

# An Efficient Algorithm for Upper Bound on the Partition Function of Nucleic Acids

Hamidreza Chitsaz  
 Department of Computer Science  
 Wayne State University  
 Detroit, MI 48202  
 chitsaz@wayne.edu

Elmirasadat Forouzmand  
 Department of Computer Science  
 Wayne State University  
 Detroit, MI 48202  
 elmira@wayne.edu

Gholamreza Haffari  
 Faculty of Information Technology  
 Monash University  
 Clayton, VIC 3800  
 Australia  
 reza@monash.edu

August 25, 2018

## Abstract

It has been shown that minimum free energy structure for RNAs and RNA-RNA interaction is often incorrect due to inaccuracies in the energy parameters and inherent limitations of the energy model. In contrast, ensemble based quantities such as melting temperature and equilibrium concentrations can be more reliably predicted. Even structure prediction by sampling from the ensemble and clustering those structures by Sfold [7] has proven to be more reliable than minimum free energy structure prediction. The main obstacle for ensemble based approaches is the computational complexity of the partition function and base pairing probabilities. For instance, the space complexity of the partition function for RNA-RNA interaction is  $O(n^4)$  and the time complexity is  $O(n^6)$  which are prohibitively large [4, 12]. Our goal in this paper is to give a fast algorithm, based on sparse folding, to calculate an upper bound on the partition function. Our work is based on the recent algorithm of Hazan and Jaakkola [10]. The space complexity of our algorithm is the same as that of sparse folding algorithms, and the time complexity of our algorithm is  $O(MFE(n)\ell)$  for single RNA and  $O(MFE(m, n)\ell)$  for RNA-RNA interaction in practice, in which  $MFE$  is the running time of sparse folding and  $\ell \leq n$  ( $\ell \leq n + m$ ) is a sequence dependent parameter.

## 1 Introduction

Since the turn of the millennium and the advent of high throughput biology in the post-genome era, startling discoveries have redefined the role of RNA as a key player in the cellular arena. The ribosome and spliceosome are essentially two major RNA machines that together with other structural RNAs such as microRNAs (miRNA), long intergenic non-coding RNAs (lincRNA), small bacterial non-coding RNAs, and many other categories of structural RNAs run the cell at an extent that is comparable to that of protein machinery. For instance, lincRNAs have been recently shown to play sophisticated regulatory roles in mammalian cells, and miRNAs play a significant role in

the development of cancer. These discoveries have put RNA together with proteins in the center of focus for research and therapeutic purposes, including personalized medicine. Humanity has just begun to unravel RNA’s complicated roles in living cells, and RNA is no longer considered a mere information medium from DNA to proteins, but it is rather in the center of attention in molecular and cellular biology research. RNA molecules often function through interaction with other RNAs. In the absence of high throughput experimental assays to observe RNA structure and RNA-RNA interactions, the problems of RNA structure prediction and RNA-RNA interaction prediction gain the highest priority in bioinformatics.

RNA structure and RNA-RNA interaction prediction have recently received significant attention. The majority of algorithms that have been developed predict the minimum free energy structure [1] or binding sites [21]. However, it is a well-known fact that minimum free energy structure is often incorrect due to inaccuracies in the energy parameters and inherent limitations of the energy model. On the other hand, it has been shown that thermodynamic quantities, such as melting temperature and equilibrium concentrations, that are derived from the partition function which captures the properties of the whole Boltzmann ensemble rather than those of the most likely structure, can be more reliably predicted [4]. Even structure prediction by sampling from the ensemble and clustering those structures by Sfold [7] has proven to be more reliable than minimum free energy structure prediction [5, 6].

The main obstacle for ensemble based approaches is the computational complexity of the partition function and base pairing probabilities. For instance, the space complexity of computing the partition function for RNA-RNA interaction is  $O(n^4)$  and the time complexity is  $O(n^6)$  which are prohibitively large [4, 12]. On the other hand, recent progress in sparse folding algorithms has provided fast algorithms for the prediction of the most likely (minimum free energy) structure [2, 20, 23]. Although the partition function cannot be calculated exactly using sparsification ideas, it may be approximated. Our goal in this paper is to give a fast algorithm, based on sparse folding, to calculate an upper bound on the partition function. Our work is based on the recent algorithm of Hazan and Jaakkola [10].

