

LETTERS

Crystal structure of the RNA component of bacterial ribonuclease P

Alfredo Torres-Larios¹, Kerren K. Swinger¹, Andrey S. Krailnikov^{1†}, Tao Pan² & Alfonso Mondragón¹

Transfer RNA (tRNA) is produced as a precursor molecule that needs to be processed at its 3' and 5' ends. Ribonuclease P is the sole endonuclease responsible for processing the 5' end of tRNA by cleaving the precursor and leading to tRNA maturation. It was one of the first catalytic RNA molecules identified¹ and consists of a single RNA component in all organisms and only one protein component in bacteria. It is a true multi-turnover ribozyme and one of only two ribozymes (the other being the ribosome) that are conserved in all kingdoms of life. Here we show the crystal structure at 3.85 Å resolution of the RNA component of *Thermotoga maritima* ribonuclease P. The entire RNA catalytic component is revealed, as well as the arrangement of the two structural domains. The structure shows the general architecture of the RNA molecule, the inter- and intra-domain interactions, the location of the universally conserved regions, the regions involved in pre-tRNA recognition and the location of the active site. A model with bound tRNA is in agreement with all existing data and suggests the general basis for RNA–RNA recognition by this ribozyme.

Ribonuclease P (RNase P) processes pre-tRNA, some viral RNAs, 4.5S RNA, tmRNA, C4 antisense RNA from some bacteriophages and a polycistronic pre-messenger RNA². In the case of pre-tRNA maturation, RNase P hydrolyzes the phosphodiester bond immediately 5' of the first nucleotide of mature tRNA, leaving 5' phosphate and 3' hydroxyl termini. Recognition of pre-tRNA by bacterial RNase P involves the acceptor stem, the T↓C stem³ and the CCA sequence at the 3' end⁴ of the pre-tRNA. The involvement of magnesium in the binding and cleavage of precursor tRNA has been established^{5–8}, and the regions involved in cleavage and recognition have been mapped^{2,9}. Unlike eukaryotic RNase P, in eubacteria and some archaeabacteria the RNA component alone can catalyse pre-tRNA cleavage *in vitro*^{10,11}.

In bacteria, the RNase P RNA component (or P RNA) consists of 300–450 nucleotides. Bacterial P RNAs can be divided into two major types (A or B). Eubacterial P RNAs consist of two independently folded domains: a specificity domain (S-domain), and a catalytic domain (C-domain)¹². The S-domain has been structurally characterized^{13,14}. RNase Ps from all organisms share many sequence similarities in their RNA component, implying that the P RNAs have several universally conserved structural features¹⁵.

The crystal structure of the ancestral, A-type RNA component of *T. maritima* RNase P was solved at a resolution of 3.85 Å. The electron density map allowed tracing of most of the molecule (Fig. 1; see also Methods and Supplementary Information). The molecule is approximately 125 × 90 × 50 Å³ and is made up of two layers, each one-helix thick. Figure 1a depicts a face-on view of the larger of the two layers composed of paired P1–P12 and P15–P17, and including the junctions J5/15, J11/12–J12/11 and loop L15 regions. The second layer

comprises helical stems P13, P14 and P18 (Fig. 1a, lower panel). The two structural domains are visible and interact extensively with each other through P8 and P9 in the S-domain and P1, P4 and P18 in the C-domain. The larger layer of the structure contains most of the universally conserved regions, the areas expected to contact the pre-tRNA substrate and the putative active site (see below). The S- and C-domains represent independent structural units that form extensive inter-domain interactions.

The S-domain structure is largely identical to the one of *Thermus thermophilus* RNase P (ref. 14), another A-type molecule, and underscores the structural stability of the domain. The slightly concave C-domain has several distinctive features: (1) a long stem formed by the co-axially stacked P1, P4 and P5 helices; (2) a second long stem formed by the co-axially arranged P2 and P3 helices; (3) a third stem, P18, which runs almost perpendicular to the other two helices and links P2/P3 to the P15/P17 stem, and is the only large structural element not in the same layer; (4) a stem that folds on itself (P15/P17) and whose end forms a pseudoknot with P6; and (5) a non-canonical region that links these structural elements and is located between P2/P3 and P4.

