Published in final edited form as:

J Mol Biol. 2007 April 13; 367(5): 1459–1470.

The Crystal Structure of *E. coli* rRNA Pseudouridine Synthase RluE

Hu Pan^{a,c}, Joseph D. Ho^b, Robert M. Stroud^a, and Janet Finer-Moore^{*,a}

Department of Biochemistry and Biophysics and Graduate Group in Chemistry and Chemical Biology, University of California at San Francisco, San Francisco CA 94143

SUMMARY

Pseudouridine synthase RluE modifies U2457 in a stem of 23S RNA in *E. coli*. This modification is located in the peptidyl transferase center of the ribosome. We determined the crystal structures of the C-terminal, catalytic domain of *E. coli* RluE at 1.2Å resolution and of full-length RluE at 1.6Å resolution. The crystals of the full-length enzyme contain two molecules in the asymmetric unit and in both molecules the N-terminal domain is disordered. The protein has an active site cleft, conserved in all other pseudouridine synthases, that contains invariant Asp and Tyr residues implicated in catalysis. An electropositive surface patch that covers the active site cleft is just wide enough to accommodate an RNA stem. The RNA substrate stem can be docked to this surface such that the catalytic Asp is adjacent to the target base, and a conserved Arg is positioned to help flip the target base out of the stem into the enzyme active site. A flexible RluE specific loop lies close to the conserved region of the stem in the model, and may contribute to substrate specificity. The stem alone is not a good RluE substrate, suggesting RluE makes additional interactions with other regions in the ribosome.

Keywords

RNA modifying enzyme; X-ray crystallography; protein-RNA complex; ribosome; substrate specificity

INTRODUCTION

In all kingdoms of life, most stable cellular RNAs undergo extensive and specific post-transcriptional modification by enzymes in the cell¹. Isomerization of uridine to pseudouridine (Ψ) (Scheme 1) is one of the most common RNA modifications, and it occurs in most types of RNA^{2; 3; 4}. The role of pseudouridylation is still unclear, although some evidence suggests that it may help stabilize RNA tertiary structure^{5; 6}. Gene deletion experiments have shown that none of the Ψ synthases that target *E. coli* rRNA are essential and only deletion of the gene for the ribosomal Ψ synthase RluD has a significant impact on cell growth². Similarly, elimination of single Ψs in the yeast ribosome does not affect growth^{7; 8}. However, several point mutations in Cbf5, the Ψ synthase that is responsible for the whole set of Ψ modifications

*Address correspondence to: Janet Finer-Moore (finer@msg.ucsf.edu), S412B UCSF-GENENTECH HALL, 600 16th Street, San Francisco, California 94143-2240, Tel: 415 502-5426, Fax: 415 476 1902

^aDepartment of Biochemistry and Biophysics

^bGraduate Group in Chemistry and Chemical Biology

^cCurrent address: Plexxikon, Inc., 91 Bolivar Drive, Berkeley, CA, 94710

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

in yeast rRNA, were strongly detrimental to yeast cell growth⁹. These results suggest that while individually the Ψ s in rRNA may be dispensable, the sum of their contributions to stability and function of rRNA is critically important in cells.

In prokaryotes, distinct pseudouridine synthases modify different RNA sites, often with exquisite specificity. For example, TruB catalyzes Ψ formation only at U55 in transfer RNAs¹⁰. Other prokaryotic Ψ synthases specifically modify two or more sites that are not in highly similar environments and may even lie on different types of RNA^{11; 12}. Thus a challenge for structural biology is to determine how substrate specificity is determined by Ψ synthase structure and plasticity.

Prokaryotic Ψ synthases have been grouped into five subfamilies based on sequence conservation^{13; 14; 15}. The subfamilies, named for their first discovered representative member, are TruA, TruB, RluA, RsuA, and TruD^{15; 16}. Crystal structures of representative members of each family have been determined and reveal that even though sequence conservation between any two subfamilies is low, all Ψ synthases have a structurally conserved catalytic core^{16; 17; 18; 19; 20; 21; 22; 23; 24; 25; 26; 27; 28; 29; 30}. The only structures of Ψ synthase-substrate complexes published to date are those of TruB and RluA co-crystallized with small RNA stem-loops that are minimal substrates of the enzymes^{21; 22; 26; 27}. These structures show that the substrate RNA binds to a positively charged surface on the protein with the target base flipped out of the stem-loop and into an active site cleft next to an invariant catalytic Asp.

E. coli RluE is a member of the RsuA family³¹. Like other RsuA family members, it is a highly specific enzyme, modifying the single site Ψ 2457 on a stem of 23S RNA. Ψ 2457 is one of five Ψ s at the peptidyl transferase center (PTC) of the *E. coli* ribosome. We have not yet been able to identify a small RNA substrate of RluE suitable for co-crystallization. Therefore, in order to shed light on the substrate specificity of this enzyme we have determined the crystal structure of the apo enzyme and docked the ribosomal stem containing Ψ 2457 into the likely active site cleft of the enzyme.

RESULTS AND DISCUSSION

Domain structure

RluE was expressed with a His₆-tag at the N-terminus of the protein. A standard His-tag cleavage reaction by thrombin yielded two products. One product had a molecular weight equivalent to that of the full-length protein, indicating that the cleavage had occurred at the expected site, while the other product was about 3 KDa smaller than the full-length product. The likely second cleavage site that yielded the smaller protein was between Arg-28 and Ser-29. There is precedent for thrombin cleavage between Arg and Ser³² and the MW of residues 1–28 is 3.5 KDa.

