Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure

Supplementary Material

Importance of inclusion of weak reflections in extension of data to high resolutions

As discussed elsewhere (Wang and Boisevert, 2003), inclusion of weak high-resolution data in structure refinement provides strong experimental constraints so that calculated amplitudes for those reflections from properly refined models will also be kept relatively small. With such refinement, individual B-factors for atomic coordinates can be more reliably and accurately refined than models built and refined without using these data. With accurately refined coordinates and individual B-factors, other modeling errors can be readily revealed and corrected using residual difference Fourier maps. However, the residual difference Fourier maps could become uninformative when systematic errors were massive distributed throughout the entire structure.

A question was also raised as to how inclusion of weak high-resolution data actually contributed to the improvement of experimental electron density maps. These reflections do not directly contribute to these maps in Fourier synthesis per se, because their amplitudes are often vanishingly small even if their phases may have been determined accurately. We believe, however, that they contribute indirectly through a solvent model during density modification. After the Fourier inversion of modified experimental densities, the calculated amplitudes for those reflections should also be kept relatively small if a solvent model was accurate and the density modification was successful. Otherwise, one would see inconsistencies between the observed and the Fourier-inverted data. By including these data in scaling between the calculated and

observed amplitudes, we indirectly constrained the solvent model and modified densities. By including weak high-resolution data beyond 3.1 Å during density modifications, the quality of experimental phases for reflections in all resolution ranges was improved. This was particularly noticeable for medium and low resolution reflections whose phases combined anomalous signals from derivative data contributed to increase peak heights for known heavy atom sites in anomalous difference Fourier maps. Thus, inclusion of certain amount of experimental observed near-zero-intensity reflections during density modification helps to improve experimental electron density maps.

Another question was how many the observed near-zero-intensity reflections should be included in density modification and in structure determination, or how high should the resolution be extended. In a previous study (Wang and Boisvert, 2003), we have analyzed the detailed distribution of reflections in term of F/σF at the highest resolution shell where their averaged <F/σF>=1.16. The distribution was asymmetric around the mean value in which the majority reflections had values near the mean value or slightly below it; and with increasing <F/σF> ratios, the number of reflections decreased exponentially. In this study, we reprocessed the native data again at 2.85 and 2.95 Å, respectively, and repeated density modification procedures. A comparison of all experimental maps after density modification showed that the quality of maps was nearly identical between the 2.85 and 2.80 Å resolutions, and that it was clearly better than the 2.90 Å resolution. If we use the F/σF=2 cut-off (or I/σI=1.0) as a criterion at the highest resolution shell, the resolution of the intron structure is 2.95 Å; and if we use the F/σF=1 or I/σI=0.5 cut-off criterion, it is 2.85 Å. An initial structure refinement resulted in an intron model with a free R-factor of 23.5% at 2.90 Å resolution where <I/σI>=0.82 or

<F/σF>=1.64 at the highest resolution shell. The free R-factor was further improved to 22.7% when refinement extended to include all data at 2.80 Å resolution. Thus, these refinement statistics fully support the validity of resolution extension. Further attempts to include even higher resolution data beyond 2.8 Å were not successful. Thus, weak high-resolution data with <F/σF>=0.76 or <I/σI>=0.38 have important structural information in them and they should be included in structure determination (Table 2).

In addition to accurate atomic coordinates and temperature B factor for the group II intron structure, inclusion of high-resolution weak data and sharpening of amplitudes revealed many ordered water molecules, which were invisible in the early stage of refinement when free R-factor was about 31%. Although single water molecule was insignificant, many of them have accumulatively helped improve refinement statistics. Amplitude-sharpening was an effective method to reveal the missing ordered water molecules in the structure, which often had relatively high temperature B-factors.

When no modifications were made to existing models, amplitude-sharpening could result in poor unweighted refinement statistics, because it also proportionally increased measurement errors in the data. When amplitude-sharpening helped to reveal errors in models or new ordered water molecules, it could result in better refinement statistics. A special amplitude-sharpening procedure has recently been developed to demodulate the amplitudes that are affected by lattice-translocation defects for selected indexes in the reciprocal space (Wang *et al*., 2005). After demodulation sharpening, overlapping structures are resolved, and each can be properly refined. Amplitude-sharpening would have little or no effects on weighted refinement statistics when no modifications were made to existing models.

