Near Native Structure in an RNA Collapsed State†

Karen L. Buchmueller and Kevin M. Weeks*

Department of Chemistry, University of North Carolina, Chapel Hill, North Carolina 27599-3290

Received August 18, 2003; Revised Manuscript Received September 17, 2003

ABSTRACT: Many large RNAs form conformationally collapsed, but non-native, states prior to folding to the native state or assembling with protein cofactors. Although RNA collapsed states play fundamental roles in RNA folding and ribonucleoprotein assembly processes, their structures have been poorly understood. We obtained 12 high-quality structural constraints for the collapsed state formed by the catalytic core of the bI5 intron RNA using site-specific cross-linking mediated by a short-lived reactant. RNA tertiary structures in the collapsed and native states are indistinguishable, even though only the native state forms a solvent-inaccessible core. Thus, structural neighbors in the collapsed state, including several long-range tertiary interactions, are approximately as close in space as in the native state, but RNA packing is sufficiently loose or dynamic to allow access by solvent. Binding by the obligate CBP2 protein cofactor has almost no effect on structural neighbors reported by cross-linking, even though protein binding chases the RNA from the collapsed state to the native state. Protein binding thus appears to promote only the final few angstroms of RNA folding rather than mediate global conformational rearrangements in the catalytic core. The bI5 RNA collapsed state functions to self-chaperone ribonucleoprotein assembly because this conformationally restrained structure lies very near that of the native state and excludes structures that otherwise misassemble efficiently.

Structural biogenesis of cellular RNA−protein complexes is a multistep process and must reflect the extent to which the nascently transcribed RNA has preexisting structure. The bI5 (*1*), bI3 (*2*), and *Tetrahymena* (*3*) group I intron RNAs, the catalytic domain of RNase P RNA (*4*), and domains of *Escherichia coli* 16S ribosomal RNA (*2*) form collapsed states prior to folding to the native state or forming native ribonucleoprotein (RNP) complexes. These RNA collapsed states are, on average, 5−15% larger than the native state (*1*, *3*, *4*) and form more rapidly than rate-limiting RNA folding (*5*, *6*) or RNP assembly (*7*) steps under roughly physiological ion conditions. Thus, current evidence suggests that the propensity to form a conformationally collapsed, but non-native state is an important, but completely unremarkable, property of large RNAs under a roughly physiological ion environment.

Two general classes of large RNA collapsed states have been identified. For RNAs that form their native tertiary structure under physiological conditions, such as the *Tetrahymena* ribozyme and the catalytic domain of RNase P, the collapsed state is observed as a transient kinetic intermediate, whose backbone is inferred to be accessible to solvent (*5*, *6*, *8*). In contrast, the bI5 and bI3 group I introns (*1*, *2*) and domains of *E. coli* 16S RNA (*2*, *9*) are examples of the presumably more typical case in which the collapsed state predominates under near physiological conditions prior to assembly with obligate protein cofactors.

A good model for the RNA collapsed state is the catalytic core of the bI5 group I intron (bI5core RNA; Figure 1). The bI5core RNA exists in an expanded state at low Mg^{2+} concentrations, and the collapsed state predominates by 7 mM Mg^{2+} (*1*). Like the parent catalytic RNA (*1*, *10*), the bI5core RNA folds to the native state only upon binding by the CBP2 protein cofactor (at 7 mM Mg^{2+}) or upon addition of Mg^{2+} to 40 mM.

Although a prerequisite for RNA collapse is the screening of negative charges on the RNA backbone, modest divalent ion concentrations (∼1 mM Mg^{2+}) are expected to be sufficient to neutralize a large fraction of the effective charge on many phosphodiester groups (*11*, *12*). Several groups have thus reasonably emphasized that formation of the RNA collapsed state may reflect a largely nonspecific electrostatic collapse (*6*, *13*−*15*). However, nonspecific contributions cannot fully account for two biologically important features of the bI5core RNA collapsed state. First, mutations that disrupt tertiary interactions in the bI5 RNA yield a collapsed state that is less compact than that formed by the native RNA (*1*). Second, under conditions where the collapsed and expanded states are in equilibrium, folding to the collapsed state functions to prevent formation of a kinetically trapped bI5 RNA−CBP2 complex (*7*). Folding from the collapsed to the native state for RNase P RNA is proposed to involve small-scale structural rearrangements (*5*), a view that similarly implies higher order structure in this RNA collapsed state. The fundamental observation that disrupting native tertiary interactions changes the structure of the collapsed state, first demonstrated for the bI5 RNA (*1*), has also recently been confirmed for the *Tetrahymena* ribozyme (*16*).

† This work was supported by the Searle Scholars Program of the Chicago Community Trust and by a grant from the NIH (GM56222) to K.M.W.

* Corresponding author. E-mail: weeks@unc.edu.

10.1021/bi035476k CCC: $25.00
Published on Web 11/05/2003

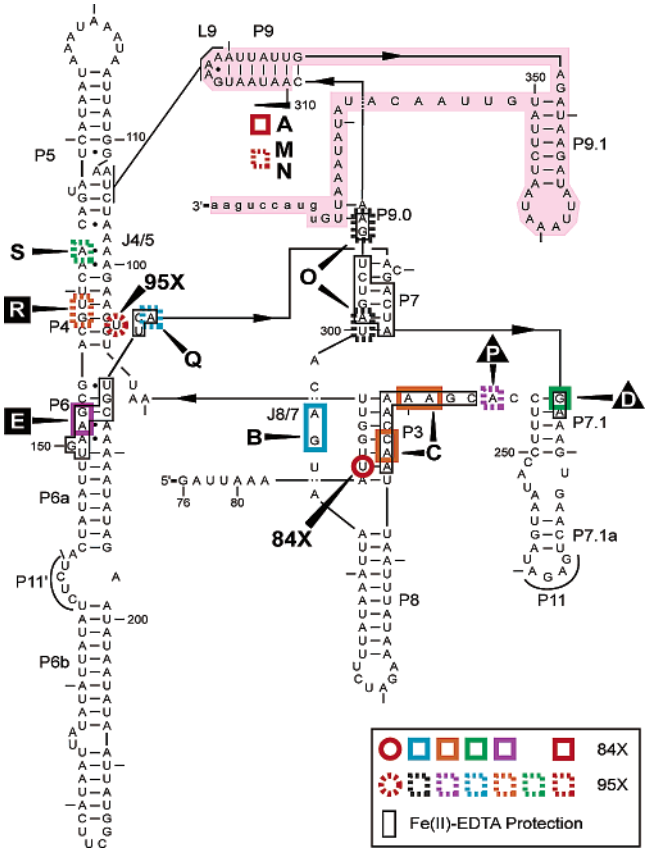

FIGURE 1: bI5core RNA secondary structure and structural neighbors. Photoactivatable cross-linking groups were introduced at U84 in the P7−P3−P8 domain (84X RNA, solid circle) or at U95 in the P5−P4−P6 domain (95X RNA, dashed circle). 84X and 95X cross-links are shown as solid and dashed rectangles, respectively. Cross-links A, M, and N lie 3′ of nt 310 (pink box). Black letters and white letters on black boxes denote cross-links that predominate under native and expanded state conditions, respectively. White letters on black triangles indicate transient cross-links. Selected regions protected from hydroxyl radical cleavage are shown by thin boxes. It is possible that P7.1a may form a loop E motif (*31*) (not shown); either local conformation is compatible with the material conclusions of this work.

Because the collapsed state forms rapidly and has been found for all large RNAs for which it has been sought, the structures populated in this state are likely to have a profound effect on biological folding and ribonucleoprotein assembly reactions. Other than the observation that these collapsed states are conformationally compact and form rapidly, there only existed sparse structural information to guide thinking about the ubiquitous RNA collapsed state.

