The 1.3 Å Crystal Structure of a Biotin-binding Pseudoknot and the Basis for RNA Molecular Recognition

Jay Nix, Django Sussman and Charles Wilson*

Department of Biology and Center for the Molecular Biology of RNA, Sinsheimer Laboratories, University of California at Santa Cruz, Santa Cruz, CA 95064, USA

A pseudoknot-containing aptamer isolated from a pool of random sequence molecules has been shown previously to represent an optimal RNA solution to the problem of binding biotin. The affinity of this RNA molecule is nonetheless orders of magnitude weaker than that of its highly evolved protein analogs, avidin and streptavidin. To understand the structural basis for biotin binding and to compare directly strategies for ligand recognition available to proteins and RNA molecules, we have determined the 1.3 Å crystal structure of the aptamer complexed with its ligand. Biotin is bound at the interface between the pseudoknot's stacked helices in a pocket defined almost entirely by base-paired nucleotides. In comparison to the protein avidin, the aptamer packs more tightly around the biotin headgroup and makes fewer contacts with its fatty acid tail. Whereas biotin is deeply buried within the hydrophobic core in the avidin complex, the aptamer relies on a combination of hydrated magnesium ions and immobilized water molecules to surround its ligand. In addition to demonstrating fundamentally different approaches to molecular recognition by proteins and RNA, the structure provides general insight into the mechanisms by which RNA function is mediated by divalent metals.

© 2000 Academic Press

Keywords: in vitro selection; aptamer; magnesium; avidin; multiple wavelength anomalous dispersion

Introduction

In vitro selection from pools of random sequence RNA molecules has yielded molecules exhibiting a range of ligand-binding and catalytic functions. Their isolation has bolstered origin of life theories that postulate a pivotal role for RNA (Benner et al., 1989). However, with the exception of certain elements of the splicing and translational machinery (Kruger et al., 1982; Noller et al., 1992), biologically extant RNA molecules do not actually perform these functions, which are carried out instead by proteins. To explore the structural basis for RNA function and to understand RNA’s intrinsic limitations as a functional macromolecule, we have determined the three-dimensional structure of an RNA pseudoknot evolved in vitro to bind the carboxylation cofactor, biotin. Biotin (Figure 1(a)) lacks both planar aromatic rings and positively charged functional groups, moieties that have served as the major determinants for recognition in most other characterized aptamer-ligand complexes (Jiang et al., 1996, 1997; Zimmerman et al., 1997). The proteins avidin and streptavidin have been similarly optimized through in vivo evolution to recognize biotin and bind it with phenomenally high affinity ($K_D \sim 10^{-15}$ M) (Green, 1975). Comparison of the structures of RNA-biotin and the protein-biotin complexes makes it possible to compare directly the strategies used by each type of biopolymer to create a binding site and to carry out molecular recognition.

All biotin-binding aptamers isolated from a pool of $5 \times 10^{14}$ RNA molecules contain a pseudoknot that conforms to the primary and secondary structure consensus shown in Figure 1(a) (Wilson et al., 1998). Analysis of the pseudoknot has shown that it specifically recognizes multiple functional groups on the biotin headgroup and that it binds with relatively high affinity compared with other apta-
mer-ligand interactions ($K_D$ approximately $6 \mu M$) (Wilson et al., 1998). Subsequent efforts to isolate tighter binders by de novo selection or by mutagenesis have been unsuccessful, suggesting that this motif represents the highest-affinity biotin aptamer accessible within the sequence space of 28 nt RNA molecules. Both specific base-pairs within the pseudoknot helices, as well as the length and sequence of intervening loops, are preserved among independently derived molecules. The RNA used in these studies perfectly matches this globally optimal biotin-binding motif.

Here, we describe the high-resolution structure of the biotin aptamer complexed with its ligand as determined by X-ray crystallography. Analysis of the structure shows how a ligand binding site can be constructed largely from base-paired nucleotides and how solvent molecules, including both water molecules and metals, can be positioned to facilitate RNA folding and function.