Having demonstrated that weak reflections are important in structure determination, we want to emphasize that poorly measured weak reflections can become problems for locating initial heavy atom sites. For an initial determination of heavy atom sites, the highest quality of data is often more important than the highest resolution. Interestingly, high-redundant heavy atom derivative data could make certain heavy atom sites to be smeared out through improper averaging of radiation damaged data, which sites may be readily located in initial determination of heavy atom structures.

Finally, we recommend that the publication of macromolecular crystal structures should provide detailed crystallographic tables as online supplementary materials, which should include statistics of the highest-resolution data used for structure determination as well as statistics of data that were truncated back to the resolution shell with $F/\sigma F=2$ or $I/\sigma I=2$ for the definition of resolutions of the structures.

References

Wang, J., & Boisvert, D. C. (2003). *J. Mol. Biol.* **327**, 843-855.

Wang, J., Kamtekar, S., Berman, A. J., & Steitz, T. A. (2005). *Acta Cryst*. **D61**, 67-74.

Figures

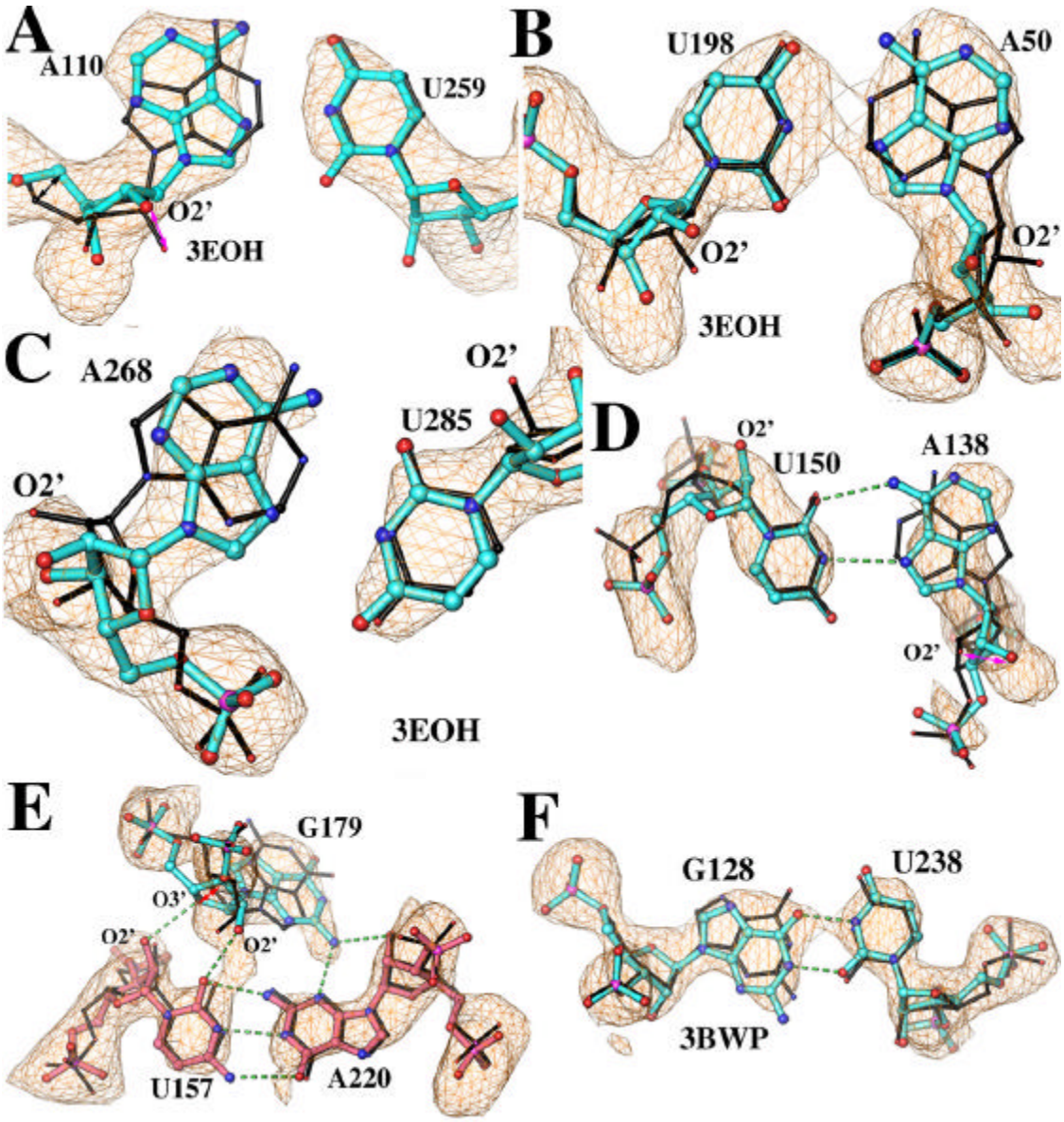

Figure S1. Hoogsteen and wobble base pairs. (A-D) Our new experimental maps (3G78) as well as the original maps (3BWP, not shown) have clearly shown that A-U Hoogsteen base pairs, whereas they were built as Watson-Crick like base pairs in all previous models (3BWP). (E) Hoogsteen face of G179 involves in a base triplet. (F) Our experimental maps (3G78) as well as the original experimental maps (3BWP, not shown)

have also clearly suggested that G and U form wobble base pairs. Our revised models (3G78) are in cyan, and the original or previously revised models (3BWP, 3EOH or 3EOG) are in gray. Maps were contoured at 1 (golden).

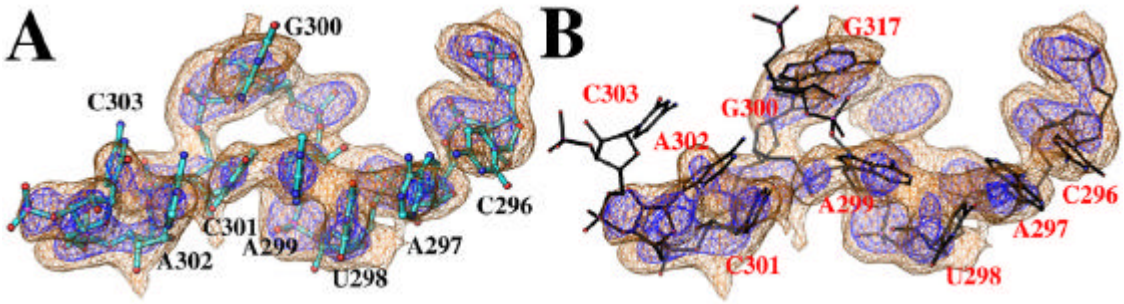

Figure S2. A single-nucleotide bulge at G300 of DIII. (A) Our new experimental maps (3G78) superimposed onto our model (cyan). (B) Our maps (3G78) superimposed onto the original model (3BWP, black). Maps were contoured at 1 (golden) and 3 (blue) σ.