We developed a photo-cross-linking approach that allows us to map structural neighbors in large RNAs in a way that requires minimal preconceived assumptions about global RNA architecture. We then use this approach to map structural neighbors in the bI5core RNA collapsed state and to explore the functional consequences of the RNA collapsed state on subsequent ribonucleoprotein assembly.

EXPERIMENTAL PROCEDURES

General. bI5 RNAs were refolded (*1*), and all reactions were carried out in folding buffer [50 mM Hepes (pH 7.6), 50 mM KCl, and $MgCl_2$] at 22 or 25 °C. Duplex (84D and 95D) and single-stranded (84S) RNAs were synthesized (*17*), and native gel electrophoresis, K_d determinations, and hydroxyl radical footprinting were performed (*1*, *18*) as described previously.

3-Hydroxyphenyl Azide-Modified bI5core RNAs. bI5core RNAs containing single, site-specific, photoactivatable cross-linking groups were obtained via a continuous semisynthesis. A 33 nt oligoribonucleotide (synthesized by Dharmacon or the Nucleic Acids Facility at North Carolina State University, 120 μmol, positions 76−108 in Figure 1) containing a 2′-amino uridine at position 84 or 95 was 5′-^{32}P-end-labeled at low specific activity to facilitate detection during processing, treated three times with 50 mM *N*-hydroxysuccinimidyl-4-azidosalicylic acid (Pierce Scientific) [1 h each, 50% formamide, 100 mM Hepes (pH 7.6), 22 °C, in amber tubes with overhead lights off], and purified by G-50 size exclusion chromatography. The remaining 270 nt of the bI5core RNA were transcribed (*1*) in the presence of 10 mM GMP and 0.15 mM GTP to yield RNAs carrying a 5′-monophosphate. Transcripts and phenyl azide-modified oligoribonucleotide were annealed to a DNA splint (5′-CAT AGT ATT ATT TAT TTA TTA ATA CCT TAG ATT TTC TTG ACA ATT, 25 °C, 10 min; 1:3:3 ratio), joined using DNA ligase (*19*) (3 h, 22 °C), purified by denaturing electrophoresis (5% polyacrylamide, 7 M urea), recovered from the gel in 500 mM sodium acetate (pH 6) and 1 mM EDTA, stored in TE [10 mM Tris (pH 7.5), 1 mM EDTA], and 5′-end-labeled as needed.

Photoinduced Cross-Linking. Renatured phenyl azide-modified 84X or 95X bI5core (0.1−50 nM) and duplex (0.4 nM) RNAs were irradiated with polystyrene filtered UV-B light (*17*) (λ_{max} 312 nm). Cross-linking reactions (10 μL, 4.4 cm from the light source, 15 min) were quenched by addition of 1 μL of 1 M DTT and 9 μL of 80% (v/v) formamide and 50 mM EDTA. bI5core RNA products were resolved on 8% or 10% gels. Cross-linking efficiencies were calculated by subtracting a background band taken near the cross-linked species and dividing by the total intensity of all bands in the lane, including un-cross-linked material. Gel-purified cross-linked duplex products, generated using a 5′-^{32}P-labeled complementary oligo, were subjected to alkaline hydrolysis [25 mM NaOH, 1.5 min at 90 °C, neutralized with $^2/_5$ volume of 1 M Tris (pH 7.5) at 4 °C] and resolved on a 15% gel. Cross-linking sites in the bI5core RNAs were identified by primer extension (*20*) using gel-purified cross-linking products and primers (5′-GAT TAT TTA ATA TCT TAT CTC or 5′-TTC AGG TAC ACA ATT TAT).

Cross-Linking with CBP2 and Mg^{2+} Jumps. Phenyl azide-modified RNA (0.4−1.0 nM) was incubated with 50 nM CBP2 (10 μL, 22 °C, 30 min) and UV-irradiated (10 min). In Mg^{2+}-jump experiments, solution conditions were changed by 10-fold dilution (from 5 to 50 μL).

RESULTS

Cross-Linking in Model RNAs. We used photo-cross-linking to obtain high-quality structural constraints for the bI5core RNA collapsed state in experiments designed to be substantially free of preexisting assumptions about tertiary interactions in the RNA. A 3-hydroxyphenyl azide photo-cross-linking group (Figure 2) was incorporated at solvent-accessible (*1*, *21*) backbone positions in either of the two pseudohelical structures that comprise the bI5core RNA: in the P7−P3−P8 domain at position 84 (84X RNA) or in the

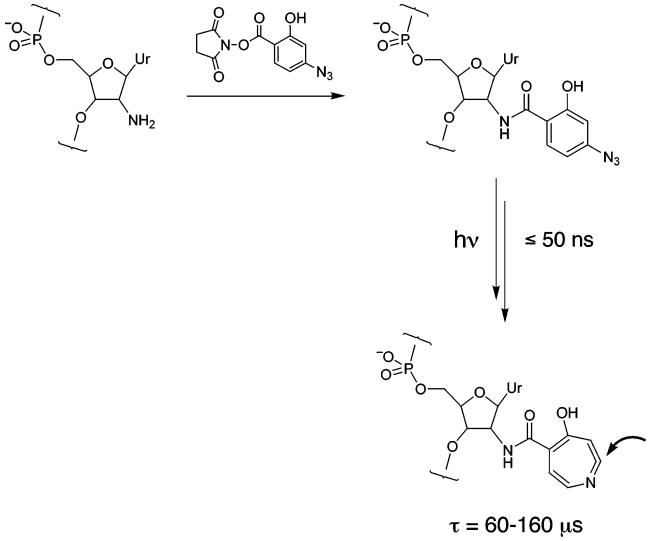

FIGURE 2: Introduction of a 3-hydroxyphenyl azide cross-linking group via a 2′-amido linkage. UV irradiation forms a ketenimine (shown) or ketenimine-derived reactive intermediate, whose lifetime is 60−160 μs (*17*).

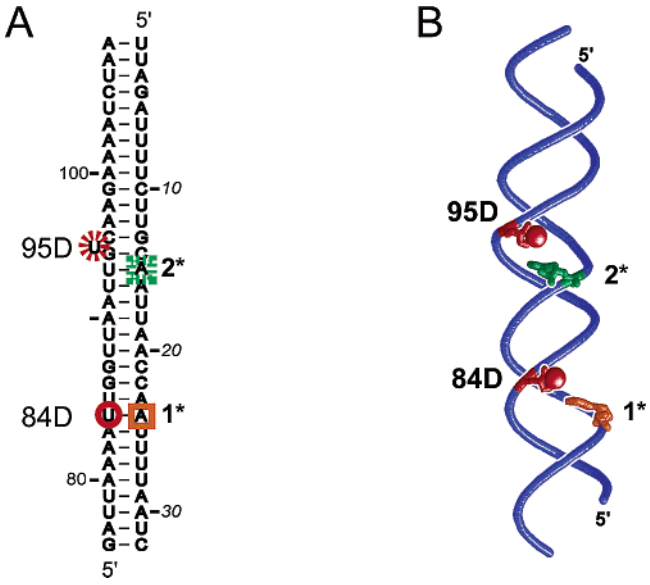

FIGURE 3: 84D and 95D model duplexes used to explore structural selectivity of photo-cross-linking. (A) Duplexes contain a 3-hydroxyphenyl azide group at either position 84 or position 95 (circles) which yield cross-links 1* or 2* (squares), respectively. (B) Cross-linker positions (84D and 95D, in red) and their cross-linking partners (1* and 2*, orange and green) visualized in three dimensions.

P5−P4−P6 domain at position 95 (95X RNA) (red circles in Figure 1).

