[12] Solving Large RNA Structures by X-Ray Crystallography

By JAMIE H. CATE and JENNIFER A. DOUDNA

Introduction

Structured RNAs play key roles in many aspects of biology, from RNA processing to protein synthesis and transport. Advances in the large-scale synthesis and purification of RNA have fostered structural studies by X-ray crystallography. In this article, we discuss those aspects of solving large RNA crystal structures that differ from the approaches developed for protein crystallography. Common strategies in crystallography are found in Volumes 276 and 277 of this series.

Preparation of Well-Diffracting Crystals

The first hurdle to solving an RNA structure by X-ray crystallography is obtaining crystals that diffract X rays to biochemically useful resolution. The electron-dense phosphate backbone of RNA is readily detectable at 5-Å resolution,[1] but to identify individual bases and registration of the sequence, 3-Å resolution or better is generally necessary. As for proteins, high-quality RNA crystals form by utilizing exposed tertiary contacts. These contacts are sometimes inherent in the RNA, as for the tRNA anticodon,[2,3] the three major crystal contacts in the *Tetrahymena* ribozyme P4–P6 domain crystals,[4] the GAAA tetraloop in the hammerhead ribozyme structures,[5,6] and loop E in the 5S rRNA fragment structure.[7] Recent work on engineering RNA–RNA crystal contacts showed that RNAs with GAAA tetraloop–tetraloop receptor motifs included for intermolecular contacts yielded many more crystal forms than controls.[8] However, the most successful design

[1] B. L. Golden, A. R. Gooding, E. R. Podell, and T. R. Cech, *Science* **282,** 259 (1989).
[2] F. L. Suddath, G. J. Quigley, A. McPherson, D. Sneden, J. J. Kim, S. H. Kim, and A. Rich, *Nature* **248,** 20 (1974).
[3] J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. Clark, and A. Klug, *Nature* **250,** 546 (1974).
[4] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna, *Science* **273,** 1678 (1996).
[5] H. W. Pley, K. M. Flaherty, and D. B. McKay, *Nature* (*Lond.*) **372,** 68 (1994).
[6] W. G. Scott, J. T. Finch, and A. Klug, *Cell* **81,** 991 (1995).
[7] C. C. Correll, B. Freeborn, P. B. Moore, and T. A. Steitz, *Cell* **91,** 705 (1997).
[8] A. R. Ferre-D'Amare, K. Zhou, and J. A. Doudna, *J. Mol. Biol.* **279,** 621 (1998).