## 2 Related Work

Methods to *approximate* the partition function for interacting RNAs have *not* been proposed in the literature. Instead, methods for *exact* computation of the partition function have been developed, having high both time and space complexity. Most notably, [4] developed an  $O(n^6)$ -time and  $O(n^4)$ -space dynamic programming algorithm that computes the partition function of RNA-RNA interaction complexes, thereby providing detailed insights into their thermodynamic properties. [13] has developed a *sampling* algorithm that produces a Boltzmann weighted ensemble of RNA-RNA interaction structures for the calculation of *interaction probabilities* (and not the partition function) for any given interval on the target RNAs.

In the context of *single* RNA secondary structure prediction, [16] devised a Metropolis Monte Carlo algorithm, called “Wang and Landau” algorithm [27], to approximate the partition function as well as density of states. Although the computation of the partition function over all secondary structures and over all pseudoknot-free hybridizations can be done by efficient dynamic programming algorithms, the real advantage of [16] is in approximating the partition function where pseudoknotted structures are allowed; a context known to be NP-complete [17].

In the machine learning community, there has been extensive research on obtaining *non-deterministic* and *deterministic* approximations of the log-partition function. Firstly, sampling and Monte Carlo methods (e.g. Gibbs Sampling and Monte Carlo Markov Chain) have been devel-

oped as non-deterministic approaches for estimating the partition function (cf. [15] and references therein). In high dimensions, obtaining independent samples from a given distribution is difficult since the mixing time is typically exponential in the size of the problem. Therefore, these methods are computationally very demanding, and in practice, they are rarely applied to large-scale problems. Secondly, variational techniques have been extensively developed as a deterministic approach to efficiently estimate the partition function in large-scale problems. In this approach, a simpler distribution is optimized as an approximation to the *true* distribution in a KL-divergence sense. However, the difficulty of this approach comes from: (i) Non-convexity of the set of feasible distributions (e.g. in “mean field” approximation [14]), and/or (ii) Hardness of computing the entropy embedded in the KL objective. Variational upper bounds on the other hand are convex, usually derived by replacing the entropy term in the KL objective with a simpler surrogate function and relaxing constraints on sufficient statistics hence convexifying the set of feasible distributions [26].

The basis of our work is [10] which provides a framework for approximating and bounding the partition function using MAP<sup>1</sup> inference (i.e. prediction of the most likely structure) on randomly perturbed models. Particularly, they propose to estimate the partition function as the max-statistics of collections of random variables, which is a major topic in *extreme value statistics* (e.g. see [3]). More broadly, there is an emerging body of work on perturbation methods, showing the benefits of explicitly adding noise into the modeling, learning, and inference pipelines [22, 25].

### 3 Preliminaries

#### 3.1 Notation

The input nucleic acid sequences are denoted by  $\mathbf{R}$  and  $\mathbf{S}$  throughout this paper. Function  $L$  denotes the length of the input sequence, and  $\mathbf{R}$  is indexed from 1 to  $L(R)$ , and  $\mathbf{S}$  is indexed from 1 to  $L(S)$  both in 5' to 3' direction. We refer to the  $i^{\text{th}}$  nucleotide in  $\mathbf{R}$  and  $\mathbf{S}$  by  $i_R$  and  $i_S$  respectively. The subsequence from the  $i^{\text{th}}$  nucleotide to the  $j^{\text{th}}$  nucleotide in a strand is denoted by  $[i, j]$ . An intramolecular base pair between the nucleotides  $i$  and  $j$  in a strand is called an *arc* and denoted by a bullet  $i \bullet j$ . An intermolecular base pair between the nucleotides  $i_R$  and  $i_S$  is called a *bond* and denoted by a circle  $i_R \circ i_S$ . We denote an RNA (RNA-RNA interaction) secondary structure by  $\mathfrak{s}$ , which is mathematically a set of constituent base pairs (arcs and bonds). The collection of all such feasible structures is denoted by  $\mathfrak{S}$ .