Both products were crystallized, and the crystal structures were solved by molecular replacement. Density for the N-terminal 35 amino acids of RluE was not visible in electron density maps calculated for either truncated or full-length RluE. A large single crystal of full-length RluE was carefully washed in protein-free mother liquor, dissolved in water and run on an SDS-PAGE gel in order to find out if full-length RluE had undergone proteolysis during crystallization. The gel showed a single protein band with the molecular mass of full-length RluE. Therefore, the N-terminal 35 residues of RluE must be disordered in the crystal.

The disordered N-terminal residues may comprise a separate protein domain that vibrates as a rigid body. Other Ψ synthases in the RsuA family contain an S4-like domain at their N-termini

that is postulated to bind nonspecifically to RNA substrate^{29; 30}. Similar RNA-binding domains are bound nonspecifically to RNA in crystal structures of TruB-RNA^{22; 26; 27} and in an RNA complex of the uridine methyltransferase RmA³³. The RNA-binding domains of Ψ synthases are usually attached to the rest of the protein by a flexible tether, and so are frequently disordered in crystal structures of the apo enzymes. The N-terminal 35 residues of *E. coli* RluE contain six arginines and three lysines. The basic nature of the N-terminal region is consistent with its putative role in RNA binding.

The cleaved N-terminal fragment of *E. coli* RluE is considerably smaller than the S4-like domains of other members of the RsuA family. However, Sunita et al. have detected weak homology between the N-terminal residues of RluE and the residues comprising the S4-like domain of RsuA³⁰. Nine of the first 28 residues are hydrophobic and predicted to be less than 25% solvent accessible³⁴. Secondary structure analysis predicts that residues 2–24 are ordered and contain a pair of β -strands and possibly an α -helix^{35; 36; 37}. Thus it is possible that the missing 35 residues in RluE form a compact RNA-binding domain, but not one with a canonical S4 fold.

All residues after Glu-35 in truncated RluE or after Gln-37 in full-length RluE are clearly defined in electron density maps except for Glu-159, which lies in a long loop connecting β -strands 7 and 8. Residues 41–217 are homologous to the catalytic domains of other Ψ synthases, especially to other members of the RsuA family (Fig. 1a). The sequence identity between the *E. coli* RluE and *E. coli* RsuA catalytic domains is approximately 30%. RluE sequences have a conserved 16-amino acid insert with respect to other family members after residue 149 (Fig. 1a).

The refined structures of residues 38–217 are the same in truncated RluE and full-length RluE except for small conformational differences reflecting the flexibility of the structure, which will be discussed later. Therefore, the higher resolution, truncated RluE structure has been used for analysis and for comparison to other Ψ synthases.

Active site structure

The RluE catalytic domain adopts the same α/β fold as RsuA²⁹ and RluF³⁰ (Fig. 1b). This fold features a bifurcated, mostly antiparallel β -sheet, which is also present in all other Ψ synthase structures, and four conserved helices (three α -helices and one 3_{10} helix) that pack against the sheet. The central strands of the β -sheet, strands $\beta 3$, $\beta 4$, $\beta 9$, $\beta 10$ in Fig. 1b, form the floor of a cleft, which is highly conserved in all Ψ synthases. This cleft almost certainly contains the active site of the enzyme. Five sequence motifs that characterize the Ψ synthase family and that contain invariant residues shown to be essential for catalysis¹⁸ lie in or adjacent to the cleft (Figs. 1a, 2). Asp-79, the essential catalytic nucleophile, lies in a loop forming one side of the cleft. Invariant Tyr-109, which has also been implicated in catalysis³⁸, points into the cleft from a β -strand on the cleft floor.

In the truncated RluE crystals, a sulfate ion of crystallization makes hydrogen bonds with the catalytic Asp and Tyr. O1 forms a hydrogen bond with the main chain amide group of Asp-79, and to the side chain of Ser-82; oxygen atom O2 hydrogen bonds to the hydroxyl group of Tyr-109; O3 makes a hydrogen bond to the side chain oxygen of Asp-79 and O4 accepts a hydrogen bonds from the main chain amide group of Asp-79. The presence of this ion explains why we were unable to introduce small molecules such as uracil or UMP in the active site of RluE by soaking or co-crystallization. Even when we reduced the ammonium sulfate concentration in the mother liquor from 200 mM to about 20 mM and increased the small molecule concentration to 200mM during co-crystallization, a sulfate ion was still invariably seen in the active site. Therefore, the binding affinity for these substrate mimics must be low.

In order to assess the degree of structural conservation of the RluE active site cleft with respect to those of other Ψ synthases, we overlapped the structure of a *Thermatoga maritima* TruB-RNA complex (PDB accession code **1R3F**) and the structure of the *E. coli* RsuA-UMP complex (PDB accession code **1KSK**) with the high resolution RluE crystal structure. When the Cas in the common cores of RluE and the other two proteins are superimposed, the conserved motifs closely align and the SO_4^{2-} in RluE approximately overlaps both the UMP in RsuA-UMP (Fig. 2b), and the target base in TruB-RNA. In the TruB complex, hydrogen bonds between the catalytic residues Asp-48 and Tyr-76 and the base and ribose of the target uridine are in structurally equivalent positions to the hydrogen bonds between Asp-79 and Tyr-109 and the sulfate oxygens in RluE. Thus, SO_4^{2-} appears to be a substrate mimic in RluE. The close structural similarity of the active sites of RluE, TruB, and RsuA suggests that the three proteins use the same enzyme mechanism and that their target uridines are positioned similarly during catalysis. This result puts a powerful constraint on models for how RluE binds its RNA substrate.