Results and Discussion

Previously reported conditions for crystallization of the biotin aptamer yielded tetragonal crystals with weak, anisotropic diffraction (Nix et al., 1999). Crystallization trials based on the Berger nucleic acid miniscreen (Berger et al., 1996) ultimately identified a new set of conditions which provided small but strongly diffracting monoclinic crystals. Attempts to stabilize these crystals prior to data collection showed that biotin was absolutely required in the stabilizing solution, suggesting that biotin was bound to the crystallized RNA and that the cofactor was in equilibrium with biotin in solution. Crystallographic phases were determined from a single-crystal multiple wavelength anomalous dispersion (MAD) experiment using an aptamer-biotin complex crystal which had been extensively equilibrated with selenobiotin, a biotin analog in which the thiophene sulfur atom is substituted by selenium. Crystallographic phases were determined from a single-crystal multiple wavelength anomalous dispersion (MAD) experiment using an aptamer-biotin complex crystal which had been extensively equilibrated with selenobiotin, a biotin analog in which the thiophene sulfur atom is substituted by selenium. A single strong site for the bound selenium was identified in the anomalous Patterson (with a peak height of 21 $\sigma$). MAD phases determined using the selenium differences and improved by solvent flattening yielded an easily interpretable map from which the initial model was built (Table 1). This model was refined against single-wavelength data from another crystal and the final 1.3 $\AA$ structure, containing 29 nt, six bound magnesium ions, and 168 water molecules, has been refined to an $R$-factor of 19.9% and an $R_{free}$ value of 24.7% (Table 1).

The overall structure of the aptamer loosely resembles that observed in previous NMR and crystallographic analyses of synthetic pseudoknots and viral pseudoknots that promote ribosomal frame-shifting (Puglisi et al., 1990; Shen et al., 1992; Kolk et al., 1998). The two helices defining the pseudoknot are directly stacked upon each other with a short connecting loop 1 lying in the major groove of helix 2, and a longer loop 2 lying in the minor groove of helix 1. Biotin is bound at the interface between the two helices and also contacts the last nucleotide of loop 2. The aptamer folds as anticipated with two notable exceptions. A9 and U33, predicted to form the last base-pair in helix 2, are clearly unpaired with A9 stacking on C8 to...
Table 1. X-ray data collection, MAD phasing, and structure refinement

| Data set          | $\lambda$ (Å) | $d_{\text{min}}$ (Å) | Observed reflections | Unique reflections | % Complete | $\langle I \rangle / \langle \sigma \rangle$ | $R_{\text{sym}}^a$ (%) | Centric (PP)$^b$ | Acentric (PP)$^b$ |
|-------------------|----------------|-----------------------|----------------------|-------------------|------------|----------------------------------------|----------------|----------------|----------------|----------------|
| Native (crystal 1) | 1.54           | 2.0                   | 32,759               | 5945              | 88.8       | 20.71                                  | 4.2            | Iso           | Iso           | Anom          |
| MAD Experiment    |                |                       |                      |                   |            |                                        |                |                |                |                |
| crystal 2         | 0.97954        | 1.5                   | 83,182               | 13,920            | 88.4       | 13.67                                  | 3.7            | 2.57          | 2.10          | 2.50          |
| Inflection point  | 0.97969        | 1.5                   | 83,441               | 13,863            | 89.5       | 15.92                                  | 3.6            | 2.87          | 2.34          | 2.23          |
| Low-energy        | 0.98734        | 1.5                   | 82,615               | 13,925            | 87.9       | 14.08                                  | 3.5            | 0             | 0             | 0.75          |
| remote            | 0.97186        | 1.5                   | 77,926               | 13,978            | 88.6       | 14.07                                  | 3.6            | 0.38          | 0.26          | 2.03          |
| High-energy       |                |                       |                      |                   |            |                                        |                |                |                |                |
| remote            | 0.9650         | 1.3                   | 130,824              | 20,852            | 91.7       | 9.00                                   | 7.8            | -             | -             | -             |
| Crystal 3         |                |                       |                      |                   |            |                                        |                |                |                |                |

Completeness %

- Native and MAD data: 93.4 (combined)
- Crystal 3: 91.7

Figure of Merit to 1.5 Å

- MAD: 0.57
- Solvent flattened: 0.82

$R$-factor$^c$

- $R_{\text{cryst}}$ (%): 21.4
- $R_{\text{free}}$ (%): 25.4

RMSD from ideality$^d$

- Bond lengths (Å): 0.020
- Bond angles (deg.): 2.18

---

$^a$ $R_{\text{sym}} = S_j S_i |I_i(h)| - \langle |I_i(h)| \rangle / S_j |I_i(h)|$, where $I_i(h)$ is the $i$th measurement and $\langle |I_i(h)| \rangle$ is the weighted mean of all measurements of $|I_i(h)|$.