We first evaluated the quality of structural information obtainable by 3-hydroxyphenyl azide photolysis (Figure 2) using model RNA duplexes, termed 84D and 95D (Figure 3). Duplex RNAs were designed so that the local environments at the phenyl azide group were similar to their respective positions in the bI5core RNA (see circled U's in Figures 1 and 3). Position U84 participates in an U-A pair and position U95 forms a single nucleotide bulge.

Photolysis of 3-hydroxyphenyl azide yields a ketenimine or ketenimine-derived electrophile (see Figure 2) (*17*). The electrophilic intermediate reacts broadly with RNA nucleotides and has a reactive lifetime in our aqueous buffer of 60−160 μs. This lifetime is much shorter than the fastest large-scale folding events observed for RNA (reviewed in Figure 12 of ref *17*); thus, only RNA structures that are physically close in space at the time of photoactivation will form a cross-link.

UV irradiation of the phenyl azide duplexes, 84D or 95D, yielded a single low-mobility cross-linking product, as resolved by denaturing gel electrophoresis. Cross-linking partners for each duplex were mapped as the first gap in an alkaline hydrolysis ladder (Figure 4B). Cross-linking partners for the 84D and 95D duplexes are superimposed on the RNA secondary structure in Figure 3A (orange and green boxes, respectively). Both duplexes form cross-strand cross-links that lie within 8 Å of the 2′-derivitized position (cross-links 1* and 2* in Figure 3B).

In contrast, no distinct cross-linking products were observed with a single-stranded RNA derivatized at position 84 (Figure 4A, formed by omitting the complementary strand). The single-stranded RNA is competent for cross-linking because it formed an efficient adduct with adenosine 5′-monophosphate added in trans (see +AMP lane in Figure 4A and also ref *17*). Cross-linking thus rigorously requires a partner RNA substrate in the immediate vicinity of the photoactivated species. The intensity of the duplex cross-links 1* and 2* decreases in a log-linear manner with increasing Mg^{2+} (Figure 4C), even though the lifetime of the photogenerated reactant is intrinsically independent of $MgCl_2$ concentration (*17*). Less efficient cross-linking is consistent with increased rigidity of RNA duplexes or with reduced base flexibility with increasing ionic strength (*22*−*24*).

These model studies thus show that 3-hydroxyphenyl azide cross-linking proceeds via a short-lived, broadly reactive electrophile (*17*) and provides an accurate, high-resolution view of structural neighbors in RNA (Figure 3).

Structural Integrity of Phenyl Azide-Modified bI5core RNAs. Two independent approaches were used to confirm that incorporation of the photo-cross-linking group at position 84 or 95 does not disrupt folding of the bI5core RNA. First, we monitored conformational collapse as a function of Mg^{2+} concentration by native gel electrophoresis (*1*). The 84X and 95X RNAs undergo Mg^{2+}-dependent collapse via transitions that are essentially identical to those of the native bI5core RNA (open symbols in Figure 4D,E). The extent of collapse for the native, 84X, and 95X RNAs is greater than for a previously described (*1*) mutant (in L9) that disrupts the L9−P5 tertiary interaction (see Figure 1) or for an RNA complementary to the native RNA (solid symbols in Figure 4E). Second, we tested the ability of the 84X and 95X RNAs to bind CBP2 at 40 mM $MgCl_2$, a stringent condition that discriminates against nonspecific protein−RNA interactions (*18*). The derivatized RNAs bind CBP2 with affinities within 2- and 5-fold of the native RNA, respectively (Figure 4F). Incorporating a 3-hydroxyphenyl azide group at position 84 or 95 via a 2′-amido linkage is compatible with native folding and specific protein binding by the bI5core RNA.

Identification of Structural Neighbors in the bI5core RNA. The 84X and 95X RNAs were subjected to UV irradiation as a function of $MgCl_2$ concentration. For the 84X RNA, we observe five distinct cross-links (A−E; Figure 5A). Cross-links A, B, and C predominate at Mg^{2+} concentrations (40 mM) that favor native RNA structure. Cross-link E (white