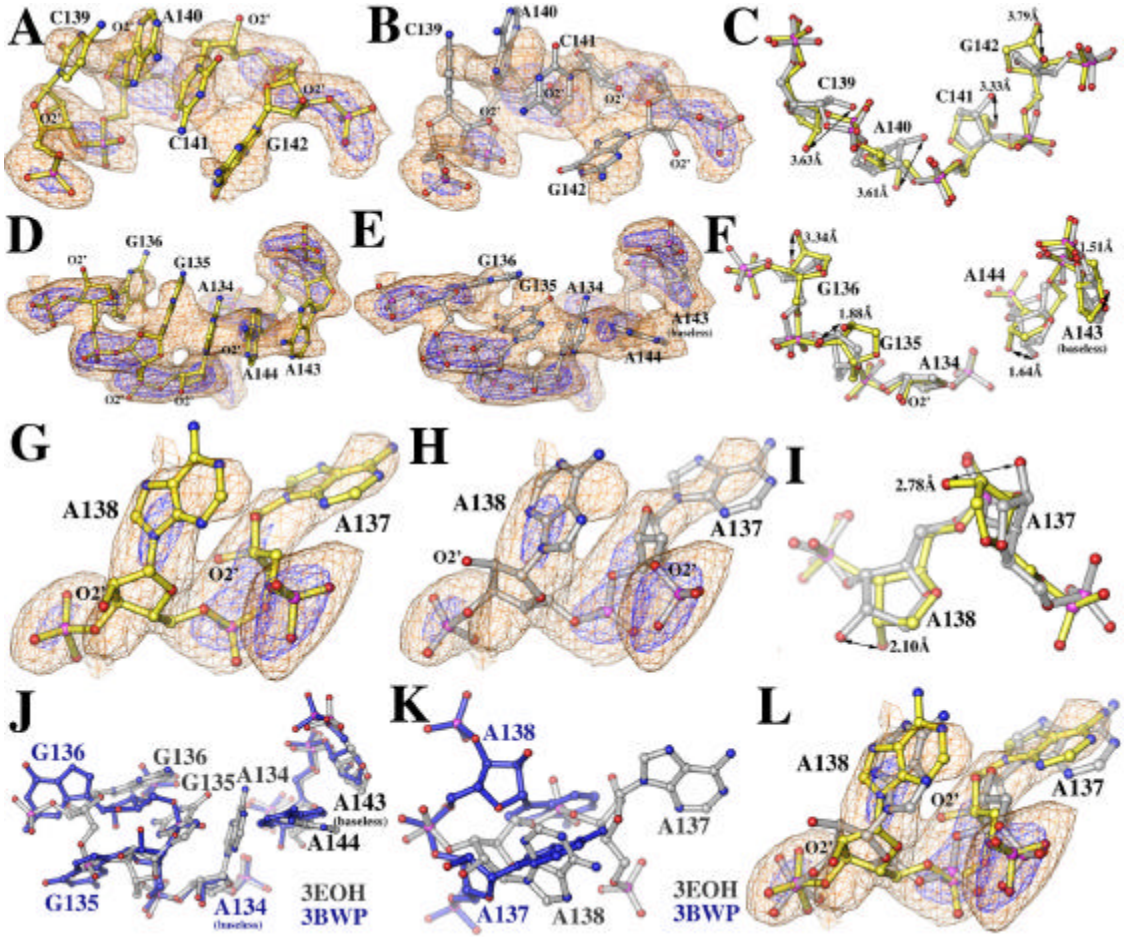

Figure S3. The revised k **region.** (A, D, G) Our new experimental maps (3G78) superimposed onto our model. (B, E, H) Our experimental maps (3G78) superimposed onto the previously revised model (3EOH). (C, F, I) Comparison of our model (yellow, 3G78) with the previously revised model (gray, 3EOH) with bases omitted. (J, K) Comparison of the original (blue, 3BWP) with the previously revised (gray, 3EOH) models. The original model (3BWP) had two "baseless" nucleotides A134 and A143, and the previously revised model (3EOH) had only one. (L) Our experimental maps (3G78) superimposed onto our model (yellow, 3G78) and their revised model (gray, 3EOH). Maps were contoured at 1 (golden) and 3 (blue) σ.