$^b$ Phasing power is the mean value of the heavy atom structure factor amplitude divided by the residual lack of isomorphous (iso) and anomalous (anom) closure error for centric and acentric reflections.

$^c$ $R$-factor = $S[F_o - F_c]/S[F_o]$, where $F_o$ and $F_c$ are the observed and calculated structure factors, respectively. $R_{\text{free}}$ is the cross-validated $R$-factor calculated for 5% of the reflections omitted in the refinement process.

$^d$ RMSD is the root-mean-square deviation.
form loop 1 and U33 lacking a unique conformation defined by the experimental electron density. The linkage between nucleotides U21 and A22 is clearly broken in the crystallized RNA, allowing formation of an intermolecular base triple involving U21 and the A3:U20 base-pair of a symmetry-related molecule. This interaction is one of only two direct RNA-RNA contacts within the crystal lattice, and thus we presume that cleavage at the hydrolysis-prone U-A dinucleotide precedes crystallization.

Weak density corresponding to the connecting phosphate group is distributed near both the 3' and 2'-hydroxyl groups of U21, although it is clearly not connected simultaneously to both as a cyclic 2',3'-phosphate group.

The overall structure of helix 1 and its minor groove-associated loop 2 is strikingly similar to that of the equivalent domain from the beet western yellow virus (BWYV) frameshifting pseudoknot (backbone atoms can be superimposed with an rms deviation of 1.3 Å). Helix 2 in the aptamer pivots by approximately 20° about the helical junction (Figure 1(c)) but coaxial stacking is maintained. As shown schematically in Figure 1(c), biotin is wedged between nucleotide A26 of loop 2 and G27 in helix 2 to effectively force this altered helical orientation and to provide continuous stacking interactions across the helical junction. In contrast to the BWYV pseudoknot and the pseudoknots characterized by NMR, base-pairing at the helical junction is maintained with only a slight tilt (15°) between the principal axes of the two helices.

As outlined below, most conserved nucleotides within the consensus structure act to promote folding of the pseudoknot, either through direct RNA-RNA contacts or by positioning magnesium ions and water molecules. With few exceptions, absolutely conserved nucleotides do not form direct sequence-specific contacts with the biotin. The preferred nucleotide requirements for the aptamer can be explained in several instances by sequence-specific tertiary interactions between nucleotide bases and the RNA phosphate backbone. For example, the absolutely conserved C4:G18 central base-pair of helix 1 forms a direct hydrogen bond via the guanosine exocyclic amine group to the phosphate backbone of A24. This interaction is one of several RNA-RNA contacts that help to position the adenosine-rich loop and thus fix both the overall pseudoknot structure and the biotin binding pocket.

The loop connecting the two strands of helix 1 in all biotin aptamers contains eight nucleotides (Wilson et al., 1998), a geometrically ideal length for the stacking of two A-form RNA helices (Du & Hoffman, 1997). The first nucleotide in the loop is invariably an unpaired cytidine residue (C8). As shown in Figure 2(b), this cytidine residue bridges the major groove of helix 2 forming a sequence-specific hydrogen bond from its N4 to the C29 pro-Rp phosphate oxygen atom on the opposing helical strand. This cytidine interaction closely resembles that observed previously in NMR structures of the ultra-stable UUCG tetraloop (Cheong et al., 1990; Allain & Varani, 1995). Remarkably, the entire nucleotide is flipped upside down relative to its immediate 5' and 3'-neighbors. This flip brings the Watson-Crick face of the nucleotide back into contact with the rest of the pseudoknot and facilitates formation of an extensive hydrogen bond network involving helix 2 nucleotides and bound water molecules. The absolutely conserved G13:C29 base-pair adopts a somewhat distorted geometry (with an 18° propeller twist and a 10° buckle) which allows the N4 of C29 to form a bifurcated hydrogen bond with both its Watson-Crick partner and the O2 of C8. The O6 of G13 forms a water bridge to position the 2'-hydroxyl group of C8. Direct coordination by the 5' and 3'-phosphate groups of C8 to magnesium ions, and base stacking with an adjacent adenosine residue (A9) combine to hold the nucleotide in place and thereby stabilize the arrangement of the two pseudoknot helices.

