Crystal Structures of the Catalytic Domains of Pseudouridine Synthases RluC and RluD from *Escherichia coli*

Kenji Mizutani, Yoshitaka Machida, Satoru Unzai, Sam-Yong Park, and Jeremy R. H. Tame*

Protein Design Laboratory, Yokohama City University, Suehiro 1-7-29, Tsurumi, Yokohama 230-0045, Japan

Received November 20, 2003; Revised Manuscript Received February 11, 2004

ABSTRACT: The most frequent modification of RNA, the conversion of uridine bases to pseudouridines, is found in all living organisms and often in highly conserved locations in ribosomal and transfer RNA. RluC and RluD are homologous enzymes which each convert three specific uridine bases in *Escherichia coli* ribosomal 23S RNA to pseudouridine: bases 955, 2504, and 2580 in the case of RluC and 1911, 1915, and 1917 in the case of RluD. Both have an N-terminal S4 RNA binding domain. While the loss of RluC has little phenotypic effect, loss of RluD results in a much reduced growth rate. We have determined the crystal structures of the catalytic domain of RluC, and full-length RluD. The S4 domain of RluD appears to be highly flexible or unfolded and is completely invisible in the electron density map. Despite the conserved topology shared by the two proteins, the surface shape and charge distribution are very different. The models suggest significant differences in substrate binding by different pseudouridine synthases.

The isomerization of uridine to pseudouridine (Ψ) within RNA is a ubiquitous process evidently present from very early in the evolution of life (*1*−*3*). The mechanism of the reaction of pseudouridine synthase I has been the subject of detailed chemical analysis strongly favoring Michael addition (*4*), but the issue of whether the catalytic aspartate attacks the ribose or the uridine is not yet finally settled (*5*). The net result of either proposed mechanism is the detachment of the base from the ribose and reattachment through a carbon atom (C5), freeing a nitrogen atom (N1) in the base to participate in hydrogen bonding. This may explain why Ψ is only found in structured RNA such as rRNA or tRNA, and not in mRNA or viral genomes (*6*). Only specific bases are converted to Ψ, and in each case, a single enzyme appears to be responsible. Each pseudouridine synthase, however, may act at more than one site within its target RNA molecule(s). Currently, only one eubacterial pseudouridine synthase (RluA) which acts on more than one substrate RNA is known (*7*). Until recently, four families of pseudouridine synthases were known, named after the TruA, TruB, RsuA, and RluA genes of *Escherichia coli* (*8*, *9*). All four families share a short motif ("motif II"), but only the last three share motif I. A novel fifth family whose sequence is only slightly similar to any of these has now also been described (*10*). The crystal structures of TruA, TruB, and RsuA have all been described previously (*11*−*13*). Shortly before submission of this paper, the crystallization of RluD was reported (*14*), and two independent structures have now been published (*15*, *16*); however, there is so far no published structure for RluC. These are the first structures of members of the RluA family. Within the Pfam database (*17*), the RsuA and TruB families are grouped together since they are structurally similar, and different from the TruA family. All pseudouridine synthases appear to share a common catalytic domain however, suggesting they are all descended from a single ancestor (*8*). There are 10 known Ψs in *E. coli* 23S rRNA within the large ribosomal subunit, and 10 enzymes responsible for introducing them (*18*). RluD creates three Ψs, all within the same stem−loop motif of *E. coli* 23S rRNA (*7*, *19*, *20*), implying that once the protein has recognized the local RNA structure, it modifies all uridines within reach. RluC also modifies three uridines in the same RNA, but these are much more widely spread along the sequence (*21*).

Two of the Ψs created by RluD, Ψ1915 and Ψ1917, are very highly conserved (*22*), although Ψ1915 is also N3 methylated in *E. coli*. Since the sites of RluD activity are close to bridges between the 50S and 30S subunits of the ribosome, the loop of helix 69 contacting the decoding site of 16S rRNA (*23*), it is highly plausible that the Ψs generated by RluD are important for ribosome function or fidelity. It has been shown, however, that RluD missing the essential catalytic aspartate residue can rescue RluD$^-$ mutants (*19*). The protein seems to act as an RNA chaperone which is important for ribosome assembly. RluC apparently has no such function and can be deleted without significantly impairing growth (*21*). The sequences of both RluC and RluD are significantly similar to that of RsuA, which acts on a single base in the rRNA of the small ribosome subunit. RsuA includes an N-terminal S4 RNA binding domain, first found in ribosomal protein S4 (*24*). RluC and RluD both have an equivalent domain, but RluA does not. An alignment of the catalytic domains of RluC and RluD with RsuA is shown in Figure 1. While sequence alignments provide tempting glimpses of evolutionary relationships and cellular functions, it is often found that protein function and structure may each be conserved while the other changes. Hemoglobin perhaps provides some of the best examples (*25*). Mutations

* To whom correspondence should be addressed. E-mail: jtame@tsurumi.yokohama-cu.ac.jp. Telephone: +81 (0)45 508 7228. Fax: +81(0)45 508 7366.