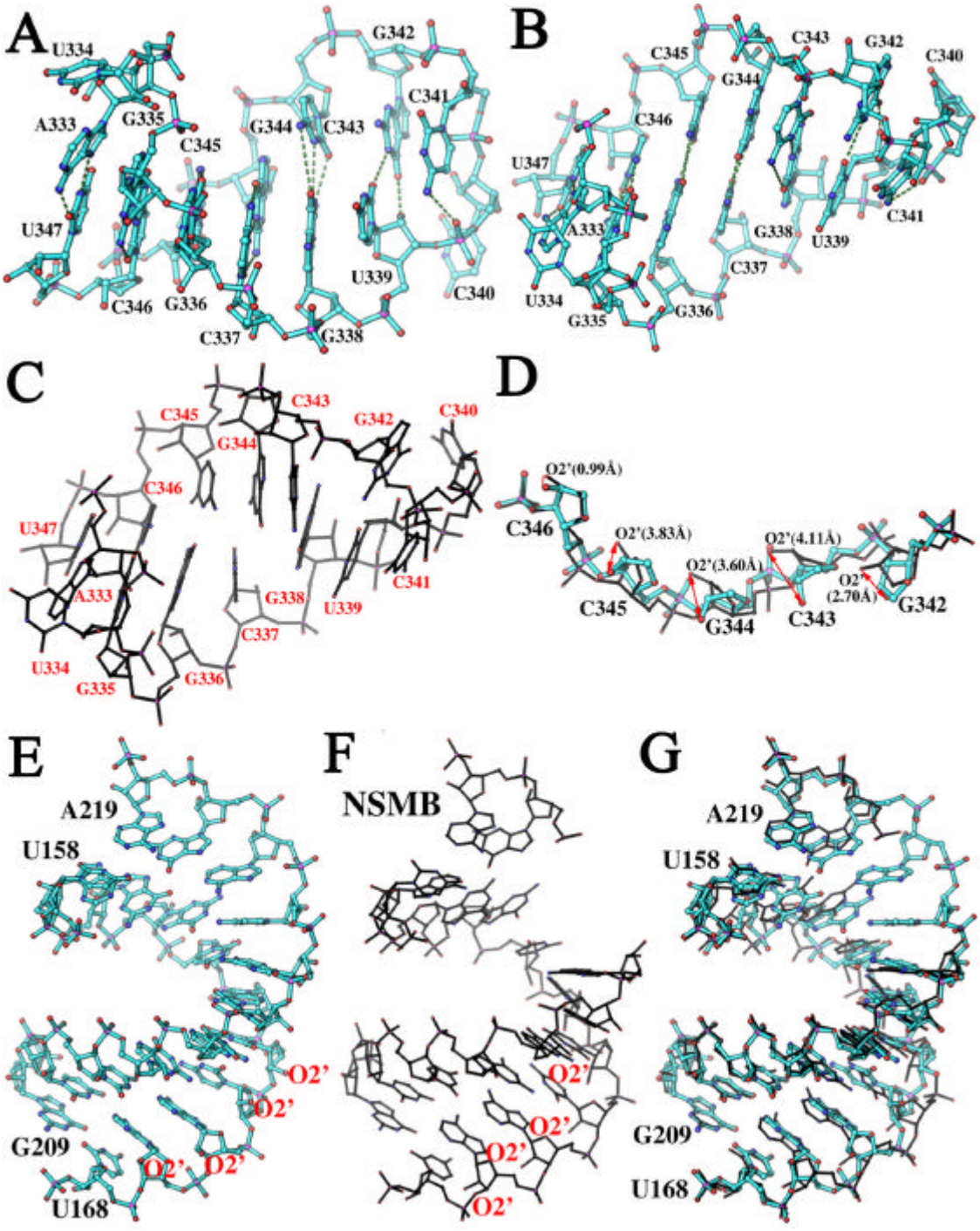

Figure S4. A-form RNA duplexes. (A-B) Two orthogonal views of our idealized DIV RNA duplex (3G78) flanked by the A333-U347 Hoogsteen base pair and a UCCG tetraloop. (C) The original model (3BWP). (D) Comparison of backbones of our duplex

(cyan, 3G78) with the original model (thin, black, 3BWP) with repositioned O2' indicated and bases omitted. (E) Our idealized ID2 duplex (3G78). (F) The original model (3BWP). (G) Comparison of our new model (cyan, 3G78) with the original model (thin, black, 3BWP).