Loop 2 contains a string of four absolutely conserved adenosine bases which, with the exception of A23, form a continuous stack running the length of the helix 1 minor groove (Figure 2(d)). A23 is flipped out of the stack, and by bridging across the face of the helix, helps pin the loop in position. As shown in Figure 2(d), extensive hydrogen bonding by both the ribose and the base stabilize the nucleotide in this conformation, which chemical probing suggests is adopted prior to biotin binding (Wilson et al., 1998). Additional stabilization is conferred by hydrophobic stacking interactions of the base with the ribose of C5. A virtually identical arrangement is also observed for an equivalent adenosine base (A21) in the BWYV pseudoknot (Su et al., 1999).

With the exception of A23, the adenosine bases in the A-rich loop interact predominantly with the phosphate backbone of helix 1 rather than with its nucleotide bases. Interactions between the major groove face of these adenosine bases and the 3'-strand of helix 1 guide the loop down the minor groove. The N6 and N7 of each adenosine base in the loop are positioned to hydrogen bond with the Watson-Crick face of the nucleotide back into contact with the various isolated biotin aptamer clones. The stacking energy for adenosine dinucleotides is stronger than that for any other nucleotide pair. It is possible that the ability to form a stable, continuous stack rather than the ability to interact specifi-
Magnesium ions stabilize the pseudoknot

Binding by six hydrated magnesium ions helps to stabilize the pseudoknot (Figure 2(a)). All of these metal atoms have clearly defined octahedral coordination geometry comprised by up to a single RNA oxygen ligand and four to six water molecules. Four of the six magnesium ions visible in the electron density (Mg1-Mg4) form a stripe down the deep major groove of the stacked pseudoknot helices. The remaining magnesium ions (Mg5, Mg6) lie on the opposite face of the pseudoknot and participate directly in biotin binding.

Magnesium ions are found along the entire length of helix 1 (Figure 2(a)). In every case, these metal ions interact predominantly via their directly coordinated water molecules with pairs of stacked base-pairs. Magnesium-coordinated water molecules allow for Mg1 to interact with the O4 atoms of U20 and U21 and for Mg2 to interact with the N7 and O6 atoms of G18 and G19. An additional magnesium atom (Mg3) lies at the junction between the two helices in the major groove. Surprisingly, this metal ion is directly coordinated to the O4 atom of U7. It is unclear as to which features of this site favor a direct magnesium-carbonyl interaction by Mg3, while magnesium ions at adjacent sites (e.g. Mg1, Mg2) choose to interact indirectly via bound water molecules. It is obvious that the direct coordination is augmented by a number of both inner and outer-shell water molecules which combine effectively to coordinate Mg3 to both strands of helix 1.

The remaining magnesium ions are involved in the stabilization of loop 1 or the formation of the binding pocket. Mg4 is the final magnesium ion lining the major groove of the aptamer and is directly coordinated to the pro-Rp oxygen atom of A9 and to five water molecules. Mg5 and Mg6 are important in the stabilization of the bases and water molecules involved in the recognition and binding of biotin. Mg5 is directly bound to the pro-Sp oxygen atom of C8 of loop 1 and contains five inner-sphere water molecules. Interactions via water bridges help position the phosphate backbone of A26 and G27 on either side of the binding pocket. Mg6 is partially buried in the binding pocket attached to the pro-Rp oxygen atom of A25 at one side and directly to the carbonyl group of biotin at the other. In addition to the role of Mg6 in ligand recognition, numerous interactions with the backbone of loop 2 are observed. There is direct outer-sphere coordination to the phosphate...
backbone of A24-A26 along with the N7 of A26 to position this important base of the binding pocket.