10.1021/bi036079c CCC: $27.50
Published on Web 03/24/2004

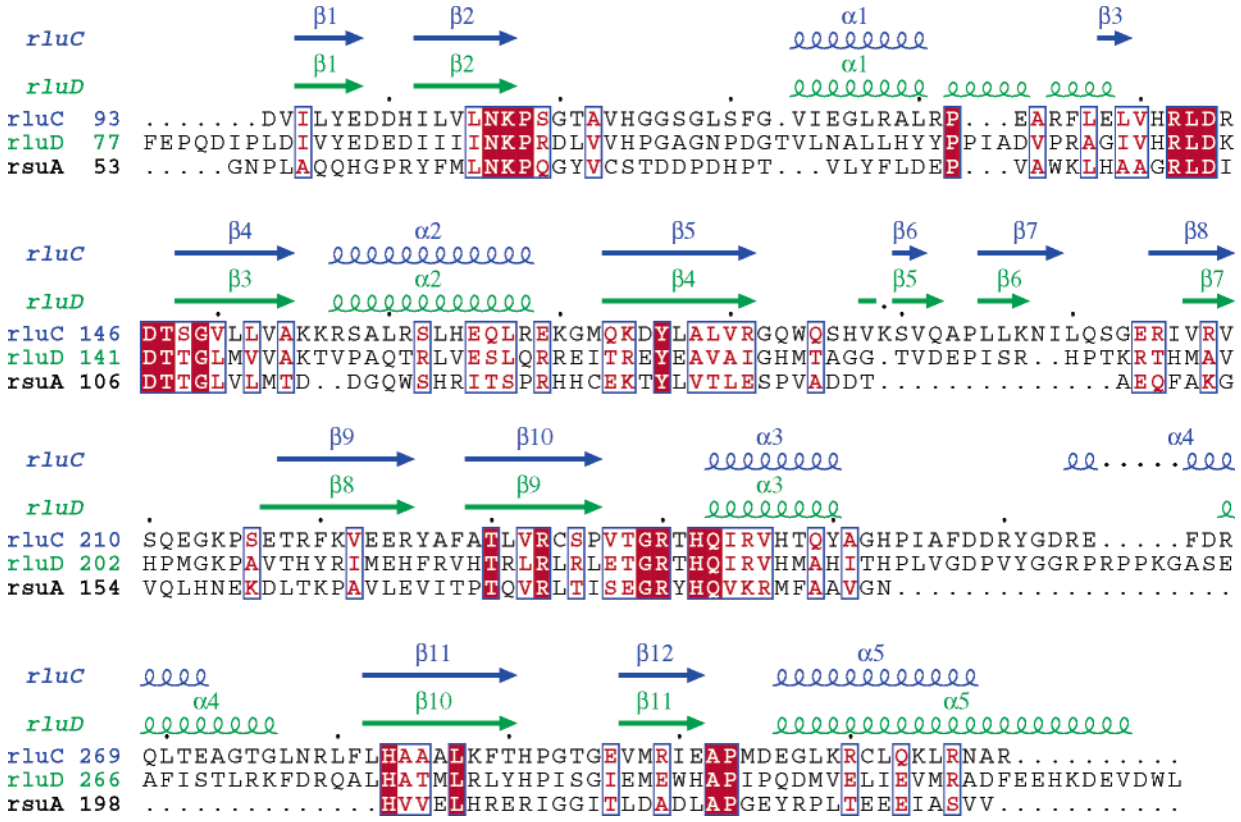

FIGURE 1: Sequence alignment of the catalytic domains of RluC and RluD with *E. coli* RsuA. The secondary structure elements of RluC and RluD are shown in blue and green, respectively. Residues conserved in all three proteins are shown with white letters on a red background. The sequence immediately following helix 1 in RluD has a slightly helical nature and is considered α-helical by some secondary structure assignment programs.

of well-conserved lysine and proline residues within motif I of RluA and TruB were expected to reduce enzyme activity substantially, but in fact had little effect (*26*). We have purified and crystallized truncated RluC (missing the S4 domain) and full-length RluD. In this paper, the refined X-ray structures are compared with other RNA binding proteins and pseudouridine synthases.

EXPERIMENTAL PROCEDURES

RluC was cloned by PCR from *E. coli* genomic DNA using the TTTGCGTCATATGGCGGACGTCATCATG-TATGAAGATGATCA and TTTGCGCTCGAGTTAGC-GCGCGTTACGCATCTTTTGCAAACAACG oligonucleotides, allowing the product to be inserted in frame into the pET21b expression vector using *Nde*I and *Xho*I restriction sites. These primers generate two Leu → Met mutations at positions 96 and 315, increasing the number of methionine residues in the expressed fragment to five. RluD was similarly cloned using TTTGCGTCCATGGCACAACGAG-TACAGCTCACTGC and TTTGCTAAGCTTTTATAAC-CAGTCCACTTCATCCTTAT. The PCR product was cloned into pET28b using *Nco*I and *Hin*dIII restriction sites. The protein product in each case has no affinity tag. The expression vectors were transformed into *E. coli* strain BL21-(DE3) for protein expression. The cells were grown in Luria-Bertani broth at 37 °C.

Protein expression was induced by the addition of 1 mM IPTG to the cells. In the case of RluD, the culture medium was first cooled to 25 °C and growth continued for 6 h after induction. RluC(92−319) was purified using a HiTrap Heparin HP column (Amersham Biosciences) equilibrated with 50 mM Tris (pH 8.1) and eluted with a linear salt gradient up to 1 M sodium chloride. Finally, the protein was purified by gel filtration in the same buffer containing 1 M salt, before being dialyzed into fresh buffer containing 300 mM salt. RluD was purified by anion exchange using a HiLoad Q (Amersham Biosciences) column washed with 50 mM potassium phosphate (pH 7.0) and a salt gradient up to 1 M sodium chloride, followed by an SP column run with the same buffer and salt gradient. The protein was then applied to a hydroxyapatite column equilibrated with 25 mM potassium phosphate (pH 7.0) and 100 mM sodium chloride, and eluted with 3% ammonium sulfate in the same buffer. The purified protein was dialyzed against 25 mM potassium phosphate (pH 7.0) and 100 mM sodium chloride. The selenomethionine-substituted protein was produced by standard protocols (*27*).

Crystals of both RluC(92−319) and RluD were grown at 20 °C using the hanging drop method. RluC(92−319) was crystallized using a 10 mg/mL protein stock solution in 50 mM Tris-HCl (pH 8.1) and 300 mM NaCl. The reservoir solution consisted of 1.4−1.6 M ammonium sulfate, 0.1 M NaCl, and 0.1 M HEPES (pH 7.5). Crystals grew over the course of 3 days. RluD was crystallized from a 40 mg/mL stock solution containing 25 mM potassium phosphate (pH 7.0) and 100 mM sodium chloride. The reservoir solution was 35% ethylene glycol, 5% PEG 3000, and 100 mM HEPES (pH 7.5). In the case of the selenomethionine-substituted protein, a reservoir solution of 46% ethylene glycol, 6% PEG 3000, and 100 mM HEPES (pH 8.0) was

Table 1: Data Collection and Model Refinement Statistics

	RluC			RluD		
space group, unit cell dimensions (Å)	$P3_221$, $a = b = 96.73$ and $c = 86.88$			$P6_522$, $a = b = 74.51$ and $c = 265.01$		
wavelength (Å)	peak	inflection	remote	peak	inflection	remote
resolution range (Å)	20.0−2.5	20.0−2.5	20.0−2.5	20.0−2.5	20.0−2.5	20.0−2.5
no. of reflections (measured/unique)	227046/16604	230334/16612	228470/16642	175559/15795	174219/15802	174376/15809
completeness (%) (overall/outer shell)[a]	100.0/100.0	100.0/100.0	99.9/99.9	99.4/99.6	99.3/99.7	99.3/99.7
R_{merge} (%) (overall/outer shell)[a]	6.2/10.9	5.3/11.9	4.8/12.3	7.9/10.6	6.1/8.0	5.8/7.8
redundancy (overall)	13.7	13.9	13.7	11.1	11.0	11.0
mean $\langle I/\sigma(I)\rangle$ (overall)	22.5	16.0	15.5	15.0	16.5	15.9
phasing (20.0−2.5 Å), mean FOM[b] after RESOLVE phasing		0.51			0.57	

Refinement Statistics

	RluC	RluD
refinement resolution (Å)	20.0−2.0	29.5−1.70
no. of reflections (measured/unique)	430252/35502	758261/47799
completeness (%) (overall/outer shell)[a]	98.0/83.5	98.7/91.0
R_{merge} (%) (overall/outer shell)[a,c]	7.2/29.6	5.8/24.1
redundancy (overall)	12.1	15.9
mean $\langle I/\sigma(I)\rangle$ (overall)	14.0	18.1
σ cutoff/no. of reflections used	0.0/35502	0.0/47799
R-factor[d]/R_{free} (%)[e]	20.9/27.6	20.0/22.4
rmsd for bond lengths (Å)/bond angles (deg)	0.022/2.1	0.013/1.45
no. of water atoms	171	264
average B-factor (Å^2) (protein/water)	37/36	24/33
Ramachandran plot		
residues in most favorable regions (%)	90.7	90.1
residues in additional allowed regions (%)	9.3	9.9