0076-6879/00 $30.00

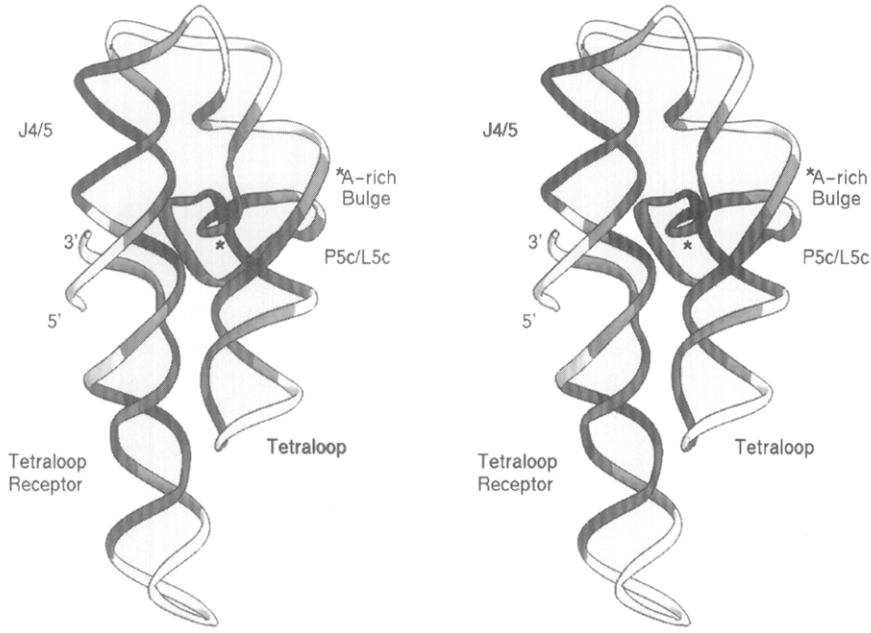

FIG. 1. Temperature factor (B factor) distribution in the P4–P4 domain crystal structure. A stereo view of one of the two molecules in the asymmetric unit of the crystal is shown, shaded according to B factor ranges as follows: black, $B \leq 24$ Å^2; dark gray, $24 < B \leq 36$ Å^2; light gray, $36 < B \leq 48$ Å^2; white, $B > 48$ Å^2. The other molecule in the asymmetric unit has similar temperature factor distributions, due to noncrystallographic symmetry (NCS) restraints. Note that the darker regions correspond to intramolecular tertiary structure and/or crystal packing contacts. The B factors can change dramatically within a few residues, for example, in the tetraloop strand and above J4/5.

incorporated a binding site for the well-characterized RNA binding domain of the U1A protein[9,10] into the hepatitis delta virus (HDV) ribozyme, thus providing stable protein–protein contacts in the crystal lattice.[11]

One problem with RNA is that regions of the molecule not involved in crystal contacts tend to have very high temperature factors (B factors), indicating motion or disorder (Fig. 1). This may explain in part why most RNA crystals do not diffract X rays well. If an RNA molecule of interest fails to crystallize or yields poorly ordered crystals, the introduction of crystallization modules into solvent-exposed regions of the RNA may be helpful.[8] The key is to insert the modules so that they do not disrupt or change the structure of the RNA. Phylogenetic covariation analysis, when

[9] K. Nagai, C. Oubridge, T. H. Jessen, J. Li, and P. R. Evans, *Nature* **348,** 515 (1990).
[10] C. Oubridge, N. Ito, P. R. Evans, C. H. Teo, and K. Nagai, *Nature* **372,** 432 (1994).
[11] A. R. Ferre-D'Amare, K. Zhou, and J. A. Doudna, *Nature* (*Lond.*) **395,** 567 (1998).

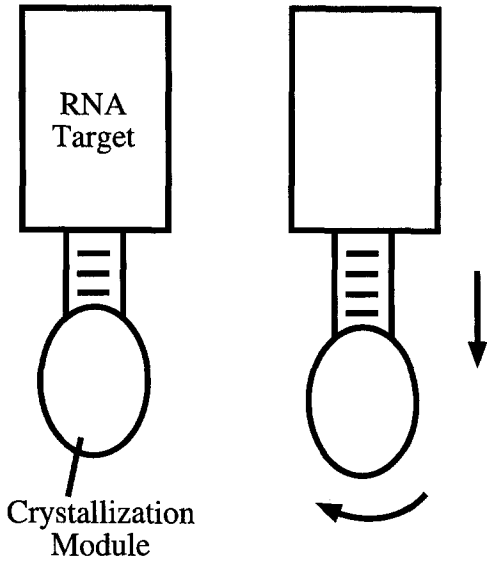

FIG. 2. Design of crystal packing interactions in RNA. The position of the crystallization module can be varied by inserting different length helical segments between the RNA of interest and the module. Each additional base pair increases the distance between the two as well as rotating their relative orientations by ~30°.

it is available, often reveals variable regions of the RNA where insertions, deletions, and sequence changes are allowed while maintaining function. Chemical probes of RNA structure, including Fe(II)-EDTA[12,13] and base modification reagents such as DMS, kethoxal, and diethyl pyrocarbonate (DEPC)[14] are also useful for identifying exposed regions of the RNA that are not critical for function. These sites can then be altered to include, for example, a binding site for the U1A–RNA binding domain or recognition motifs for tetraloops. It is important to design multiple constructs to vary the length of the stem connecting the inserted element to the core of the RNA, which will alter their relative orientations (Fig. 2). It is also essential to test that the biochemical activity of the designed molecules is not significantly different from that of the wild-type RNA.

