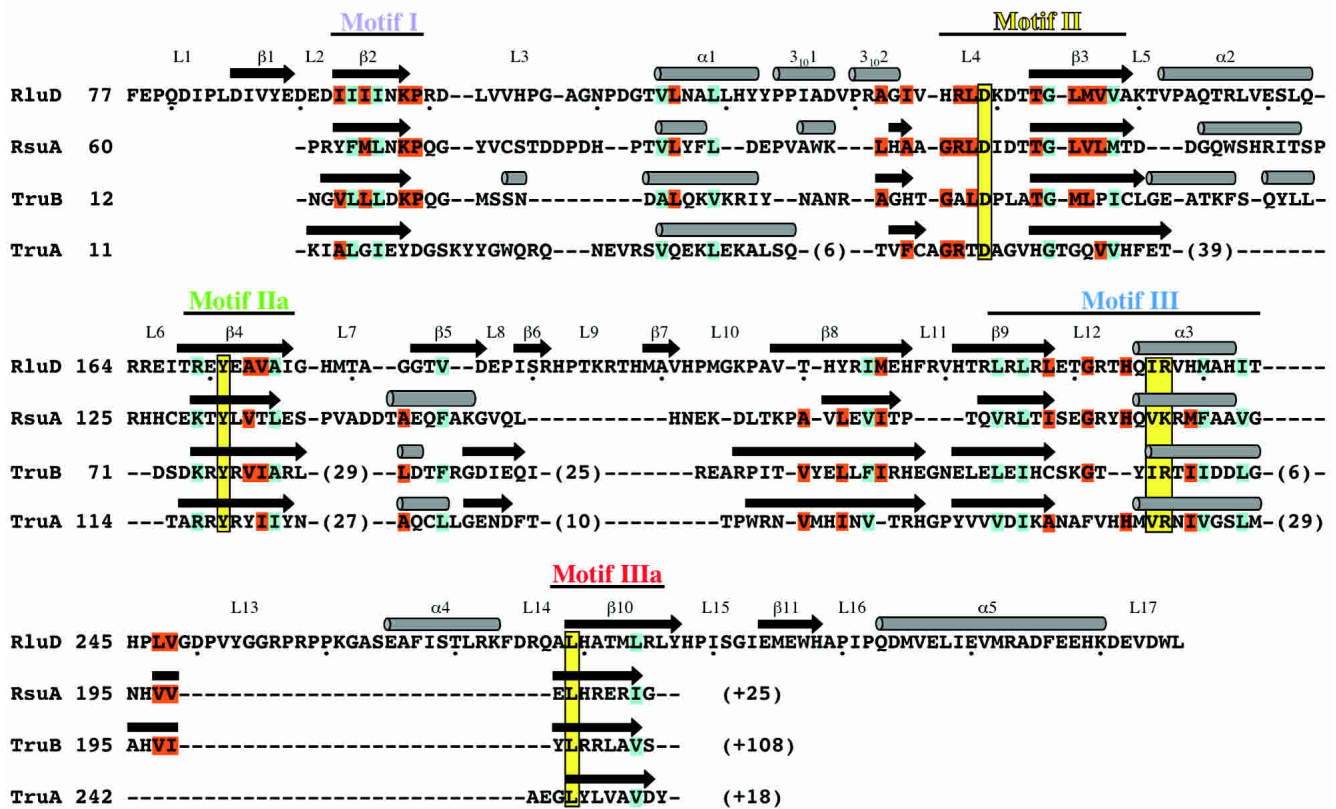
RluD must exist to allow each uridine to flip into the active site. Two conformations of H69 have already been noted (Harms et al. 2001).

### The N3-methyl- $\Psi$ at 1915

In *E. coli* 23S RNA, the naturally occurring modification at uridine 1915 is N3-methyl- $\Psi$  (Kowalak et al. 1996). This means an as yet unidentified *E. coli* methyltransferase also recognizes H69, and specifically modifies position 1915. Cells lacking RluD appear to be methylated on uridine 1915, so the methyltransferase would appear to not require  $\Psi$  for action (Raychaudhuri et al. 1998; Gutsell et al. 2001). However, it is not known if RluD can convert N3-methyluridine to N3-methyl- $\Psi$  or whether the methyltransferase can convert  $\Psi$  to N3-methyl- $\Psi$ . Perhaps both reactions are possible. If RluD can convert N3-methyluridine to N3-methyl- $\Psi$  then the active site pocket of RluD must be able to accommodate the uracil base plus the N3-methyl group. To test this, we modeled the structure of N3-methyluridine (Partridge and Pritchard 1995) into the active site pocket of RluD (data not shown). Although the active site pocket appears to be big enough to accommodate the N3-methyluracil base, no conclusion can be made regarding the orientation or placement of this modification in the absence of structural data on RNA substrate binding.