[a] Completeness and R_{merge} are given for overall data and for the highest-resolution shell. The highest-resolution shells for the RluC, RluD, and MAD data sets are 2.07−2.0, 1.76−1.70, and 2.59−2.50 Å, respectively. [b] The figure of merit (FOM) = $|F_{best}| - |F|$. [c] $R_{merge} = \sum|I_i - \langle I\rangle|/\sum|I_i|$, where I_i is the intensity of an observation, $\langle I\rangle$ is the mean value for that reflection, and the summations are over all equivalents. [d] R-factor = $\sum_h||F_o(h)| - |F_c(h)||/\sum hF_o(h)$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. [e] R_{free} was calculated with 5% of the data excluded from the refinement.

used. Crystals grew over the course of 2 days to roughly 500 μm in length.

High-resolution native diffraction data and moderate-resolution (2.5 Å) multiwavelength data were obtained from cryocooled (100 K) crystals using synchrotron radiation at the Photon Factory beamline NW12 station (Tsukuba, Japan). Intensity data were collected with an ADSC detector. Diffraction data were integrated and scaled with HKL2000 and SCALEPACK (*28*). The positions of selenium atoms and initial phases were determined using SOLVE (*29*), and density modification was carried out with RESOLVE (*30*). Molecular replacement (to determine RluD from the RluC model) was carried out with MOLREP (*31*). For both RluC and RluD, the experimental phases produced very clear electron density maps; in both cases, the selenomethionine-substituted protein gave isomorphous crystals to the native protein. General data manipulation was performed using the CCP4 suite of programs (*32*). Models were built using TURBO (*33*) and XTALVIEW (*34*). Structural refinement against the native data was performed using ARPwARP (*35*), X-PLOR version 3.851 (*36*), and REFMAC5 (*37*). A *B*-factor cutoff of 50.0 was applied to water molecules, and any refining to higher values were removed from the models. Structural analysis of the final models of RluC(92−319) and RluD using PROCHECK (*38*) indicated that ~90% of the residues are in the most favorable regions of the Ramachandran plot, with no residues in "disallowed" regions. A summary of the data collection and refinement statistics is given in Table 1. In the case of RluC, there are two molecules in the asymmetric unit, but no noncrystallographic symmetry restraints were applied during the refinement. The 908 main chain atoms have an rms deviation of 0.69 Å. Apart from the C-terminal residues of the model, the largest difference between these two copies of RluC is found at Pro215. This is in a surface region close to crystal contacts in one molecule, and apparently rather flexible in the other which has weak electron density over residues Val209−Ser216.

RESULTS

Expression systems for large-scale production of RluC and RluD were produced by PCR. In a previous report about the crystallization of RluC, it was found that the N-terminal domain was missing, and the crystallized protein was proteolytically truncated by 88 residues (*39*). We therefore chose to express only the C-terminal catalytic domain of this 319-residue protein, from residue 92, and the crystals of truncated RluC that were obtained were essentially identical to those described in 1999 (*39*). Full-length RluD (326 amino acid residues) was purified in the hope of observing the N-terminal domain. Analytical ultracentrifugation sedimentation velocity experiments showed both RluC(92−319) and RluD are monomeric in solution (data not shown). X-ray phases for RluC were determined by the method of multiple-wavelength anomalous dispersion (MAD) using the selenomethionine-substituted protein. The introduction of two Leu → Met mutations, when the gene was cloned, facilitated this step so that after density modification the RluC structure could be refined very rapidly. This model was used to determine the RluD structure by molecular replacement. Upon examination of the RluD electron density maps, it was evident that the N-terminal domain did not appear. The molecular mass of the crystallized protein was checked by