Throughout this paper, we denote the partition function by

$$Q := \sum_{\mathfrak{s} \in \mathfrak{S}} e^{-\frac{G(\mathfrak{s})}{RT}}, \tag{1}$$

in which  $G(\mathfrak{s})$  is the free energy of  $\mathfrak{s}$ ,  $R$  is the gas constant, and  $T$  is temperature. For a more detailed presentation of the partition function see [19, 4]. We use the Turner energy model for single RNAs [18] and our energy model for RNA-RNA interaction [4].

In this paper, we consider only the canonical base pairing system, i.e. each nucleotide is Watson-Crick paired with at most one nucleotide. We also assume there are no pseudoknots in individual secondary structures of  $\mathbf{R}$  and  $\mathbf{S}$ , and there are no crossing bonds and zigzags between  $\mathbf{R}$  and  $\mathbf{S}$  [4]. However, an extension of our ideas to non-canonical base pairing systems [11] and pseudoknotted (crossing and zigzagged) structures [24] is straight forward. The key requirement for such an extension is the existence of a fast minimum free energy structure prediction algorithm that can incorporate per base-pair energy contributions for the considered class of structures.

---

<sup>1</sup>Maximum a posteriori

### 3.2 Energy Perturbations

Following the Hazan-Jaakkola's approach [10], let  $\{\gamma_{i_R \bullet j_R}\}$ ,  $\{\bar{\gamma}_{i_R \bullet j_R}\}$ ,  $\{\gamma_{i_S \bullet j_S}\}$ ,  $\{\bar{\gamma}_{i_S \bullet j_S}\}$ ,  $\{\gamma_{i_R o i_S}\}$ , and  $\{\bar{\gamma}_{i_R o i_S}\}$  be six families of independent and identically distributed (i.i.d.) random variables, which are energy perturbations corresponding to presence and absence of base pairs, following the Gumbel distribution whose cumulative distribution function is

$$F(x) := P[\gamma \leq x] = e^{-e^{-(x+C)}}. \quad (2)$$

Above,  $C$  is the Euler's constant, so that the mean of our Gumbel distribution defined in (2) is zero. For every structure  $\mathfrak{s} \in \mathfrak{S}$ , let the energy perturbation of a structure be

$$\gamma(\mathfrak{s}) = \sum_{i \bullet j \in \mathfrak{s}} \gamma_{i \bullet j} + \sum_{i \bullet j \notin \mathfrak{s}} \bar{\gamma}_{i \bullet j} \quad (3)$$

if  $\mathfrak{s}$  is single RNA structure and

$$\begin{aligned} \gamma(\mathfrak{s}) = & \sum_{i_R \bullet j_R \in \mathfrak{s}} \gamma_{i_R \bullet j_R} + \sum_{i_R \bullet j_R \notin \mathfrak{s}} \bar{\gamma}_{i_R \bullet j_R} + \sum_{i_S \bullet j_S \in \mathfrak{s}} \gamma_{i_S \bullet j_S} + \sum_{i_S \bullet j_S \notin \mathfrak{s}} \bar{\gamma}_{i_S \bullet j_S} \\ & + \sum_{i_R o i_S \in \mathfrak{s}} \gamma_{i_R o i_S} + \sum_{i_R o i_S \notin \mathfrak{s}} \bar{\gamma}_{i_R o i_S} \end{aligned} \quad (4)$$

if  $\mathfrak{s}$  is RNA-RNA interaction structure.

## 4 Upper Bound on the Partition Function

Corollary 1 in [10] states that

$$\begin{aligned} \log Q & \leq E_\gamma \left[ \max_{\mathfrak{s} \in \mathfrak{S}} \{-G(\mathfrak{s})/RT + \gamma(\mathfrak{s})\} \right] \\ & = -E_\gamma \left[ \min_{\mathfrak{s} \in \mathfrak{S}} \{G(\mathfrak{s})/RT - \gamma(\mathfrak{s})\} \right], \end{aligned} \quad (5)$$

in which  $E$  is the expectation with respect to  $\gamma$ 's. The perturbations  $\gamma$  include terms that depend on base pairs ( $\gamma$ .) and terms that depend on lack of base pairs ( $\bar{\gamma}$ .) To simplify the incorporation of  $\bar{\gamma}$  terms into the energy model, let