The P1/P4/P5 and P2/P3 stems are at a ~15° angle from each other, closer to the parallel orientation¹⁶ than the perpendicular orientation¹⁷ which was predicted in earlier models. P18 is linked to P15 and P2 and runs almost perpendicular to the other helices of the C-domain. No canonical tertiary interactions are seen between the helices forming the C-domain, but the four major structural elements (P1/P4/P5, P2/P3, P15/P16 and P18) are held together by linkers that include several highly conserved nucleotides that form complex interactions.

The molecule crystallized as a dimer via formation of an intermolecular version of the pseudoknot P6 (see Supplementary Information). From the structure, the functional significance of this dimer cannot be assessed, but an intramolecular P6 pseudoknot (Fig. 1) can be considered by including coordinates for P16/P17/P6 from a symmetry-related molecule; although this places P15 and P16 too far apart to interact through the short L15 internal loop. For intramolecular pseudoknot formation, there has to be a dramatic kink in the P15/P16/P17 helices around the L15 region, which has previously been suggested¹⁶. The intermolecular pseudoknot also indicates that the P15/L15/P16 region is dynamic and that the orientation of P15 and P16/P17/P6 could easily change, bringing L15 closer to the main body of the molecule. In subsequent discussions, the intramolecular pseudoknot is assumed.

The structure of the C-domain is formed mostly by helices, with several non-helical regions joining them. At the centre of the P1/P4/P5 helix there is a region where several of these non-helical regions converge. The nucleotides joining P2/P3 to P4 at the P1/P4 junction and P2 to P18 at the P4/P5 junction come close in space and

¹Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA. ²Department of Biochemistry and Molecular Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA. [†]Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

face the outside of the molecule. These nucleotides are all in the same general region, filling the space between the P1/P4/P5 and P2/P3 helices. The nucleotides linking P15 to P18 and P15 to P5 are also in the vicinity. At the resolution of the structure it is not possible to assign a position for these bases, but the phosphate backbone could be traced with confidence.

P3 is the only element in the C-domain that is far from the core of the molecule. Its role could be to link the universally conserved regions (Fig. 2). P3 can be replaced and shortened but not deleted¹⁸. P1 is essential for the proper folding of the P RNA due to its interaction with P9¹⁹. Both P4 and P18 are essential, with P18 exhibiting covariation with P8 and P14²⁰. The role of P4 has been

extensively analysed due to its sequence conservation, proximity to the catalytic site and its potential to bind the metal ions involved in catalysis. Nevertheless, the structure suggests that its role could be structural and its sequence conservation could be due to constraints imposed by its proximity and interaction with the connectors in the C-domain.

The molecule forms a concave structure with the S- and C-domains intimately interacting with each other. The central organizer of the domains is the P8/P9 stack. Tetraloop–tetraloop receptor interactions occur between P8 and L14, P8 and L18 and L9 and P1 (ref. 21). An A-minor interaction is found between L8 and P4 (see Supplementary Information). All these interactions ensure that the