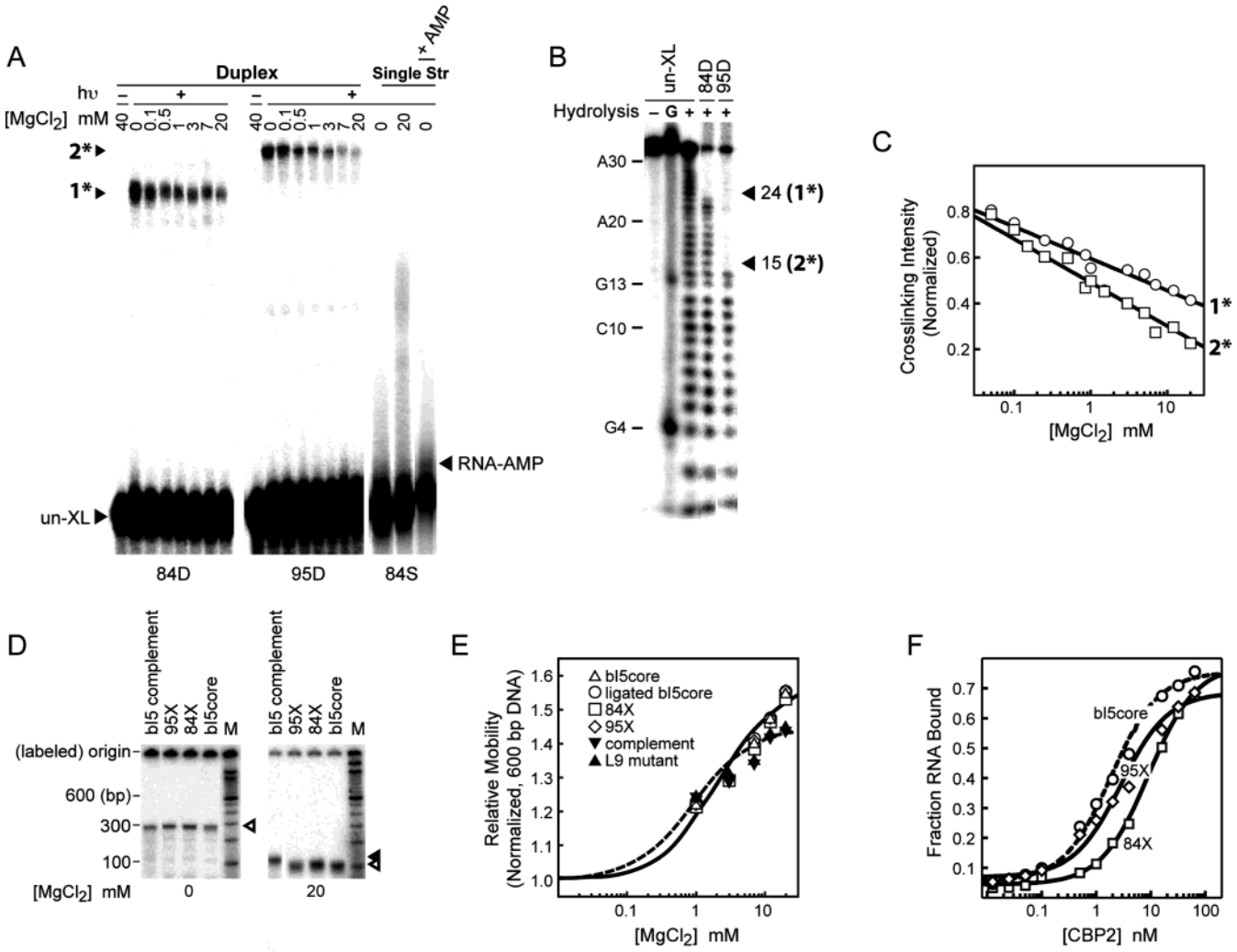

FIGURE 4: Cross-linking with model duplexes and effect of 2′-derivitization on native bI5core RNA structure and folding. (A) Cross-linking of 5′-end-labeled 84D and 95D duplex RNAs as a function of $[MgCl_2]$. No cross-linked products are observed upon irradiation of the single-stranded 84S RNA or in the absence of UV light ($-h\nu$). Cross-linked products with mobilities faster than the un-cross-linked RNAs were not observed. Cross-linking efficiencies for the 84D and 95D duplexes (at 0 mM $MgCl_2$) are 7% and 2%, respectively. The +AMP lane is shown at a lower exposure, for clarity; adenosine 5′-monophosphate was present at 100 mM. (B) Mapping cross-links 1* and 2* by alkaline hydrolysis using a 5′-^{32}P-end-labeled complementary strand. The guanosine ladder (G) was generated by T1 RNase cleavage of un-cross-linked RNA. (C) Normalized efficiency of cross-links 1* and 2* (obtained using 84D and 95D RNAs, respectively) as a function of Mg^{2+}. (D) Mobilities for the native and phenyl azide-modified bI5core RNAs (open triangle) and an RNA complementary to the bI5core (see ref *1*; closed triangle) monitored by native gel electrophoresis. The marker (M) is a double-stranded DNA ladder, in 100 bp increments. Origins were marked by brief electrophoresis of ^{32}P-labeled RNA, just prior to dismantling the gel. (E) The 84X and 95X RNAs have Mg^{2+}-dependent mobilities and transition midpoints ($Mg_{1/2}$'s) for collapse that are similar to the native bI5core RNA (solid line). The extent of collapse for the native, 84X, and 95X RNAs is greater than for a previously described (*1*) mutant (in L9) that disrupts the L9−P5 tertiary interaction (see Figure 1) or for an RNA complementary to the native RNA (dashed line). Data are fit (*1*) assuming noncooperative Mg^{2+} binding; transition midpoints for solid and dashed lines are 2.3 and 1.1 mM, respectively. (F) Equilibrium binding by the native, 84X, and 95X RNAs to CBP2. Fraction of RNA bound $= \epsilon[CBP2]/(K_d + [CBP2])$, where ϵ is a binding efficiency and K_d is the equilibrium dissociation constant; K_d's are 1.9, 2.9, and 9.4 nM (±50%), respectively.

letter on black square) is observed at low Mg^{2+} concentrations, where the expanded state predominates (*1*). The fifth cross-link (D, black triangle) has a maximum intensity at ∼1 mM Mg^{2+}, where the expanded state is favored over the collapsed state (*1*), and then decreases in intensity with increasing Mg^{2+} concentration.

Cross-linking with the 95X RNA (Figure 5B) yielded seven cross-links (M−S). Four (M, N, O, and Q) are most efficient at Mg^{2+} concentrations that stabilize the native state. Cross-link R (black box) reports the expanded state. Cross-link S is approximately constant and reports a structural neighbor found in all folding states. Cross-link P (black triangle) is most efficient at ∼1.5 mM Mg^{2+} and is less efficient at both lower and higher Mg^{2+} concentrations (Figure 5B).

UV-irradiated RNAs containing a 2′-amine at position 84 or 95, but lacking the phenyl azide group, and derivatized RNAs, not exposed to UV irradiation, do not yield cross-links (leftmost lanes, Figure 5A,B). The cross-linking pattern is unaffected by adding a 100-fold excess of unlabeled bI5core RNA (100× lanes, Figure 5A,B). Cross-linking thus requires the phenyl azide and UV irradiation and reports (concentration-independent) intramolecular interactions.

Four of five cross-linking partners for the 84X RNA and five of seven partners for the 95X RNA were mapped by primer extension (summarized as solid and dashed boxes,

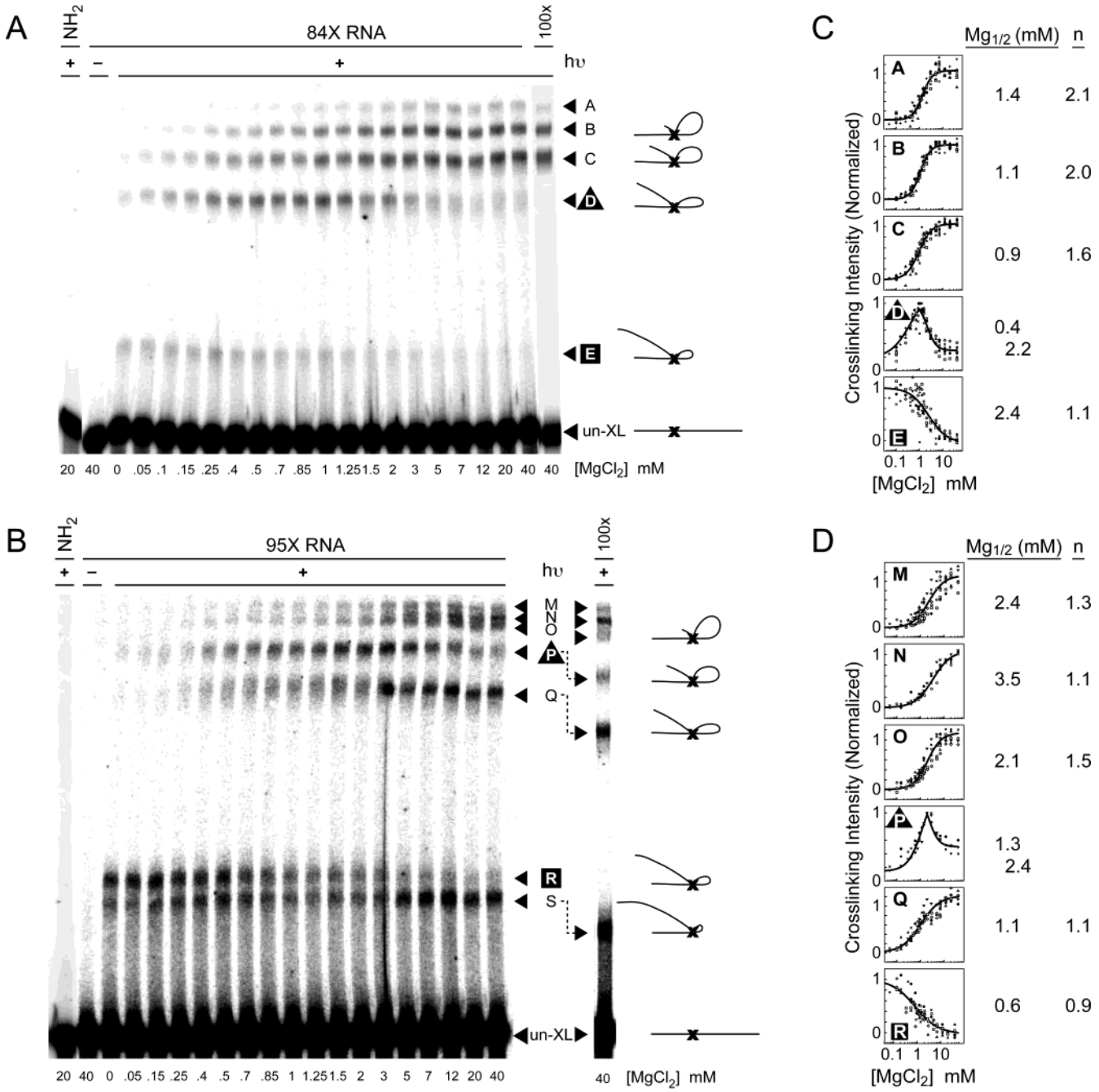

FIGURE 5: Structural neighbors in the bI5core RNA. (A, B) Five distinct cross-links (A−E) are resolved for the 84X RNA. Seven cross-links (M−S) are observed for the 95X RNA. Cross-linking requires UV illumination (+*hν*) and 2′-NH_2 derivatization (2′-NH_2) and is unaffected by the addition of a 100-fold (100×) concentration of excess, unmodified bI5core RNA. Cartoons at right illustrate lariat structures; x is the phenyl azide. (C, D) Cross-linking efficiency as a function of [$MgCl_2$]. Different symbols in each panel indicate data from up to seven independent experiments. Analysis of multiple experiments shows that the intensity of cross-link S is independent of [$MgCl_2$], after normalizing for modest loading differences. Data were fit to cross-linking intensity = $[Mg^{2+}]^n/(Mg_{1/2}{}^n + [Mg^{2+}]^n)$, where $Mg_{1/2}$ is the transition midpoint and *n* is the Hill coefficient. Expanded state cross-links were fit to 1 minus this equation.

