

# Supplements: Predictions of RNA secondary structure by combining homologous sequence information

Michiaki Hamada<sup>abc\*</sup>; Kengo Sato<sup>dbe</sup>, Hisanori Kiryu<sup>b</sup>, Toutai Mituyama<sup>b</sup> and Kiyoshi Asai<sup>fb</sup>

<sup>a</sup>Mizuho Information & Research Institute, Inc,  
2-3 Kanda-Nishikicho, Chiyoda-ku, Tokyo 101-8443, Japan,

<sup>b</sup>Computational Biology Research Center,  
National Institute of Advanced Industrial Science and Technology (AIST),  
2-41-6, Aomi, Koto-ku, Tokyo 135-0064, Japan,

<sup>c</sup>Department of Computational Intelligence and System Science, Tokyo Institute of Technology,  
4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan,

<sup>d</sup>Japan Biological Informatics Consortium (JBIC),  
2-45 Aomi, Koto-ku, Tokyo 135-8073, Japan,

<sup>e</sup>Department of Biosciences and Informatics, Keio University,  
3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8522, Japan,

<sup>f</sup>Graduate School of Frontier Sciences, University of Tokyo,  
5-1-5 Kashiwanoha, Kashiwa 277-8562, Japan

---

\*To whom correspondence should be addressed. E-mail: [hamada-michiaki@aist.go.jp](mailto:hamada-michiaki@aist.go.jp)