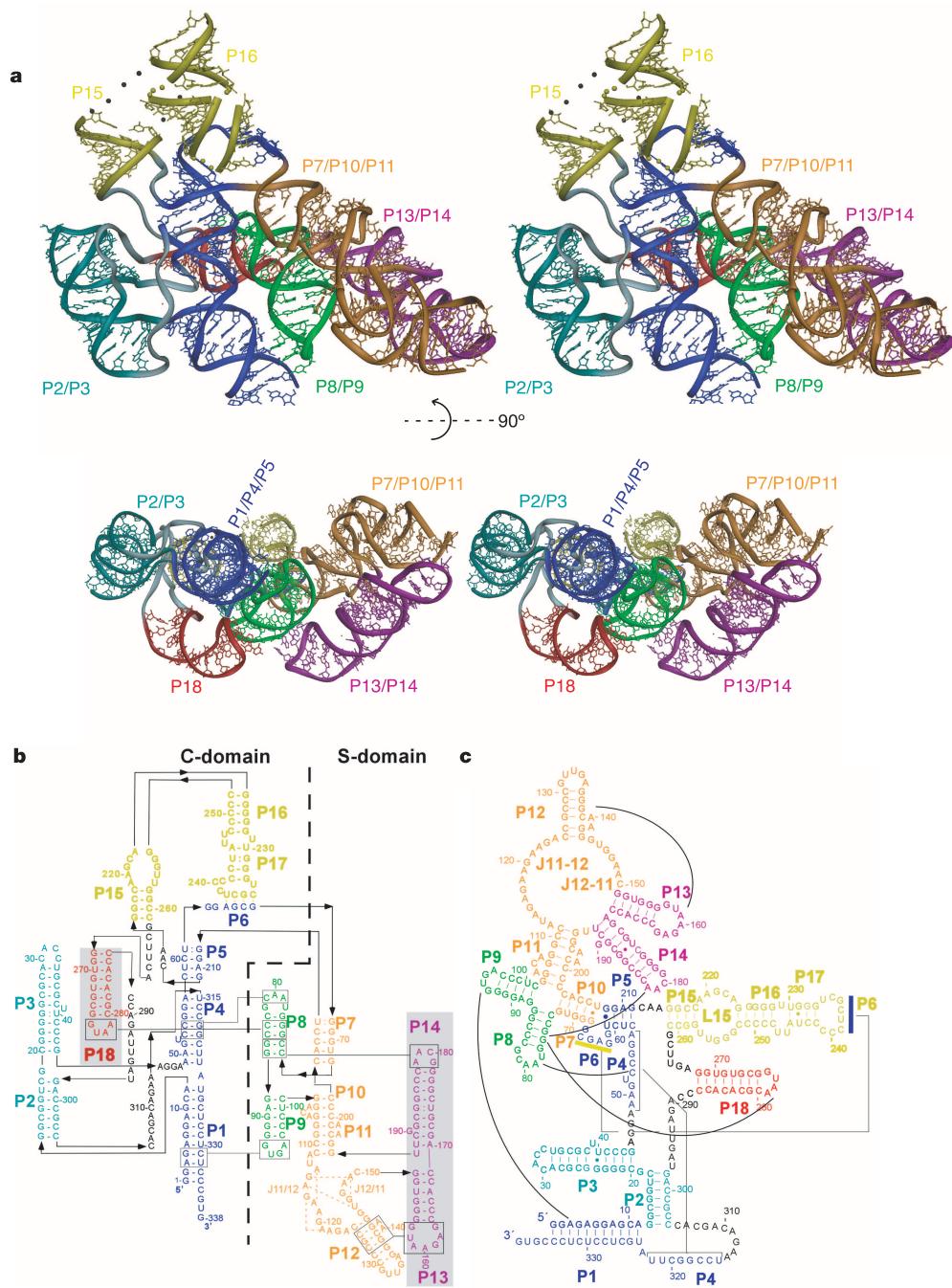


Figure 1 | Schematic diagram of *T. maritima* RNase P RNA. **a**, Stereo ribbon diagram of two orthogonal views of the structure. **b, c**, Sequence and secondary structure diagrams of the P RNA showing major interactions observed in the crystal structure. Filled circles represent non-canonical base

pairing; dashed lines correspond to other types of interactions; arrows indicate 5' to 3' direction; and lines linking boxed nucleotides represent tertiary interactions. Shaded boxes correspond to regions in the smaller of the two layers.

two domains are positioned correctly and oriented in the manner appropriate for tRNA binding. On the tertiary structure level, interactions between hairpin loops and helices dominate. Overall, all the interactions between domains predicted by a variety of approaches are observed, and the structure agrees well with earlier models and with secondary structure predictions^{16,17}.

Sequence analysis suggests that all P RNAs have a common core that includes five conserved regions (CR I–V; refs 15, 22) (Fig. 2a; see also Supplementary Information). CR II and III are located in the J11/12–J12/11 module of the S-domain. This region has an unusual fold and contains the highly conserved G147 involved in tRNA recognition^{23,24}. CR IV is located in the linker joining stems P2 to P18 (Fig. 2), in between P2/P3 and P1/P4/P5. The CR IV region neighbours other highly conserved nucleotides and it may have a structural role by helping to confer a specific organization to the active site. The two other conserved regions, CR I and CR V, are also non-helical and are clustered together and located adjacent to the P4 stem. The presence of several conserved regions of nucleotides lacking an easily identifiable secondary or tertiary structure suggests that sequence conservation may reflect structural constraints.