RNA binding

The surface electrostatic potential of RluE shows that on the same side of the molecule as the active site cleft there is a patch of positively charged surface that encompasses the cleft and extends the length of the protein (Fig. 3). Loop L_{1-2} and the β_4 - β_5 ribbon form walls on either side of this surface, defining a 20Å-wide shallow trough. The surface has a prominent depression near the catalytic Asp that could accommodate the target base, similar to the depression seen in RluF³⁰. We were able to manually dock a fragment of *E. coli* rRNA (G2450-C2499) containing the target U2457 onto the trough such that the target U was within 5Å of the catalytic Asp-79, adjacent to the depression on the surface. (Fig. 3, Fig. 4). The trough was just wide enough to accommodate the RNA stem in either of two, two-fold related orientations. The RNA configuration shown in Fig. 4 is the more attractive of the two models because the RluE-specific insert, loop L_{7-8} , is close to the upper part of the RNA, which is conserved, while the putative N-terminal RNA-binding domain would be positioned to interact with the nonconserved end of the rRNA stem via nonspecific interactions.

As is the case for other Ψ synthases, the target uridine base is buried in the substrate RNA, where it base-pairs to G2529, and it would need to be flipped out of the RNA stem in order to enter the enzyme active site. Arg-77, which is conserved in the members of the RluA, RsuA and TruA families, is adjacent to the target U in the docked complex. In this position it could assume its proposed role in flipping the target U out of the RNA stem into the enzyme active site²¹. In RluA, the guanidinium group of the Arg corresponding to Arg-77 helps to flip the target U into the active site of the enzyme by intercalating between layers of stacked bases in the substrate RNA, essentially substituting for the target uridine in the stem-loop²¹.

The walls of the RNA-binding trough are relatively flat and the RNA stem shown in Fig. 4 is fairly symmetrical in shape, thus it does not appear that shape complementarity alone could determine the vertical alignment of the 43-mer in the binding trough. The 43-mer in the docked complex is not a substrate for RluE, nor have we identified any analogs of the stem that are minimal substrates for the enzyme; therefore interactions between RluE and other regions of the fully or partially assembled ribosome may be required for alignment of the target U with the active site. The RluE-specific insert, L_{7-8} , (residues 149–164) which is conserved among RluE species (Fig. 1A) may contribute some of these interactions. The loop contains three arginines and one lysine, all highly conserved (Fig. 1a) that could contribute to RNA binding. The loop has a well-defined structure that is conserved in the three crystallographically independent copies of RluE seen in the two crystal structures reported here. Residues 157–163 in the C-terminal half of the loop have isotropic temperature factors that are approximately

twice as high as the average atomic B-factor for the protein, indicating they are more mobile than the rest of the protein and may be able to adapt their conformation to bound substrate. This mobile region of the loop is bounded by cis-Pro-155 and Pro-156 on one side and Pro-164 on the other. Analysis of elastic network models^{39; 40} suggests there are hinge residues at the base of the loop that allow it to undergo rigid body motions large enough to bring it into contact with the top of the docked 43-mer (as viewed in Fig. 4).

Conformational differences between RluE crystal structures indicate flexibility

The unit cell of full-length RluE crystals contains two protein molecules in an asymmetric unit, referred to as molecules A and B. Thus the two crystal structures reported here show the RluE catalytic domain in three different crystal packing environments. Conformational differences between these three independent structures indicate the flexible regions of the molecule and suggest protein motions that may be important in catalysis.

In the full-length RluE crystal structure, one side of molecule A packs against the putative RNA binding trough of molecule B (Fig. 5a). The overall rms deviation for the C α s of the two molecules after they are aligned is large for identical proteins, 1.55Å, which indicates there are significant conformational differences between the molecules. Difference-distance analysis identifies two large subdomains of conserved structure on either side of the active site (Fig. 5b). The rms deviation in the positions of 58 C α s within the larger subdomain, after the two molecules are aligned on these atoms, is 0.166Å. The rms deviation in positions of 28 C α s in the smaller subdomain is 0.45Å after they are superimposed. The subdomains have different relative orientations, leading to a more closed active site cleft in molecule B than in molecule A. Truncated RluE contains the same two subdomains of conserved structure, and its cleft is slightly less open than the cleft in molecule A (Fig. 5c). These conformational differences between RluE molecules in different crystal packing environments suggests the protein uses a hinge motion to open and close the active site cleft. Elastic network model analysis predicts such a hinge motion in RluE and identifies two residues at the base of the cleft that would mediate this motion^{39; 40} (HingeProt web server, <http://bioinfo3d.cs.tau.ac.il/HingeProt/index.html>). The hinge motion would bring the β 4- β 5 ribbon and loop L₁₋₂ closer to the RNA substrate, and therefore may play a role in RNA binding. A similar hinge motion has been detected in apo TruB¹⁷ and this motion helps to close the TruB active site cleft upon substrate RNA binding²⁶.

