SUPPLEMENTARY INFORMATION

Crystal structure of the RNA component of bacterial Ribonuclease P

Alfredo Torres-Larios*, Kerren K. Swinger*, Andrey S. Krasilnikov⁎, Tao Pan†, and Alfonso Mondragón⁎‡

⁎*Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, IL 60208, USA*

†*Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA*

‡Corresponding author

Tel: (847)491-7726

FAX: (847)467-6489

email: a-mondragon@northwestern.edu

Methods.

Crystallization and data collection

The 338 nucleotide RNA component of *T. maritima* RNase P was prepared by *in vitro* transcription and purified by standard procedures. The *in vitro* produced RNA had the correct molecular weight, as verified by mass spectrometry, with no 5' or 3' terminal additions. To fold the RNA, the purified molecule was resuspended in 50 mM MES pH 6.5 and warmed up to 92 °C. It was then cooled down to 55 °C over 2.5 hours, followed by addition of $MgCl_2$ to a final 10 mM concentration and finally cooled down over a 12 hour period to 4 °C . This folding procedure was essential for crystallization as it leads to a homogeneously folded population as assayed by native gel electrophoresis. The RNA folded in this way was as active as RNA folded in more standard procedures[1] (kcat/Km: 0.28 μM^{-1} min^{-1} and 0.26 μM^{-1} min^{-1} for the standard and slow folding protocols respectively under the following reaction conditions: 50 mM Tris, pH 8.1, 25 mM $MgCl_2$, 0.5 M KCl at 37°C). Crystals were obtained using the hanging-drop vapor diffusion technique at 30°C. RNA (5 mg/ml) was mixed in a 1:1 ratio with reservoir solution consisting of 25 % (v/v) MPD, 200 mM KCl, 5 mM $SrCl_2$,, 50 mM MES pH 6.5. Large crystals were obtained by standard micro-seeding techniques. For data collection, crystals were soaked in 1mM osmium hexammine or cobalt hexammine for 4 hours and then instantly frozen in liquid nitrogen. Crystals belong to space group C2 (a=179.3Å, b=85.2Å, c=101.9Å, β=96.8°) and contain 1 molecule per asymmetric unit. Diffraction data (Table I) were collected at DND and LS CAT at the APS using a MAR CCD detector, integrated with the program XDS,[2] and scaled with SCALA.[3] Crystals were very sensitive to radiation damage, so data on only one wavelength were collected per crystal.

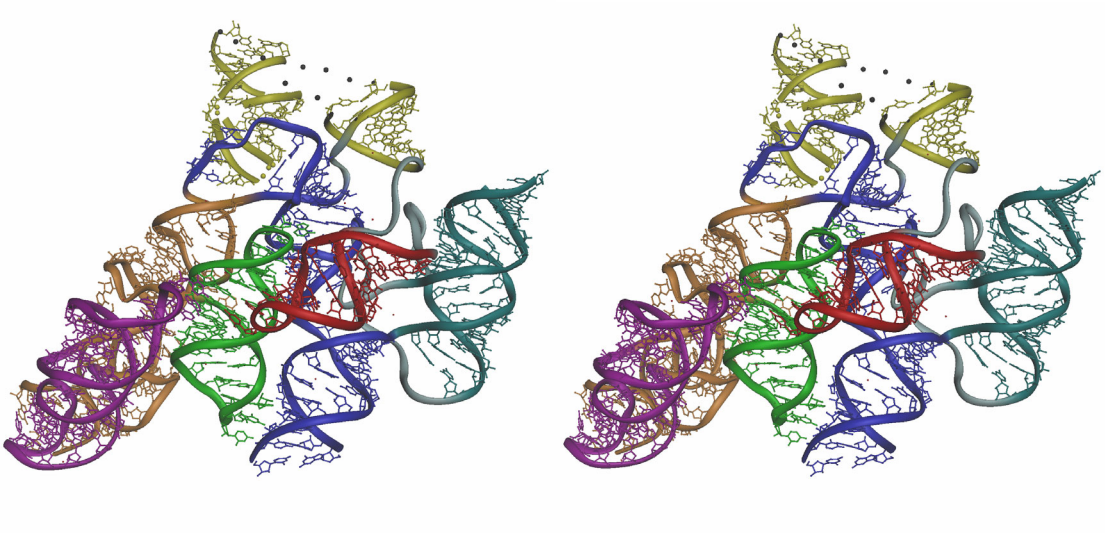
Phase determination and structural refinement