### The N-terminal S4 domain

The S4 domain, named after the rRNA-binding domain of ribosomal protein S4, is a small, modular domain found in many proteins either known or predicted to bind RNA (Aravind and Koonin 1999; Staker et al. 2000). It is found



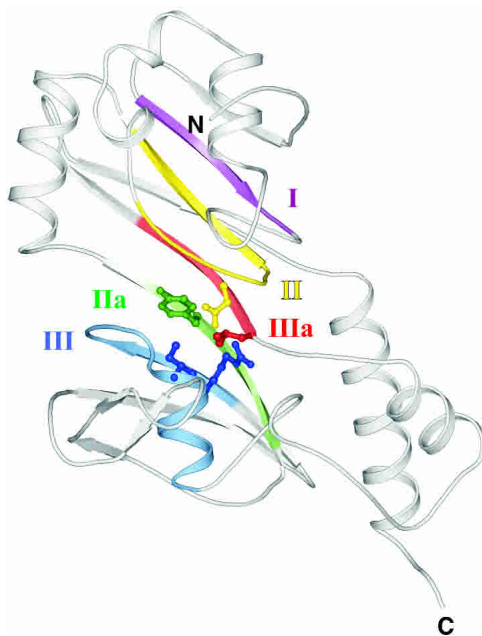
**FIGURE 5.** Structural sequence alignment of RluD, RsuA, TruB, and TruA. The entire sequence of RluD (residues 76–326) is shown with every 10th residue marked *underneath*. Parts of RsuA, TruB, and TruA that were superimposed onto RluD by the Dali server (Holm and Sander 1993) are shown *underneath*. Gaps in sequences are noted by dashes. The number of residues left out because they do not align to RluD is shown in parentheses. Secondary structure elements are shown *above* their respective sequences, and are labeled for RluD. Positions of the classical  $\Psi$  synthase motifs I, II, and III and two new motifs IIa and IIIa are indicated. Conserved active site residues are shown boxed and highlighted in yellow; other conserved residues are highlighted light blue if they occur in all four sequences or orange if they occur in three sequences.

at the N terminus of some RsuA and RluA family members that modify rRNA, and it has been shown that RluD residues Gln 18-Asn 68 fit the consensus for an S4 domain (Sivaraman et al. 2002). Hence, the N-terminal domain of RluD most likely folds into a similar structure. The S4 domain should be connected to the catalytic domain of RluD through the end of loop L1 (Fig. 2B). Because extended loops are known to be substrates for proteases, as was noted for the linker between RsuA and its S4 domain (Fig. 4A; Sivaraman et al. 2002), it is possible to think that the S4 domain of RluD in our crystals is not disordered but rather has been removed by a protease. Three lines of evidence argue that the domain is present but disordered. First, RluD crystals analyzed by gel electrophoresis showed migration consistent with full-length protein (data not shown). Second, there is electron density extending past F77, but it is too poor in quality to be certain in which direction the backbone is going. Third, there is more than adequate space around each molecule of RluD, in the asymmetric unit, for an S4 domain to be accommodated (Fig. 3D). The current solvent content is on the high end, ~75% (average protein crystals are 40%–60 solvent), whereas the estimated solvent

content for full-length RluD including the purification tag is 60% (Del Campo et al. 2003).

## DISCUSSION

The structure of RluD further emphasizes that the RluA, RsuA, TruB, and TruA families of  $\Psi$  synthases arose by divergent evolution from a common ancestor. Currently, the only structures with overall significant similarity to RluD in the Protein Data Bank are RsuA, TruB, and TruA. It seems that during the course of evolution, the  $\Psi$  synthase core fold has been reserved for carrying out the uridine-to- $\Psi$  conversion. In addition, we have identified even more sequence similarities in the core between these families. The assertion that motifs I, II, and III exist in all families (Ofengand and Rudd 2000) has now been shown by a structural alignment (Fig. 5). Two novel additional motifs, IIa and IIIa, are also shared among these families. The five shared motifs unite the RluA, RsuA, TruB, and TruA families into one superfamily. All five motifs come together in space to form the core of the  $\Psi$  synthase fold (Fig. 6). Whereas motif I appears to be solely structural in nature based on bio-