respectively, in Figure 1; mapping data are shown in Figure 6). Cross-links C and O each map to two regions, separated by three to four nucleotides. Cross-links form to all four ribonucleotides (Figure 1), as expected given the broad reactivity of the photogenerated intermediate (*17*). There is a monotonic correlation between cross-link lariat size and relative gel mobility, with larger loops (such as cross-links B and O) migrating more slowly (Figure 5). The slowly migrating cross-links A, M, and N, which are too close to the 3′ end of the RNA to be resolved by primer extension, are assigned to be 3′ of nt 310 (pink box in Figure 1).

The Meaning of Cross-Linking Efficiency. The total efficiency of all 84X or 95X cross-links at Mg^{2+} concentrations between 1 and 40 mM is constant at 5 ± 2% and is the same as the cross-linking efficiency of the (native) model duplexes (see legend to Figure 4). Thus, 5% cross-link formation in the bI5core RNA native state (at 40 mM Mg^{2+}) reflects approximately complete folding to this state.

Low, but reproducible, cross-linking efficiencies are a valuable feature of these cross-linking experiments. Model studies show that cross-linking efficiency reflects competition between covalent adduct formation with RNA and rapid

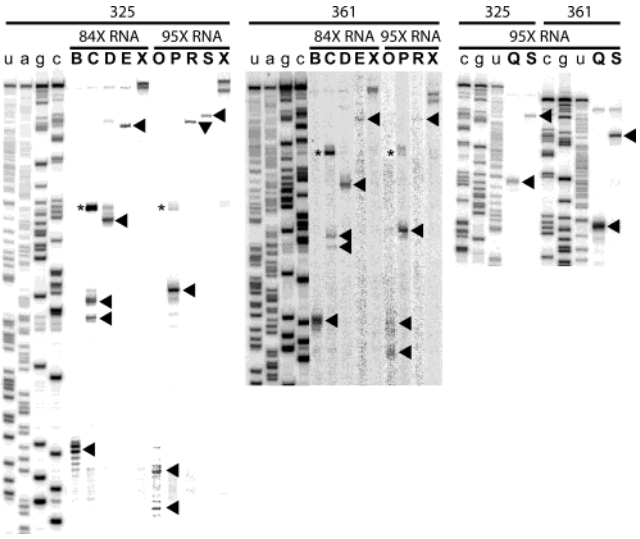

FIGURE 6: Mapping cross-linking partners by primer extension (*20*) and comparison with dideoxy sequencing ladders (u, a, g, c). Cross-link positions (arrows) were confirmed using two independent primers (3′ nt complementary to positions 325 and 361). Asterisks indicate background stops (at positions 216 and 198, respectively) observed with only one of the two primers. Cross-links form to A, C, G, and U ribonucleotides (Figure 1), which reflects the broad reactivity of the photogenerated intermediate (*17*).

(60−160 μs) pseudo-first-order decay of the electrophilic reactive intermediate (*17*). In the absence of a nearby cross-linking partner, the activated intermediate simply decays via reaction with buffer. The degree to which this decay occurs (95%) is not surprising given that the aryl azide cross-linking groups lie at positions designed to be accessible to solvent in the native state. Total cross-linking efficiencies are less (1−2%) at low Mg^{2+} concentrations (≤1 mM) due to reduced proximity of RNA structural neighbors in the bI5core expanded state.

Nucleotide Resolution View of RNA Folding. Cross-linking partners were visualized using a good-quality, three-dimensional model of the bI5 catalytic core (*21*, *25*, *26*), consistent with a large body of biochemical and crystallographic information. Residues mapping to cross-links B and C are close to the 84X cross-linker position (X) in the three-dimensional model (Figure 7A), even though they are well separated in the primary sequence (Figure 1). Similarly, cross-links O, Q, and S, observed with the 95X RNA, also map to a compact region in the three-dimensional model (Figure 7B). The 84X cross-links are localized in the lower, near region of the RNA in the orientation shown (Figure 7A), while 95X cross-linking products occur in the upper, far region (Figure 7B). Cross-linking with the 84X and 95X RNAs thus reports a self-consistent view of tertiary structural neighbors in the bI5core native state.

In several regions, the cross-links provide a high-resolution view of the conformational changes that accompany folding from the expanded to the collapsed and native states. The 95X RNA forms two cross-links, Q and R, in the P4 region of the bI5core RNA (Figure 7C). In the expanded state, the aryl azide at position 95 (in red) forms a cross-strand cross-link, R (orange) in P4. Native state cross-link Q (teal) is also close to position 95 and lies in J6/7, which forms a triple helix with P4 in the native state (*25*, *26*) (Figure 7C). These structural constraints suggest that folding from the expanded state (cross-link R) to the native state (cross-link Q) involves binding of the J6/7 strand at the minor groove to form the P4 triple helix tertiary interaction (Figure 7C). Cross-linking at R is lost in this transition, due to occlusion by the triple helix.

Similarly, position 84 is close to both cross-links C and E (Figure 7D). Expanded state cross-link E likely becomes occluded from the cross-linking group at residue 84 when the P3 helix assumes a native-like tertiary structure (*25*, *27*), as reported by formation of the cross-strand cross-link C.

Transient cross-links D (green) and P (purple) (Figure 7E), obtained with the 84X and 95X RNAs, respectively, form with low efficiency in the absence of Mg^{2+}, exhibit a maximum intensity at ~1 mM Mg^{2+}, and are then disfavored with increasing Mg^{2+} concentrations (Figure 5A,B). These data are consistent with a folding state, at 1 mM Mg^{2+}, involving an alternative packing of P7.1 that places cross-links D and P near helices P3 and P4. In folding to the native state, cross-link D becomes physically occluded from position 84 by the P7.1 helix and J7.1/3 linking region (green backbone in Figure 7E), while cross-link P becomes separated from U95 by P7 and J6/7 (purple backbone in Figure 7E).

Near Native Structure in the Collapsed State. In aggregate, the 12 cross-links provide a comprehensive snapshot for structure in the bI5core RNA (Figure 1). Cross-links A, B, M, N, and O report long-range tertiary interactions. Cross-link Q and the 5′ portion of C report tertiary interactions requiring precise local positioning of RNA structures. Cross-links E, S, and the 3′ component of C are cross-strand cross-links similar to those detected in the model duplexes. Expanded state cross-links E and R report non-native interactions, whose structural neighbors are in fact close in space to, but physically occluded from, the aryl azide cross-linking group in the native state (Figure 7C,D). The non-native structural neighbors reported by cross-links E and R are almost compatible with native folding but lack formation of a key structural element, the P4−J6/7 triple helix or the P6−P3 tertiary interaction, that constrains folding in the native state.