The structure was solved by using an osmium hexammine derivative, which changes the unit cell considerably. For this reason, a crystal soaked in cobalt hexammine was used as a native. In either case, the crystals were soaked in 1 mM osmium or cobalt hexammine for 4 hours before freezing. Molecular Replacement trials with the program PHASER[4] using the S-domain structure of *T. thermophilus* RNase P[5] yielded one weak solution (Z-score = 5.9). Phases calculated from this Molecular Replacement solution were used to locate three Os atom sites in an anomalous difference Fourier map. 14 additional sites used in phase determination were added during the refinement of the heavy atom model with the program SHARP.[6] For heavy atom refinement, 3 osmium hexammine data sets collected at different energies and a cobalt hexammine data set were used. Following phasing, density modification was done using the programs Solomon[7] and dm[8] as implemented in SHARP. The initial solvent-flattened electron density map allowed recognition and building of most of the stems in the C-domain and the S-domain using the program O.[9] The placing of the sequence was mostly based on the predicted secondary structure of the molecule[10] using the length of the helices as a guide, and on the structure of the *T. thermophilus* S-domain.[5]

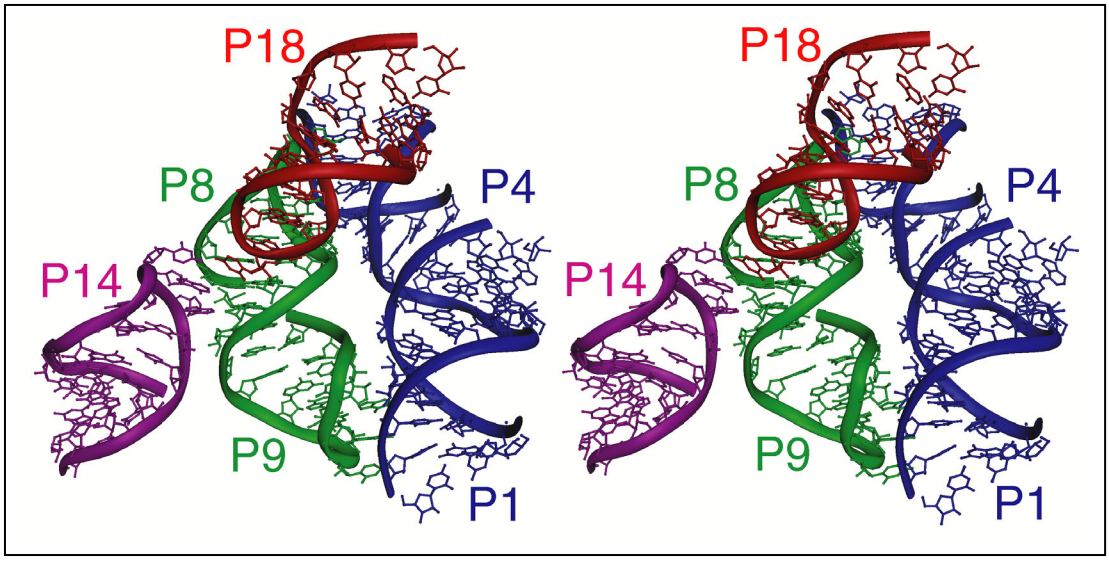
Structural refinement was done using Refmac5[11] and CNS.[12] Due to the resolution of the data, no attempts were made to place nucleotides not belonging to helical regions, except for the S-domain, where the structure was built using the *T. thermophilus* structure[5] as a guide. Of the 338 nucleotides in the molecule, 309 were built. Most nucleotides are found in secondary structure elements and could be built with ease. For nucleotides 39, 45-48, 52, 62, 68, 207, 212-214, 262-267, 288-297, and 305-314 only the phosphate backbone was visible and only phosphates were built. Density was not found

for 29 of the 338 nucleotides (30-33, 120-122, 131-135, 221-223, 229, 234, 239-241, 247-248, 253-255, and 335-338). Phasing information from the refined model was used to confirm the position of the Os sites, and 9 new weak Os sites were identified. Two temperature factors were assigned per nucleotide. The final R_{factor} and R_{free} for the structure were 35% and 36.6% respectively, which is consistent with expected values for structures at this resolution.[13] Figures were prepared with Dino (http://www.dino3d.org) and the accessible surface area was calculated with Grasp.[14]