Once suitable crystals are in hand, two primary factors determine how rapidly the structure can be solved. First, highly reproducible conditions for growing the crystals must be found. It may take hundreds of crystals

[12] D. W. Celander and T. R. Cech, *Biochemistry* **29,** 1355 (1990).
[13] D. W. Celander and T. R. Cech, *Science* **251,** 401 (1991).
[14] S. Stern, D. Moazed, and H. F. Noller, *Methods Enzymol.* **164,** 481 (1988).

to find appropriate stabilizing conditions and the heavy atom derivatives needed to solve the structure. Several approaches may increase the yield of usable crystals. The length and conformational purity of the sample should be verified by denaturing and native gel electrophoresis, and by dynamic light scattering.[15] Although 2- to 4-μl crystallization drop volumes are useful for initially screening conditions, drop volumes as large as reasonably possible should be used to obtain large crystals for heavy atom derivatization. The kinetics of equilibration are retarded in sitting drop volumes of 10–30 μl, resulting in more reproducible crystal nucleation and growth.[16] Furthermore, such drops are convenient for macro- and microseeding.[17]

RNA crystals tend to decay rapidly when exposed to X rays. Thus, diffraction data are generally collected from RNA crystals at cryogenic temperatures. Compounds used in protein crystal stabilization are useful for RNA crystals as well.[18] These include sugars and nonvolatile alcohols such as glucose, glycerol, 2-methyl-2,4-pentanediol (MPD) and low molecular weight polyethylene glycols (PEGs). In addition, volatile alcohols such as ethanol and isopropanol have been useful in RNA crystal cryostabilization. Typically the cryoprotectant should constitute at least 15–20% (v/v) of the soaking solution for the crystals. Addition of the cryoprotectant should be in 1–2% increments by adding solution to the drop containing the crystal and resealing the coverslip to prevent evaporation. If the crystal appears unchanged after 2 min, another addition of the protectant can be made. Once the final concentration of cryoprotectant in the drop has been adjusted to high enough concentrations, the reservoir solution within the crystallization setup should be adjusted to the same composition, and the system resealed to equilibrate.

RNA crystals are highly sensitive to mono- and divalent salts and polyamines, and these additives must be adjusted in concert to find optimal stabilizing conditions. Often, as the concentration of divalent metal ions is decreased, the concentrations of polyamines such as spermine and spermidine need to be increased. Appropriate concentrations to use are highly empirical: the P4–P6 domain crystals cracked at concentrations of spermine higher than ~0.5 m*M*, whereas the HDV ribozyme crystals were optimal in 6–25 m*M* spermine. When searching for cryostabilizing conditions, one should consider the length of time required for later manipulations, i.e.,

[15] A. R. Ferre-d'Amare and J. A. Doudna, *Methods Mol. Biol.* **74,** 371 (1997).
[16] J. R. Luft and G. T. DeTitta, *Methods Enzymol.* **276,** 110 (1997).
[17] A. McPherson, *Methods Enzymol.* **114,** 112 (1985).
[18] D. W. Rodgers, *Methods Enzymol.* **276,** 183 (1997).

heavy atom soaks. The best conditions should not degrade crystal quality over time.

Note that RNA crystals are often highly susceptible to nonisomorphism, or crystal-to-crystal variation. In the case of the P4–P6 domain crystals, for example, one unit cell dimension fluctuated in length from 128 to 132 Å depending on the stabilizing conditions. This hinders all stages of phasing the structure factors because it changes the periodic sampling of the electron density in reciprocal space. In the case of P4–P6, the change in cell dimensions correlated with the percentage of isopropanol included in the cryostabilizer, up to approximately 5% 2-propanol. By increasing 2-propanol to 10%, we eliminated this variability.