$$\lambda(\mathfrak{s}) = \sum_{i \bullet j \in \mathfrak{s}} (\gamma_{i \bullet j} - \bar{\gamma}_{i \bullet j}) \quad (6)$$

for single RNA structure and

$$\begin{aligned} \lambda(\mathfrak{s}) = & \sum_{i_R \bullet j_R \in \mathfrak{s}} (\gamma_{i_R \bullet j_R} - \bar{\gamma}_{i_R \bullet j_R}) + \sum_{i_S \bullet j_S \in \mathfrak{s}} (\gamma_{i_S \bullet j_S} - \bar{\gamma}_{i_S \bullet j_S}) \\ & + \sum_{i_R o i_S \in \mathfrak{s}} (\gamma_{i_R o i_S} - \bar{\gamma}_{i_R o i_S}) \end{aligned} \quad (7)$$

for RNA-RNA interaction structure. Since the difference between two random variables following the Gumbel distribution follows the logistic distribution, we can rewrite the single RNA  $\lambda(\mathfrak{s})$  as

$$\lambda(\mathfrak{s}) = \sum_{i \bullet j \in \mathfrak{s}} \lambda_{i \bullet j} \quad (8)$$

and the RNA-RNA interaction one as

$$\lambda(\mathfrak{s}) = \sum_{i_R \bullet j_R \in \mathfrak{s}} \lambda_{i_R \bullet j_R} + \sum_{i_S \bullet j_S \in \mathfrak{s}} \lambda_{i_S \bullet j_S} + \sum_{i_R \circ i_S \in \mathfrak{s}} \lambda_{i_R \circ i_S}, \quad (9)$$

where  $\lambda$ 's are independent identically distributed random variables following the logistic distribution. In that case, (3) and (6) imply

$$\gamma(\mathfrak{s}) = \sum_{i \bullet j} \bar{\gamma}_{i \bullet j} + \lambda(\mathfrak{s}) \quad (10)$$

and (4) and (7) imply

$$\gamma(\mathfrak{s}) = \sum_{i_R \circ i_S} \bar{\gamma}_{i_R \circ i_S} + \sum_{i_R \bullet j_R} \bar{\gamma}_{i_R \bullet j_R} + \sum_{i_S \bullet j_S} \bar{\gamma}_{i_S \bullet j_S} + \lambda(\mathfrak{s}). \quad (11)$$

Since  $\sum_{i \bullet j} \bar{\gamma}_{i \bullet j}$  and  $\sum_{i \circ j} \bar{\gamma}_{i \circ j}$  are constants inside the minimization in (5), whose expectations are zero because of (2), we can rewrite (5) as follows where  $\lambda$  follows the logistic distribution:

$$\log Q \leq -E_\lambda \left[ \min_{\mathfrak{s} \in \mathfrak{S}} \{G(\mathfrak{s})/RT - \lambda(\mathfrak{s})\} \right]. \quad (12)$$

Our algorithm computes the right hand side of (12) and calculates the upper bound  $Q_{ub} = \exp(-E_\lambda [\min_{\mathfrak{s} \in \mathfrak{S}} \{G(\mathfrak{s})/RT - \lambda(\mathfrak{s})\}])$ . The minimization inside the expectation is essentially minimum free energy prediction, albeit with a perturbed energy. Recall that the energy perturbation  $-\lambda(\mathfrak{s})$  is the sum of individual base-pair perturbations for all base-pairs in  $\mathfrak{s}$ ; therefore, incorporation of such a perturbation in fast minimum free energy prediction algorithms, such as [2] which exploits sparsity, is straight forward. Particularly, we only need to add  $-\lambda_{i \bullet j}$  to the calculation of  $L^c(i, j)$  in [2] when  $i$  and  $j$  can form a base-pair. Additionally, the scaling of energy by  $RT$  in (12) has to be carefully applied to the sparse algorithm. Similarly, we only need to add  $-\lambda_{i_R \circ i_S}$  to the calculation of hybrid components in [23], in addition to proper handling of  $-\lambda_{i_R \bullet j_R}$  and  $-\lambda_{i_S \bullet j_S}$  in intramolecular base-pairings.