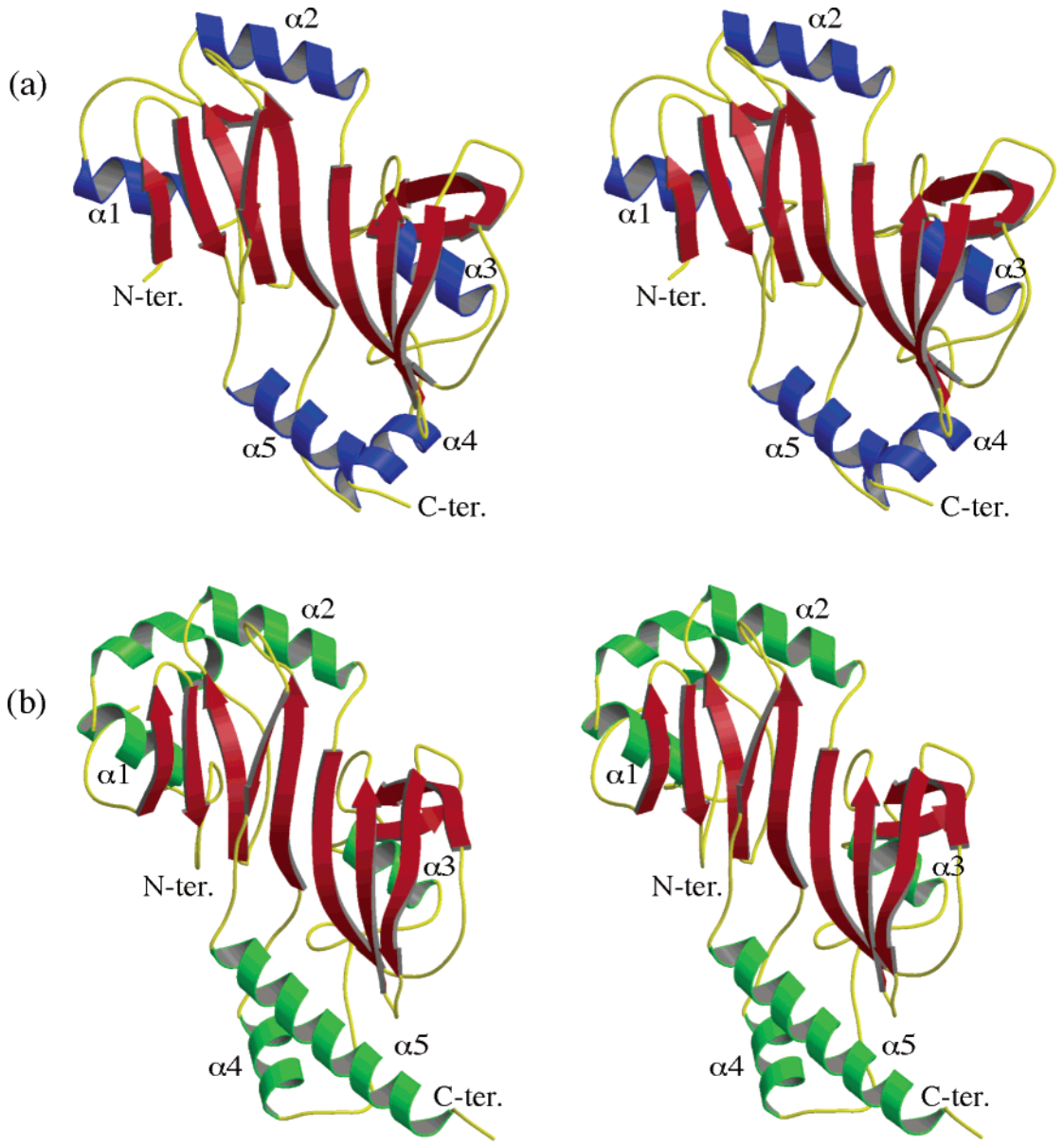

FIGURE 2: Stereo ribbon diagrams of (a) RluC(92−319) and (b) RluD. Both proteins form ordered C-termini, the final β-strand and α-helix adopting similar conformations in each case. This figure was made with MOLSCRIPT (*53*).

gel electrophoresis and found to be the same as that of the purified protein (37.1 kDa), indicating no proteolysis had occurred (data not shown). Independent experimental phases were therefore determined for RluD using the SeMet MAD method. This map confirms that no electron density corresponding to the N-terminal domain can be seen, despite the high quality and resolution of the data (Table 1). The final models include all residues of RluC from Asp93 to Arg319, refined to 2.0 Å resolution, and residues of RluD from Phe77 to Leu326, refined to 1.7 Å resolution. These models have been deposited in the Protein Data Bank as entries 1V9K (RluC) and 1V9F (RluD).

On revising this paper, we had the opportunity to compare our RluD structure with those refined by two other groups to 1.8 Å (space group $P2_12_12_1$) and 2.0 Å resolution (space group $P4_32_12$) [PDB entries 1PRZ (*15*) and 1QYU (*16*)], respectively. In all three cases, the asymmetric unit contains one monomer, and only the C-terminal domain is present in the model. Both 1PRZ and 1QYU have selenomethionine residues in place of methionine, but comparison with our native model shows this has caused no major structural changes. Superimposing the models with a least-squares fitting shows that they diverge considerably at the last five residues of the C-terminus. In our model, the terminal carboxyl group of Leu326 interacts with Arg169′ (prime indicating a neighboring molecule), and Trp325 is sandwiched between the side chains of Arg131′ and Arg137′, suggesting that the two C-terminal residues may be mimicking the RNA substrate of the symmetry-related neighbor. The C-terminus of 1QYU is much closer to the helical structure found in our model, but in all three models, the protruding C-terminus is involved in crystal contacts, and is presumably flexible in the unbound protein in solution. In a comparison of the main chain atoms of residues 77−320, the rms deviations of our model from 1PRZ and 1QYU are 0.833 and 0.713 Å, respectively. Within this region, the largest difference from 1PRZ is at a short surface coil region around Gly254−Pro259, just after motif III. Our electron density maps are very clear in this region, showing our model agrees with 1QYU. 1PRZ has a crystal contact at this position. The

largest difference from 1QYU is at the carbonyl oxygen atom of Ala108, which is possibly due to a pep-flip error in that model (the residue is in a mobile loop). Our data set has rather more reflections (45 392 working and 2406 free) versus 27 504 working reflections for 1PRZ and 32 009 for 1QYU, largely due to the different sizes of the asymmetric unit in each case.

The models for RluC(92−319) and RluD exhibit highly similar topology (Figure 2). Searching the Protein Data Bank for structures related to RluC(92−319) using DALI (*40*) gave significant but not especially strong matches to three entries, 1KSK (RsuA, *Z* score = 12.4), 1K8W (TruB, *Z* score = 9.9), and 1DJ0 (TruA, *Z* score = 8.5), all of which are representative pseudouridine synthases of families defined on the basis of sequence. In each case, the match included roughly 150 residues with an overall rms deviation of 2.7−3.5 Å. The fit to RsuA is best, with 23% of the sequence being identical over the 147 residues which were aligned. RluD gave similar results but also found a weak match to formiminotransferase cyclodeaminase (PDB entry 1QD1). Since the *Z* score for this last match was only 2.7, and the 97 Cα atoms compared gave an rms deviation of 7.2 Å, it is clearly not significant. RluC and RluD match each other rather better, with a *Z* score of 27.3 over 220 residues with 33% of the sequence being identical. In a comparison of RluC and RluD with RsuA, the overall topology is very similar, but with significant differences at the N- and C-termini of the RluC and RluD models (Figure 2). From Gly206 onward, RsuA has a fold quite different from that of RluC or RluD, and the last strand of *β*-sheet 1 has the opposite orientation. RsuA is also a much shorter protein, with 231 amino acid residues. RluC has an extra α-helix inserted between residues Ala256 and Leu280, and an extra *β*-hairpin between Ala193 and Gly213. RluD has loops in these regions, and a longer C-terminal tail (much longer than that of RsuA). These surface differences are close to the active site and almost certainly related to substrate recognition. The RsuA model (PDB entry 1KSK) includes the N-terminal S4 domain which is invisible in our RluD structure, but it appears the S4 domain of RluC and RluD adopts a different position relative to the catalytic domain since the N-terminal regions do not follow that of RsuA.

The reason for the absence of any electron density for the S4 domain of RluD is unclear. Using the Instability Index (*41*), the first 75 amino acids of RluD score 39.4 (a score of ≥40 indicates an unstable protein). However, the remaining 251 residues in the sequence score 37.3, which is not significantly lower. The N-terminal sequence of RluD does not have the low level of sequence complexity characteristic of naturally unfolded proteins which adopt an ordered structure only on ligand binding (*42*). Sequence searches suggest up to 2069 proteins in the SwissProt database may contain an S4 domain. The PROSITE database (*43*) uses a stricter definition which matches 214 known sequences, which do not include RluC or RluD. The first 75 residues of RluC have a predicted pI of 10.06, with 12% lysine and 10% arginine, an excess of positively charged residues which might be expected for an RNA binding domain. The first 75 residues of RluD have a predicted pI of 4.88, 12% of the residues being glutamate. The N-terminal domain does not therefore appear to be a simple bolt-on structure which confers RNA binding. Presumably, flexibility of the N-terminal domains of RluC and RluD allows highly specific binding to the correct molecular targets without producing very high affinity. In this way, the enzymes can modify the appropriate sites within the rRNA efficiently and avoid becoming trapped in unproductive enzyme−product complexes.