Phasing by Heavy Atom Substitution

In protein crystallography, many of the methods for solving the phase problem are highly tuned. For proteins and protein–DNA complexes, brominated DNA and selenomethionyl-substituted proteins are used routinely for phasing the structure factors.[19] These methods apply to RNA as well: brominated RNAs were used to solve the structures of the hammerhead and loop E of 5S rRNA,[6,20,21] and selenomethionine was used to determine the structure of the HDV ribozyme–U1A complex.[11] We recommend using the technique of multiwavelength anomalous diffraction (MAD) for phase determination with RNA because it avoids the problems of nonisomorphism that are so often encountered (see later discussion). In this regard, iodinated RNAs are less useful because the absorption edge for iodine is at an inaccessible wavelength.[22] Larger RNAs may require a heftier signal than can be obtained from bromines or selenomethionine. For example, a molecule the size of P4–P6 (160 nucleotides) would require five or more bromine atoms to determine the structure by MAD.[22]

Two kinds of transition metal ions have been used with great success in RNA crystallography. First, lanthanide series cations contributed to solving the structures of tRNA, the hammerhead ribozyme, and the group I intron P4–P6 domain (see Ref. 23 for summary of conditions). Second, osmium(III) hexammine provided the high-resolution phasing for the P4–P6 domain structure.[24] Both types of ions are likely to bind to RNA at sites normally occupied by magnesium ions, though not exclusively to

[19] S. Doublie, *Methods Enzymol.* **276,** 523 (1997).
[20] W. G. Scott, J. T. Finch, R. Grenfell, J. Fogg, T. Smith, M. J. Gait, and A. Klug, *J. Mol. Biol.* **250,** 327 (1995).
[21] C. C. Correll, B. Freeborn, P. B. Moore, and T. A. Steitz, *J. Biomol. Struct. Dyn.* **15,** 165 (1997).
[22] W. A. Hendrickson, and C. M. Ogata, *Methods Enzymol.* **276,** 494 (1997).
[23] A. R. Ferre-D'Amare and J. A. Doudna, *Curr. Protocols Nucleic Acid Chem.*, in press (1999).
[24] J. H. Cate, and J. A. Doudna, *Structure* **4,** 1221 (1996).

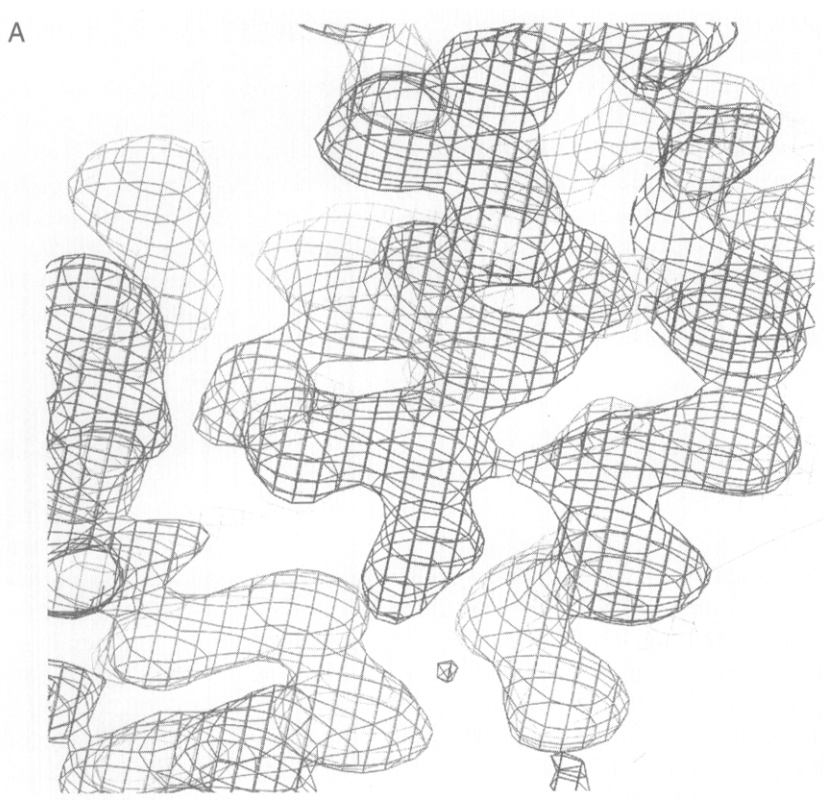

FIG. 3. Map quality of MAD versus MIR experiments. The A-rich bulge region of the P4–P6 domain electron density map is shown contoured at 1 standard deviation above the mean. Three data sets were used in the map calculations: Oshex λ1 and Oshex λ2, MAD data sets collected at the Os LIII absorption edge; Cohex 1, "native" data set.[4] (A) MAD treatment of the data. Heavy atom parameters were refined and phases were calculated with Oshex λ1 as the reference data set. (B) MIR treatment of the data. Heavy atom parameters were refined and phases were calculated with Cohex 1 as the reference data set.