Mg^{2+}-induced RNA folding reported by each cross-link is summarized (Figure 5C,D) by the transition midpoint ($Mg_{1/2}$) and Hill coefficient (*n*). As judged by cross-linking, the bI5core RNA forms structural neighbors characteristic of the native state (cross-links A, B, C, M, N, O, and Q) with $Mg_{1/2}$'s spanning 1.1−3.5 mM.

The photo-cross-linking experiments were compared to complementary information obtained from solvent-based hydroxyl radical footprinting experiments performed under identical solution conditions. Whereas cross-link formation requires that two structures be juxtaposed in three-dimensional space (Figures 3 and 7; ref *17*), protection from hydroxyl radical cleavage imposes the more stringent, but three-dimensionally less structurally informative, requirement that the RNA backbone be stably sequestered from solvent (*28*, *29*).

Cross-links C, O, and Q ($Mg_{1/2}$'s = 0.9−2.1 mM; Figure 5C,D) overlap structures protected from hydroxyl radical cleavage in the native state (see regions boxed with thin black lines in Figure 1). The $Mg_{1/2}$ values observed for overlapping hydroxyl radical protection (17, 16, and 13 mM, respectively; Figure 8A) are much higher than the transition midpoints reported by cross-linking. Similarly, expanded state and

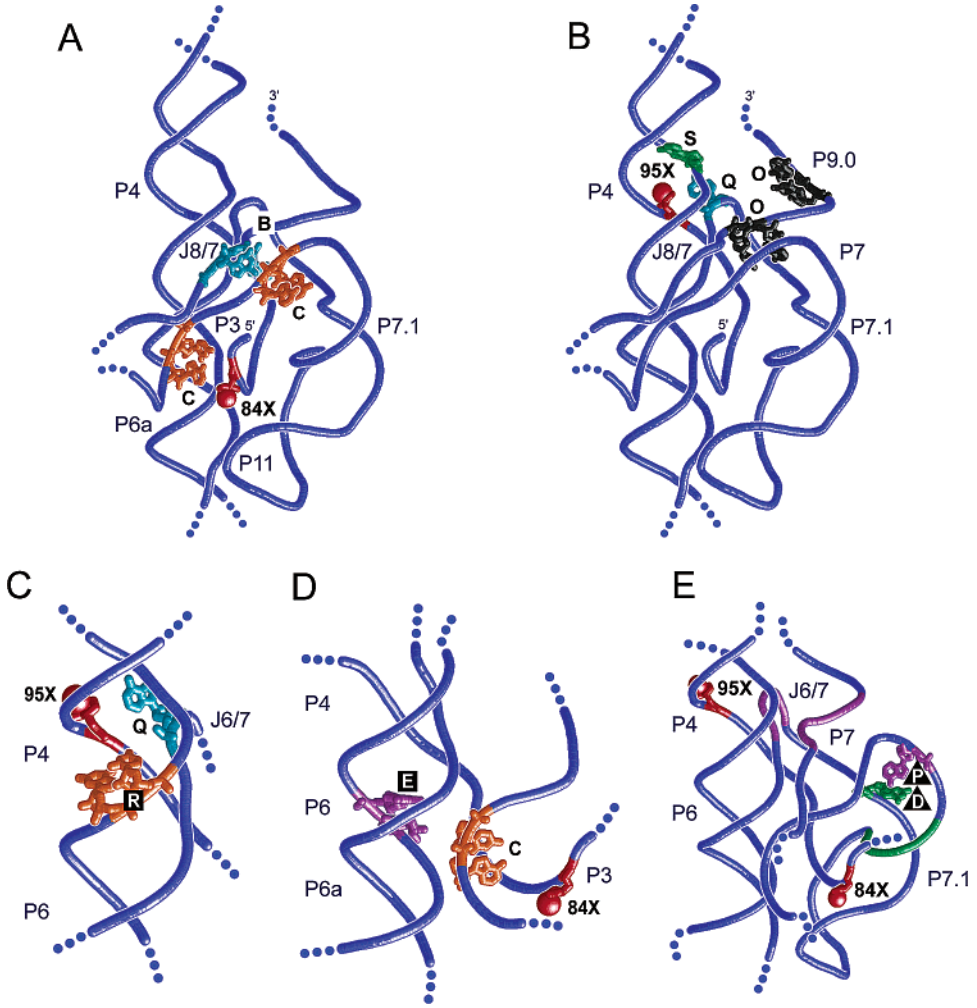

FIGURE 7: Visualizing structural neighbors in the bI5core RNA. (A) Phenyl azide group (X) at the 2′-ribose position of nucleotide 84 and native cross-linking partners B (teal) and C (orange). (B) Nucleotide 95 (X) and native cross-linking products O (gray), Q (teal), and S (green). (C) Cross-linking in the P4−J6/7 region in the 95X RNA with the expanded state (R, orange) and native state (Q, teal) cross-linking partners. (D) Tertiary structure transitions at P3−P6 observed with 84X RNA and expanded state (E, purple) and native state (3′-C, orange) cross-links. (E) Aryl azide-modified positions 84 and 95 and their transient cross-linking partners D (green) and P (purple), respectively. Structures proposed to occlude these cross-links in the collapsed/native states are shown with green and purple backbones, respectively.

transient cross-links E, D, and P ($Mg_{1/2}$'s = 2.2−2.4 mM; Figure 5C,D) are suppressed at substantially lower Mg^{2+} concentrations than overlapping regions are protected from hydroxyl radical cleavage ($Mg_{1/2}$'s = 12−17 mM).

Using methods sensitive only to global properties in prior work, folding of the bI5core RNA could be previously characterized as having an bulk $Mg_{1/2}$ of 3 mM and Hill coefficient of ~1 (*1*). $Mg_{1/2}$ values detected by cross-linking cluster in the range 0.6−3.5 mM, consistent with the prior work; however, they are not superimposable, nor do they show the same cooperativity in Mg^{2+} binding (Figure 5C,D). Four of seven cross-links (A, B, C, and O) reporting native interactions are characterized by modest cooperativity in Mg^{2+} (n = 1.5−2.1; Figure 5C,D). In contrast, expanded state cross-links E and R show significantly less cooperativity in Mg^{2+} binding ($n \leq 1.3$). Thus, formation of native-like interactions, detected either by hydroxyl radical footprinting (n = 1.5−1.8; Figure 8A) or by cross-linking (Figure 5), is generally cooperative in Mg^{2+}, whereas folding from the expanded state to intermediate states does not show this cooperativity.

The striking conclusion is that the cross-linking pattern observed for the bI5core collapsed state at 7 mM Mg^{2+} is essentially indistinguishable from that observed in the native state (20−40 mM Mg^{2+}), even though the collapsed state RNA backbone is readily accessible to solvent-based cleavage by hydroxyl radical. We infer that structural neighbors in the collapsed state, including both long-range (cross-links A, B, M, N, and O) and local (cross-links Q and 5′-C) tertiary interactions, are approximately as close in space as in the native state, but RNA packing is sufficiently loose or dynamic to allow access by solvent.

Effects of CBP2 Binding on Nearest Neighbors in the bI5core RNA. CBP2 binds stably to the bI5 RNA independent of whether the RNA is in the expanded, collapsed or native state (*7*) (K_d's ~ 2 nM). However, the nature of the final CBP2−RNA complex is extremely dependent on the starting state of the RNA. At 7 mM Mg^{2+}, CBP2 binding to the RNA collapsed state chases the RNA to the native, active state (*10*, *21*), whereas binding to the expanded state (at [Mg^{2+}] < 3 mM) produces a stable, but non-native, RNA−protein complex (*7*).