Substrate recognition involves contacts with nucleotides in the S-domain (in P4 and around P5), through base pairing of the CCA-3' end with nucleotides in the L15 loop, and with the protein com-

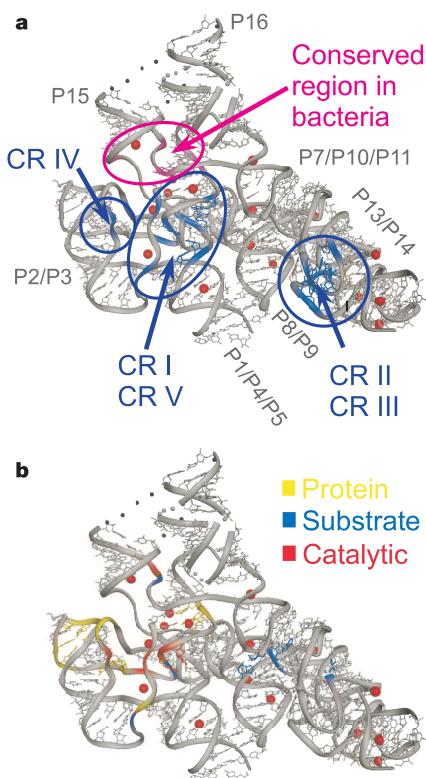


Figure 2 | Conserved and interacting regions in RNase P. **a**, Universally conserved regions¹⁵ are shown in blue and a highly conserved region in bacterial P RNAs is shown in magenta. CR II and III map to the S-domain. CR IV maps to the P2 to P18 connector. The other regions are located in nucleotides near P4. The role of these conserved regions appears to be structural, as they are found in areas that have a well-defined conformation. A highly conserved region in bacteria near A312 is located at the J5/15 junction and has been implicated in direct contacts with the substrate²⁶. This region, together with the vicinity of the universally conserved nucleotides A49 and A314, may correspond to the general location of the active site. **b**, The nucleotides that are involved in interactions with the substrate (blue) or the protein (yellow) are shown and those that are involved in catalysis are shown in red (see Supplementary Information; and refs 2, 9, 27). Red spheres are osmium atoms.

ponent. The interacting nucleotides in the S-domain have been mapped^{23,25}, and two correspond to G147 and A198 in *T. maritima*. In the case of *Bacillus subtilis* RNase P (refs 24, 25), these nucleotides interact with nucleotides 54, 56, 61 and 62 in the T ψ C stem of tRNA, and it is likely that the interactions are conserved in all P RNAs. The CCA at the 3' end of the pre-tRNA forms base pairs with nucleotides around the L15 internal loop⁴. Additionally, interactions have been mapped to the conserved regions in P4, the J5/15 linker and the J2/4 connector (summarized in refs 2, 9) (Fig. 2b; see also Supplementary Information).

To gain a better understanding of pre-tRNA recognition, tRNA was modelled into the structure (Fig. 3). In this model, the T ψ C stem loop lies in the S-domain opening and the 5' end of the tRNA lies near the P4/P5 region. The acceptor stem of the tRNA sits on the shallow groove created by the concave C-domain. The tRNA stem runs almost parallel to P1/P4/P5 and passes near the conserved nucleotides in this region. The cleavage site sits near the universally conserved residues A49 in P4 and A314 in J2/4. The cleavage site is also near J5/15, although some changes would be required to form interactions between this region and the tRNA. J5/15 contains A213, which is highly conserved in all bacteria and implicated in specific interactions with tRNA²⁶. A213 is found between the region involved in base pairing with the CCA at the 3' end⁴ and the conserved regions around P4. Conserved U52 in P4 forms an unstacked base that cannot be seen in the map. In our model the phosphate of U52 is found near the tRNA, but on the other side of the stem from where the conserved nucleotides are located. Thus it seems that A49, A213 and A314 mark the general location of the active site. Finally, in the tRNA/RNase P model, the unpaired CCA at the 3' end of tRNA is close to the L15 region. The L15 loop would have to move in order to interact with the pre-tRNA substrate. The fact that the P15/L15/P16