**FIGURE 6.** Five  $\Psi$  synthase motifs form the core of the structure of RluD. Motifs I (residues 93–99), II (residues 136–148), IIa (residues 169–175), III (residues 224–243), and IIIa (residues 279–284) are colored violet, yellow, green, blue, and red, respectively (the same colors are used for motif labels in Fig. 5). The side chains of the five conserved active site residues (highlighted yellow in Fig. 5) are shown in ball and stick representation and are colored the same as the motif to which they belong. Ribbon cartoon shown at half normal width for clarity. Generated with PyMOL (DeLano Scientific).

chemical (Spedaliere et al. 2000) and structural (Mueller 2002) evidence, motifs II, IIa, III, and IIIa not only provide hydrophobic residues to the hydrophobic core of the  $\Psi$  synthase fold, but also contribute five conserved residues to the active site pocket (Fig. 6). Three of these residues, the catalytic aspartate (motif II), a tyrosine (motif IIa), and a leucine (motif IIIa), are almost invariant among these four  $\Psi$  synthase families (M. Del Campo, unpubl. observations). Although the tyrosine has recently gained attention for possible stacking interactions with the flipped out target uridine (Hoang and Ferré-D’Amaré 2001; Ferré-D’Amaré 2003), the leucine has not been noted previously. Of the remaining two conserved residues in the active site pocket, the basic arginine/lysine and hydrophobic isoleucine/valine (motif III) both have been noted previously (Hoang and Ferré-D’Amaré 2001; Sivaraman et al. 2002).

The exact contribution of peripheral N-terminal S4 domains to the catalytic activity of  $\Psi$  synthases that modify rRNA has not yet been determined. Had the S4 domain of RluD been ordered in our structure, its exact function would still not be clear because it is at the end of a flexible linker and could adopt many conformations in the absence of bound substrate. Two (RluA and RluE) out of seven rRNA  $\Psi$  synthases in *E. coli* do not have an S4 domain, so it is not required for these enzymes to make  $\Psi$ . In fact,

plasmid-borne RluC lacking its S4 domain is able to form all three of its target  $\Psi$  in a  $\Delta rluC::kan$  mutant (*S. Jean-Charles and J. Ofengand, unpubl. results*). Whether this is true for RluD is currently being investigated.

The RluA family of  $\Psi$  synthases has traditionally been intriguing because of the multisite specificity of two of its members, RluC and RluD. There are many unanswered questions about how RluD is able to form  $\Psi$ 1911,  $\Psi$ 1915, and  $\Psi$ 1917 in H69 of 23S RNA. Is there a sequential order to the modification or is it random? Does RluD bind H69 once and make all three  $\Psi$  or are there three separate binding events? In one study, RluD was capable of making  $\Psi$ 1915 and  $\Psi$ 1917, but not  $\Psi$ 1911, on an in vitro transcript of 23S RNA, perhaps an indication that RluD can more easily access the uridines in the loop of H69 (Raychaudhuri et al. 1998). In other in vitro studies, RluD was capable of modifying all three sites plus additional sites in 23S RNA and even modifying some 16S RNA sites when the  $Mg^{2+}$  concentration was lowered (Huang et al. 1998a; Wrzesinski et al. 2000). Because higher order RNA structure depends on the presence of sufficient  $Mg^{2+}$ , this may indicate that RluD can access certain uridine residues on partially unfolded RNA. This is consistent with our attempts at manually docking the one available structure of H69 (Fig. 1C) into the RNA-binding cleft of RluD. Many initial orientations are possible, but for any of the three substrate uridines to get close to the pocket without any clashes, some of the RNA would have to be unfolded, presumably by RluD. In addition, to get each of three uridines into the active site pocket would require three unfolded conformations of H69, possibly three inherent conformations, as its structure is known to vary (Bashan et al. 2003). Indeed, recent cocrystal structures of RNA modifying enzymes with substrate RNAs illustrate various conformational rearrangements of the RNA in the bound state. For example, *E. coli* TruB flips three bases out of a 22mer T-loop RNA (Hoang and Ferré-D’Amaré 2001), prokaryotic tRNA guanine transglycosylase rearranges the loop of a 17mer RNA stem loop (Xie et al. 2003), and archaeal archaeosine tRNA guanine transglycosylase completely rearranges the D-arm of a bound tRNA (Ishitani et al. 2003).