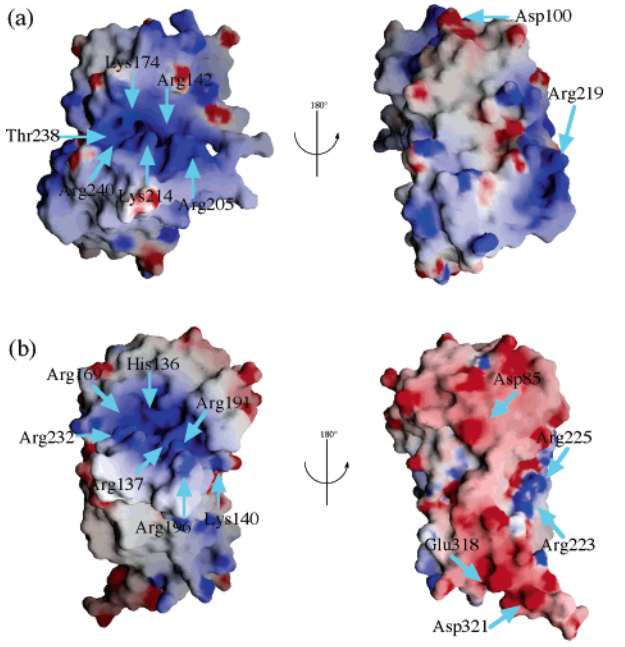

FIGURE 3: Molecular surfaces of (a) RluC(92−319) and (b) RluD colored according to electrostatic potential. Blue indicates positive charge and red negative. Maximum color saturation corresponds to an electron energy of 10 kcal/mol. On the left-hand side, the highly positively charged region around the active site can be seen. A number of positively charged residues clustered around the active site are indicated. The opposite face is also positively charged in the case of RluC, but strongly negatively charged in the case of RluD. This figure was made using GRASP (*54*).

The very different electrostatic properties of the N-terminal domains of RluC and RluD are mirrored by the different surface potentials for the two catalytic domains (Figure 3). While the active site of each protein has an overall positive charge (shown on the left side of Figure 3), the opposite faces of the molecular surfaces are quite different in the two cases (right side of Figure 3). The sequences of RluC and RluD are overall only weakly similar with that of RsuA (Figure 1). Although the central *β*-sheets are conserved, the surface α-helices and loops show considerably more variation. The C-terminal helix packs against the rest of the protein rather differently, and is much longer in RluD, creating a distinctive and highly negatively charged extension jutting into the surrounding solvent. As noted above, the five C-terminal residues of RluD appear to be highly flexible. The distinctive surfaces presumably confer on the enzymes their respective site specificities.

RluC was crystallized in the presence of sulfate ions and RluD in the presence of phosphate. One ordered anion molecule is found for each protein monomer, but not at equivalent positions (Figure 4). One sulfate ion is found for each monomer of RluC(92−319), located outside the active site and bound to surface positive charges, possibly in positions where RNA backbone phosphate groups might bind. The two sulfate ions in the asymmetric unit are not at equivalent positions relative to the protein however; one is

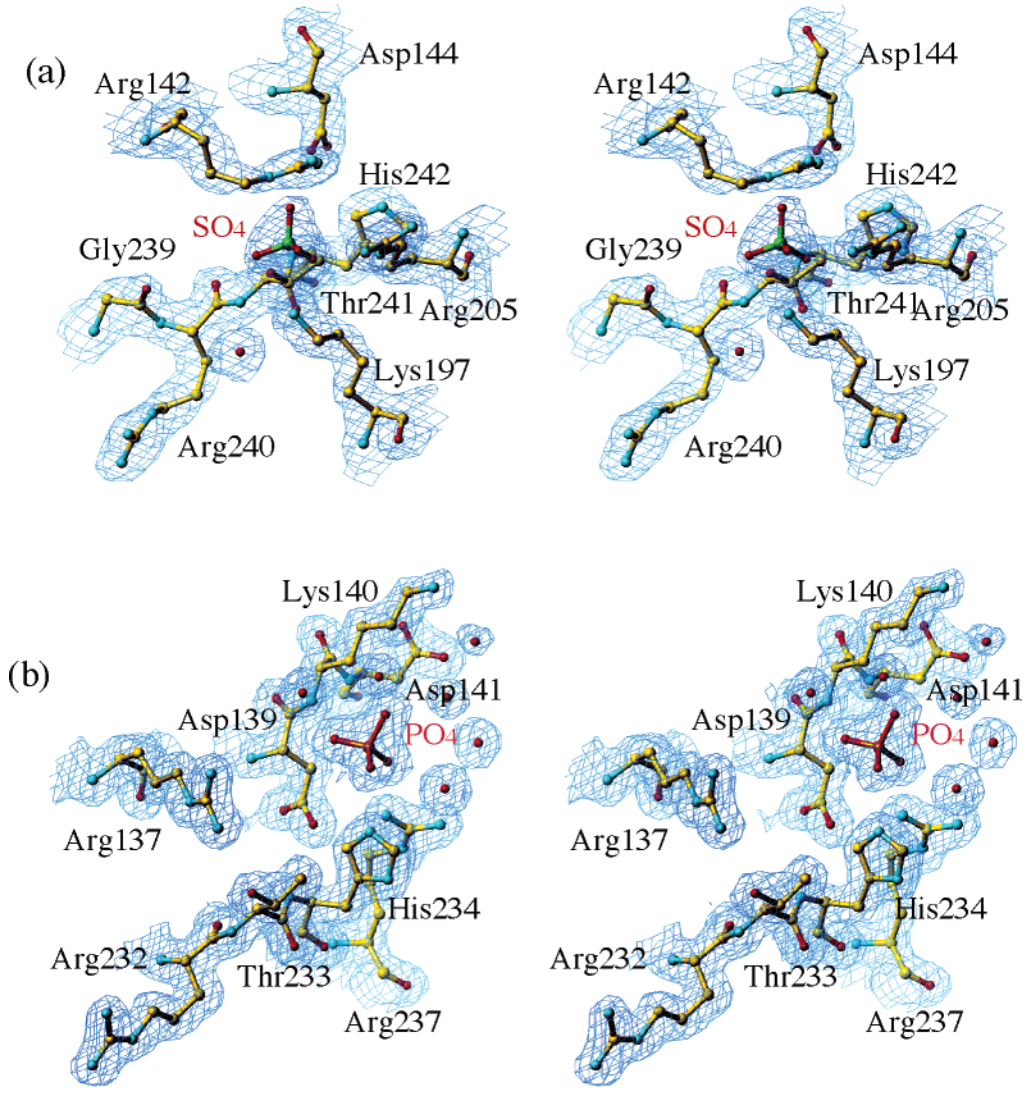

FIGURE 4: Stereo $2mF_o - DF_c$ electron density maps covering the active sites of (a) RluC(92−319) and (b) RluD. Density is contoured at 1.3σ. Water molecules are shown as red spheres. In RluC, the active site aspartate (Asp144) forms bonds to the side chains of both Arg142 and Arg245. RluD has Arg137 and Arg237 on either side of the active site aspartate (Asp139), equivalent to Arg142 and Arg245 in RluC, respectively.