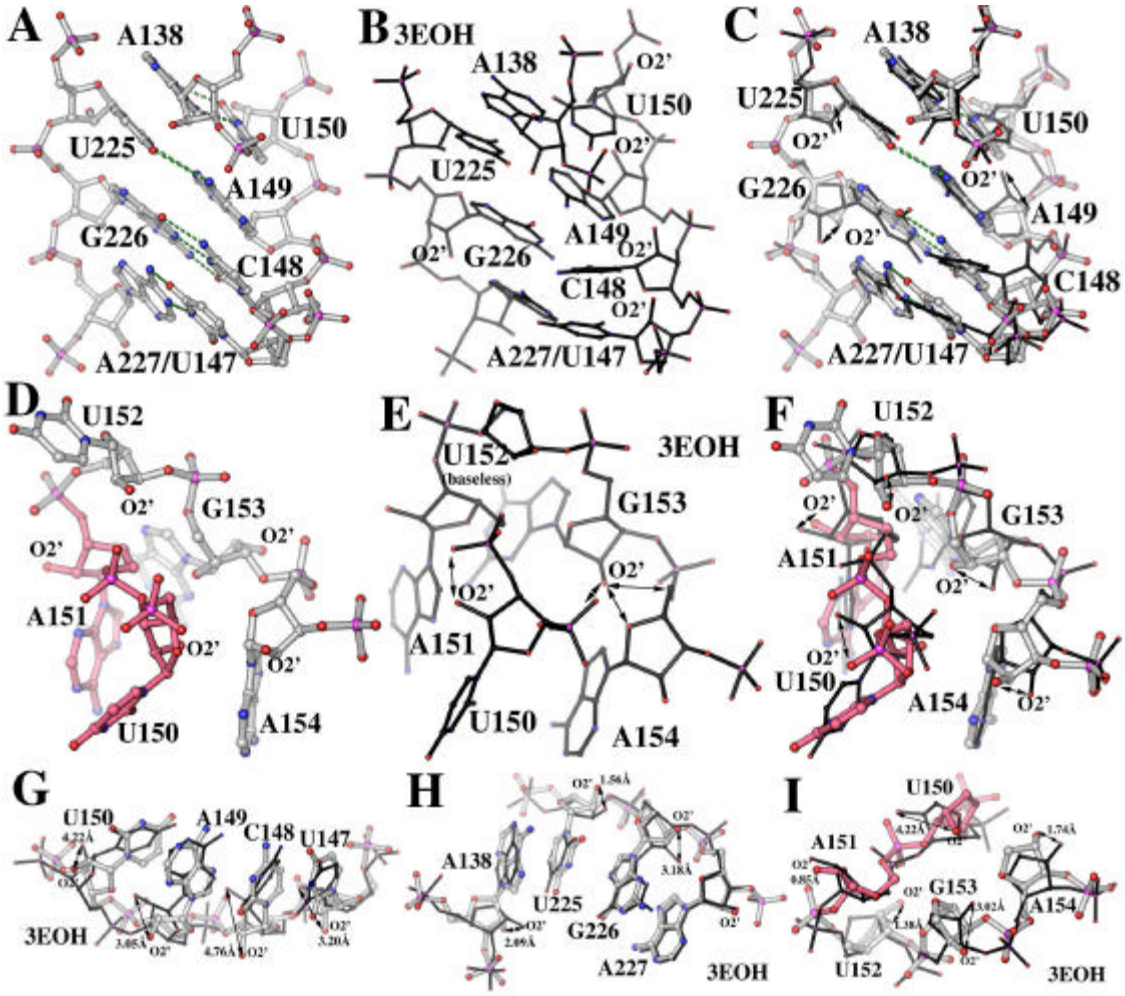

Figure S5. Coordination loop and ID1 duplex. (A-C, G, H) A RNA duplex stem next to the coordination loop. (D-F, I) Coordination loop. (A,D) Our new models (3G78). (B,E) The original models (3BWP or 3EOH). (C,F-I) Comparison of our model (thick, magenta and silver, 3G78) with the original models (thin, black, 3BWP or 3EOH). (I) Comparison of phosphate backbones with bases omitted.

