SURVEY AND SUMMARY
Comparative analysis of hairpin ribozyme structures and interference data

Sean P. Ryder and Scott A. Strobel*

Yale University, Department of Molecular Biophysics and Biochemistry, 260 Whitney Avenue, New Haven, CT 06520-8114, USA

Received November 16, 2001; Revised and Accepted January 14, 2002

ABSTRACT

Great strides in understanding the molecular underpinnings of RNA catalysis have been achieved with advances in RNA structure determination by NMR spectroscopy and X-ray crystallography. Despite these successes the functional relevance of a given structure can only be assessed upon comparison with biochemical studies performed on functioning RNA molecules. The hairpin ribozyme presents an excellent case study for such a comparison. The active site is comprised of two stems each with an internal loop that forms a series of non-canonical base pairs. These loops dock into each other to create an active site for catalysis. Recently, three independent structures have been determined for this catalytic RNA, including two NMR structures of the isolated loop A and loop B stems and a high-resolution crystal structure of both loops in a docked conformation. These structures differ significantly both in their tertiary fold and the nature of the non-canonical base pairs. These loops dock into each other to create an active site for catalysis. Recently, three independent structures have been determined for this catalytic RNA, including two NMR structures of the isolated loop A and loop B stems and a high-resolution crystal structure of both loops in a docked conformation. These structures differ significantly both in their tertiary fold and the nature of the non-canonical base pairs formed within each loop. Several of the chemical groups required to achieve a functioning hairpin ribozyme have been determined by nucleotide analog interference mapping (NAIM). Here we compare the three hairpin structures with previously published interference data to assess the convergence between the structural and functional data. While there is significant disparity between the interference data and the individual NMR loop structures, there is almost complete congruity with the X-ray structure. The only significant differences cluster around an occluded pocket adjacent to the scissile phosphate. These local differences may suggest a role for these atoms in the transition state, either directly in chemistry or via a local structural rearrangement.

INTRODUCTION

Nucleotide analog interference mapping (NAIM) is an efficient chemical group mutagenesis strategy used to identify the subset of atoms important for the function of an RNA molecule (1,2). The method identifies chemical groups involved in secondary or tertiary hydrogen bonds, ligand binding, active site chemistry or other roles required for RNA activity. These data should be useful for analyzing the functional relevance of three-dimensional RNA structures because chemical groups that display an effect on RNA activity when modified or deleted should be the same groups structurally implicated for their involvement in activity (3,4). If there is only a weak correlation between the structural and biochemical data, then it suggests that the interactions observed in the structure are not relevant to the functioning RNA or that the structure does not represent the active state. In either case, NAIM analysis is highly complementary to structure determination, as it can ascribe functional relevance to contacts observed within a structure.

The intent of this report is to compare previously published interference data with recent structures obtained for the hairpin ribozyme, a small catalytic RNA motif that performs a reversible self-cleavage reaction. In nature, the ribozyme functions in the processing of replication intermediates in the ‘life’ cycle of a tobacco ringspot nepovirus satellite RNA (5). In this context, the ribozyme folds into an extended cruciform structure (Fig. 1) (6). The minimum elements of the hairpin RNA required for catalytic activity in vitro are just two arms of this cruciform, which include four short helical elements (helix 1–4) and two internal loop regions, A and B (7,8). The ribozyme performs a cleavage reaction at a specific site within the symmetric loop A, yielding two products with a 5′-hydroxyl group and a 2′,3′-cyclic phosphate, respectively (5). The products of this reaction are similar to three other naturally occurring catalytic RNAs, the hammerhead, VS and HDV ribozymes, although significant differences in secondary structure imply that each molecule has its own distinct catalytic strategy for arriving at these products (9).

Three structures have been reported for the hairpin ribozyme. These include two NMR structures of the isolated stems containing loop A and loop B (10,11) and a high-resolution
crystal structure of the docked complex with a 2′-O-methyl inhibitor in place of the 2′-OH nucleophile (12). These structures differ significantly in both their tertiary structure fold and in the nature of the non-canonical base pairs formed in each loop. Such dramatic differences raise a significant question: which if any of these structures represent the hairpin ribozyme in its active conformation?

Previously, we performed NAIM analysis on the hairpin with several nucleotide analogs of A, C, G and U (13). These included analogs that eliminated or modified functional groups on either the heterocyclic base or the ribose sugar. Here, we utilize these data to assess the functional relevance of the available hairpin ribozyme structures.