these sites. One key difference between the lanthanides and osmium hexammine is the mode of binding to the RNA. Whereas the lanthanides can form inner-sphere coordination to the RNA, osmium hexammine can only form "outer-sphere" interactions, similar to fully hydrated magnesium ions. The best concentration ranges to use for these ions in the published structures varies from 100 μM to low millimolar. It is worthwhile screening all available lanthanides (all but Pm are commercially available) due to the change in ionic radius from Ce^{3+} to Lu^{3+}. In the case of the P4–P6 domain crystals, only Sm^{3+} provided a usable heavy atom derivative.[4]

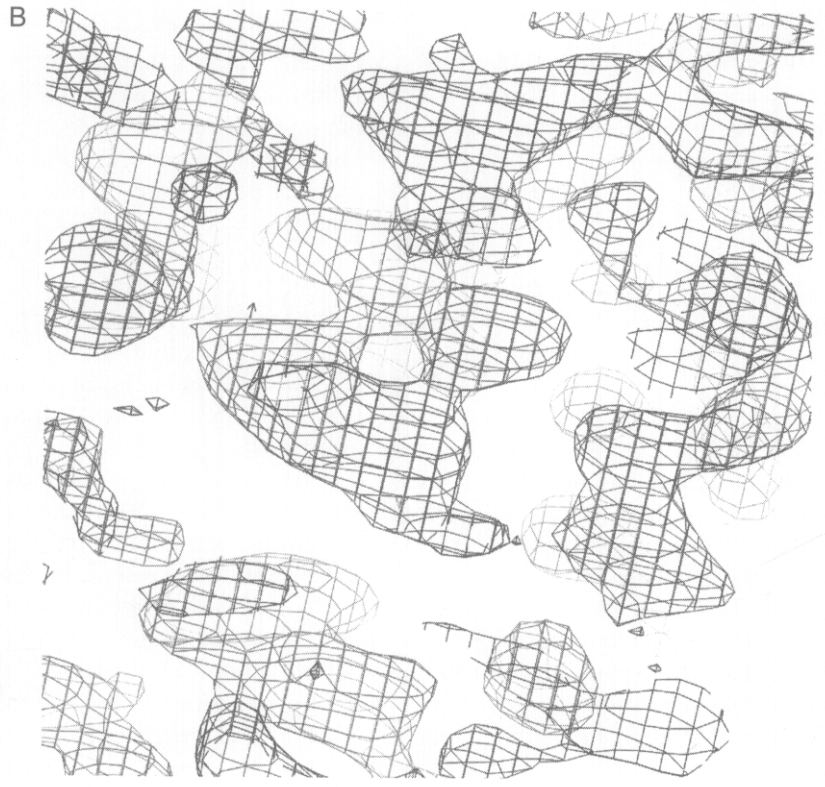

FIG. 3. (*continued*)

Osmium hexammine and the lanthanides, with a strong anomalous signal at the LIII absorption edge, have been exploited at synchrotron radiation sources to solve the structures of many proteins and the P4–P6 domain RNA by MAD phasing. Even slight nonisomorphism can interfere with multiple isomorphous replacement (MIR) phasing, while the same data sets treated as a MAD experiment yield a clean experimental electron density map. As an example, the same diffraction data sets used to produce the interpretable electron density map for the P4–P6 domain structure, when treated as MIR data, yielded a map with broken density (Fig. 3). Strategies for conducting a MAD experiment and determining phases from the heavy atom substitution are published elsewhere.[22,25]

[25] T. C. Terwilliger, *Methods Enzymol.* **276,** 530 (1997).