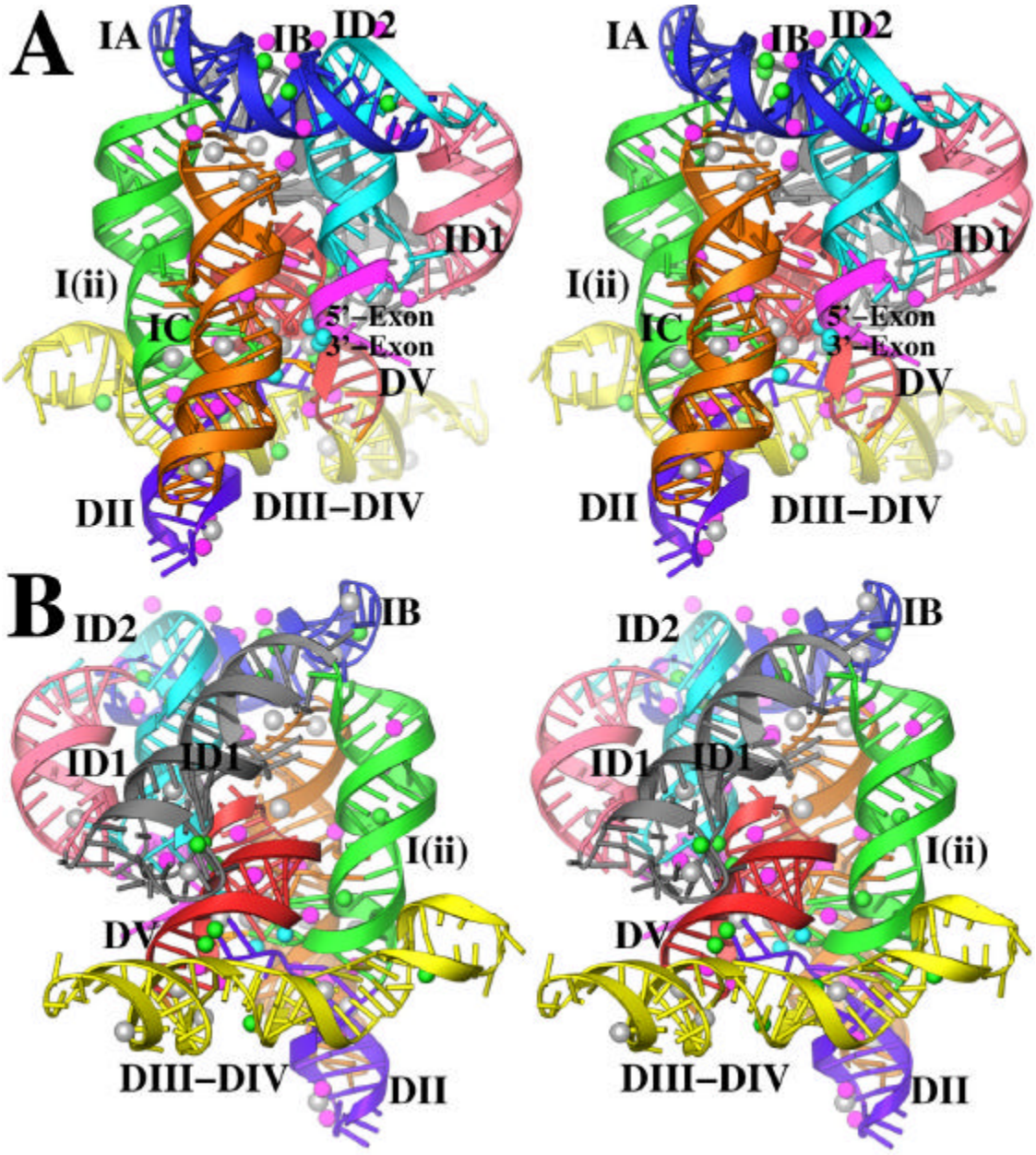

Figure S6. Metal ions in the revised intron structure. (A, B) Stereodiagram of our revised intron structure (3G78) in front and back views, respectively. Mg sites that are derived from Yb^{3+} and Ir^{3+} binding sites are shown in cyan and green spheres, respectively. Mg and K sites are shown in magenta and silver spheres, respectively.