It is likely that parts of the putative RNA-binding cleft of RluD also undergo a conformational change upon RNA binding. For instance, loop L9 (residues 192–198) is in a location equivalent to the thumb region of TruB that packs into the major groove of a bound T-loop RNA and is described as “closing” on the RNA (Hoang and Ferré-D’Amaré 2001). Because loop L9 has the highest main chain *B*-factors in the RluD structure (Fig. 3E), we speculate that it may change conformation upon RNA binding and function in a similar fashion. Loop L3 (residues 105–112) also has high *B*-factors (Fig. 3E) in the cleft and may move to accommodate RNA. The remainder of the cleft is involved in a crystal-packing interaction that may have altered the cleft in a nonphysiological way. The precise nature of the protein–

RNA interactions between RluD and H69 will require cocrystal structures of RluD with bound substrate RNA.

## MATERIALS AND METHODS

### Protein expression, purification, crystallization, and data collection

SeMet-substituted RluD expression, purification, crystallization, and data collection were carried out as previously described (Del Campo et al. 2003). In brief, full-length RluD was overexpressed with a 23 amino acid N-terminal hexa-histidine tag. SeMet-substituted RluD was purified and crystallized using ethylene glycol as a precipitant by the hanging drop technique (Del Campo et al. 2003). These crystals are tetragonal (space group  $P4_32_12$ ) with a single molecule of RluD in the asymmetric unit. A three-wavelength multiwavelength anomalous dispersion (MAD) data set was collected on a single flash-frozen crystal of SeMet RluD at the X12-C beamline, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL).

### Structure determination

Details of data collection and refinement are shown in Table 1. The diffraction data was processed using HKL (Otwinowski and Minor 1997), and MAD phases were determined using SOLVE (Terwilliger and Berendzen 1999). Of the 14 expected selenium sites, SOLVE located nine using MAD data to 2.1 Å and generated an initial electron density map with an overall figure of merit of 0.56. None of the selenomethionines in the disordered N-terminal part of RluD could be located by SOLVE. The phases were improved to a figure of merit of 0.64 using density modification and solvent flattening in RESOLVE (Terwilliger 2000). The automated building feature in RESOLVE was used to position about 150 amino acids in the electron density, and the model was then extended by several rounds of manual building in O (Jones et al. 1991). Model refinement and phase combination between rounds of model building were done with CNS (Brünger et al. 1998).

Although a cross-validation set with 5% of the reflections was used for all manual model building and refinement, the initial automated build by RESOLVE included all reflections. This is likely the reason for the small spread between  $R$  and  $R_{\text{free}}$  (Table 1).

The N-terminal purification tag and 76 amino acids could not be modeled due to poor density. The final model contains residues 77–326, and 261 water molecules in the asymmetric unit, with 88.2% of the residues in the most favored region of the Ramachandran plot and no residues in the disallowed region. Model quality was monitored using PROCHECK (Laskowski et al. 1993) and the MOLPROBITY web tool (Lovell et al. 2003). A number of ordered water molecules fill the positively charged active site cavity, although some of them appear to adopt multiple positions based on positive  $F_o - F_c$  electron density maps. We attempted to model chloride and tris ions in a large blob of density in this region, but these gave poorer  $R$  factors and  $F_o - F_c$  maps; this region is modeled in our structure with a group of water molecules.

### Coordinates

The atomic coordinates and structure factors have been deposited in the Protein Data Bank (accession code 1QYU).

**TABLE 1.** Crystallographic data, phasing, and refinement statistics

Data collection and phasing			
Space group	$P4_32_12$		
Unit cell dimensions			
$a = b$ (Å)	75.14		
$c$ (Å)	181.81		
$\alpha = \beta = \gamma$ (°)	90		
	Inflection	Peak	Remote
Wavelength (Å)	0.978795	0.978462	0.950037
Resolution (Å)	50.0–2.0	50.0–2.0	50.0–2.0
Completeness (%) <sup>a</sup>	95.1 (69.6)	88.9 (51.2)	84.7 (45.1)
Overall $I/\sigma(I)$	38.3 (9.2)	31.5 (5.8)	31.1 (4.3)
$R_{\text{merge}}$ (%) <sup>b</sup>	5.3 (13.3)	7.3 (18.2)	5.8 (18.0)
F.O.M. <sup>c</sup>	0.56		
Refinement			
Resolution range (Å)	40.0–2.0		
Number of reflections	32009		
$R$ (%) <sup>d</sup>	21.9		
$R_{\text{free}}$ (%) <sup>a</sup>	23.2		
Cross-validated			
Luzatti error	0.26		
R.m.s. deviations			
Bond length (Å)	0.006		
Bond angle (°)	1.3		
Mean $B$ -factor (Å <sup>2</sup> )	35.9		
Number of protein atoms	2007		
Number of solvent atoms	261		

<sup>a</sup>Data in parentheses are for the highest-resolution shell (2.07–2.00 Å).