Map Interpretation and Model Building

RNA electron density maps have two strong features that can be used to place an initial model. First, at a resolution of approximately 3.5 Å or better, the phosphates appear as spheres in an electron density map contoured at a high signal-to-noise ratio. Second, purine bases form elongated disks at high signal-to-noise ratios. The hardest part of the map to fit is usually the ribose. As the weakest constraints are to the 5′ side of the ribose moiety, it is better to use the O-3′–P bond length to find the ribose pucker that best matches the density. Torsion angles can be used to position the ribose, with the base and surrounding phosphates held fixed, to try to match up the O-3′–P bond length. There is some flexibility around the C-5′–O-5′–P angle to position O5′.[26] The density around the phosphate may not always be centered on the phosphorous atom; it may "streak" along the P–O-5′–C-5′ direction toward the ribose.

After first positioning the phosphates and bases as purine–pyrimidine, a useful strategy is to use a ribose library to connect the phosphates with reasonable sugar puckers (Fig. 4).[24] After an initial round of manual placement or a refinement cycle, the library of ribose conformers is superimposed on C-1′, O-4′, and N-1 or N-9 of the attached base of each ribose in question to find the pucker that best fits the density. It may take some adjusting of the χ dihedral torsion angle (around N-1 or N-9 and C-1′ bond), with the base held fixed, to "position" the library by least-squares superposition in a graphical modeling program[27] such as O. At ~3.0 E resolution, C3′-endo and C2′-endo puckers are distinguishable for all but the most disordered residues. In the P4–P6 domain maps, the density for C2′-endo riboses was usually perpendicular to the plane of the base density (Fig. 4).

Refinement of the RNA model presents one major difference from that of protein structures. Because the furanose ring can take on either C3′-endo or C2′-endo puckering, one must use a more complicated restraint for the dihedral angles within and external to the ring. For example, CNS (or X-PLOR) provides a multiwell potential function that allows the ribose to go into either dominant pucker (X-PLOR manual). The depth of the wells around the C3′-endo and C2′-endo puckering may need some adjustment, depending on the resolution of the structure. In addition, we have found it necessary to choose the pucker of each ribose based on the experimental map prior to refinement, i.e., one may not be able to "flip" the pucker simply by least-squares refinement or simulated annealing procedures. If the pucker is chosen incorrectly, positive and negative density will appear

[26] W. Saenger, "Principles of Nucleic Acid Structure." Springer-Verlag, New York, 1984.
[27] T. A. Jones, J. Y. Zou, S. W. Cowan, and Kjeldgaard, *Acta Crystallogr. A*, **47**, 110 (1991).

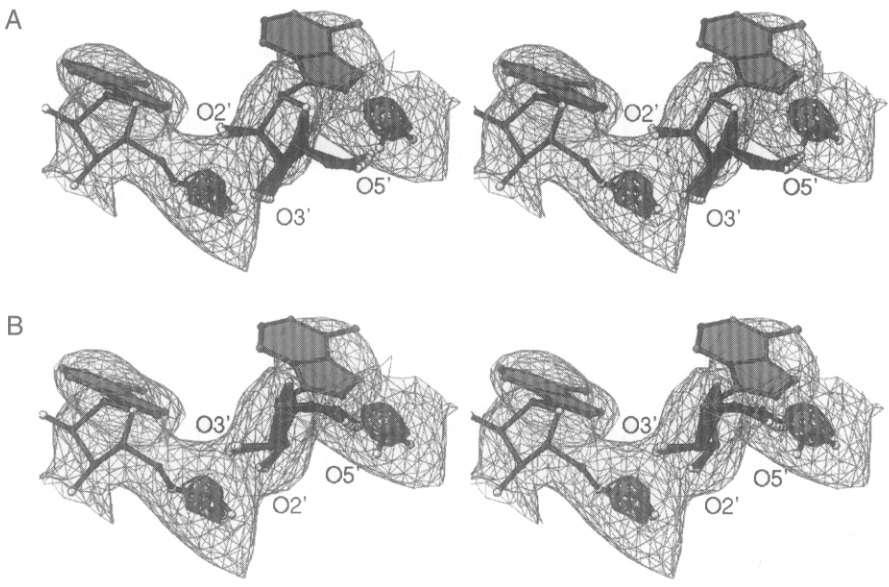

FIG. 4. Ribose rotamer library. Ribose moieties with ring puckers of 0–36 degrees (C3′-endo) or 145–181 degrees (C2′-endo), in 9-degree increments, are positioned to achieve a reasonable O3′–P bond distance. (A) C3′-Endo library positioned for nucleotide A206 in P4–P6 domain molecule A. Note that no torsional rotation of the ribose about the glycosidic bond could fit O2′, O3′, and C5′/O5′ in the density while maintaining the O3′–P distance. (b) C2′-endo library positioned to give a reasonable O3′-P bond length. All atoms of the ribose can be placed in the density.