HAIRPIN RIBOZYME STRUCTURES

Two independent NMR structures have been determined for the hairpin ribozyme. The structure of helix 1, loop A, and helix 2 (termed stem A) was solved in isolation from the rest of the ribozyme by Cai and Tinoco (11) (Fig. 2). In this structure, the bases of loop A form a series of four non-canonical pairs resulting in an extended helix. One base, U+2, flips out of the helix and is oriented toward the solvent.

The other half of the ribozyme was also determined by NMR. Butcher and coworkers (10) reported the structure of helix 3, loop B, and helix 4 (termed stem B) (Fig. 3). Similar to the loop A structure, the nucleotides in loop B form a series of non-canonical base pairs with two bulged uridines, U39 and U41. The geometry of the non-canonical pairing disrupts the continuous stack, so that helix 3 stacks upon four pairs in the top of loop B, while helix 4 stacks upon the three pairs at the bottom of loop B. As a result, this stem adopts a bent structure with a narrow minor groove that appears to have some conformational flexibility. Though stem A and stem B each form an independently folded unit, the authors suggest that conformational changes may be necessary to achieve the active, docked structure.

SUPERPOSITION OF NAIM DATA ON THE HAIRPIN STRUCTURES

In an effort to systematically compare the data obtained on active hairpin ribozymes with the available hairpin structures, we superimposed the interference data collected with all of the analogs onto the coordinates of each structure (Figs 2–4). The interferences used in this comparison are listed in Table 1. We used the following criterion to assess the correlation between each structure and the biochemical data. (i) A given interference is considered to be agreement if the analog deletes a functional group involved in an interaction within the structure. For example, the N2 amine of G21 interacts with the N7 of A22 in...
the loop B NMR structure. Both of these groups lead to interference when modified, hence this contact is consistent with the interference data. (ii) An interference resulting from incorporation of a bulky nucleotide analog matches the structure if it creates a steric clash, which implies that the structure cannot sufficiently accommodate the analog modification. For example, the N2 amine of G8 makes two contacts via its N2 amine in the crystal structure to the 2′-OH of A-1 and the O2 of U+2. Consistent with these contacts, methylation of this amine disrupts activity. (iii) An analog may affect the local conformation of the nucleotide, for example disruption of the sugar pucker. An example of this occurs at U41 with interference from 2′-fluoro substitution, an analog that favors a C-3′-endo conformation. U41 base adopts C-2′-endo in both the NMR and the crystal structure and hence 2′-fluoro interference is consistent with both structures. For interferences that meet these criteria, the affected functional groups are designated with blue colored spheres in Figures 2–4. If none of these criteria are met within the structure, the functional groups that show interference are designated with yellow colored spheres.

**STEM A AND B STRUCTURES ARE LARGELY INCONSISTENT WITH THE NAIM DATA**

Even a cursory inspection of the figures shows that the isolated stem A and stem B NMR structures are not well matched with the NAIM data (Figs 2 and 3). In the structure of loop A, most of the interferences (>50%) cannot be explained by the structure. Similarly, the loop B NMR structure does not match well with the interference pattern.

It is possible that the failure of some functional groups to appear relevant within the NMR structures could result from a lack of tertiary contacts to the opposing stem in the functional docked structure. For example, several interferences at positions A9-U12 do not reflect contacts within the secondary structure of the stem, but they cluster in the minor groove surface of the loop A secondary structure. This may reflect a pre-ordered loop B binding surface.

However, there is reason to doubt that the stem A NMR structure represents a pre-folded structural domain that utilizes this minor groove surface to interact with stem B in the functioning ribozyme. Several functional groups expected to be necessary for ribozyme activity based on the geometry of stem A fail to show interference with analogs that delete or modify them. While these are negative data, they may indicate that certain aspects of the NMR structure do not adequately reflect the docked, active structure. For example, the N2 exocyclic amine of G+1 was shown to be critical for ribozyme activity (15,16). Site-specific inosine substitution at this position reduces ribozyme cleavage activity to below detectable levels. In the NMR structure, this amine interacts with the N7 of A9. Analogs that delete or modify the N7 moiety do not interfere with ribozyme activity at A9. If the sole role of G+1 amine is to form a single H-bond interaction with the A9 N7, then analogs that disrupt this position should affect ribozyme activity at a level comparable with the G+1 inosine effect. While it is possible that the G+1 amine may form an additional contact in the active structure, these data argue that the stem A structure might not reflect that conformation.