## 4.1 Complexity Analysis

Note that the perturbed energy is such that the triangle inequality in Property 1 of [2] still holds. Therefore, the running time of all sparse folding algorithms based on the triangle inequality [2, 20, 23] is not affected by our energy perturbation. To calculate the expectation above, we sample  $\lambda$ 's independently from the logistic distribution until the estimation of the expectation by simple averaging converges. Our experimental results show that the number of samples needed is much lower than the size of the input sequence.

For a single RNA with length  $n$ , the time complexity of our upper bound algorithm is  $O([n^2 + MFE(n)]\ell)$  in which  $\ell \leq n$  is the number of samples needed for the expectation estimation to converge, and  $MFE(n)$  is the running time of minimum free energy prediction. In this case, the space complexity is  $O(n^2 + MFES(n))$ , in which  $MFES(n)$  is the memory space needed for minimum free energy prediction.

For RNA-RNA interaction with lengths  $m$  and  $n$ , the time complexity of our algorithm is  $O([m^2 + n^2 + MFE(m, n)]\ell)$  in which  $\ell \leq n + m$  is the number of samples needed for the expectation estimation to converge, and  $MFE(m, n)$  is the running time of minimum free energy prediction. In this case, the space complexity is  $O(m^2 + n^2 + MFES(m, n))$ , in which  $MFES(m, n)$

is the memory space needed for minimum free energy prediction. Usually the running time and space complexity of our upper bound are dominated by those of minimum free energy prediction; therefore in practice, the time complexity of our upper bound is  $O(MFE(n)\ell)$  for single RNA and  $O(MFE(m, n)\ell)$  for RNA-RNA interaction, and its space complexity is often the same as that of minimum free energy prediction.

## 5 Results

We implemented the upper bound algorithm in our `piRNA` package [4]. To test the overestimation of partition function and also the number of samples needed for the algorithm to converge, we randomly selected totally 273,512 RNA sequences from Rfam 11.0 database [8, 9] and computed the upper bounds for them. Since the partition functions are often large numbers, we report the ensemble energies. The ensemble energy is proportional to the minus log of the partition function and an overestimation of the partition function becomes an underestimation of the ensemble energy. Fig. 1 depicts the results of our experiment. The top plot shows the histogram of ensemble energy underestimation calculated by  $-RT \log Q + RT \log Q_{ub}$ . The middle plot shows the histogram of ensemble energy underestimation percentage calculated from the ratio  $\log Q_{ub} / \log Q - 1$ . This plot shows that for a vast majority of cases this ratio is below 40%. The bottom plot depicts the distribution of ensemble energy  $-RT \log Q$  in the input dataset. Although this distribution exhibits multiple peaks, the middle distribution, which is the underestimation percentage, has a unimodal behaviour. Out of 273,512 RNAs, for 249,622 of them the underestimation is less than 50% of their ensemble energy, and for about half of the sequences (148,762) this difference is smaller than 30%. The number of sequences for which this difference is negative is 2,356 which is less than 1% of the total number of RNAs.

Fig. 2 shows the performance of our algorithm. The top plot is the histogram of the number of samples  $\ell$  (iterations) needed for the algorithm to converge. The middle and bottom plots show the histogram of the number of iterations per input size and per the log of input size. The vast majority of sequences required less than 15% of their length iterations to converge. The number of iterations starts with 7 and almost all the sequences need less than 40 samples. 243,119 or 89% of RNAs need not more than 20 iterations, and for more than half of the sequences (153,515) the experiment has been done with less than 15 iterations. Therefore at most 40 iterations are enough for different Rfam RNAs with different lengths. The number of iterations per length ratio for most of the sequences is less than 30%. For 90% of the RNAs in this dataset, this number is between 7% and 25%. Clearly the relation between length and the number of iterations is not linear, and upper bounds for different RNAs with the same length require different number of iterations.

Recall that the space complexity of our algorithm is the same as that of sparse minimum free energy prediction. Therefore, our algorithm is both fast and space efficient in practice.

## 6 Conclusion

We gave a fast algorithm, based on the recent algorithm of Hazan and Jaakkola [10], to iteratively compute an upper bound on the partition function of nucleic acids by perturbing energy. Our upper bound algorithm uses a fast minimum free energy prediction in each iteration. Our algorithm preserves the properties on which sparsification methods rely; therefore, we minimally modified sparse folding algorithms [2, 20, 23] to obtain the required fast minimum free energy prediction.