We therefore tested the effect of CBP2 binding on RNA structural neighbors. There is a modest generalized reduction in RNA cross-linking efficiency in the presence of CBP2 which reflects partial trapping of the photoactivated species

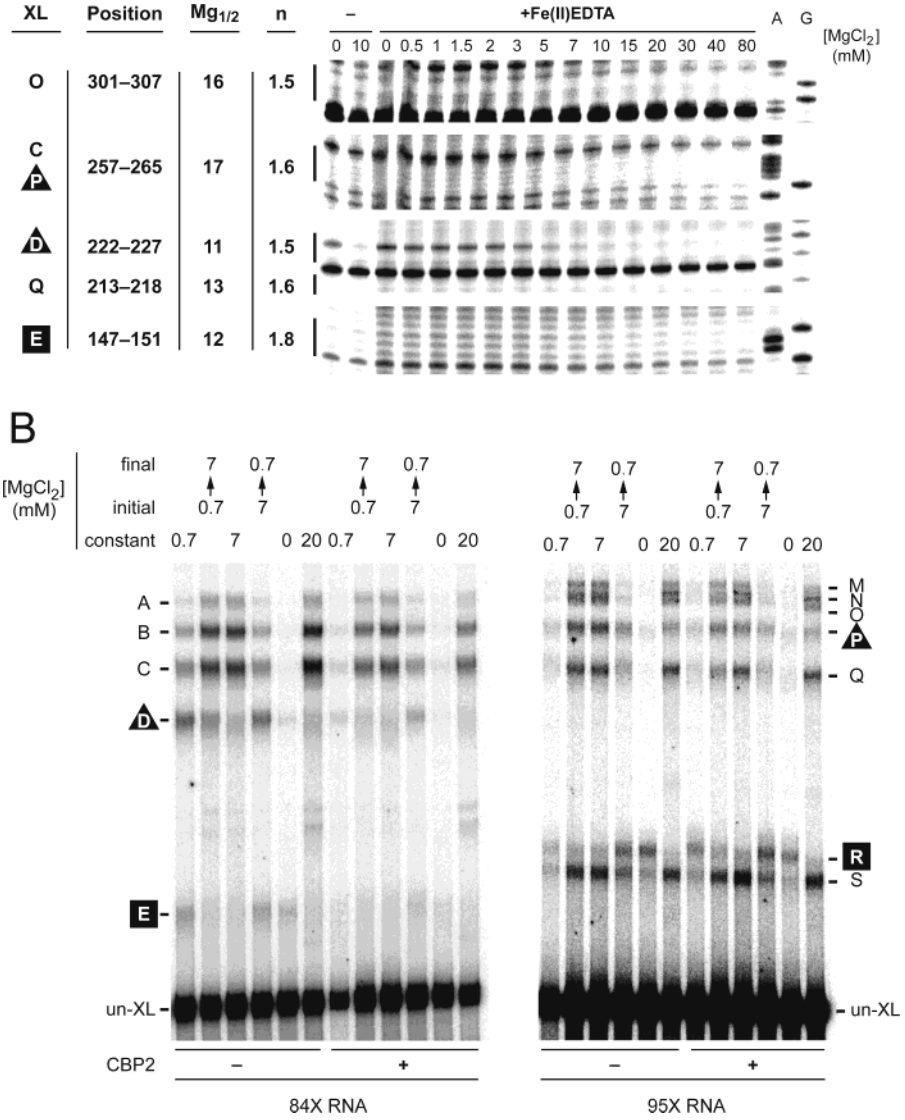

FIGURE 8: Structural neighbors, solvent accessibility, and CBP2 binding. (A) Hydroxyl radical footprinting of regions that overlap structural neighbors reported by photo-cross-linking. Regions protected, transition midpoints ($Mg_{1/2}$), and apparent cooperativity (*n*) as detected by hydroxyl radical cleavage are indicated at the left. Sequencing ladders (A and G) are iodine-treated phosphorothioate-containing transcripts. (B) CBP2 has little effect on cross-linking. 84X and 95X RNAs were incubated in either the absence (−) or presence (+) of CBP2 at an initial [$MgCl_2$] and diluted 10-fold to yield the indicated final Mg^{2+} concentration.

by RNA−CBP2 cross-links (data not shown). Strikingly, for experiments performed at 0, 0.7, 7, or 20 mM Mg^{2+} concentrations using either the 84X or 95X RNAs, CBP2 binding has almost no effect on the global structure of the bI5core RNA [compare (+) and (−) CBP2 (at constant $MgCl_2$) lanes in Figure 8B].

Binding by CBP2 also does not yield new RNA−RNA cross-links beyond those observed for the free RNA. The absence of any effect of CBP2 binding to the collapsed state on the pattern of cross-linking provides strong independent support for the conclusion drawn above that structural neighbors in the collapsed and (CBP2 stabilized) native states are substantially similar.

We tested whether binding by CBP2 to the expanded state (at 0.7 mM $MgCl_2$) constrains the RNA when solution conditions are adjusted (a Mg^{2+} jump) to favor the native ribonucleoprotein complex (at 7 mM Mg^{2+}). In the absence of CBP2, the cross-linking pattern observed following a Mg^{2+} jump is identical to reactions incubated in the presence of a constant 0.7 or 7 mM Mg^{2+} concentration [compare 0.7 (initial) → 7 mM $MgCl_2$ (final) and 7 → 0.7 lanes with constant $MgCl_2$ lanes in Figure 8B].

When the native RNP complex, formed at 7 mM Mg^{2+}, is diluted to a Mg^{2+} concentration that favors the expanded state, the final pattern of cross-linking (at 0.7 mM Mg^{2+}) is essentially identical to experiments performed at a constant 0.7 mM Mg^{2+} concentration [compare 7 (initial) → 0.7 mM $MgCl_2$ (final) and constant 0.7 mM $MgCl_2$ for (+) CBP2 lanes in Figure 8B]. Similarly, prior formation of a CBP2−RNA complex at 0.7 mM does not constrain the RNA following a Mg^{2+} jump to 7 mM because the final cross-linking pattern is exactly that expected for the native state. Thus, the innate structure of the RNA also governs its structure in a ribonucleoprotein complex with CBP2 over the experimental 10 min time frame.

DISCUSSION

Assembly of the bI5 RNP via a Near Native Collapsed State. RNA-tethered photo-cross-linking via a short-lived electrophilic intermediate (*17*) supports two fundamental and

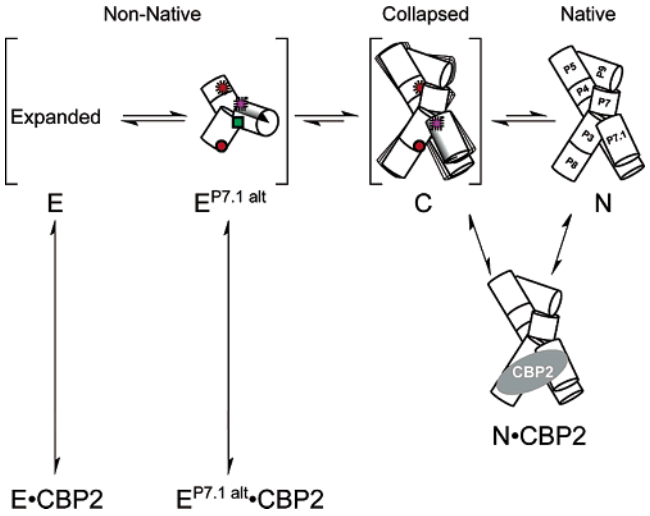

FIGURE 9: Structure of an RNA collapsed state and consequences for assembly of the bI5 ribonucleoprotein complex. The collapsed state (C) is a loose and/or dynamic ensemble, whose average structure lies very close to the native structure. Binding by CBP2 to the expanded state (E or $E^{P7.1alt}$) yields nonproductive complexes (*7*), whereas only binding to the collapsed state chases the RNA to the native state (N). Structural constraints that define $E^{P7.1alt}$ are indicated by cross-links between positions 84 and 95 (solid and dashed red circles) and partners D and P (solid green and dashed purple boxes).

unanticipated conclusions. First, the solvent-accessible bI5core RNA collapsed state is essentially indistinguishable from the native state as judged by structural neighbors in the vicinity of positions 84 and 95 (Figure 5). The collapsed state has substantial native-like structure and is unlikely to be stabilized simply by nonspecific electrostatic interactions. Second, although binding by CBP2 is required for the RNA to fold stably to the native state in a physiological ion environment (*7*) and to carry out its biological splicing function (*30*), binding by this protein cofactor makes no significant difference to the structural neighbors mapped by photo-cross-linking (Figure 8B). Instead, the photo-cross-linking data indicate that CBP2 facilitates only the final consolidation of RNA folding rather than functioning to mediate long-range structural rearrangements or differentiate among states with substantially different conformations.