There are conformational differences in the loops that are not described by rigid body motion of the conserved subdomains, indicating the loops are flexible. By analogy to TruB and RluA, some of these loops are probably involved in RNA binding^{21; 22; 26}. The conformational differences in the RluE-specific loop, L₇₋₈, are minor except at the most mobile residues, residues 157–163 (Fig. 5c).

Comparison with other Ψ synthases from different families

The structures of Ψ synthases from all five families have been solved. Although RNA-bound structures of only two of these enzymes have been determined, there are common features seen in all the structures that indicate they may use variations on the same mechanism for RNA recognition and binding. All have a catalytic core with a large electropositive surface area that would complement the shape of RNA duplex^{16; 20; 25; 26; 30}. This suggests that, as in the case of TruB, a region of the RNA substrate binds as a rigid body to the protein via relatively nonspecific electrostatic interactions. The protein-RNA interface is extended in many cases by a separate RNA-binding domain connected to the catalytic domain by a flexible tether. Most Ψ synthases have unique inserts that form flexible loops bordering the active site (Fig. 6). These loops likely contribute to substrate specificity. By analogy to TruB and RluA, the loops and

the RNA substrate probably adopt mutually complimentary conformations that maximize specific binding interactions and orient the target base in the active site. Thus in all Ψ synthases, RNA substrates may bind using a combination of rigid docking to a large electropositive surface and induced fit in the vicinity of the active site. Substrate specificity therefore may be partly determined by the accessible conformations of RNA-binding loops and/or of the regions of RNA in the local vicinity of the target uridine.

The mechanism for substrate recognition is likely different in detail for RluE, which modifies a U in an rRNA stem, than it is for TruB or RluA, which modify Us in tRNA or rRNA stem loops. Crystal structures have been solved for TruB and RluA in complexes with ~20-nucleotide long hairpins that are analogs of tRNA stem loops and also substrates for the enzymes^{21; 22; 26}. TruB and RluA both have prominent loops on one or both sides of the active site that bind in the grooves of the RNA hairpins and orient them for catalysis. In TruB a 29-amino acid loop, called the thumb loop, anchors the tRNA T loop in the active site cleft, almost completely burying it, and providing most of the specific protein-RNA contacts²⁶. RluA has a smaller thumb loop in the equivalent position, which also binds in the major groove of the substrate RNA²¹. In addition, RluA has a second loop, the “forefinger loop” that projects into the active site from the opposite side of the binding cleft and binds in the minor groove of the substrate. The corresponding loops in RluE are very small (in the case of the thumb loop) or folded back on the protein (in the case of the forefinger loop). Thus they are not able to clamp down on the substrate in the same fashion as in the former enzymes. They are also not sufficient for substrate binding and orientation since analogs of the *E. coli* rRNA stem are not minimal substrates for RluE (unpublished). RluE interactions with rRNA regions outside the stem apparently provide the affinity and orientation constraints required for catalysis.

MATERIALS AND METHODS

Protein Expression and Purification

The *E. coli* rluE gene was PCR amplified from its genomic DNA and was cloned into NdeI-NotI sites of plasmid pET-28b (Novagen). The resulting construct had an N-terminal (His)₆ tag. The protein was expressed in *E. coli* BL21(DE3) cells and was purified by metal affinity chromatography using Talon resin (Clontech). Cleavage of the His-tag using thrombin generated two products: one with a mobility on SDS-PAGE gels corresponding to the molecular weight of full length RluE (24.9 KDa) and another with a molecular weight approximately 3KDa less, corresponding to a version of the protein that was truncated after a putative N-terminal RNA-binding domain, most likely by cleavage after Arg-28. The two products were further purified on a Hi-trap S-sepharose column (Pharmacia).

Crystallization and Data Collection

The truncated *E. coli* RluE protein was concentrated to 4 mg/ml in 10 mM Tris, pH 7.5, 2 mM EDTA and 2 mM DTT. Crystals were grown by vapor phase diffusion from hanging drops at room temperature. Protein was equilibrated against reservoirs containing 22% (w/v) MME PEG 2000, 200 mM ammonium sulfate, 100 mM sodium acetate pH 5.0. Initial rod-like crystals appeared in 3–5 days, however, all the crystals were twinned and grew in clusters. The microseeding technique was used to improve the crystal quality by transferring crushed and diluted initial crystal seeds into preequilibrated drops containing 18–22% MME PEG 2000, 200mM ammonium sulfate, 100mM sodium acetate pH 5.0, and 2mg/ml protein. Single crystals appeared in the next day and developed to full size (200×100×50 μm^3) in about two weeks. The same microseeding procedure was used to obtain crystals of full-length RluE.

Crystals of full-length and truncated RluE were frozen in mother liquor (24% MME PEG 2000, 200mM ammonium sulfate, 100mM sodium acetate pH 5.0) with 15% glycerol added. Native data were collected on an ADSC Quantum4 CCD detector at the Advanced Light Source (Beamline 8.3.1) using single wavelength X-rays produced with a double crystal monochromator.