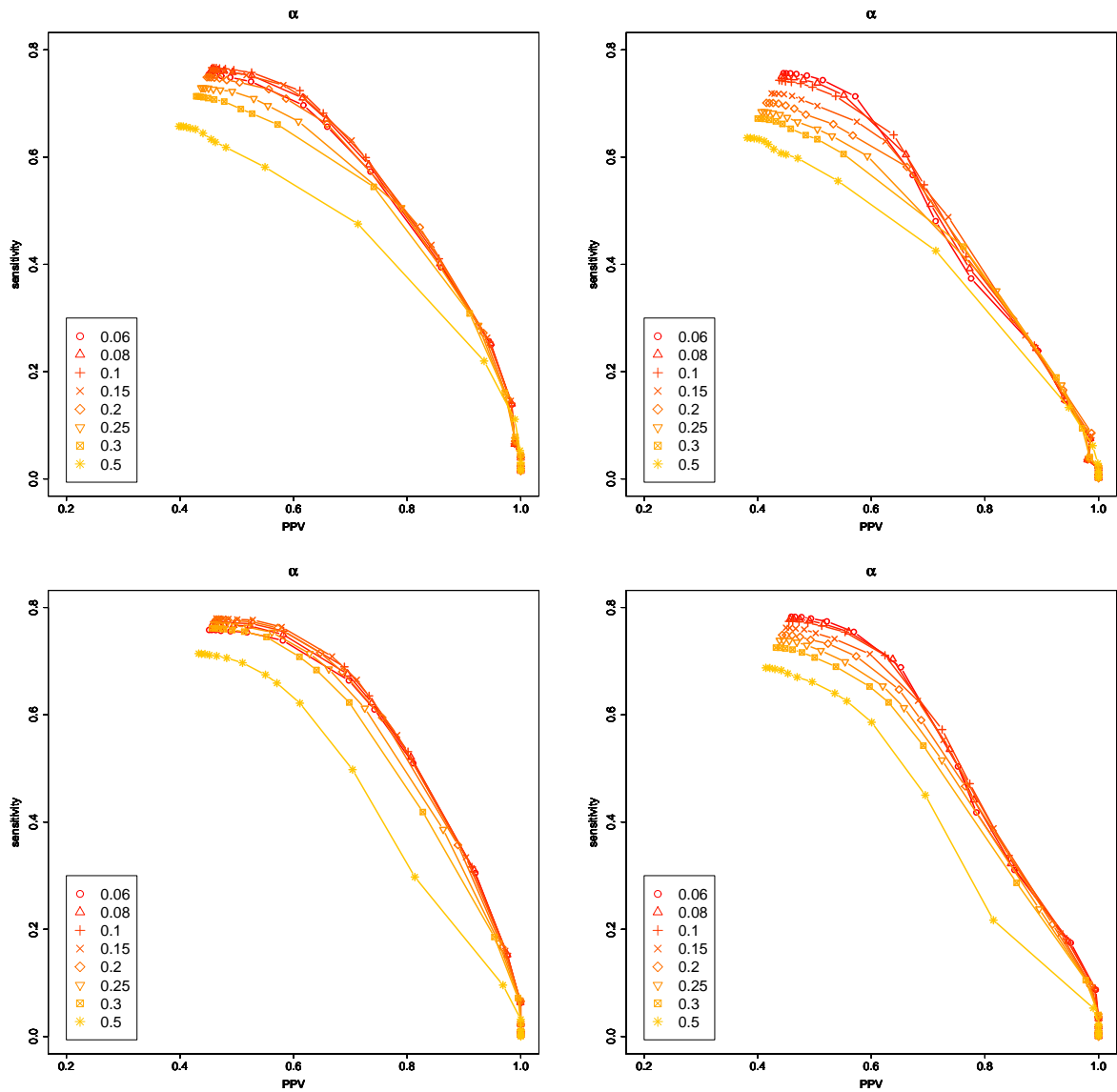


Figure S1: Performance of “appro3-1” for various  $\alpha$  parameter. We fixed  $\delta = 0$  in this experiments. We used the Probcons model [3] and the McCaskill model [6] (top, left), the Probalign model [7] and the McCaskill model (top, right), the Probcons model and the CONTRAfold model [4] (down, left) and the Probalign model ant the CONTRAfold model (down, right).