Unlike the interferences in stem A, those in stem B do not cluster into a discreet region of the structure. Instead, they are scattered across the minor and major grooves throughout the length of the molecule. As such, these interferences are not
**Table 1. NAIM analysis of the hairpin ribozyme**

<table>
<thead>
<tr>
<th>Position</th>
<th>Analog effects</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>G8</td>
<td>Ino, m^2G</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>m^2A, DAP, Pur, 2AP, c^3A, n^4A</td>
<td>n^4A is an enhancement, attributed to the N1 position (17)</td>
</tr>
<tr>
<td>A10</td>
<td>Pur, 2AP, dA, ^5A, c^1A, n^4A</td>
<td>^nA interference is pH dependent, attributed to N1 ionization (17)</td>
</tr>
<tr>
<td>G11</td>
<td>Ino, m^2G, dG, ^5G</td>
<td></td>
</tr>
<tr>
<td>U12</td>
<td>dU, ^3U</td>
<td></td>
</tr>
<tr>
<td>G21</td>
<td>Ino, m^6G</td>
<td></td>
</tr>
<tr>
<td>A20</td>
<td>m^6A</td>
<td></td>
</tr>
<tr>
<td>A22</td>
<td>7dA, m^6A</td>
<td></td>
</tr>
<tr>
<td>A23</td>
<td>m^6A, DAP, 2AP</td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td>DAP, Pur, ^5A, c^1A</td>
<td></td>
</tr>
<tr>
<td>C25</td>
<td>dC, m^5C, Zeb</td>
<td></td>
</tr>
<tr>
<td>A26</td>
<td>DAP, 2AP</td>
<td></td>
</tr>
<tr>
<td>A38</td>
<td>7dA, DAP, Pur, 2AP, dA, ^5A</td>
<td></td>
</tr>
<tr>
<td>U39</td>
<td>^3U, m^2U</td>
<td></td>
</tr>
<tr>
<td>A40</td>
<td>DAP, 2AP, c^1A</td>
<td></td>
</tr>
<tr>
<td>U41</td>
<td>^3U</td>
<td></td>
</tr>
<tr>
<td>U42</td>
<td>5^mU, dU</td>
<td>dU is an enhancement, attributed to increased conformational flexibility (13)</td>
</tr>
<tr>
<td>A43</td>
<td>7dA</td>
<td></td>
</tr>
<tr>
<td>C44</td>
<td>Zeb</td>
<td></td>
</tr>
</tbody>
</table>

Ino, inosine; m^2G, N2-methylguanosine; dG, 2′-deoxyguanosine; ^5G, 2′-deoxy-2′-fluoroguanosine; m^6A, N6-methyladenosine; DAP, 2′-aminopurine riboside; Pur, purine riboside; 2AP, 2-aminopurine riboside; c^1A, 3-deazaadenosine; n^4A, 8-azadenosine; dA, 2′-deoxyadenosine; ^5A, 2′-deoxy-2′-fluoroadenosine; 7dA, 7-deazaadenosine; dC, 2′-deoxycytidine; 5^mC, 5-methyl-2′-fluorocytidine; Zeb, Zebularine; dU, 2′-deoxyuridine; ^3U, 2′-deoxy-2′-fluorouridine; 5^mU, 5-methyluridine (ribozymide).

likely to reflect the formation of tertiary contacts missing in the isolated loop B stem structure. It is more likely that the secondary structure of loop B undergoes a rearrangement prior to forming the docked structure.

**THE CRYSTAL STRUCTURE MATCHES THE NAIM DATA EXCEPT NEAR THE SCISSILE PHOSPHATE**

In contrast to the isolated stem structures, the crystal structure of the intact ribozyme is almost fully consistent with the NAIM data (Fig. 4). Approximately 85% of the interferences can be explained by features in this structure. Interference is observed at most of the functional groups involved in the non-canonical secondary structure in stems A and B, as well as those in tertiary interface. Bulky analogs are accommodated on the surface of this structure, but not in the tightly packed interface between the internal loops.