For the lower bound, one can trivially use the single term corresponding to the minimum free energy. The lower bound algorithm of Hazan and Jaakkola [10] requires modification to be

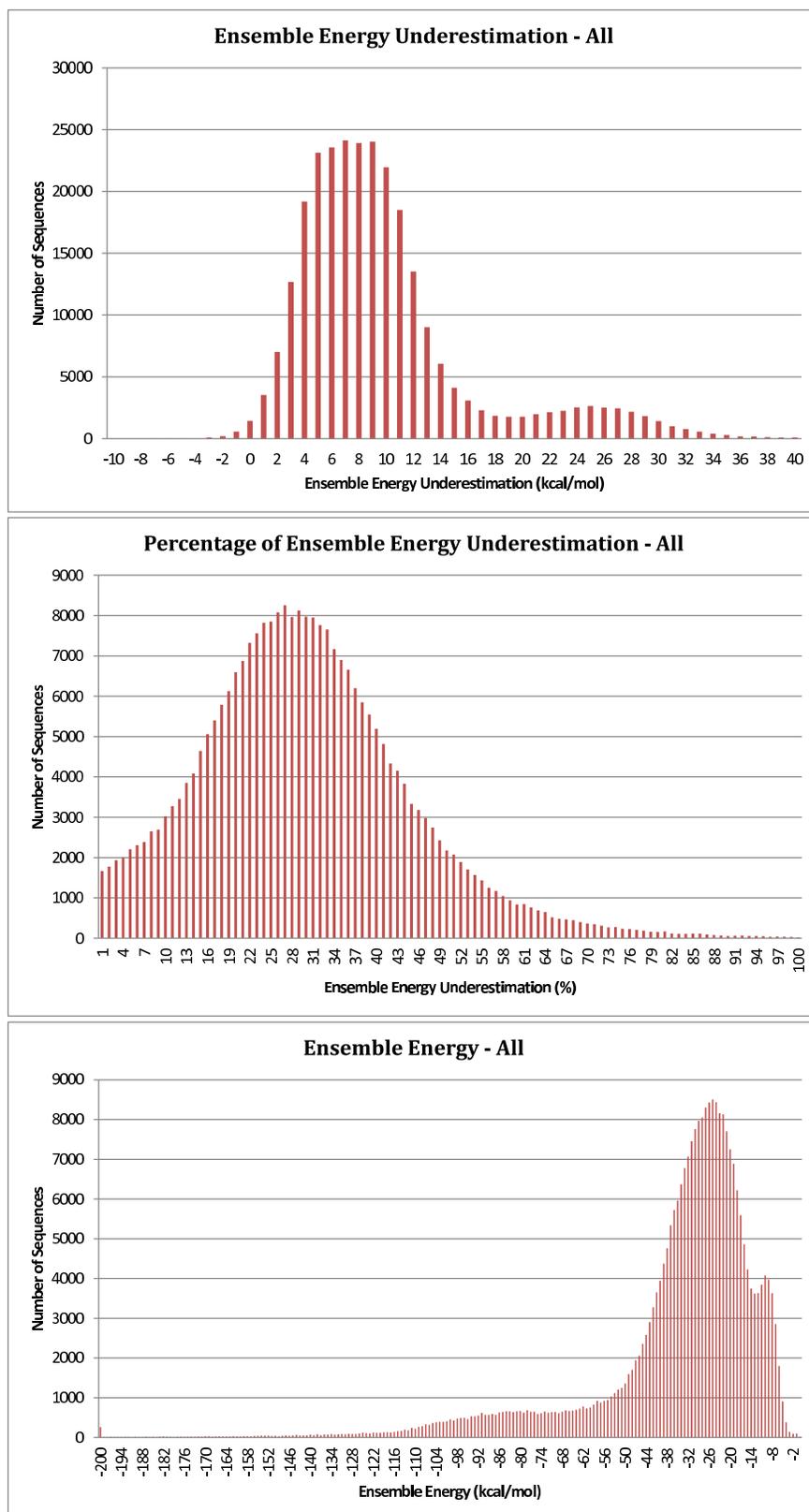


Figure 1: Histograms of (top) ensemble energy underestimation  $-RT \log Q + RT \log Q_{ub}$  for all 273,512 sequences in our dataset, (middle) percentage of ensemble energy underestimation  $\log Q_{ub} / \log Q - 1$ , and (bottom) ensemble energy  $-RT \log Q$  in the input dataset.