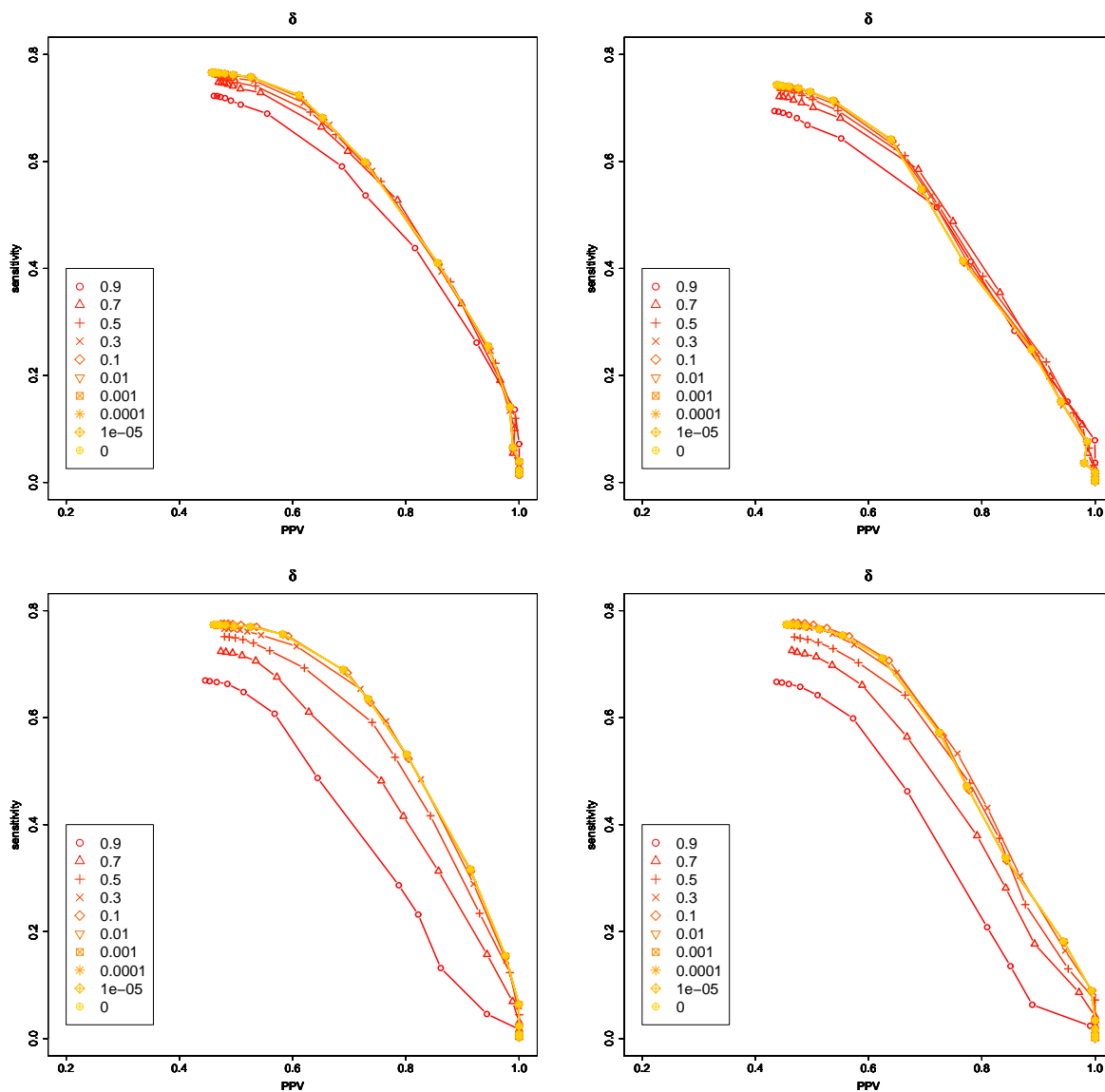


Figure S2: Performance of “appro3-1” for various  $\delta$  parameter. We fixed  $\alpha = 0.1$  in this experiments. We used the Probcons model [3] and the McCaskill model [6] (top, left), the Probalign model [7] and the McCaskill model (top, right), the Probcons model and the CONTRAfold model [4] (down, left) and the Probalign model ant the CONTRAfold model (down, right).