close to the active site, and the other is close to other surface arginines (157 and 161). Repeated attempts to introduce small substrate analogues (uridine, uracil, and UMP) into the active site, by soaking or cocrystallization, all failed, even when concentrations as high as 100 mM were used, despite the modest ionic strength of the mother liquor. Neither RluC nor RluD seems therefore to have an appreciable affinity for simple substrate analogues under the crystallization conditions that were used. In contrast, such experiments were successful in the case of RsuA (*12*), indicating that the active sites exhibit different behavior as well as being lined by different side chains. Since the core β-sheet structure appears to be rigid, it is unlikely to undergo large conformational changes on substrate binding. Surface loops around the active site may be more flexible (Figure 5). Upon comparison of the active sites of RluC and RluD, it can be seen that residues around the catalytic aspartate are conserved, but not their interactions. The active site aspartic acid in both cases has an arginine on either side; in the case of RluC, there are hydrogen bonds from Asp144 to Arg142 and Arg245. In RluD, the active site aspartate (Asp139) has turned, breaking the bond to Arg137 and instead hydrogen bonding to Thr142 (Figure 4). The bond between Asp139 and Arg237 is maintained. This arginine residue (RluC residue 245 and RluD residue 237) in helix α3 (Figure 1) is preserved in motif III of TruA and TruB, but RsuA has a lysine residue at this position. The differences in the conformations may be partly due to the presence of sulfate and phosphate ions, but suggest the active site residues are flexible. A hydrogen bond between the nearby side chains of His242 and Tyr261 in RluC is maintained in RluD between His234 and Tyr253.

The structures of RluC and RluD readily explain the mutagenesis experiments of Spedaliere and colleagues (*26*), who mutated the conserved lysine in motif I, believing it may play a catalytic role. This lysine is in fact buried within the protein at the loops formed by motifs I and II, and interacts with the carbonyls of five residues, including the catalytic aspartate. Placing arginine or methionine at this position greatly destabilized RluA and TruB (preventing measurements of any enzyme activity for the RluA mutants), but at the same time not greatly affecting the catalytic activity of TruB. Replacing the adjacent conserved proline with glycine or leucine gave only a modest drop in k_{cat} and an increase in K_M (*26*). This suggests that some increase in

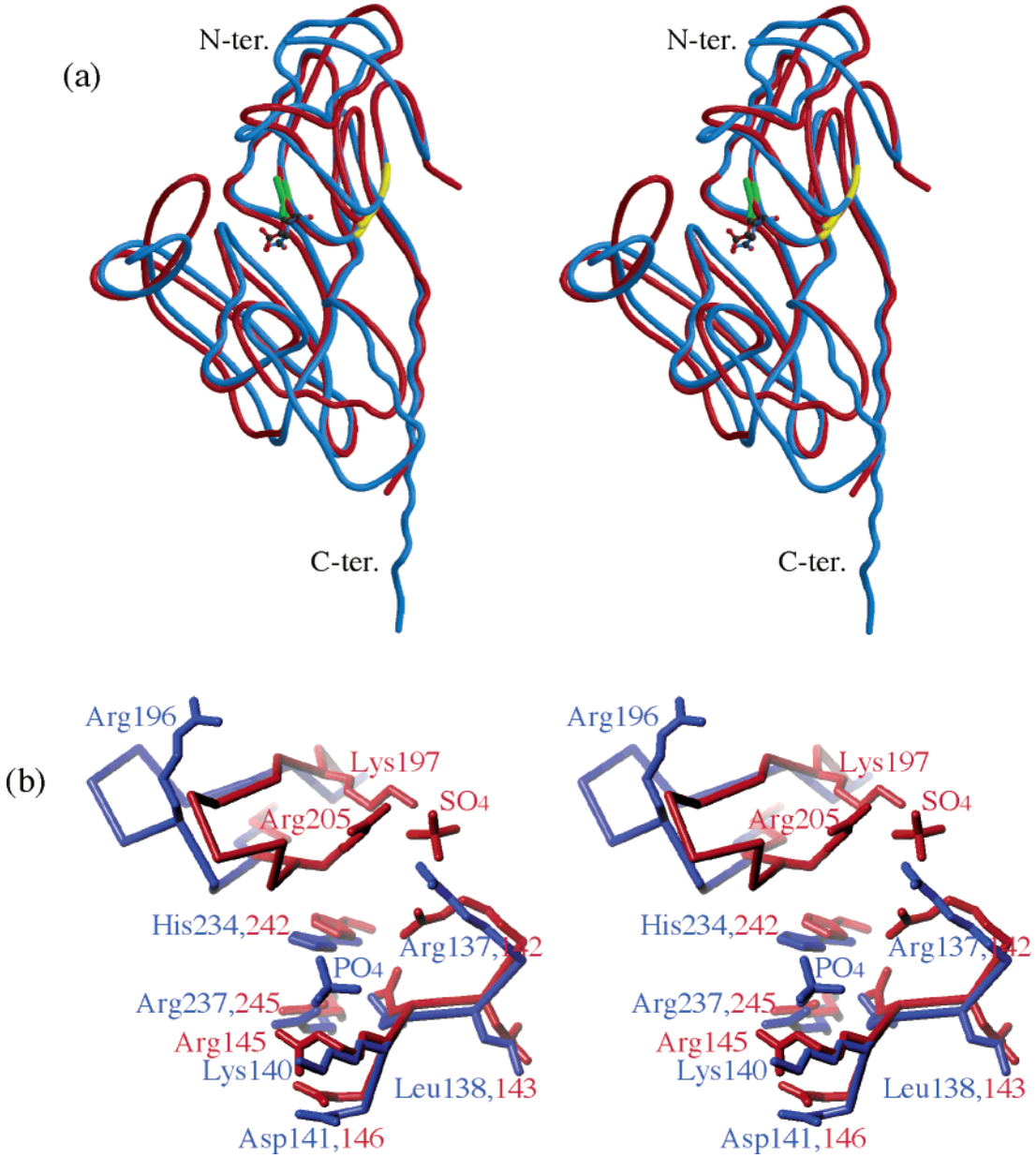

FIGURE 5: (a) Stereo superposition of the Cα traces of RluC (red) and RluD (blue) catalytic domains. Motif I is found at the end of the second β-strand; the X-Asn-Lys-Pro-Y sequence (residues 105−109 in RluC and residues 96−100 in RluD) is shown in yellow. Motif II is very close in space, between β-strands 3 and 4. The Arg-Leu-Asp sequence, including the catalytic aspartate (residues 142−144 in RluC and residues 137−139 in RluD), is shown in green. (b) Closer view around the opening of the superimposed active sites of RluD (blue) and one copy of RluC (red), showing the positively charged residues and the positions of the phosphate and sulfate anions. The sulfate ion that is shown lies outside the active site of RluC, close to Lys197, Arg205, and Arg245. The second monomer has a sulfate ion in a different location, near Arg157 and Arg161; both anions are found close to crystal contacts but may indicate the positions of phosphate groups in the bound RNA substrate. The phosphate ion in RluD is closer to the catalytic aspartate, and hydrogen bonded to the side chains of His234 and Arg237 and the main chain nitrogen of Lys140.

flexibility in the loop carrying the catalytic aspartate does not strongly impair activity, and the conservation of the lysine is to maintain protein stability.