a

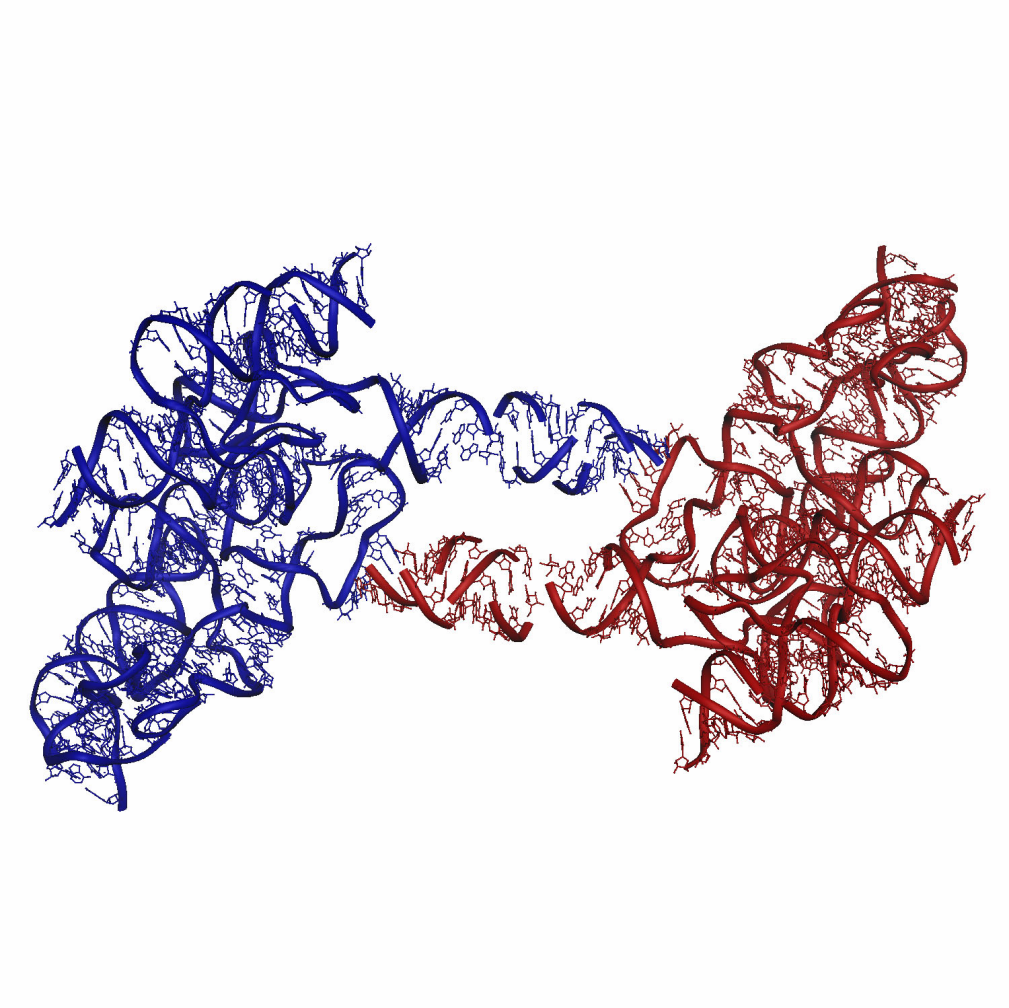


b

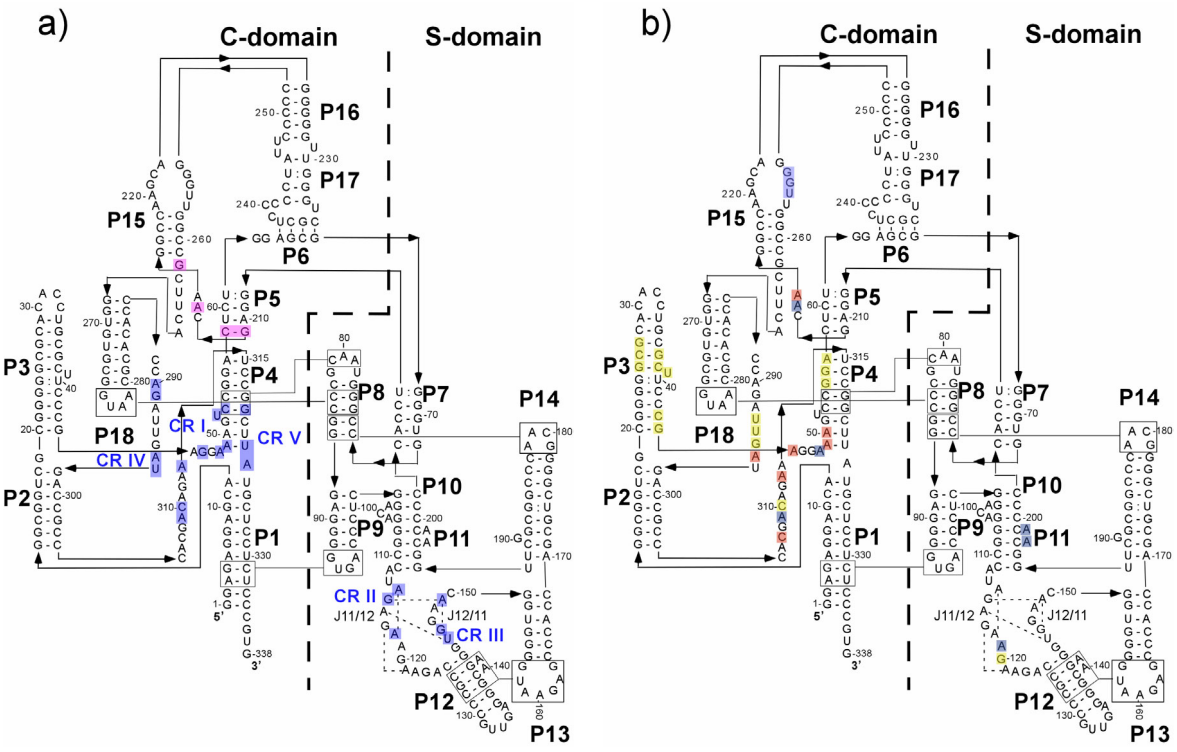


Supplementary Figure 1. Schematic diagrams of the RNA component of *T. maritima* RNase P. a) Stereo ribbon diagram of the structure. The orientation corresponds to a 180° rotation of the one in the top diagram in Figure 1A and shows a back view of the model. The secondary structure elements are coloured identically to the ones in Figure 1. b) Tertiary interactions between the specificity and the catalytic domains. The stereo diagram shows the region where stems

P1/P4, P8/P9, P14, and P18 come together. The P8 stem, in the cruciform on the S-domain, forms tertiary interactions with the loops at the end of P14 and P18 while the loops on P8 and P9 interact with P4 and P1 respectively. The two domains form extensive interactions in this region. Loop-helix interactions appear to be the dominant type of tertiary interactions in the structure.