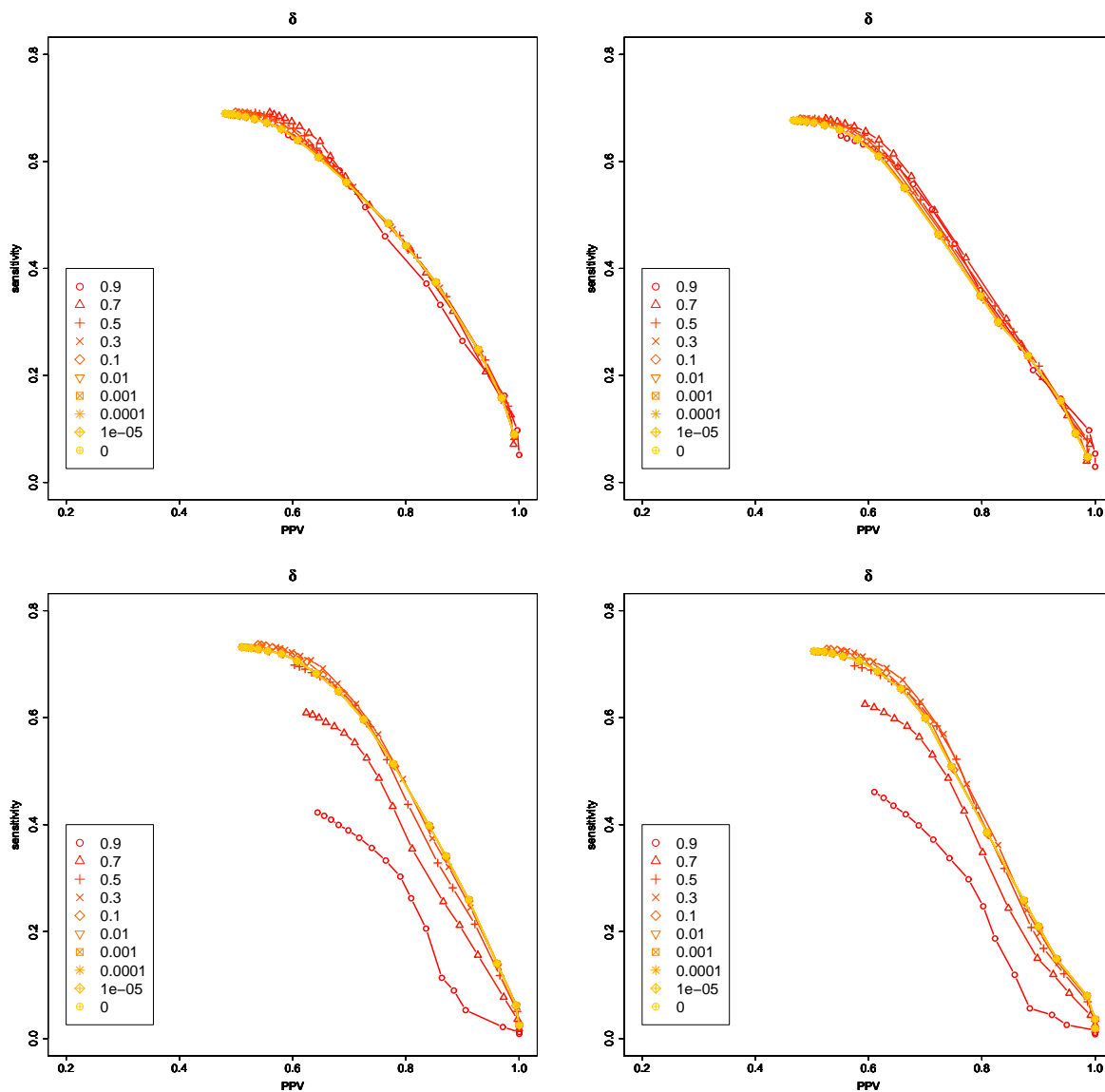


Figure S3: Performance of “appro3-2” for various  $\delta$  parameter. We fixed  $\alpha = 0.1$  in this experiments. We used the Probcons model [3] and the McCaskill model [6] (top, left), the Probalign model [7] and the McCaskill model (top, right), the Probcons model and the CONTRAfold model [4] (down, left) and the Probalign model ant the CONTRAfold model (down, right).