The crystals of truncated RluE diffracted to 1.2Å resolution and belonged to space group $P2_12_12_1$ with cell dimensions $a = 32.3\text{\AA}$, $b = 56.8\text{\AA}$ and $c = 91.3\text{\AA}$, one molecule in the asymmetric unit, and a solvent content of 37% ($V_m = 2.0\text{\AA}^3/\text{Da}$). Crystals of the full-length RluE belonged to the same space group, but had cell dimensions $a = 59.0\text{\AA}$, $b = 78.9\text{\AA}$ and $c = 84.2\text{\AA}$ and two molecules per asymmetric unit ($V_m = 2.0\text{\AA}^3/\text{Da}$), and diffracted to 1.4Å. Data were collected to 1.2Å and 1.6Å resolution for truncated and full-length RluE, respectively. Diffraction intensities were integrated and reduced using the program DENZO, and scaled using SCALEPACK⁴¹ (Table 1).

The diffraction images for the full-length RluE crystal were contaminated with sharp powder diffraction rings, so reflections in the resolution shells containing the rings (2.24Å-2.28Å and 1.85Å-1.92Å) were removed from the data set. The data collection statistics for full-length RluE in Table 1 are for the complete data set, while the refinement statistics in Table 2 are for the data set with the contaminated reflections removed (R_{crys} and R_{free} for full-length RluE refined against the complete data set were approximately 1% higher than those reported in Table 2).

Structure determination

The structures of full-length and truncated RluE were determined by molecular replacement in CNS⁴² using data to 3.5Å resolution. The search model was the crystal structure of RsaA (PDB **1KSV**) with the N-terminal RNA-binding domain, waters and ligands removed and nonconserved residues replaced by serines.

For the truncated protein structure, rounds of rebuilding and addition of solvent molecules using Quanta (Accelrys, San Diego, CA), alternated with restrained refinement of atomic positions and isotropic temperature factors in CNS⁴² were carried out until the R-factor reached 20.1% ($R_{\text{free}} = 21.3\%$). The structure was then refined by conjugate gradient minimizations using individual, restrained, anisotropic B-factors, and hydrogens included in riding positions first using SHELX-97⁴³ and finally using REFMAC5⁴⁴, until convergence at $R = 14.9\%$ and $R_{\text{free}} = 17.9\%$. All refinements in CNS and REFMAC5 were against a maximum likelihood target and used a bulk solvent correction. There was no electron density for residues before Asn-36 or for residue Glu-159 in loop L₇₋₈ (Fig. 1); these residues are either not present in the truncated protein or are disordered in the crystals. An analysis of the geometry⁴⁵ shows that all refinement statistics are well within the expected values at this resolution (Table 2).

Full-length RluE was refined in a similar manner, except that anisotropic B-factors were not refined, and only CNS and REFMAC5 were used for refinement. The two protein molecules in the asymmetric unit had slightly different conformations, so non-crystallographic symmetry restraints were not used. No density for residues 1–37 in one molecule or 1–38 in the other was present in electron density maps, and Glu-159 was also disordered in both molecules. R_{crys} and R_{free} were 21.5% and 25.5%, respectively, at the end of refinement (Table 2).

Structure analysis

For overlap of structurally homologous proteins we used Ca-Ca distance-differences to identify the largest structurally conserved core for the two proteins⁴⁶, then superimposed the core Ca atoms by least-squares refinement using LSQMAN from the DEJAVU suite of programs.

GRASP⁴⁷ was used for calculating and plotting the electrostatic potential surface of RluE. The RNA stem containing the RluE target base, Ψ2457, was taken from the crystal structure of the *E. coli* large ribosomal subunit⁴⁸ (PDB accession code **2AW4**) and manually docked into the active site of RluE using Quanta. The model of the complex was subjected to several rounds of energy minimization in CNS⁴².

Sequences were aligned using CLUSTALW⁴⁹ and plotted using ESPript⁵⁰. Figs. 1b, 2, 4, 5 and 6b were made with PyMOL⁵¹.

Protein Data Bank accession codes

Coordinates and structure factors have been deposited with RCSB Protein Data Bank with accession codes 2OML and 2OLW for truncated and full-length RluE, respectively.

Acknowledgements

We thank the staff of beamline 8.3.1 at ALS for assistance with data collection, Consuelo Castro for technical assistance, and Patricia J. Greene for critical reading of the manuscript. This research was supported by grant from the National Institutes of Health GM51232 (to R.M.S.).