A model for the structure of the bI5core RNA collapsed state and its contribution to RNP assembly with CBP2 is summarized in Figure 9. In the absence of Mg^{2+} sufficient to screen backbone charges, the RNA exists in an expanded (E or $E^{P7.1alt}$) state (*1*). Experiments designed to map local nucleotide flexibility (*24*) show that the expanded state is characterized by relatively flexible core structures but is anchored by stable peripheral stem−loops at P5, P6b, and P8. Non-native local interactions are detected in this state by cross-linking; these non-native interactions need to undergo only modest local rearrangements at P4 (Figure 7C), at P6 and P3 (Figure 7D), or at P7.1 (pie-shaped shaded region of $E^{P7.1alt}$ in Figure 9) to yield a native fold.

Under roughly physiological ionic conditions (7 mM Mg^{2+}), the bI5 RNA is in a collapsed state (C), whose overall structure lies very near the native state (Figures 5 and 7). The collapsed state is solvent accessible due to a looser packing, due to dynamic motions that allow transient access by solvent, or due to a combination of these factors (Figure 9).

CBP2 forms stable complexes with both E and $E^{P7.1alt}$ structures (*7*), but protein binding does not move these states any closer to the native state conformation (Figure 8B). In contrast, CBP2 binding to the collapsed state chases the RNA to the functional native state (N•CBP2; Figure 9) as judged by recovery of the hydroxyl radical protection pattern reflecting native intramolecular RNA−RNA interactions (*1*, *7*).

An Essential Function of the Collapsed State in Ribonucleoprotein Assembly. Near native structure in the collapsed state provides a compelling physical explanation for the observation that rapid folding to the collapsed state facilitates correct assembly of this RNA with the CBP2 protein cofactor (*7*). The RNA collapsed state functions similarly, but autonomously, as some chaperones in the protein folding world to prevent formation of misfolded or premature interactions (Figure 9). Thus, the ability of the collapsed state to prevent formation of non-native RNA−protein complexes with the expanded state is a self-chaperoning function. The RNA collapsed state self-chaperones productive ribonucleoprotein assembly because it lies very near in structure to the native state and excludes structures that otherwise misassemble efficiently.

As was initially postulated (*1*), collapsed RNA states have been identified for most RNAs for which they have been sought and there is no reason to expect that assembly of the bI5 ribonucleoprotein is a special case. The inability of CBP2 to change the global RNA structure (Figure 8) indicates that the RNA retains primary influence over folding and assembly of the bI5 ribonucleoprotein complex.

The structural information obtained in this work suggests a new view for biogenesis of cellular ribonucleoprotein complexes. Instead of starting from a loosely folded RNA, featuring only isolated secondary structure, we propose that many nascent ribosomal, messenger, and other large cellular RNAs will assemble, with protein cofactors, from rapidly formed, roughly globular, native-like, collapsed states. The ability of analogous RNA collapsed states to modulate interactions with protein cofactors may play a role in, for example, the ordered assembly of the ribosome and in selective spliceosome assembly.

REFERENCES

1. Buchmueller, K. L., Webb, A. E., Richardson, D. A., and Weeks, K. M. (2000) *Nat. Struct. Biol. 7*, 362−366.
2. Buchmueller, K. L., Richardson, D. A., and Weeks, K. M., as reported in Buchmueller, K. L. (2002) Ph.D. Dissertation, University of North Carolina at Chapel Hill, Chapel Hill, NC.
3. Russell, R., Millett, I. S., Doniach, S., and Herschlag, D. (2000) *Nat. Struct. Biol. 7*, 367−370.
4. Fang, X., Littrell, K., Yang, X., Henderson, S. J., Siefert, S., Thiyagarajan, P., Pan, T., and Sosnick, T. R. (2000) *Biochemistry 39*, 11107−11113.
5. Fang, X.-W., Thiyagarajan, P., Sosnick, T. R., and Pan, T. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 8518−8523.
6. Russell, R., et al. (2002) *Proc. Natl. Acad. Sci. U.S.A. 99*, 4266−4271.
7. Webb, A. E., and Weeks, K. M. (2001) *Nat. Struct. Biol. 8*, 135−140.
8. Takamoto, K., He, Q., Morris, S., Chance, M. R., and Brenowitz, M. (2002) *Nat. Struct. Biol. 9*, 928−933.
9. Powers, T., and Noller, H. F. (1995) *RNA 1*, 194−209.
10. Weeks, K. M., and Cech, T. R. (1995) *Cell 82*, 221−230.
11. Manning, G. S. (1979) *Acc. Chem. Res. 12*, 443−449.
12. Misra, V. K., and Draper, D. E. (1999) *Biopolymers 48*, 113−135.
13. Woodson, S. A. (2000) *Nat. Struct. Biol. 7*, 349−352.

14. Heilman-Miller, S. L., Thirumalai, D., and Woodson, S. A. (2001) *J. Mol. Biol. 306*, 1157−1166.
15. Murthy, V. L., and Rose, G. D. (2000) *Biochemistry 39*, 14365−14370.
16. Das, R., et al. (2003) *J. Mol. Biol. 332*, 311−319.
17. Buchmueller, K. L., Hill, B. T., Platz, M. S., and Weeks, K. M. (2003) *J. Am. Chem. Soc. 125*, 10850−10861.
18. Weeks, K. M., and Cech, T. R. (1995) *Biochemistry 34*, 7728−7738.
19. Moore, M. J., and Sharp, P. A. (1992) *Science 256*, 992−997.
20. Chamberlin, S. I., and Weeks, K. M. (2000) *J. Am. Chem. Soc. 122*, 216−224.
21. Webb, A. E., Rose, M. A., Westhof, E., and Weeks, K. M. (2001) *J. Mol. Biol. 309*, 1087−1100.
22. Kebbekus, P., Draper, D. E., and Hagerman, P. J. (1995) *Biochemistry 34*, 4354−4357.
23. Steger, G., Muller, H., and Riesner, D. (1980) *Biochim. Biophys. Acta 606*, 274−284.
24. Chamberlin, S. I., and Weeks, K. M. (2003) *Biochemistry 42*, 901−909.
25. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
26. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1998) *Science 282*, 259−264.
27. Rangan, P., Masquida, B., Westhof, E., and Woodson, S. A. (2003) *Proc. Natl. Acad. Sci. U.S.A 100*, 1574−1579.
28. Latham, J. A., and Cech, T. R. (1989) *Science 245*, 276−282.
29. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
30. Hill, J., McGraw, P., and Tzagoloff, A. (1985) *J. Biol. Chem. 260*, 3235−3238.
31. Leontis, N. B., and Westhof, E. (1998) *J. Mol. Biol. 283*, 571−583.

BI035476K