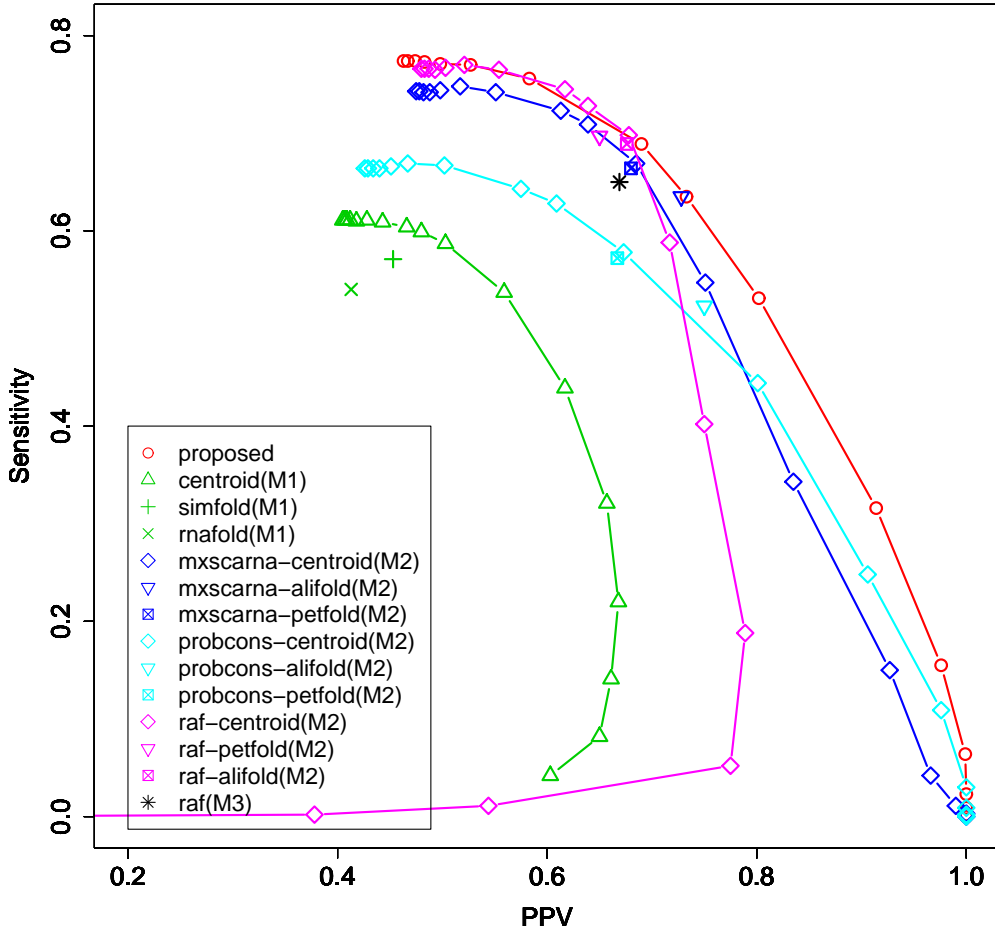


Figure S4: Results of the experiments using dataset1. “proposed” means “appro3-1 with CONTRAfold model and ProbCons model for  $\alpha = \frac{1}{1+|D|}$  and  $\delta = 0.01$ ”. “X-Y (M2)” means the method that predict common secondary structure by X after aligning  $\{x\} \cup D$  by Y. “raf (M3)” means the method that directly predict the common secondary structure by using RAF [2]. “centroid” means the  $\gamma$ -centroid estimators [5] for the single structure prediction and the averaged  $\gamma$ -centroid estimators [5] for the common secondary structure prediction, respectively.