By comparing the environments surrounding the magnesium ions and associated water molecules, we can identify structural motifs that may define metal binding sites in other RNA molecules. Both Mg2 and Mg4 interact via bound water molecules with guanosine dinucleotides in stacked Watson-Crick base-pairs (G12-G13 and G18-G19). Previous characterization of the heavy-atom binding sites in the Tetrahymena pre-rRNA group I intron identified stacked G-U wobble pairs as preferred metal sites, a result later confirmed by NMR analysis of an intron-derived hairpin (Kieft & Tinoco, 1997). The limited resolution of these previous studies, however, made it impossible to identify the specific interactions that stabilize metal binding at these sites. As shown in Figure 2(c), Mg2 and Mg4 occupy virtually identical positions relative to the RNA major groove and are held in place by related sets of water-mediated hydrogen bonds, including three interactions that are preserved in both sites. In both cases, one water ligand hydrogen bonds to the N7 of the 5’-guanosine base, a perpendicularly orientated water ligand hydrogen-bonds to the N7 of the 3’-guanosine base, and a third water molecule, orthogonal to the first two, hydrogen-bonds to a water molecule that solvates the phosphate linkage between the guanosine pair. In addition to these interactions, the O6 atom of each guanosine base is involved in hydrogen bonding, although the specific partners vary at the two different metal sites. A water molecule from the Mg2 coordination shell simultaneously hydrogen-bonds to both the G18 and G19 major groove carbonyl groups. At the Mg4 site, however, an additional water molecule is recruited by the magnesium hydration shell to hydrogen bond to these groups. This outer-sphere water molecule is placed in a position superimposable with that of an inner-sphere water molecule of Mg2. The specific arrangement of hydrogen bonding groups presented by the stacked guanosine bases thus provides a defined but adaptable binding site for hydrated magnesium ions. It is interesting to note that while neither the G12-G13 nor G18-G19 stacks are absolutely conserved among the biotin aptamer clones, all of the aptamers contain G-G stacks in both helices, suggesting that the magnesium sites are preserved in every helix, although their relative position varies. A virtually identical arrangement of water-mediated hydrogen bonds to a guanosine stack has been previously observed in the structure of a 5 S rRNA fragment (Correll et al., 1997).

It is striking that with only one exception (Mg4), single water molecules do not bridge directly between magnesium ions and phosphate oxygen atoms. Instead, we observe that magnesium ions are either directly coordinated to the phosphate backbone (with no bridging water molecules) or, alternatively, both magnesium ions and phosphate oxygen atoms retain separate hydration shells and thus interact with each other across two bridging water molecules. We can speculate that the cost of partially desolvating a magnesium ion or phosphate group can be re-couped only if the water molecule is replaced with a direct electrostatic interaction.

Specific recognition of biotin

The biotin binding pocket is formed at the junction of the two pseudoknot helices and surrounds the ligand on all sides except for its N1 edge (Figure 3(a)). If one considers bound metal and water molecules to be part of the active site, the biotin headgroup is inserted into a very tight, well-defined binding pocket. The puckered thiophene ring is sandwiched between the last nucleotide in the adenosine loop (A26) and the first nucleotide of helix 2 (G27). The backbone connecting these nucleotides contacts the edge of the thiophene to further enhance binding. The ureido ring is similarly locked between the major groove edge of A26 and the U7 ribose. Watson-Crick-paired bases at the bottom of helix 1 (G6:C17 and U7:A16) complete the binding pocket, forming contacts across the biotin N2 side. The degree of steric complementarity between the RNA and the ligand is remarkable (Figure 4(a)), especially considering the unusual bent arrangement of the rings in the biotin headgroup and the fact that the binding site is formed largely from nucleotides tied up in conventional Watson-Crick base-pairing.

Specific recognition of the ligand is accomplished at the ureido head group. Two axial magnesium inner-shell contacts bridge the biotin ureido carbonyl group and the RNA phosphate backbone (Figure 3(b)). By hydrogen bonding to other active site nucleotides, the magnesium-coordinated water molecules help to orient the metal with respect to the ligand. The biotin carbonyl group also appears to hydrogen bond directly to RNA via the N2 of G6. While the N1 side of biotin does not interact directly with the RNA, a pair of bound magnesium ions and associated water molecules combine to block its accessibility to bulk solvent (Figure 3(a)). A tetrahedral network of hydrogen bond donors and acceptors positions three water molecules between the solvated magnesium ions, ultimately constraining one of them to serve as an acceptor for the biotin N1 amide proton. Surprisingly, no hydrogen bond acceptor is positioned to accept the buried N2 amide proton on the opposite side of the ureido ring.

Characterization of several self-splicing and self-cleaving ribozymes has implicated metal ions in direct stabilization of their reaction transition states (Dahm & Uhlenbeck, 1991; Grosshans & Cech, 1989; Piccirilli et al., 1993), but no previous structural studies have revealed exactly how such metal ions are positioned within RNA active sites and how they interact with reactants. The biotin aptamer provides a high-resolution view of how a divalent metal can be exploited by an RNA, and
thus may serve as a useful model for understanding ribozyme function.