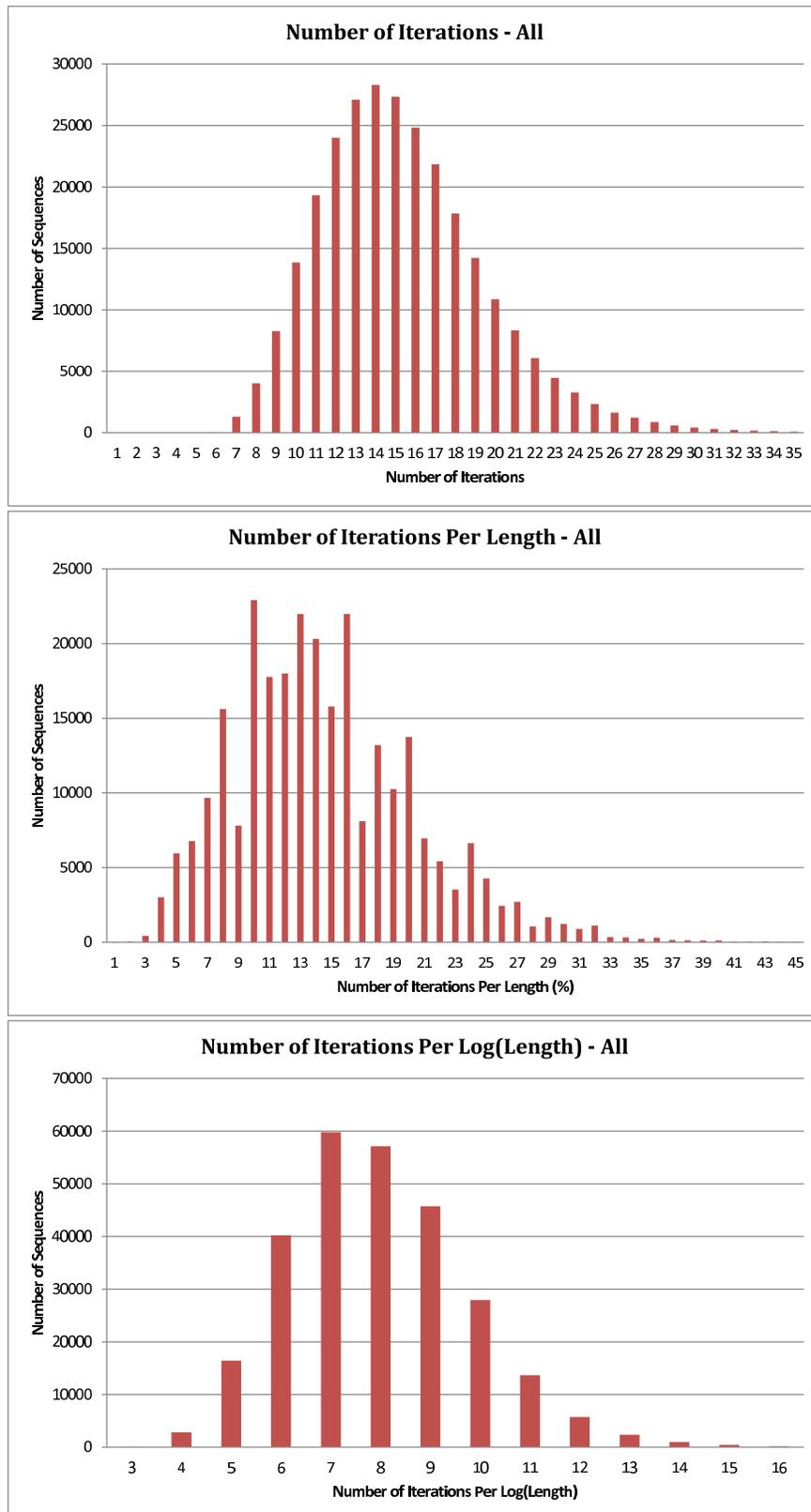


Figure 2: Histograms of (top) number of iterations  $\ell$  needed for all sequences in our dataset, (middle) number of iterations as a percentage of the input sequence length, and (bottom) number of iterations per the log of the input sequence length.

applicable to our problem. We leave such more accurate lower bound algorithms for future work.