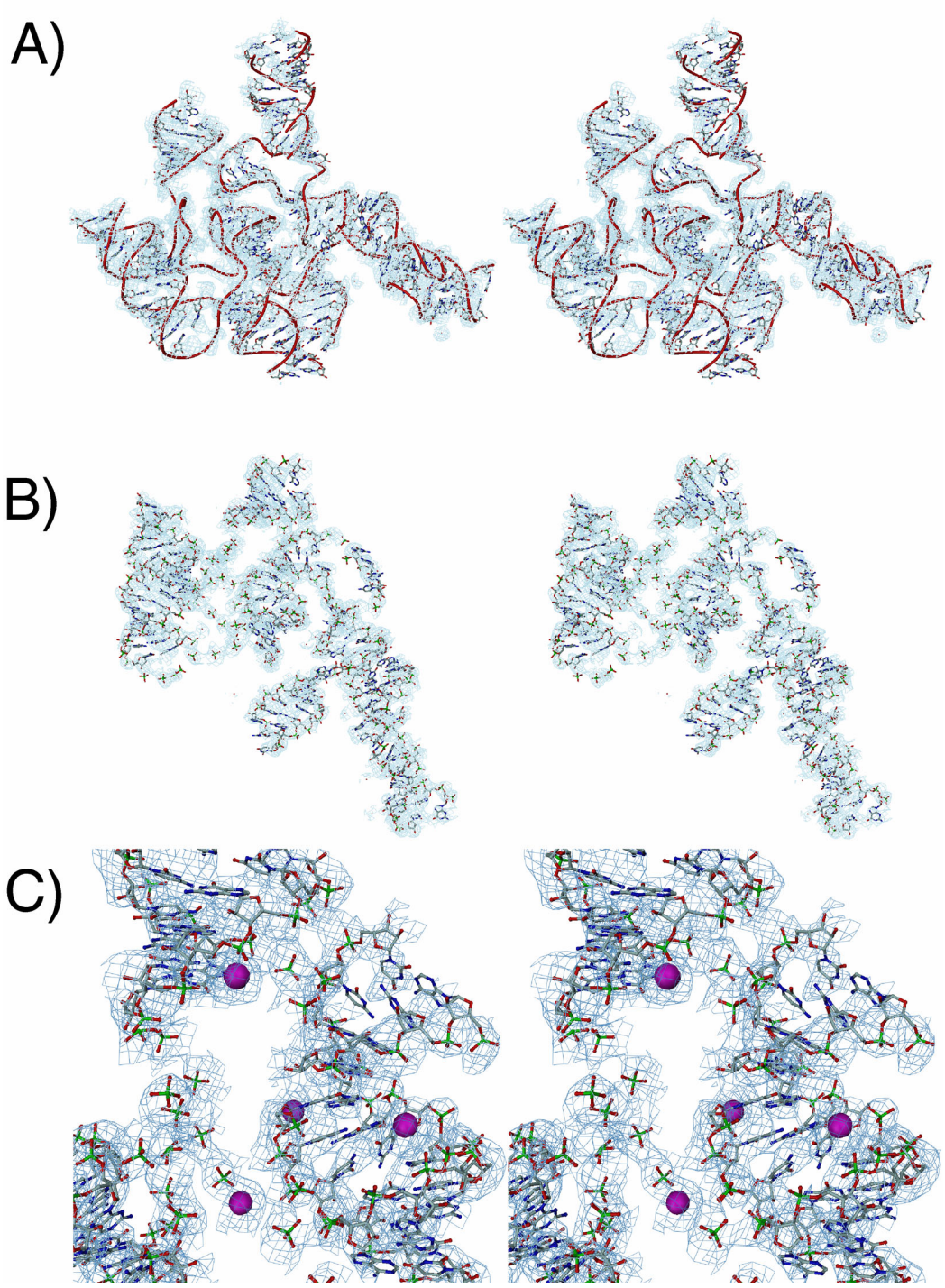


Supplementary Figure 2. Diagram of the crystallographic dimer. Two molecules in the unit cell related by a crystallographic two-fold axis are shown. One molecule is shown in red and the symmetry related one in blue. As can be seen, the P15/P16/P17 stems from one molecule extend towards the other molecule and form a pseudoknot with the symmetry related molecule. The pseudoknot helix, P6, is formed by nucleotides at the end of the P17 stem base paired with nucleotides adjacent to the P7 and P5 stems of the other molecule. It is not clear whether this intermolecular pseudoknot is preferred over the intramolecular pseudoknot *in vivo*.