**RNA versus protein**

Biotin is bound to the protein avidin with a phenomenally high affinity \( (K_D \approx 10^{-15} \text{ M}, 10^9\text{-fold tighter than the biotin aptamer}) \). The structure of the avidin-biotin complex has been previously determined by X-ray crystallography (Pugliese et al., 1993; Livnah, et al. 1993), making it possible to compare directly protein and RNA-based approaches to biotin recognition. While some superficial similarities are apparent, fundamental differences in the mechanisms for recognition extend beyond the obvious distinctions between amino acids and nucleotides. Biotin is buried deep in the hydrophobic core of avidin and is contacted by several aromatic side-chains, some of which pack analogously to the nucleotide bases in the aptamer structure (Figure 4(b)). In contrast, the aptamer buries only a portion of its ligand in a relatively solvent-accessible pocket. Whereas polar interactions with the biotin headgroup are mediated by hydroxyl and amide-bearing amino acid side-chains in the avidin complex, bound solvent molecules serve this role in the RNA complex. Non-polar interactions extend the full length of the biotin fatty acid tail in the avidin complex while the biotin headgroup itself is relatively loosely packed. In contrast, the aptamer contacts only the proximal end of the fatty acid tail but (treating bound solvent as part of the binding pocket) packing around the head group is significantly tighter (Figure 4(c)). The aptamer was originally isolated using biotin immobilized via its carboxylate tail to

**Figure 3. Biotin binding site.** (a) The aromatic bases of A26 and G27 and the backbone connecting them interact tightly with the biotin thiophene ring. Packing with the A26 base and the U7 ribose moiety, direct coordination to a buried, solvated magnesium ion (Mg6), and hydrogen bonds to the G6 exocyclic amine and a frozen water molecule stabilize the ureido ring. (b) \( \sigma \), weighted \( 2F_o - F_c \) electron density map of a portion of the ligand binding pocket at 1.3 Å contoured at 1.0 \( \sigma \) above the mean. Figures prepared using Conic (Huang, 1991) and RASTER3D (Merritt & Murphy, 1994).

Structure of a Biotin-binding Pseudoknot

1241
an agarose support. Steric constraints may have thus prevented significant interactions between the RNA and the fatty acid tail and thus account for the relatively poor recognition observed for this part of the molecule.

Whereas the biotin-aptamer complex dissociates in a matter of seconds (as judged by rate at which specifically bound RNA can be competitively eluted from biotin agarose), the half-life for the avidin complex is measured in months ($t_{1/2} \approx 120$ days; Green, 1975). Why does the aptamer bind biotin orders of magnitude weaker? Excluding the end of the fatty acid tail, the RNA makes largely the same complement of ligand interactions as the protein, suggesting enthalpic stabilization of the bound state may not be the crucial difference. In contrast to avidin, the aptamer relies on several solvent cofactors whose positioning entails some entropic cost. In addition, many interactions observed in the aptamer-biotin complex may replace similar interactions in the unbound state, reducing their net contribution to binding. For example, stacking of biotin with A26 and G27 likely substitutes for direct stacking between the two nucleotides in the absence of biotin. Similarly, magnesium coordination of the biotin carbonyl group must compensate for release of a water ligand from a fully solvated free magnesium ion in the unbound state. In contrast, amino acid side-chains in the avidin binding pocket are locked in position by the protein’s rigid tertiary and quaternary structure, effectively preventing them from forming stable interresidue contacts (Pugliese et al., 1994). It is interesting to note that the monomeric form of avidin, lacking both a binding pocket tryptophan residue and the rigid structure of the tetrameric form, has an affinity comparable to that of the biotin aptamer (Kohanski & Lane, 1990).

The aptamer structure reveals how a highly complementary binding pocket for an unusual ligand can be built by an RNA using mostly Watson-Crick base-paired nucleotides and a collection of bound solvent molecules. Structural analysis of mutant pseudoknots and biotin analogs combined with quantitative measurements of ligand binding is set to reveal the energetic basis for molecular recognition by RNA and to pave the way for future efforts at RNA-targeted rational drug design.