References

1. Bjork, G. Stable RNA Modification. In: Neidhardt, FC.; Curtiss, R., editors. *Escherichia coli and Salmonella: cellular and molecular biology*. Washington D.C.; ASM Press: 1996.
2. Ofengand J. Ribosomal RNA pseudouridines and pseudouridine synthases. *FEBS Lett* 2002;514:17–25. [PubMed: 11904174]
3. Patton JR. Multiple pseudouridine synthase activities for small nuclear RNAs. *Biochem J* 1993;290 (Pt 2):595–600. [PubMed: 8452551]
4. Samuelsson T, Olsson M. Transfer RNA pseudouridine synthases in *Saccharomyces cerevisiae*. *J Biol Chem* 1990;265:8782–7. [PubMed: 2187870]
5. Durant PC, Davis DR. Stabilization of the anticodon stem-loop of tRNA^{Lys},3 by an A+-C base-pair and by pseudouridine. *J Mol Biol* 1999;285:115–31. [PubMed: 9878393]
6. Yarian CS, Basti MM, Cain RJ, Ansari G, Guenther RH, Sochacka E, Czerwinska G, Malkiewicz A, Agris PF. Structural and functional roles of the N1- and N3-protons of psi at tRNA's position 39. *Nucleic Acids Res* 1999;27:3543–9. [PubMed: 10446245]
7. Lowe TM, Eddy SR. A computational screen for methylation guide snoRNAs in yeast. *Science* 1999;283:1168–71. [PubMed: 10024243]
8. Samarsky DA, Fournier MJ. A comprehensive database for the small nucleolar RNAs from *Saccharomyces cerevisiae*. *Nucleic Acids Res* 1999;27:161–4. [PubMed: 9847166]
9. Zebardjian Y, King T, Fournier MJ, Clarke L, Carbon J. Point mutations in yeast CBF5 can abolish in vivo pseudouridylation of rRNA. *Mol Cell Biol* 1999;19:7461–72. [PubMed: 10523634]
10. Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J. Purification, cloning, and properties of the tRNA psi 55 synthase from *Escherichia coli*. *Rna* 1995;1:102–12. [PubMed: 7489483]
11. Huang L, Pookanjanatavip M, Gu X, Santi DV. A conserved aspartate of tRNA pseudouridine synthase is essential for activity and a probable nucleophilic catalyst. *Biochemistry* 1998;37:344–51. [PubMed: 9425056]

12. Wrzesinski J, Nurse K, Bakin A, Lane BG, Ofengand J. A dual-specificity pseudouridine synthase: an *Escherichia coli* synthase purified and cloned on the basis of its specificity for psi 746 in 23S RNA is also specific for psi 32 in tRNA(phe). *Rna* 1995;1:437–48. [PubMed: 7493321]
13. Gustafsson C, Reid R, Greene PJ, Santi DV. Identification of new RNA modifying enzymes by iterative genome search using known modifying enzymes as probes. *Nucleic Acids Res* 1996;24:3756–62. [PubMed: 8871555]
14. Kaya Y, Ofengand J. A novel unanticipated type of pseudouridine synthase with homologs in bacteria, archaea, and eukarya. *Rna* 2003;9:711–21. [PubMed: 12756329]
15. Koonin EV. Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. *Nucleic Acids Res* 1996;24:2411–5. [PubMed: 8710514]
16. Kaya Y, Del Campo M, Ofengand J, Malhotra A. Crystal structure of TruD, a novel pseudouridine synthase with a new protein fold. *J Biol Chem* 2004;279:18107–10. [PubMed: 14999002]
17. Chaudhuri BN, Chan S, Perry LJ, Yeates TO. Crystal structure of the apo forms of psi 55 tRNA pseudouridine synthase from *Mycobacterium tuberculosis*: a hinge at the base of the catalytic cleft. *J Biol Chem* 2004;279:24585–91. [PubMed: 15028724]
18. Del Campo M, Ofengand J, Malhotra A. Crystal structure of the catalytic domain of RluD, the only rRNA pseudouridine synthase required for normal growth of *Escherichia coli*. *Rna* 2004;10:231–9. [PubMed: 14730022]
19. Ericsson UB, Nordlund P, Hallberg BM. X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold. *FEBS Lett* 2004;565:59–64. [PubMed: 15135053]
20. Foster PG, Huang L, Santi DV, Stroud RM. The structural basis for tRNA recognition and pseudouridine formation by pseudouridine synthase I. *Nat Struct Biol* 2000;7:23–7. [PubMed: 10625422]
21. Hoang C, Chen J, Vizthum CA, Kandel JM, Hamilton CS, Mueller EG, Ferre-D'Amare AR. Crystal structure of pseudouridine synthase RluA: indirect sequence readout through protein-induced RNA structure. *Mol Cell* 2006;24:535–45. [PubMed: 17188032]
22. Hoang C, Ferre-D'Amare AR. Cocrystal structure of a tRNA Psi55 pseudouridine synthase: nucleotide flipping by an RNA-modifying enzyme. *Cell* 2001;107:929–39. [PubMed: 11779468]
23. Hoang C, Ferre-D'Amare AR. Crystal structure of the highly divergent pseudouridine synthase TruD reveals a circular permutation of a conserved fold. *Rna* 2004;10:1026–33. [PubMed: 15208439]
24. Matte A, Louie GV, Sivaraman J, Cygler M, Burley SK. Structure of the pseudouridine synthase RsuA from *Haemophilus influenzae*. *Acta Crystallograph Sect F Struct Biol Cryst Commun* 2005;61:350–4.
25. Mizutani K, Machida Y, Unzai S, Park SY, Tame JR. Crystal structures of the catalytic domains of pseudouridine synthases RluC and RluD from *Escherichia coli*. *Biochemistry* 2004;43:4454–63. [PubMed: 15078091]
26. Pan H, Agarwalla S, Moustakas DT, Finer-Moore J, Stroud RM. 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* 2003;100:12648–53. [PubMed: 14566049]
27. Phannachet K, Huang RH. Conformational change of pseudouridine 55 synthase upon its association with RNA substrate. *Nucleic Acids Res* 2004;32:1422–9. [PubMed: 14990747]
28. Sivaraman J, Iannuzzi P, Cygler M, Matte A. 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* 2004;335:87–101. [PubMed: 14659742]
29. Sivaraman J, Sauve V, Larocque R, Stura EA, Schrag JD, Cygler M, Matte A. Structure of the 16S rRNA pseudouridine synthase RsuA bound to uracil and UMP. *Nat Struct Biol* 2002;9:353–8. [PubMed: 11953756]
30. Sunita S, Zhenxing H, Swaathi J, Cygler M, Matte A, Sivaraman J. Domain organization and crystal structure of the catalytic domain of *E.coli* RluF, a pseudouridine synthase that acts on 23S rRNA. *J Mol Biol* 2006;359:998–1009. [PubMed: 16712869]
31. Del Campo M, Kaya Y, Ofengand J. Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *Rna* 2001;7:1603–15. [PubMed: 11720289]