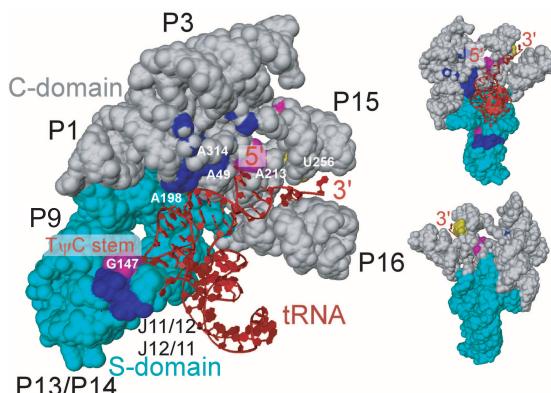


Figure 3 | Model of RNase P-tRNA interactions. The diagram shows the accessible surface area of P RNA and a diagram of tRNA^{Phe}. The model was built using the higher resolution structure of *T. thermophilus* S-domain as a guide, as the interactions between the S-domain nucleotides (G147 and A198) and the tRNA T ψ C stem are well characterized^{23,25}. In this model, the T ψ C stem loop lies in the S-domain opening and the 5' end of the tRNA lies near the P4/P5 region and putative active site. The distance between the more distal points in these two regions is around 46 Å, whereas the distance between the T ψ C stem loop and the cleavage site in mature tRNA^{Phe} is around 41 Å. The 3' CCA sequence is close to the L15 loop, where base pairing interactions occur. Colour coding is as follows: dark blue, nucleotides in the universally conserved regions; pink, highly conserved nucleotides in bacteria; and yellow, the L15 region that interacts with the pre-tRNA CCA 3' end. The views are as follows: the concave side of RNase P (left); down the anti-codon arm of tRNA (top right); and backside of the molecule (bottom right). The general active site location is marked by the 5' end of the tRNA molecule, in the vicinity of universally conserved nucleotides A49, A314 and the highly conserved bacterial nucleotide A213. The model conveys the general arrangement of the two molecules and the excellent fit between the tRNA and the two major regions of interaction with P RNA.

region seems to be quite dynamic indicates that this may be a region where conformational changes occur to accommodate the substrate. The model, although built independently, has an overall similarity to an RNA/holoenzyme model²⁷, suggesting that the holoenzyme model captures the general location of the protein correctly.

RNA–RNA recognition involves the presence of a structurally well defined and conserved specificity site that is responsible for selecting tRNA molecules through interactions with the TΨC stem loop. The catalytic active site appears to be fully formed even in the absence of the pre-tRNA substrate. Other RNA–RNA interactions appear to be non-specific and may involve metal ions. They concentrate along the acceptor stem of tRNA and the P4 stem of RNase P. Thus, RNA–RNA recognition involves regions that make specific contacts with the substrate, a wide range of non-specific interactions between the two molecules, and a fully or almost fully assembled active site. A fuller understanding of tRNA recognition and RNase P catalysis awaits further structural work at atomic resolution.

METHODS

The crystal structure of the 338-nucleotide-long *T. maritima* RNase P molecule was solved by SIRAS using osmium hexammine and cobalt hexammine soaked crystals. The structure was phased with SHARP²⁸ and refined with Refmac5 (ref. 29) and CNS³⁰. Two temperature factors were assigned per nucleotide. Due to the resolution of the data, no attempts were made to place nucleotides not belonging to helical regions, except in the S-domain, where the structure was built using the *T. thermophilus* structure as a guide. Of the 338 nucleotides in the molecule, 309 were built. For nucleotides 39, 45–48, 52, 62, 68, 207, 212–214, 262–267, 288–297 and 305–314 only the phosphate backbone was built. Density was not found for nucleotides 30–33, 120–122, 131–135, 221–223, 229, 234, 239–241, 247–248, 253–255 and 335–338. The final *R*_{factor} and *R*_{free} for the structure were 35% and 36.6% respectively. Figures were prepared with Dino (<http://www.dino3d.org>). For more complete details see the Supplementary Information.

Received 15 June; accepted 28 July 2005.

Published online 21 August 2005.

1. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
2. 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).
3. McClain, W. H., Guerrier-Takada, C. & Altman, S. Model substrates for an RNA enzyme. *Science* **238**, 527–530 (1987).
4. Kirsebom, L. A. & Svard, S. G. Base pairing between *Escherichia coli* RNase P RNA and its substrate. *EMBO J.* **13**, 4870–4876 (1994).
5. Christian, E. L., Kaye, N. M. & Harris, M. E. Evidence for a polynuclear metal ion binding site in the catalytic domain of ribonuclease P RNA. *EMBO J.* **21**, 2253–2262 (2002).
6. Beebe, J. A., Kurz, J. C. & Fierke, C. A. Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNAAsp. *Biochemistry* **35**, 10493–10505 (1996).
7. Perreault, J. P. & Altman, S. Pathway of activation by magnesium ions of substrates for the catalytic subunit of RNase P from *Escherichia coli*. *J. Mol. Biol.* **230**, 750–756 (1993).
8. Smith, D. & Pace, N. R. Multiple magnesium ions in the ribonuclease P reaction mechanism. *Biochemistry* **32**, 5273–5281 (1993).
9. Kurz, J. C. & Fierke, C. A. Ribonuclease P: a ribonucleoprotein enzyme. *Curr. Opin. Chem. Biol.* **4**, 553–558 (2000).
10. Pannucci, J. A., Haas, E. S., Hall, T. A., Harris, J. K. & Brown, J. W. RNase P RNAs from some Archaea are catalytically active. *Proc. Natl Acad. Sci. USA* **96**, 7803–7808 (1999).
11. Reich, C., Olsen, G. J., Pace, B. & Pace, N. R. Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme. *Science* **239**, 178–181 (1988).

12. Loria, A. & Pan, T. Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA* **2**, 551–563 (1996).
13. Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760–764 (2003).
14. Krasilnikov, A. S., Xiao, Y., Pan, T. & Mondragón, A. Basis for structural diversity in homologous RNAs. *Science* **306**, 104–107 (2004).
15. Chen, J.-L. & Pace, N. R. Identification of the universally conserved core of ribonuclease P RNA. *RNA* **3**, 557–560 (1997).
16. 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).
17. Harris, M. E. et al. Use of photoaffinity crosslinking and molecular modeling to analyse the global architecture of ribonuclease P RNA. *EMBO J.* **13**, 3953–3963 (1994).
18. Guerrier-Takada, C. & Altman, S. Reconstitution of enzymatic activity from fragments of M1 RNA. *Proc. Natl Acad. Sci. USA* **89**, 1266–1270 (1992).
19. Harris, M. E., Kazantsev, A. V., Chen, J. L. & Pace, N. R. Analysis of the tertiary structure of the ribonuclease P ribozyme–substrate complex by site-specific photoaffinity crosslinking. *RNA* **3**, 561–576 (1997).
20. Brown, J. W. et al. Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial populations reveals tertiary structural elements. *Proc. Natl Acad. Sci. USA* **93**, 3001–3006 (1996).
21. Massire, C., Jaeger, L. & Westhof, E. Phylogenetic evidence for a new tertiary interaction in bacterial RNase P RNAs. *RNA* **3**, 553–556 (1997).
22. Siegel, R. W., Banta, A. B., Haas, E. S., Brown, J. W. & Pace, N. R. Mycoplasma fermentans simplifies our view of the catalytic core of ribonuclease P RNA. *RNA* **2**, 452–462 (1996).
23. LaGrande, T. E., Huttenhofer, A., Noller, H. F. & Pace, N. R. Phylogenetic comparative chemical footprint analysis of the interaction between ribonuclease P RNA and tRNA. *EMBO J.* **13**, 3945–3952 (1994).
24. Odell, L., Huang, V., Jakacka, M. & Pan, T. Interaction of structural modules in substrate binding by the ribozyme from *Bacillus subtilis* RNase P. *Nucleic Acids Res.* **26**, 3717–3723 (1998).
25. Loria, A. & Pan, T. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **36**, 6317–6325 (1997).
26. 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).
27. Tsai, H. Y., Masquida, B., Biswas, R., Westhof, E. & Gopalan, V. Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme. *J. Mol. Biol.* **325**, 661–675 (2003).
28. delaForte, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
29. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
30. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Yusupov, B. Golden and J. Cate for their gift of osmium and iridium hexamine, and Y. Zhang for mass spectrometry. We thank A.-C. Dock-Bregeon, E. Sontheimer and O. Uhlenbeck for comments and suggestions, P. Nissen for advice and A. Vega-Miranda for help with the figures. Research was supported by an NIH grant to A.M. Support from the R.H. Lurie Cancer Center of Northwestern University to the Structural Biology Center is acknowledged. Portions of this work were performed at the DND, LS, IMCA and SER Collaborative Access Teams at the Advanced Photon Source. We thank members of these teams for their help and support. DND-CAT is supported by DuPont, Dow, and the NSF, and use of the Advanced Photon Source is supported by the DOE.

Author Information Coordinates and structure factors have been deposited in the Protein Data Bank under the accession code 2A2E. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.M. (a-mondragon@northwestern.edu).