<sup>b</sup> $R_{\text{merge}} = \sum (|I_j - \langle I \rangle|) / \sum I_j$ , where  $I_j$  is the observed intensity of reflection  $j$  and  $\langle I \rangle$  is the average intensity of multiple observations.

<sup>c</sup>Figure of merit of the MAD solution in SOLVE (Terwilliger and Berendzen 1999).

<sup>d</sup> $R = \sum (|F_{\text{obs}}| - |F_{\text{calc}}|) / \sum |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes, respectively.

<sup>e</sup> $R_{\text{free}}$  is the  $R$  value calculated for a test set of reflections, comprising a randomly selected 5% of the data not used during refinement.

## ACKNOWLEDGMENTS

We thank Kenneth E. Rudd and Tristan F. Fiedler for helpful discussions. We thank Anand Saxena at the X12-C beamline, NSLS, BNL, for assistance with data collection. This work was supported in part by NIH grant GM58879 (to J.O.), NIH fellowship GM66374 (to M.D.), Florida Biomedical Research Foundation grant BM030 (to A.M.), and an American Heart Association Florida/Puerto Rico Affiliate grant SDG-0130456B (to A.M.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

Received September 15, 2003; accepted October 22, 2003.

## NOTE ADDED IN PROOF

While this paper was in proof, a structure of TruB with and without substrate has appeared (Pan et al. 2003), revealing significant movements in TruB upon binding RNA that provide support for our prediction that RluD must undergo a conformational change



to bind helix 69 of 23S rRNA. Also, a recent publication (Sivaraman et al. 2004) describes a structure of residues 75–326 of *E. coli* RluD that agrees well with the structure described here despite different crystal packing parameters.

