Responses to Reviewers’ Comments

We thank the referees for their many constructive recommendations. Below we delineate changes that made in response.

1. Reviewer I:

We regret the oversight of comparisons with more comparable methods. In this new version of the paper, we present comparisons with other extant methods that both predict structure and align sequences using data sets from the literature in Figures 1 and 2. We included the comparison in the previous version in the supplement as the two new comparisons are more appropriate.

As we say in the conclusions of the paper, to date we have made no effort to improve the speed of RNAG. We believe that the algorithms improved ability to predict structures, the principled approach taken using Liu’s theorem that permits an effective exploration of the full posterior space, and the added ability of this sampling algorithm to characterize the full posterior spaces justify publication.

1. Reviewer II:

We regret that our previous description of the SAM riboswitch example was not adequately clear. We have modified this section to indicate that RNAG does predict the Xray structure of the unbound, SAM-off, form well as do other methods, but RNAG and previous efforts to predict the bound form have not been successful.

We thank the referee for the constructive suggestion to be more specific about the reference structures. In the supplement we now give a table that indicates the method used for each reference structure in the 17 families in the Kiryu data set. We added a few sentences to section 4.3 to briefly describe the data set of Kiryu, which shows that thirteen of the structures in this data set were determined by Xray crystallography or NMR, and only two by covariation analysis. As noted by Kiryu et.al. (2007), the assembly of a good comparison data set from the Rfam data base is a challenge. We believe it would be folly to build a larger data set without the involvement of individuals who are very familiar with the Rfam database such as those working at Janelia Farms. But we agree with the comment that using only 17 families, especially when they aren’t a random sample, is limiting. Thus we have added to the discussion to point out the limitations in existing comparison data sets. We are always skeptical for comparisons that depend on the authors of new methods to select comparison data sets and perform analysis with others’ software that often requires some tuning. Thus, we have specifically chosen three comparisons sets directly from the literature, and compared RNAG to the published results for these data sets. In this way we seek to avoid self-serving selection and biased application of others’ methods. As now pointed out in the discussion, in this implementation we did very little tuning. As now indicated in the discussion we used the default options and default parameters of each of the algorithms we employed. Also, we used exclusively the Kiryu et.al. (2007) data set to choose among the small number of available algorithms we used for this implementation of RNAG. Thus, as stated in the discussion the MASTR and the BRAliBASE II data sets are test data sets.

1. Reviewer III:

The major problems about the comparison with other algorithms and the Rfam dataset are given above and we here to answer the uncovered questions. We appreciate suggestions to improve the clarity of our presentation, and we have changed our submission to address each of the questions on clarification as follows:

1. We added text in the abstract to indicate that RNAG conducts global structural alignments.
2. We define RNA secondary structure as a binary matrix, with aij=1 if position i and position j in the sequence are paired and defined the ensemble as the set of all feasible such matrices in order to clarify the words “high-dimensional space of structures”.
3. As indicated in the response to referee II, there are many parameters in the covariance model, which are described in the Infernal package and associated publications. As pointed out in our submission estimates for these are obtained using the EM algorithm in Infernal using default settings. Also as indicated in the submission RNAalifold has only one free parameter, which is set at its default value in our analysis. We added the requested flow diagram of the algorithm to the supplement. We would not object to moving it to the text if the referees and/or the editor see this as appropriate.
4. We added sentences to the introduction describing the similarities of our approach to that taken by the related works of Eddy and Durbin (1994) and Yao el .al (2006) in CMfindiner.
5. We have modified the description of the separation index to clarify the definitions of Hamming distance and credibility limits.
6. Following the referee’s constructive suggestion, we re-did the experiment in section 3.1.2 on the group of the sequences with less than 60% identity and that more than 60%. As reported in the revision of this section we found that the sequences with less similarity will gain more from the increase of number of sequences in the alignment. Detailed of results with the two groups are now in the supplementary material, and summarized in the text, and percent identify is now given in Table 2.
7. RNAG records the consensus structure and multiple alignment sampled at each iteration separately. The consensus structure is given in .str format, which records the sampled structure at each iteration. The alignment is given in .aln format, which specifies the sampled alignment at each iteration . With these two files, we can get the projected structures for each sequence and do clustering analysis on the last 1000 samples. We have included an example output in the supplement, which specifies these formats.
8. The flow chart of the algorithm is included in the supplementary material. (Figure S4)
9. We found this referee’s other recommendations for clarity constructive and we made changes to capitalize on nearly all of these useful suggestions.

Also, we thank referees for pointing our minor changes and typos, which we corrected.

Yours, Sincerly

Donglai Wei, Charles Lawrence

2011.4.4