Responses to Reviewers’ Comments

1. Major Changes:
2. Reviewer I:

We regret the oversight of comparisons with more comparable methods. In this new version of the paper, we present comparisons with several 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 are sufficient to 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 references structures. We now delineate the evidence supporting each structure in the 17 families in the Kiryu data set. We added a few sentences to section 3.2 to briefly describe the data set of Kiryu, and pointed the readers to their paper for the details. As noted by Kiryu et.al. (2007), the assembly of a good comparison data set from the Ram 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.

1. Reviewer III:

We regret that our previous manuscript is not clear in many places. The major problems about the comparison with other algorithms and the Rfam dataset have been issued above and we here to answer the uncovered questions.

1. We add the notion that RNAG conducts global structural alignments in the introduction.
2. We define RNA secondary structure as a binary matrix, with aij=1 if position i and position j in the sequence are paired, which clarifies the notion of “high-dimensional space of structures”. However, we do not further elaborate on the parameter space which depends on different modeling approach, where our major interest lies in creating the framework to make use of various extant software.
3. CMfinder is a algorithm to finding an instances of an RNA motif given a structure, and thus differs from the align-fold algorithms described here. Our final goal is to build fully Bayesian RNA ab-initio motif finder, we expect to use the experience gained in CMfinder to achieve this end. We added a citation to CMfindiner.
4. We have modified the description of the separation index to clarify the definitions of Hamming distance and credibility limits.

5) Following the referee’s suggestion, we redo the experiment in section 3.1.2 on the group of the sequences with less than 60% identity and that more than 60%. We found that the sequences with less similarity will gain more from the increase of number of sequences in the alignment. Detailed description is added in the supplementary material. **6)** RNAG records the consensus structure and multiple alignment sampled at each iteration seperately. The structure is given in .str format, which specifies the iteration count followed by the consensus structure sampled at that point. The alignment is given in .aln format, which specifies the iteration count followed by the alignment sampled at that point. 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.

1. Clear definition:
2. Page 1: Most current algotithms ... : needs reference ->We delete the phrase
3. page 2: Numerous algorithms: ditto ->We delete the phraseThe discussion of Sankoff-like algorithms stops 2006. Clearly there are significant more recent Sankoff-like methods that need to be mentioned and compared with the presented approach. ->We add the citations of extant algorithms.page 2: right column. Define t for completeness. ->We add the definition.
4. The reference to RNAalifold is old. ->We add the lastest citation.
5. page 3:  In this discrete setting, the mean ... please define->We elaborate on it
6. Fig 3: please consider to visualize which point corresponds to which family ->We modify the figure 4 as suggested.
7. page 5: in recent publications: please provide references->We add the citation.
8. page 5: minimum departure ? ->We rephrase it as “close to”.
9. The caption text of Table 1 and 2 should be improvement to carefully state what's in the them.-> We elaborate on them
10. Typo Corrected:
11. Page 1, right col., line 40:  
    ``accuracy(MEA)'' => ``accuracy (MEA)''
12. Page 1, ``,et al.'', ``et.al.'' => ``et al.''
13. Page 1, left col., line 55: ``structure(S)'' => ``structure (S)''
14. Inferno => Infernal
15. Page 2, left col., line 30: ``Sankoff (1985), described'' => ``Sankoff (1985) described''
16. Page 2, left col., line 52: ``Geman [et.al](http://et.al" \t "_blank). (1984)'' => ``Geman and Geman (1984)''
17. Page 3, left col., line 1: ``structures ,and credibility'' => ````structures and credibility''
18. Page 3, left col., line 16: ``in (Hamada [et.al](http://et.al" \t "_blank).,2010), projecting'' => ``in Hamada et al. (2010), projecting''
19. Page 3, right col., after (1): Give ``D'' (Hamming distance) in italics
20. Page 3, section 3.1.2: Give ``N'' (number of sequences) in italics
21. page 3: because the => because of the
22. page 3:caparison => comparison?
23. page 3: We plot PPV (positive prediction value) => We plot positive prediction value (PPV) ... same for SEN
24. page 3: comparison the => comparison of the
25. page 4: caption Fig 1: SEN=...=TP(TP+TN) => SEN=...=TP(TP+FN)
26. page 4: Riboswitch is part => A riboswitch is part
27. page 4: that that => that
28. page 4: predicted for this ri   (hyphenation)
29. Page 4, left col., line 55: ``table 2'' => ``Table 2''
30. Page 4, right col., line 54:``reference that that'' => ``reference than that''
31. Title of section 3.2: ``Detecting Riboswitch:'' => ``Detecting the structure of a Riboswitch''
32. Page 6, References:  
    ''Ji Y, et al (2004)'' => ''Ji. Y,, Xu, X. and Stormo, G.D. (2004)''  
    ''Kiryu H, et al (2007)'' => ''Kiryu, H., Tabei, Y., Kin, T. and Asai, K. (2007)''
33. Header of Table 1: ''Table 1.Eeffects'' => ''Table 1. Effects''