## REFERENCES

- Aravind, L. and Koonin, E.V. 1999. Novel predicted RNA-binding domains associated with the translation machinery. *J. Mol. Evol.* **48**: 291–302.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**: 905–920.
- Bashan, A., Agmon, I., Zarivach, R., Schluenzen, F., Harms, J., Berisio, R., Bartels, H., Franceschi, F., Auerbach, T., Hansen, H.A.S., et al. 2003. Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol. Cell* **11**: 91–102.
- Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst. D* **54**: 905–921.
- Del Campo, M., Kaya, Y., and Ofengand, J. 2001. Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA* **7**: 1603–1615.
- Del Campo, M., Ofengand, J., and Malhotra, A. 2003. Purification and crystallization of *Escherichia coli* pseudouridine synthase RluD. *Acta Cryst. D* **59**: 1871–1873.
- Ferré-D'Amaré, A.R. 2003. RNA-modifying enzymes. *Curr. Opin. Struct. Biol.* **13**: 49–55.
- Foster, P.G., Huang, L., Santi, D.V., and Stroud, R.M. 2000. The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. *Nat. Struct. Biol.* **7**: 23–27.
- Gustafsson, C., Reid, R., Greene, P.J., and Santi, D.V. 1996. Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucleic Acids Res.* **24**: 3756–3762.
- Gutgsell, N.S., Del Campo, M., Raychaudhuri, S., and Ofengand, J. 2001. A second function for pseudouridine synthases: A point mutant of RluD unable to form pseudouridines 1911, 1915, and 1917 in *Escherichia coli* 23S ribosomal RNA restores growth to an RluD-minus strain. *RNA* **7**: 990–998.
- Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. 2001. High resolution structure of the large ribosomal subunit from a mesophilic eubacterium. *Cell* **107**: 679–688.
- Hoang, C. and Ferré-D'Amaré, A.R. 2001. Cocrystal structure of a tRNA Ψ55 pseudouridine synthase: Nucleotide flipping by an RNA-modifying enzyme. *Cell* **107**: 929–939.
- Holm, L. and Sander, C. 1993. Protein structure analysis by alignment of distance matrices. *J. Mol. Biol.* **233**: 123–138.
- Huang, L., Ku, J., Pookanjanatavip, M., Gu, X., Wang, D., Greene, P.J., and Santi, D.V. 1998a. Identification of two *Escherichia coli* pseudouridine synthases that show multisite specificity for 23S RNA. *Biochemistry* **37**: 15951–15957.
- Huang, L., Pookanjanatavip, M., Gu, X., and Santi, D.V. 1998b. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* **37**: 344–351.
- Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. 2003. Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. *Cell* **113**: 383–394.
- Jones, T.A., Zou, J.-Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**: 110–119.
- Kaya, Y. and Ofengand, J. 2003. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *RNA* **9**: 711–721.
- Kleywegt, G.J. 1999. Experimental assessment of differences between related protein crystal structures. *Acta Cryst. D* **55**: 1878–1884.
- Koonin, E.V. 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.
- Kowalak, J.A., Bruenger, E., Hasizume, T., Peltier, J.M., Ofengand, J., and McCloskey, J.A. 1996. Structural characterization of U\*<sup>1915</sup> in domain IV from *Escherichia coli* 23S ribosomal RNA as 3-methylpseudouridine. *Nucleic Acids Res.* **24**: 688–693.
- Laskowski, R.A., MacArthur, M.V., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–291.
- Lovell, S.C., Davis, I.W., Arendall 3rd, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by C-α geometry: φ, ψ, and C-β deviation. *Proteins Struct. Funct. Genet.* **50**: 437–450.
- Mueller, E.G. 2002. Chips off the old block. *Nat. Struct. Biol.* **9**: 320–322.
- Newby, M.I. and Greenbaum, N.L. 2002. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nat. Struct. Biol.* **9**: 958–965.
- Nicholls, A., Sharp, K.A., and Honig, B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**: 281–296.
- Ofengand, J. 2002. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett.* **514**: 17–25.
- Ofengand, J. and Rudd, K.E. 2000. Bacterial, archaeal, and organellar rRNA pseudouridines and methylated nucleosides and their enzymes. In *The ribosome: Structure, function, antibiotics, and cellular interactions* (eds. R.A. Garrett et al.), pp. 175–189. ASM Press, Washington, DC.
- Ofengand, J., Malhotra, A., Remme, J., Gutgsell, N.S., Del Campo, M., Jean-Charles, S., Peil, L., and Kaya, Y. 2001. Pseudouridines and pseudouridine synthases of the ribosome. *Cold Spring Harbor Symp. Quant. Biol.* **66**: 147–159.
- Otwindowski, Z. and Minor, W. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Pan, H., Agarwalla, S., Moustakas, D.T., Finer-Moore, J., and Stroud, R.M. 2003. Structure of tRNA pseudouridine synthase TruB and its RNA complex: RNA recognition through a combination of rigid docking and induced fit. *100*: 12648–12653.
- Partridge, B.L. and Pritchard, C.E. 1995. Two methylated ribonucleosides: 3-methyluridine and 1-methylinosine. *Acta Cryst. C* **51**: 1929–1932.
- Raychaudhuri, S., Conrad, J., Hall, B.G., and 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.
- Sivaraman, J., Sauv e, V., Larocque, R., Stura, E.A., Schrag, J.D., Cygler, M., and Matte, A. 2002. Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nat. Struct. Biol.* **9**: 353–358.
- Sivaraman, J., Iannuzzi, P., Cygler, M., and Matte, A. 2004. Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23 S rRNA and is essential for normal cell growth of *Escherichia coli*. *J. Mol. Biol.* **335**: 87–101.
- Spedaliere, C.J., Hamilton, C.S., and Mueller, E.G. 2000. Functional importance of motif I of pseudouridine synthases: Mutagenesis of aligned lysine and proline residues. *Biochemistry* **39**: 9459–9465.
- Staker, B.L., Korber, P., Bardwell, J.C.A., and Saper, M.A. 2000. Structure of Hsp15 reveals a novel RNA-binding motif. *EMBO J.* **19**: 749–757.
- Terwilliger, T.C. 2000. Maximum likelihood density modification. *Acta Cryst. D* **56**: 965–972.
- Terwilliger, T.C. and Berendzen, J. 1999. Automated structure solution for MIR and MAD. *Acta Cryst. D* **55**: 849–861.
- Wrzesinski, J., Bakin, A., Ofengand, J., and Lane, B.G. 2000. Isolation and properties of *Escherichia coli* 23S-RNA pseudouridine 1911, 1915, 1917 synthase (RluD). *IUBMB Life* **50**: 33–37.
- Xie, W., Liu, X., and Huang, R.H. 2003. Chemical trapping and crystal structure of a catalytic tRNA guanine transglycosylase covalent intermediate. *Nat. Struct. Biol.* **10**: 781–788.
- Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. 2001. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**: 883–896.