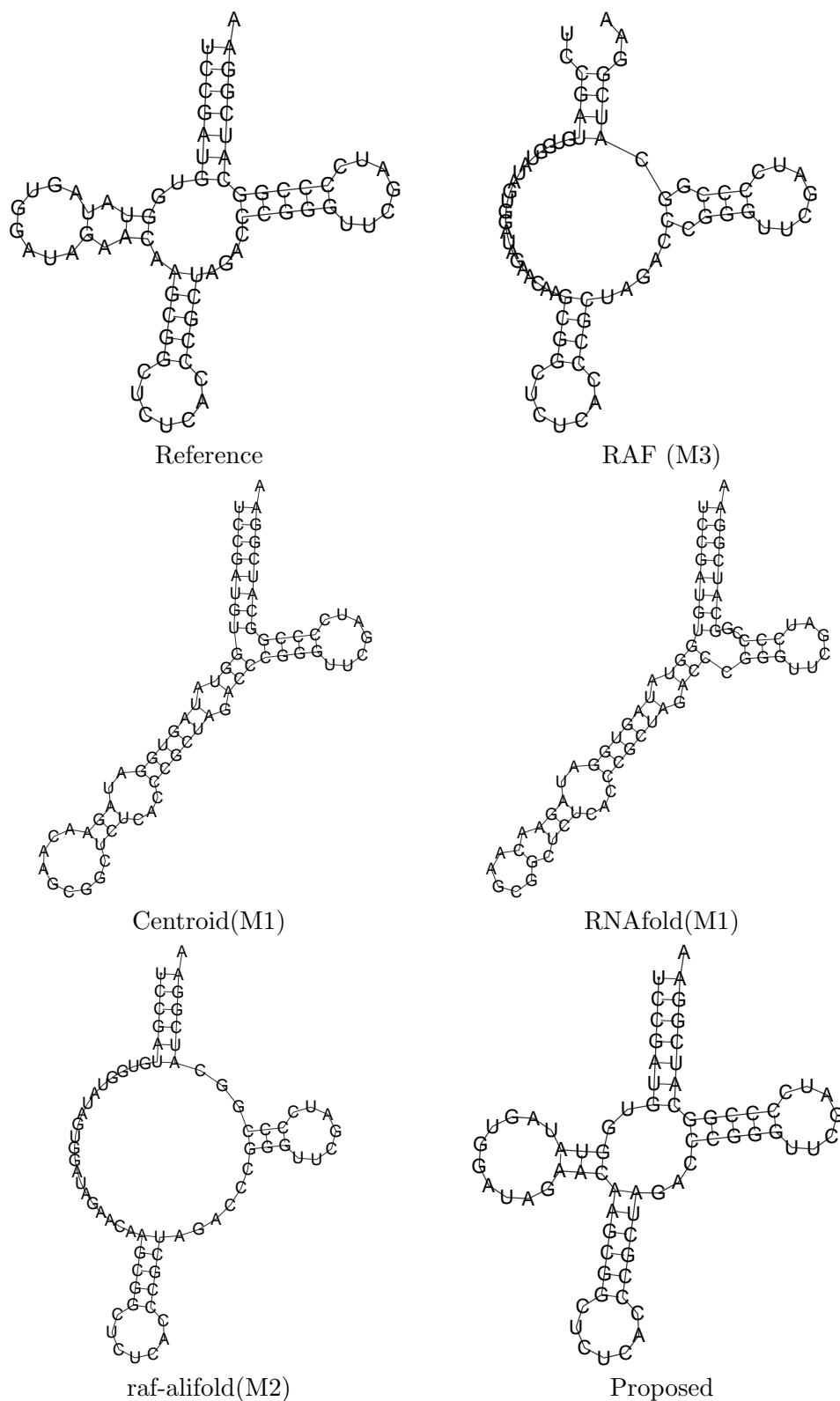


Figure S5: An example of predicted secondary structures of tRNA (ID:SPR\_00633) in the RNA STRAND database [1]. We used eight homologous sequences containing four unusual tRNA sequences (SPR\_00397,00629,00832,00938) whose secondary structures lack the D-domain stem-loop.

## References

- [1] M. Andronescu, V. Berez, H. Hoos, and A. Condon. RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinformatics*, 9:340, 2008.
- [2] C. Do, C. Foo, and S. Batzoglou. A max-margin model for efficient simultaneous alignment and folding of RNA sequences. *Bioinformatics*, 24:i68–76, Jul 2008.
- [3] C. Do, M. Mahabhashyam, M. Brudno, and S. Batzoglou. ProbCons: Probabilistic consistency-based multiple sequence alignment. *Genome Res.*, 15:330–340, Feb 2005.
- [4] C. Do, D. Woods, and S. Batzoglou. CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, 22:e90–98, Jul 2006.
- [5] M. Hamada, H. Kiryu, K. Sato, T. Kin, and K. Asai. Prediction of RNA secondary structure using generalized centroid estimators. *Bioinformatics Advance Access published on December 18*, 2008.
- [6] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7):1105–1119, May 1990.
- [7] U. Roshan and D. Livesay. Probalign: multiple sequence alignment using partition function posterior probabilities. *Bioinformatics*, 22:2715–2721, Nov 2006.