above and below the ribose "plane" near the O4′ and C2′/O2′ positions in $(F_o - F_c)\varphi$ difference maps, where F_o and F_c are observed and calculated amplitudes, and φ are either experimental or calculated phases, depending on the stage of refinement. An additional clue comes from phosphate-to-phosphate distances. Phosphates are often approximately 5 Å apart when separated by a C3′-endo puckered ribose, whereas phosphates are often about 7 Å apart when separated by a C2′-endo puckered ribose.[26]

Interpretation of Solvent Electron Density Peaks

Metal ions play key roles in the folding and function of structured RNAs.[28] One of the best ways to identify these metal sites comes from RNA crystal structures. However, at the resolution of most of the known RNA structures (~3 Å), positively identifying solvent peaks in the electron

[28] A. M. Pyle, *Science,* **261,** 709 (1993).

density map proves difficult. Metal ions in the experimental electron density map of the P4–P6 domain crystal structure appeared at close to the same density level as the phosphates, initially complicating model building. To identify magnesium ions and other ions among the solvent electron density peaks may take a two-pronged approach, coupling crystallography and RNA biochemistry.

From the experimental electron density or refined model, indirect data can be obtained to identify metal ions. Although water molecules, magnesium ions, and sodium ions all have similar scattering properties, the coordination of the solvent peak to the RNA may provide clues.[29] Magnesium ions have octahedral coordination geometry and bond lengths of approximately 2.1 Å. Sodium ions have bond lengths of 2.4 Å, but may have variable geometry.[30,31] Water molecules can be identified by their involvement in hydrogen-bond length interactions with the RNA (2.7–3.0 Å). The differences in these geometries should be compared to the coordinate error for the RNA model, which may be on the order of 0.3–0.4 Å for a 3-Å resolution structure.[32] In practical terms, due to the close coordination of water molecules to magnesium and sodium, these hydrated metal ions may appear as larger blobs of electron density than that observed for single waters. Fully hydrated ions may appear as electron density peaks centered approximately 4 Å from the nearest RNA ligand.

Potassium ions, while having variable coordination geometry and bond lengths similar to those of water (2.7 Å),[30] scatter electrons more strongly than water or the other ions. This scattering difference may appear in the refined model as follows: if the solvent peak is originally modeled as water, its refined temperature factor may end up much lower than that of the surrounding RNA. In the case of the P4–P6 group I intron domain, putative waters in the major groove of AA platforms refined with temperature factors of 5–10 $Å^2$, whereas the temperature factors of the surrounding RNA hovered around 30–40 $Å^2$. However, at modest (~3-Å) resolution, unrestrained temperature factor refinement may be suspect and lead to model bias.

Crystallography provides a direct approach for identifying ions in a structure. Transition metal ions that can functionally substitute for lighter ions can be exploited for their anomalous scattering properties. In many cases, manganese ions partially or completely substitute for magnesium

[29] S. R. Holbrook, J. L. Sussman, R. W. Warrant, and S. H. Kim, *J. Mol. Biol.* **123,** 631 (1978).
[30] S. R. Holbrook, J. L. Sussman, R. W. Warrant, G. M. Church, and S. H. Kim, *Nucleic Acids Res.* **4,** 2811 (1977).
[31] J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church, and S. H. Kim, *J. Mol. Biol.* **123,** 607 (1978).
[32] V. Luzzati, *Acta Crystallogr.* **5,** 802 (1952).