DISCUSSION

TruB is the enzyme responsible for converting uridine 55 in the T stem−loop motif of tRNA, an almost universally conserved modification among tRNAs (*44*−*46*). Recent studies with TruB show the protein uses both induced fit and rigid docking to bind its target RNA, the C-terminal domain moving as a rigid body to bind the substrate RNA (*47*). Since RluC and RluD have no structure equivalent to this domain, and their C-termini are found on the other side of the catalytic domain, it is clear that the pseudouridine synthases use different modes of substrate binding. TruB has no equivalent of the N-terminal domain of RluC or RluD. The crystal structure of *E. coli* TruB complexed with 22 nucleotides of RNA, including a 5-fluorouridine at the position mimicking uridine 55 (*13*), shows that the enzyme works by a nucleotide flipping mechanism (*48*). A conserved histidine (His43), close to the catalytic Asp (*48*), occupies the position of G18, the guanine base paired with U55 in the T stem−loop sequence. RluC and RluD have no equivalent histidine, but do have another nearby histidine (His114 and His105, respectively) which may serve instead. These proteins also have a conserved HRLD sequence within

motif II (Figure 1), from which RsuA misses the histidine and TruB misses both the histidine and arginine. Overlapping the TruB structure with RluC and RluD shows that the motif II arginine (Arg142 and Arg137, respectively) would clash with the RNA substrate of TruB, but may stack against the RluC or RluD RNA substrate. Some other residues lining the active site of TruB are found in RluC and RluD, including Tyr76, Leu200, Ile180, and Arg181 (TruB numbering), but overall, the active sites of RluC and RluD show rather little sequence conservation with models of other pseudouridine synthases. While the pseudouridine synthases probably all share essentially the same enzymatic mechanism, there is no very strongly conserved constellation of amino acid residues at the active site such as the catalytic triad of the serine proteases. In the absence of suitable complexes with which to compare RluC and RluD, the precise details of the interactions with substrates cannot be inferred from the apo structures alone. Complexes with larger RNA molecules will be required to observe the role of individual residues, especially the catalytic aspartate.

It was originally proposed, by analogy with thymidylate synthase, that the nucleophile used to attack the uridine was a cysteine side chain, but this proposal was disproved by mutating the conserved Asp within motif II of TruB, TruA, and RluA (*49*, *50*). Aspartic acid is much less nucleophilic, and enzymes using carboxyl groups as nucleophiles (such as glycosidases) generally use pairs of carboxyl groups to achieve sufficient reactivity, one carboxyl acting as a proton donor and one as the nucleophile. Clan GH-A glycosidases have an arginine close to the nucleophile glutamate. Pseudouridine synthases have only one conserved aspartate, accompanied by a close-by arginine or histidine, at the active site. It is not clear why a more chemically obvious thiol group is not used, but aspartic acid may be a better choice of nucleophile since it is a better leaving group (*50*). It may also be that a strong nucleophile is not necessary. In the case of glycosidases, the Brønsted β value is low (i.e., the reaction rate depends only weakly on the pK_a of the attacking nucleophile), implying the transition state is close to the starting materials (*51*). The catalytic Asp forms a salt bridge with Arg181, which Hoang and Ferré-D'Amaré suggest may increase the nucleophilicity of the carboxyl group (*13*). The double hydrogen bond between Asp48 and Arg181 found in TruB−RNA complexes (*13*, *47*) has no counterpart among the glycosidases.

The structures of RluC and RluD are the first of the rlu (ribosome large subunit) family of pseudouridine synthases to be determined. Not unexpectedly, the catalytic domains are topologically similar to the previously determined structure of RsuA (*12*), but RluC and RluD are readily distinguished from each other and RsuA by their molecular masses and charges at neutral pH. RsuA has a predicted pI of 5.8, much closer to that of RluD (6.3) than to that of RluC (9.9), but the 68 N-terminal residues of RsuA, which form the S4 domain, have a predicted pI of 9.8. It is interesting to note that RluD has a highly negatively charged surface away from its active site, since catalytically incompetent mutants can still serve as RNA chaperones (*19*). A similar effect has been demonstrated for TruB (*45*). Other proteins, either transiently associated with or incorporated into the ribosome, possibly help mediate RluD−RNA interactions. Although the presence within the active site of a number of conserved residues implies a similar mechanism among pseudouridine synthases, significant differences also exist. The active site of RsuA is much more open than those of TruB, RluC, and RluD which have extended β-hairpins and loops occluding the substrate pocket. Access to the uridine targeted for conversion to pseudouridine requires the base to be "flipped" out from the local RNA tertiary structure, and RluC and RluD probably use means broadly similar to those of TruB and RsuA to achieve this. As with all enzymes, however, catalysis and recognition are closely related, and it is clear that the interactions with RNA are different. The role of the positive charges close to the catalytic aspartate is unclear. They possibly bring the aspartate into a catalytically active position on correct substrate binding, and may hold it in an inactive conformation otherwise. This would help ensure the chemically simple transformation to pseudouridine (which requires no input energy or cofactors) is restricted to the intended sites. Since uracil and UMP evidently bind RluC and RluD extremely weakly, binding energy from other parts of the substrate is needed to drive the reaction.

The absence of any hint of electron density for the N-terminal domain of RluD was disappointing, but suggests that this protein (and possibly RluC) adopts the familiar tactic of using disordered regions to provide specific binding of moderate affinity (*52*). Interestingly, Sivaraman and colleagues were unable to obtain well-ordered crystals of full-length RluD, presumably because of attachment of an N-terminal histidine tag. They therefore used proteolysis to identify a folded domain, and recloned RluD missing the first 68 residues (*15*). A 23-residue N-terminal tag did not prevent crystallization in space group $P4_32_12$ (*16*). We did not use histidine tags since both native RluD and RluC(92−319) proved to be stable, soluble proteins that could be readily purified by conventional column chromatography, making tags (and recloning) unnecessary. A 29-amino acid disordered stretch of TruB was found to become ordered on RNA binding (*47*), so this may be a general feature of these enzymes. The sequences of the N-terminal regions of RluC and RluD give no suggestion of a low level of complexity. Further experiments are underway to determine the structure of these proteins in the presence of substrate and analogues.

ACKNOWLEDGMENT

We are grateful to Dr. K. Nagai for help with crystallization, Prof. S. Wakatsuki, Drs. M. Suzuki, N. Igarashi, N. Matsunaga, and T. Yokoyama for help with data collection at the Photon Factory, and to Prof. G. J. Davies for helpful discussions.