All of the unexplained interferences cluster around the scissile phosphate in a pocket formed by the conserved residues G+1, A9, A10, C25 and A38 (12). All of the functional groups that point into this cavity are required for catalysis, including the N2 amine of G+1, the N6 amine of A38, and the N1 imino group of A10 (2,13,15). Furthermore, occupation of this cavity by methylation of the N6 amine of A9 significantly disrupts activity (13). Previous work indicated that ionization of A10 is important for ribozyme catalysis, that the N6 amine of A38 is indispensable for catalysis, and that these groups are likely to be involved in the stabilization of the catalytic transition state (17). These few exceptions may reflect local disturbances in the structure necessary to accommodate the 2′-O-methyl substitution at the cleavage site or transition state specific effects on ribozyme chemistry. They may provide clues into the mechanism of catalysis by the hairpin ribozyme.

**INTERPRETATION AND ANALYSIS**

Several lines of evidence suggest that the crystal structure of the docked complex more accurately reflects the biologically relevant structure than either NMR structure. (i) The strong correlation between the NAIM data and the crystal structure evident from Figure 4 in this analysis. (ii) The nucleophile, reactive phosphate, and 5′-oxy leaving group are in an in-line configuration (12). This specific geometry is necessary for the SN2 phosphodiester reaction, but is not the geometry of the scissile phosphate within the loop A structure (11). (iii) The crystal structure matches the artificial phylogenetic data on the ribozyme. All of the conserved nucleotides in the hairpin sequence are involved in specific secondary and tertiary structure contacts (18). (iv) Several specific tertiary interactions predicted by crosslinking and mutagenesis experiments are observed in the crystal structure (8,14,19). (v) The structure is consistent with the majority of solution modification experiments and site-specific analog substitution data (8,16,20–24). (vi) Hampel and Burke have shown that a highly reactive photo-crosslink indicative of the ‘loop E-like’ motif present in the loop B NMR structure is inhibitory to docking and ribozyme catalysis and that the crosslink is lost upon formation of the active structure (21).

Although the biochemical data suggest that the NMR structures of stem A and stem B do not reflect the active conformation, it does not mean that these structures are incorrect. It is likely that each structure reflects the accurate, undocked conformation. In fact, this interpretation is suggested by crosslinks observed in stem B that are indicative of features observed in the NMR structure (21). Furthermore, the lack of observable imino NOEs for the loop region of stem B, as well as intermediate J-couplings for the majority of the bases, suggest that the structure of this molecule is unusually dynamic (10). If true, the hairpin ribozyme must undergo a complex series of rearrangements in order to form the docked complex. This is corroborated by FRET studies that suggest docking limits the rate of ligation chemistry (25).

While the NAIM data strongly support the functional relevance of the crystal structure, a handful of the most interesting interferences cluster in a cavity surrounding the scissile phosphate and are not explained by the structure. These groups do not appear to be making ground state contacts within the tertiary structure, but instead may participate in promoting the chemical transition. One possibility is that the cavity is occupied by a reaction molecule, such as a water, and that these groups are in position to coordinate and activate it for catalysis. While no water molecule was observed in the cavity within the...
hairpin crystal structure, the authors note that there is sufficient room for one to be accommodated (12). Alternately, ionization of A10 could stabilize charge in the pre-ligation ground state or in the transition state by interacting with a non-bridging phosphate oxygen of the scissile phosphate, while a subtle change in geometry could allow the amine of A38 to hydrogen bond to the 5′-hydroxyl nucleophile in the ligation reaction. Although this model would require alterations in the geometry of the bases that surround the reactive phosphate, it is possible that removal of the 2′-O-methyl inhibitor or subtle differences between the cleaved and ligated forms of the ribozyme could lead to these changes. In either case, movement of A10 and A38 by <3 Å would position these bases for a direct role in catalysis.

In summary, the NAIM data argue that the docked crystal structure very closely reflects the functioning state of the hairpin ribozyme. However, the crystal structure does not immediately elucidate the mechanism of hairpin ribozyme catalysis. Strikingly, the only positions of interference that do not correlate to the crystal structure lie in the vicinity of the catalytic core of a group I intron. Nature Struct. Biol., 6, 1037–1042.

We would like to thank Michael Recht and Pete Funke for assistance with preparing figures. This work was supported by NSF grant CHE-0100057 to S.A.S.

REFERENCES