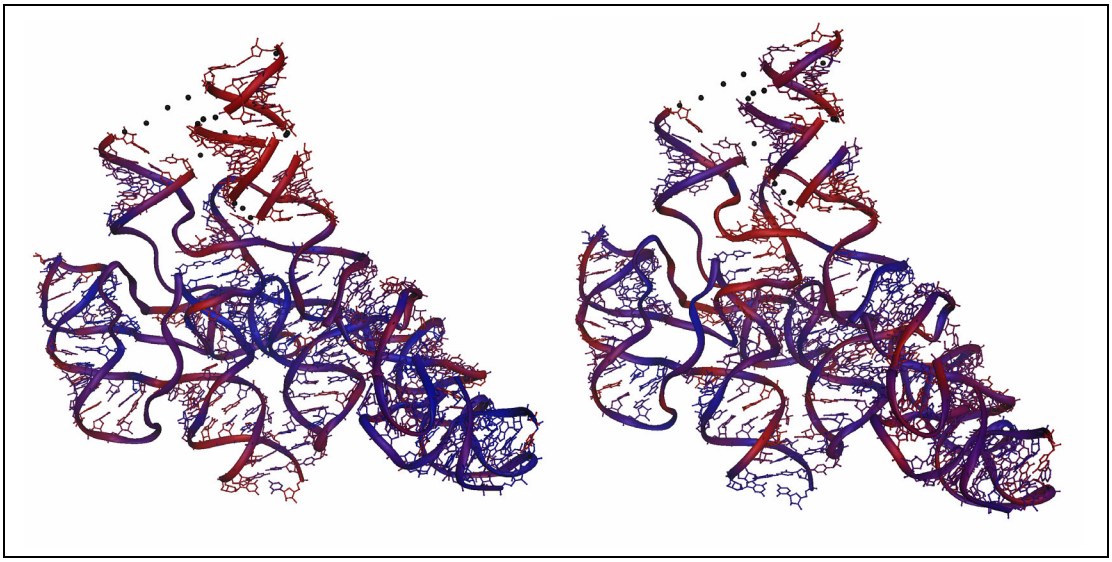
Supplementary Figure 3. Conserved and interaction regions in RNase P. a) The location of the universally conserved regions in all RNase P molecules[15] are blue and a very highly conserved region on bacterial RNase P is purple. Universally conserved regions CR II and CR III map to the S-domain and form part of an area that has a very specific structure. CR IV maps to a loop connecting P2 to P18. The other regions are located near P4 in loops connecting secondary structure elements. The role of these conserved regions appears to be structural, as they are found in areas that have a very well defined conformation. In *T. maritima* RNase P, the sequence of the universally conserved regions corresponds to: CR I ^{44}GA**G**G**AA**AG**UC**CG^{55}; CR II ^{112}AU**AG**AG**A**AG^{120}; CR III ^{142}GG**UG**GA**A**CG^{151} CR IV ^{288}CC**AG**AUUG**AU**GA^{299}; CR V ^{307}CG**AC**AGA**A**UCCG**G**CU**UA**UG^{325} (bold, underlined nucleotides correspond to those that are universally conserved). A highly conserved region in bacteria (pink), centered around A213, is located at the J5-P15 junction and has been

implicated in direct contacts with the substrate.[16] This region together with the vicinity of the universally conserved nucleotides A49 and A313 may correspond to the general location of the active site. b) Nucleotides involved in interactions with the substrate (blue) or the protein(yellow) and in catalysis (red). For a compilation of the interactions see 17,18.



Supplementary Figure 4. Experimental electron density and model of the *T. maritima* P RNA. A) The entire model is shown with the experimental, density-modified map around. B) Large slab of the experimental, density-modified map showing the fit of the model. The connecting loops were traced by phosphates as

it was not possible to orient the bases with confidence. C) Close-up of the P4, P5 and P15 helices in the catalytic domain showing the fit of the final model to the initial experimental map. The osmium atoms in the region are shown as spheres. The experimental map is shown at the 1.0 σ level. The diagram corresponds to the final, refined model.



Supplementary Figure 5. Diagrams illustrating the distribution of the temperature factors and real space R-factor. The diagram on the left is coloured by temperature factor using a blue to red ramp from 40 $Å^2$ to 150 $Å^2$. In the diagram, blue corresponds to more ordered regions and red to more disordered regions. The diagram on the right is coloured by real space R-factor using a blue to red ramp from 10% to 40%. Blue corresponds to better determined regions while red to more poorly determined regions. As expected, there is an overlap between high temperature regions and poorly determined regions.

Supplementary Table I. Data collection and refinement statistics

Data collection

Crystal	*Os-I*	*Os-III*	*Os-IIII*	*Co-Hex*
Wavelength (Å)	1.13974	1.14079	1.10500	1.13974
Resolution (Å)	20 - 4.0	20 - 4.0	20 - 4.0	20 – 3.85
Measured reflections[a]	63,740 (5,878)	64,390 (6,006)	53,138 (2,998)	37,285 (2,712)
Unique reflections[a]	12,957 (1,274)	12,933 (1,266)	12,565 (793)	13,927 (1,046)
Completeness[a] (%)	99.1 (100)	99.1 (100)	98.7 (95)	97.2 (98.9)
$\langle I \rangle / \sigma(I)$ [a]	11.9 (4.7)	17.5 (6.0)	11.7 (3.7)	14.9 (4.9)
R_{sym}[a,b] (%)	9.1 (25.7)	6.1 (21.3)	8.4 (28.6)	6.0 (23.1)
R_{meas}[a,c] (%)	16.1 (33.5)	7.7 (27.0)	11.0 (37.4)	7.4 (28.5)
MFID[a,d] (%)	-	9.1 (15.9)	10.1 (17.3)	17.0 (26.7)
Phasing				
Number of sites	17			
Phasing Power[e]				
Dispersive (centric/acentric)	-	1.094/0.855	0.544/0.477	1.735/1.366
Anomalous	1.218	0.289	0.515	
FOM				
Initial (centric/acentric)	0.45/0.39			
After density mod.	0.717			
Refinement[f]				
Resolution (Å)	15 – 3.85			
Number of reflections				
Working set	11,707			