32. Yang Y, Thomas WR, Chong P, Loosmore SM, Klein MH. A 20-kilodalton N-terminal fragment of the D15 protein contains a protective epitope(s) against Haemophilus influenzae type a and type b. *Infect Immun* 1998;66:3349–54. [PubMed: 9632604]
33. Lee TT, Agarwalla S, Stroud RM. A unique RNA Fold in the RumA-RNA-cofactor ternary complex contributes to substrate selectivity and enzymatic function. *Cell* 2005;120:599–611. [PubMed: 15766524]
34. Cuff JA, Barton GJ. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* 2000;40:502–11. [PubMed: 10861942]
35. Cheng J, Baldi P. Three-stage prediction of protein beta-sheets by neural networks, alignments and graph algorithms. *Bioinformatics* 2005;21(Suppl 1):i75–84. [PubMed: 15961501]
36. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. *Bioinformatics* 1998;14:892–3. [PubMed: 9927721]
37. Garnier J, Gibrat JF, Robson B. GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol* 1996;266:540–53. [PubMed: 8743705]
38. Phannachet K, Elias Y, Huang RH. Dissecting the roles of a strictly conserved tyrosine in substrate recognition and catalysis by pseudouridine 55 synthase. *Biochemistry* 2005;44:15488–94. [PubMed: 16300397]
39. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys J* 2001;80:505–15. [PubMed: 11159421]
40. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997;2:173–81. [PubMed: 9218955]
41. Otwinowski Z, Minor W. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997;276:307–326.
42. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54 (Pt 5):905–921. [PubMed: 9757107]
43. Sheldrick GM, Schneider TR. High Resolution Refinement. *Methods Enzymol* 1997;277:319–343.
44. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–55. [PubMed: 15299926]
45. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1983;26:283–291.
46. Perry KM, Fauman EB, Finer-Moore JS, Montfort WR, Maley GF, Maley F, Stroud RM. Plastic adaptation toward mutations in proteins: structural comparison of thymidylate synthases. *Proteins* 1990;8:315–33. [PubMed: 2128651]
47. Nicholls A, Sharp KA, Honig B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 1991;11:281–96. [PubMed: 1758883]
48. Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* 2005;310:827–34. [PubMed: 16272117]
49. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–80. [PubMed: 7984417]
50. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 1999;15:305–8. [PubMed: 10320398]
51. Delano, WL. Delano Scientific; San Carlos, CA: 2002.

Abbreviations

Ψ

5-ribosyluracil (pseudouridine)

UMP	uridine-5'-monophosphate
Cα	alpha carbon
DTT	dithiothreitol
PEG	polyethyleneglycol

Figure 1.

Figure 1a) Aligned sequences of the catalytic domains of Ψ synthases from the RsuA family. The RluE sequences are from *Escherichia coli* (ECOLI), *Haemophilus influenzae* (HAEIN), *Xanthomonas axonopodis* (XANAC), *Salmonella typhi* (SALTI), *Yersinia pestis* (YERPR), *Vibrio cholerae* (VIBCH), *Pseudomonas aeruginosa* (PSEAE), and *Shigella flexneri* (SHIFL), and the RsuA, RluB, and RluF sequences are from *Escherichia coli*. Invariant residues are shown as white letters on a red background, and other conserved residues are shown in red letters. The sequence numbers and secondary structure of *E. coli* RluE are shown above its sequence (arrows represent β -strands, coils represent helices). The locations of five sequence motifs that characterize the Ψ synthase family are shown with colored bars under the last sequence.

b) Stereo ribbon drawing of the RluE structure. The broken line in loop L7-8 indicates the location of a disordered residue that could not be fit to density. The side chain of the catalytic Asp is shown with blue sticks. Labels on the secondary structure elements match the labels in (a).

Figure 2.

Figure 2a) 2Fo-Fc map for the active site cleft of truncated RluE with the structure overlaid (protein in orange, sulfate in yellow, waters in green)

b) A cartoon rendition of the superimposed active site clefts of *E. coli* RluE (dark grey) and *E. coli* RsuA (light grey). The five sequence motifs for the Ψ synthases are labeled and their colors correspond to the colored bars under the motifs in Fig. 1a. Conserved residues, including the catalytic Asp-79 and Tyr-109, are plotted with stick bonds. The uridine 5'-monophosphate from the RsuA structure and the sulfate ion from the RluE structure are also shown in light pink and magenta, respectively.

Figure 3.

The plot on the left shows the surface of RluE, as seen looking into the active site cleft, and colored by electrostatic potential, ranging from red (−7 kT) to blue (10 kT). The surface plot on the right shows the rRNA stem containing the RluE target uridine docked into the RNA binding trough.