**Methods**

**Crystal growth and preparation**

Aptamer RNA was prepared and purified as described (Nix et al., 1999). The crystallization solution contained RNA at a concentration of 1 mg/ml in 3.3% (v/v) methyl-2,4-pentanediol (MPD), 40 mM KCl, 13.3 mM potassium cacodylate-6.6 mM potassium Hepes (pH 7.4), 0.1 mM spermine, 3.5 mM MgCl₂ and 3.5 mM biotin. Crystallization was performed using the hanging drop method by equilibration of the RNA solution over a reservoir of 22.5% MPD at 28 C. Crystals appeared
within three days and grew to final size (250 μm × 180 μm × 120 μm) within one week. Crystals were prepared for data collection by transferring them into a solution containing 40 mM potassium cacodylate (pH 7.4), 0.1 mM spermine, 50 mM MgCl₂, 5 mM biotin, 100 mM KCl, 22.5% MPD and 10% (v/v) glycerol for no less than 15 minutes and no longer than 45 minutes. They were then transferred to a nylon loop, flash frozen in liquid propane and stored in liquid nitrogen.

Introduction of selenobiocin was accomplished by soaking crystals in a solution containing 40 mM potassium cacodylate (pH 7.4), 0.1 mM spermine, 50 mM MgCl₂, 2.5 mM selenobiocin, 100 mM KCl, and 22.5% MPD for 24 hours. These crystals were cryostabilized using the above selenobiocin solution with the addition of 10% glycerol and treated in the same manner as the native crystals.

Data collection on all crystals was performed at −170°C. The intensities for the native crystal (crystal 1) were collected as 1° oscillation images using CuKα radiation produced from a Rigaku RU200 rotating anode generator operating at 50 kV and 180 mA and collected on a R-AXIS IIC imaging plate system (Molecular Structure Corporation, Houston, Texas). Data for the two selenobiocin containing crystals were collected at the Advanced Light Source Beamline 5.0.2. A four-wavelength anomalous dispersion experiment was performed on crystal 2 using the selenium atom of the RNA-bound selenobiocin as the anomalous scatterer. Wavelengths were determined from an in situ X-ray fluorescence scan of the crystal at a right-angle to the beam. Data were collected using 1° oscillations in 10° blocks and the Bijvoet pairs were then measured by rotating the crystal 180°.

The final 1.3 Å resolution data were collected on crystal 3 by changing the 2-theta angle of the detector to 12°. For all crystals, the data were reduced and scaled using the packages DENZO and SCALEPACK (Otwinowski, 1993) (Table 1). The crystals belong to the monoclinic space group C2 and have dimensions a = 47.99 Å, b = 31.09 Å, c = 62.08 Å, β = 111.88° with one molecule in the asymmetric unit.

Structure determination

An anomalous Patterson map (calculated using the MADSYS package; Hendrickson, 1991) contains a single selenium peak (21 σ) from which starting phases were calculated using the program SHARP (La Fortelle et al., 1997). A total of 721 reflections missing from the MAD experiment were merged from the native data set and phased by extension using histogram matching, and a solvent content of 27% in the CCP4 DM program (Collaborative Computational Project No. 4, 1994). The combined data yield a readily interpretable electron density map for both the sugar-phosphate backbone and bases from which an initial model was built using XtalView (McRee, 1992). The model was subsequently refined by multiple rounds of simulated annealing in CNS (Brüger et al., 1998) and manual rebuilding using difference Fourier maps. Solvent molecules were introduced using the Water utility of XtalView once the entire RNA-biotin molecule was modeled. Water molecules displaying geometries inconsistent with hydrogen-bonding or containing B-factors greater than 55 Å² were removed. The 1.5 Å MAD data were then replaced with the 1.3 Å data from crystal 3 and the model was further refined by energy minimization (with restrained individual atomic temperature factors and occupancies) using CNS for the final structure.

Acknowledgments

We thank Dr W. Hendrickson for providing selenobiocin, Dr T. Earnest for assistance with MAD experiments at ALS and Dr J. Cate for help in data collection. Supported by awards to C.W. from the NIH, NSF, and David and Lucile Packard Foundation. The Center for Molecular Biology of RNA at UCSC is supported by a grant from the W.M. Keck Foundation.

References


Edited by I. Tinoco

(Received 24 November 1999; received in revised form 14 January 2000; accepted 17 January 2000)