ions in RNA folding and activity.[33] Ion channel studies have shown that thallium ions (Tl^+) substitute for potassium ions.[34] Note that thallium is toxic and should be handled with gloves in a fume hood. Both Mn^{2+} and Tl^+ exhibit strong anomalous scattering that can be observed using a rotating anode X-ray generator and Cu-K_α radiation. To measure the anomalous data required to identify metal ion sites, the best approach is to use "inverse beam" geometry. From a single crystal, measure two complete data sets from starting rotations separated by 180°. In this way one is guaranteed of observing both Friedel mates for each unique reflection, while at the same time reducing systematic errors, i.e., in absorption. If crystal decay is significant, the data sets may have to be collected in 10° wedges. For example, the data might be collected from 0 to 10°, followed by 180 to 190°, and then back to 10 to 20°, and so on. We find it useful to exploit the Strategy option in the program MOSFLM to identify the rotations necessary to collect a complete data set, then repeat the suggested range ±180°.[35] See the *Methods in Enzymology* chapter by Hendrickson and Ogata for a more detailed discussion of anomalous data collection in the context of MAD experiments.[22]

Substitution of metal ions in RNA crystals requires some consideration of the competing effects of different salts. Manganese ions, for example, tend to bind soft ligands with higher affinity than Mg^{2+} ions do, i.e., to the N-7 position of guanosine. Thallium ions also bind soft ligands well. To overcome artifactual binding of these transition metals and to enhance their specificity, it may be necessary to drastically change crystal stabilization conditions. We found that by increasing the ratio of polyamine (spermine) to divalent cation, we could decrease the concentration of Mg^{2+} 10-fold while still maintaining reasonable diffraction from the P4–P6 domain crystals. With a much lower requirement for divalent cation, the crystals could then be transferred directly into Mn^{2+} with the complete removal of Mg^{2+}. Thallium tends to be insoluble as the chloride salt, thus requiring a change in stabilizer to one containing only acetate counterions, for example.

Conclusions

Although few RNA crystal structures have been determined, some general trends for solving RNA structures by crystallography have emerged. There are now proven methods for obtaining well-diffracting crystals of large RNAs. MAD phasing has become the most efficient approach for

[33] C. A. Grosshans and T. R. Cech, *Biochemistry* **28,** 6888 (1989).
[34] B. Hille, *J. Gen. Physiol.* **61,** 669 (1973).
[35] *Acta Crystallogr. D,* **50,** 760 (1994).

phasing structure factors and obtaining high-quality electron density maps. Crystallographic experiments coupled with biochemical tests have allowed the functional interpretation of RNA structures. As occurred in the protein field, the number of RNA crystal structures solved per year is likely to grow exponentially.

Acknowledgment

We thank Adrian Ferré-D'Amaré for helpful discussions about modeling ribose puckers in electron density maps. J.H.C. is a postdoctoral fellow of the Damon Runyon-Walter Winchell Cancer Research Fund; J.A.D. is an assistant investigator of the Howard Hughes Medical Institute, a Searle Scholar, a Beckman Young Investigator, and a Fellow of the David and Lucile Packard Foundation; this work was supported in part by a grant from the NIH.

[13] Conventional and Time-Resolved Ribozyme X-Ray Crystallography

By WILLIAM G. SCOTT and JAMES B. MURRAY

Introduction

Can X-ray crystallography help us to understand how ribozymes work? Although a number of biochemical techniques, perhaps most notably *in vitro* selection methods, have provided much needed insight into the nature and potentially broad spectrum of the catalytic capabilities of RNA, the problem of deducing the correlations between RNA three-dimensional structure and enzymatic activity remains a formidable one. Recent advances in crystallization methodologies and in RNA synthetic strategies have enabled us to begin to address this problem. In particular, the hammerhead ribozyme, hepatitis delta virus ribozyme, group I intron, hairpin ribozyme, RNase P, and other catalytic RNAs have been crystallized, and crystal structures of the first three have now been published. In addition, application of time-resolved crystallographic techniques to the hammerhead ribozyme allows us to begin to correlate the structure of a ribozyme directly with the chemistry of its catalysis. Here we discuss the methods used for both conventional and time-resolved crystallographic analyses of the hammerhead ribozyme. This ribozyme serves as a prototype for other analyses, because the techniques employed should be fairly generalizable for future work on other ribozyme systems.

0076-6879/00 $30.00