Figure 4.

Figure 4a) Stereo ribbon drawing of a model of an RluE - RNA stem complex in which the RNA is docked as shown in Fig. 3b. Loops that form the walls of the RNA binding trough are colored orange. The target uridine, catalytic Asp, and a conserved Arg postulated to flip the target base into the enzyme active site are plotted in yellow, orange and red stick bonds, respectively.

b) View of the same RluE-RNA model shown in a) rotated by $\sim 90^\circ$ about a horizontal axis in the plane of the paper.

Figure 5.

Figure 5a) The two molecules of RluE comprising the asymmetric unit of the full-length RluE crystals are plotted as ribbons. Molecule A (green) packs against the RNA binding trough of molecule B (blue). The RluE-specific loops that may contribute to RNA binding are labeled and colored red.

b) Conserved structural core regions of full-length RluE, which were identified using difference distance matrices for the two molecules in the asymmetric unit of the full-length RluE crystals, are colored red and yellow on the ribbon drawing. Comparison of the full-length structures to the truncated RluE structure by the same method identifies approximately the same conserved cores.

c) Stereo ribbon drawing of molecules A and B from the full-length RluE crystals (green and blue, respectively) and truncated RluE (purple), overlapped by alignment of the C α s in their largest conserved core (red residues in b).

Figure 6.

Figure 6a) Schematic of a structural alignment of catalytic domains from five pseudouridine synthases. Red ellipses denote regions of conserved three-dimensional structure that are part of the protein core, yellow rectangles denote regions of nonconserved structure, and black lines denote gaps. Long N- and C-terminal regions on either side of the catalytic core are not plotted; in TruB, RsuA, and RluD these regions include separate RNA binding domains. Blue lines above the TruB diagram indicate location of loops in TruB that were in contact with a co-crystallized stem-loop.

b) On top is shown a stereo ribbon drawing of the proteins in Fig. 5a, which have been superimposed by alignment of the C α s of their conserved catalytic cores. Underneath the stereo plot, the five proteins are shown individually in the same orientations as in the superposed view. The catalytic cores are colored red. Nonconserved regions are colored according to protein, as follows: green – RluE, dark blue – RsuA (PDB **1KSK**), cyan – RluD (PDB **2IST**), pink – TruB (PDB **1R3E**), and yellow/pale yellow – TruA dimer (PDB **1DJ0**). The RNA stem-loop from the TruB-RNA structure of Pan et al.²⁶ is plotted as a black ribbon, with the target uridine shown as sticks.

Table 1

Crystallographic Data Collection Statistics

	Truncated RluE	RluE
Unit cell lengths (Å)	32.3,56.8,91.3	59.0,78.9,84.2
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Z	4	8
V _m (Å ³ /Da), %solvent	2.0, 37	2.0, 38
Temperature (°K)	100	100
Beam size (μM)	100	100
Wavelength (Å)	0.954	1.127
Resolution (Å)	30.5-1.2 (1.24-1.2) ¹	32.8-1.6 (1.66-1.6)
Mosaicity (°)	0.4	0.5
Wilson B-factor (Å ²)	10.3	15.5
No. of unique reflections	52188 (4929)	49168 (5102)
Redundancy	4.6(3.5)	3.1(3.3)
% Completeness	98.1 (94.0)	93.2 (98.4)
I/σ (I)	18.0 (2.1)	12.0 (2.9)
R _{merge} ² (%)	6.3 (71.3)	9.6 (77.8)

¹ numbers in parentheses refer to the final resolution shell

² R_{merge} = $\sum |I - \bar{I}| / \sum I$; negative intensities included as zero

Table 2

Crystallographic Refinement Statistics

	Truncated RluE	RluE
Resolution (Å)	30.5-1.2 (1.23-1.2) ¹	32.8-1.6 (1.66-1.6)
No. reflections	44594 (2651)	45609 (4775)
Data cutoff (σ)	0	0
Protein atoms refined	1435	2797
Statistically disordered residues	3	3
No. waters	204	262
No. ions (SO ₄ or acetate)	5	6
B-factor model	Restrained anisotropic	Restrained isotropic
No. reflns, R _{free} test set	4982	2311
R _{crys} (%) ²	14.9 (24.3)	21.5 (21.3)
R _{free} (%) ³	17.9 (28.3)	25.5 (24.1)
RMSD bond lengths (Å)	0.014	0.013
RMSD angles (°)	1.67	1.46
RMSD B _{iso} , main chain (Å ²)	2.1	1.2
RMSD B _{iso} , side chain (Å ²)	3.8	2.8
Ramachandran plot		
Most favored (%)	93.1	90.3
Favored (%)	6.9	9.7
Generously allowed (%)	0	0
Disallowed (%)	0	0
B-factor, protein (Å ²)	13.1	18.8
B-factor, waters (Å ²)	30.7	29.9

¹ numbers in parentheses refer to the final resolution shell

² $R_{crys} = \sum ||F_O| - |F_C|| / \sum |F_O|$, where F_O and F_C are the observed and calculated structure factors used in refinement

³ $R_{free} = \sum ||F_O| - |F_C|| / \sum |F_O|$, where F_O and F_C are the observed and calculated structure factors in the 5–10% of the data omitted from refinement