REFERENCES

1. Ofengand, J., and Fournier, M. (1998) in *Modification and Editing of RNA* (Grosjean, H., and Benne, R., Eds.) pp 229−253, ASM Press, Washington, DC.
2. 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.
3. Charette, M., and Gray, M. W. (2000) Pseudouridine in RNA: what, where, how, and why, *IUBMB Life 49*, 341−351.
4. Gu, X., Liu, Y., and Santi, D. V. (1999) The mechanism of pseudouridine synthase I as deduced from its interaction with 5-fluorouracil-tRNA, *Proc. Natl. Acad. Sci. U.S.A. 96*, 14270−14275.

5. Mueller, E. G. (2002) Chips off the old block, *Nat. Struct. Biol. 9*, 320−322.
6. Newby, M. I., and Greenbaum, N. L. (2002) Investigation of Overhauser effects between pseudouridine and water protons in RNA helices, *Proc. Natl. Acad. Sci. U.S.A. 99*, 12697−12702.
7. Raychaudhuri, S., Niu, L., Conrad, J., Lane, B. G., and Ofengand, J. (1999) Functional effect of deletion and mutation of the *Escherichia coli* ribosomal RNA and tRNA pseudouridine synthase RluA, *J. Biol. Chem. 274*, 18880−18886.
8. 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.
9. 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.
10. Kaya, Y., and Ofengand, J. (2003) A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya, *RNA 9*, 711−721.
11. 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.
12. Sivaraman, J., Sauve, 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.
13. Hoang, C., and Ferre-D'Amare, A. R. (2001) Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme, *Cell 107*, 929−939.
14. Del Campo, M., Ofengand, J., and Malhotra, A. (2003) Purification and crystallization of *Escherichia coli* pseudouridine synthase RluD, *Acta Crystallogr. D59*, 1871−1873.
15. Sivaraman, J., Iannuzzi, P., Cygler, M., and Matte, A. (2004) Crystal structure of the RluD pseudouridine synthase catalytic module, an enzyme that modifies 23S rRNA and is essential for normal cell growth of *Escherichia coli*, *J. Mol. Biol. 335*, 87−101.
16. Del Campo, M., Ofengand, J., and Malhotra, A. (2004) Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*, *RNA 10*, 231−239.
17. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam Protein Families Database, *Nucleic Acids Res. 30*, 276−280.
18. 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.
19. Gutgsell, N. S., Del Campo, M. D., 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 normal growth to an RluD-minus strain, *RNA 7*, 990−998.
20. 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.
21. Conrad, J., Sun, D., Englund, N., and Ofengand, J. (1998) The rluC gene of *Escherichia coli* codes for a pseudouridine synthase that is solely responsible for synthesis of pseudouridine at positions 955, 2504, and 2580 in 23S ribosomal RNA, *J. Biol. Chem. 273*, 18562−18566.
22. Ofengand, J., and 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.
23. Yusupov, M. M., et al. (2001) Crystal structure of the ribosome at 5.5 Å resolution, *Science 292*, 883−896.
24. Davies, C., Gerstner, R. B., Draper, D. E., Ramakrishnan, V., and White, S. W. (1998) The crystal structure of ribosomal protein S4 reveals a two domain molecule with an extensive RNA-binding surface, *EMBO J. 17*, 4545−4558.
25. Naoi, Y., Chong, K. T., Yoshimatsu, K., Miyazaki, G., Tame, J. R. H., Park, S.-Y., Adachi, S., and Morimoto, H. (2001) The functional similarity and structural diversity of human and cartilaginous fish haemoglobins, *J. Mol. Biol. 307*, 259−270.
26. 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.
27. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure, *EMBO J. 9*, 1665−1672.
28. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol. 276*, 307−326.
29. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Crystallogr. D55*, 849−861.
30. Terwilliger, T. C. (2001) Maximum-likelihood density modification using pattern recognition of structural motifs, *Acta Crystallogr. D57*, 1755−1762.
31. Vagin, A. A., and Teplyakov, A. (2000) An approach to multi-copy search in molecular replacement, *Acta Crystallogr. D56*, 1622−1624.
32. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
33. Roussel, A., and Cambillau, C. (1989) in *Silicon Graphics Geometry Partners Directory*, Silicon Graphics, Mountain View, CA.
34. McRee, D. E. (1999) XtalView/Xfit: A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol. 125*, 156−165.
35. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
36. Brunger, A. T. (1996) *X-PLOR*, version 3.85, Yale University Press, New Haven, CT.
37. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K., and Dodson, E. J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT, *Acta Crystallogr. D55*, 247−255.
38. Laskowski, R., MacArthur, M. W., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
39. Corollo, D., Blair-Johnson, M., Conrad, J., Fiedler, T., Sun, D., Wang, L., Ofengand, J., and Fenna, R. (1999) Crystallization and characterization of a fragment of pseudouridine synthase RluC from *Escherichia coli*, *Acta Crystallogr. D55*, 302−304.
40. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.
41. Guruprasad, K., Reddy, B. V. B., and Pandit, M. W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, *Protein Eng. 4*, 155−161.
42. Wright, P. E., and Dyson, H. J. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54−60.
43. Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A., and Bucher, P. (2002) PROSITE: a documented database using patterns and profiles as motif descriptors, *Briefings Bioinf. 3*, 265−274.
44. Nurse, K., Wrzesinski, J., Bakin, A., Lane, B. G., and Ofengand, J. (1995) Purification, cloning, and properties of the tRNA psi 55 synthase from *Escherichia coli*, *RNA 1*, 102−112.
45. Gutgsell, N., Englund, N., Niu, L., Kaya, Y., Lane, B. G., and Ofengand, J. (2000) Deletion of the *Escherichia coli* pseudouridine synthase gene truB blocks formation of pseudouridine 55 in tRNA in vivo, does not affect exponential growth, but confers a strong selective disadvantage in competition with wild-type cells, *RNA 6*, 1870−1881.
46. Kinghorn, S. M., O'Byrne, C. P., Booth, I. R., and Stansfield, I. (2002) Physiological analysis of the role of truB in *Escherichia coli*: a role for tRNA modification in extreme temperature resistance, *Microbiology 148*, 3511−3520.
47. 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, *Proc. Natl. Acad. Sci. U.S.A. 100*, 12648−12653.
48. Cheng, X., and Blumenthal, R. M. (2002) Cytosines do it, thymines do it, even pseudouridines do it: base flipping by an enzyme that acts on RNA, *Structure 10*, 127−129.
49. Ramamurthy, V., Swann, S. L., Spedaliere, C. J., and Mueller, E. G. (1999) Role of cysteine residues in pseudouridine synthases of different families, *Biochemistry 38*, 13106−13111.

50. Huang, L., Pookanjanatavip, M., Gu, X., and Santi, D. V. (1998) A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst, *Biochemistry 37*, 344−351.
51. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman, New York.
52. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321−331.
53. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
54. Nicholls, A., Sharp, K. A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins: Struct., Funct., Genet. 11*, 281−296.

BI036079C