Test set	602
R_{factor}[g] (%)	34.96
R_{free}[h] (%)	36.59
Number of atoms	5,920
Bond length r.m.s.d.	0.01 Å
Bond angles r.m.s.d.	1.87°
Average B-factor	88.9 $Å^2$
B-factor r.m.s.d.	32.4 $Å^2$
Cross-validated r.m.s. coordinate error for the final model[i]	0.52 Å
Mean real-space R-factor from a composite simulated annealing map for the final model[j]	21.7%

[a] Numbers in parenthesis correspond to the highest resolution shell.

[b] $R_{sym}=\Sigma|I-\langle I\rangle|/\Sigma I$, where I=observed intensity, and <I>=average intensity obtained from multiple measurements.

[c] R_{meas} as defined by Diederichs & Karplus.[19]

[d] MFID (mean fractional isomorphous difference) = $\Sigma||F1| - F2||/\Sigma|F1|$, where $|F1|$ = reference structure factor amplitude and $|F2|$ = compared structure factor amplitude.

[e] Phasing power = r.m.s. ($|Fh|/E$) and $|Fh|$ = heavy atom structure factor amplitude and E = residual lack of closure error, reported for all acentric reflections.

[f] Reference data set was used for all refinement calculations.

[g] $R_{factor}=\Sigma||Fo|-|Fc||/\Sigma|Fo|$, where |Fo|=observed structure factor amplitude and |Fc|=calculated structure factor amplitude.

[h] R_{free}: R_{factor} based on 5% of the data excluded from refinement.

[I] Calculated with SigmaA.[3,20]

[j] Calculated with CNS.[12]

References

1. Qin, H., Sosnick, T. R. & Pan, T. Modular construction of a tertiary RNA structure: the specificity domain of the Bacillus subtilis RNase P RNA. *Biochemistry* **40**, 11202-11210 (2001).

2. Kabsch, W. Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J. Appl. Crystallogr.* **26**, 795-800 (1993).

3. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **D50**, 760-763 (1994).

4. Read, R. J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta. Crystallogr.* **D57**, 1373-82 (2001).

5. Krasilnikov, A. S., Xiao, Y., Pan, T. & Mondragon, A. Basis for structural diversity in homologous RNAs. *Science* **306**, 104-107 (2004).

6. delaFortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472-494 (1997).

7. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr.* **D52**, 30-42 (1996).

8. Cowtan, K. in *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 34-38 (1994).

9. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110-119 (1991).

10. Massire, C., Jaeger, L. & Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* **279**, 773-793 (1998).

11. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53**, 240-255 (1997).

12. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905-921 (1998).

13. Kleywegt, G. J. Validation of protein crystal structures. *Acta Crystallogr.* **D56**, 249-265 (2000).

14. Gilson, M. K., Sharp, K. A. & Honig, B. H. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.* **9**, 327-335 (1987).

15. Chen, J.-L. & Pace, N. R. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3**, 557-560 (1997).

16. Zahler, N. H., Christian, E. L. & Harris, M. E. Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P. *RNA* **9**, 734-745 (2003).

17. Kurz, J. C. & Fierke, C. A. Ribonuclease P: a ribonucleoprotein enzyme. *Curr. Opin. Chem. Biol.* **4**, 553-558 (2000).

18. Christian, E. L., Zahler, N. H., Kaye, N. M. & Harris, M. E. Analysis of substrate recognition by the ribonucleoprotein endonuclease RNase P. *Methods* **28**, 307-322 (2002).

19. Diederichs, K. & Karplus, P. A. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat. Struct. Biol.* **4**, 269-275 (1997).

20. Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149 (1986).