## References

- [1] Alkan, C., Karakoc, E., Nadeau, J.H., Sahinalp, S.C., Zhang, K.: RNA-RNA interaction prediction and antisense RNA target search. *Journal of Computational Biology* **13**(2) (2006) 267–282
- [2] Backofen, R., Tsur, D., Zakov, S., Ziv-Ukelson, M.: Sparse RNA folding: Time and space efficient algorithms. *Journal of Discrete Algorithms* **9**(1) (2011) 12 – 31 20th Anniversary Edition of the Annual Symposium on Combinatorial Pattern Matching (CPM 2009).
- [3] Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics (2004)
- [4] Chitsaz, H., Salari, R., Sahinalp, S., Backofen, R.: A partition function algorithm for interacting nucleic acid strands. *Bioinformatics* **25**(12) (2009) i365–i373
- [5] Ding, Y., Chan, C.Y., Lawrence, C.E.: RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. *RNA* **11** (Aug 2005) 1157–1166
- [6] Ding, Y., Chan, C.Y., Lawrence, C.E.: Clustering of RNA secondary structures with application to messenger RNAs. *J. Mol. Biol.* **359** (Jun 2006) 554–571
- [7] Ding, Y., Lawrence, C.E.: A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.* **31** (Dec 2003) 7280–7301
- [8] Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., Bateman, A.: Rfam: Wikipedia, clans and the decimal release. *Nucleic Acids Research* **39**(suppl 1) (2011) D141–D145
- [9] Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., Eddy, S.R.: Rfam: an rna family database. *Nucleic Acids Research* **31**(1) (2003) 439–441
- [10] Hazan, T., Jaakkola, T.: On the partition function and random maximum a-posteriori perturbations. In: *Proceedings of the International Conference on Machine Learning. ICML* (2012)
- [11] Honer zu Siederdisen, C., Bernhart, S.H., Stadler, P.F., Hofacker, I.L.: A folding algorithm for extended RNA secondary structures. *Bioinformatics* **27**(13) (Jul 2011) i129–136
- [12] Huang, F.W.D., Qin, J., Reidys, C.M., Stadler, P.F.: Partition function and base pairing probabilities for RNA-RNA interaction prediction. *Bioinformatics* **25**(20) (2009) 2646–2654
- [13] Huang, F.W.D., Qin, J., Reidys, C.M., Stadler, P.F.: Target prediction and a statistical sampling algorithm for RNA-RNA interaction. *Bioinformatics* (2010) 175–181
- [14] Jordan, M.I., Ghahramani, Z., Jaakkola, T., Saul, L.K.: An introduction to variational methods for graphical models. *Machine Learning* **37**(2) (1999) 183–233
- [15] Koller, D., Friedman, N.: *Probabilistic Graphical Models: Principles and Techniques*. MIT Press (2009)

- [16] Lou, F., Clote, P.: Thermodynamics of RNA structures by Wang-Landau sampling. *Bioinformatics (ISMB 2010)* **26**(15) (Jun 2010) i278–i286
- [17] Lyngso, R., Pedersen, C.: RNA pseudoknot prediction in energy-based models. *Journal of Computational Biology* **7** (2000) 409–427
- [18] Mathews, D., Sabina, J., Zuker, M., Turner, D.: Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288** (May 1999) 911–940
- [19] McCaskill, J.: The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29** (1990) 1105–1119
- [20] Mohl, M., Salari, R., Will, S., Backofen, R., Sahinalp, S.C.: Sparsification of rna structure prediction including pseudoknots. *Algorithms for Molecular Biology* **5**(1) (2010) 39
- [21] Mückstein, U., Tafer, H., Bernhart, S.H., Hernandez-Rosales, M., Vogel, J., Stadler, P.F., Hofacker, I.L.: Translational control by RNA-RNA interaction: Improved computation of RNA-RNA binding thermodynamics. In Elloumi, M., Küng, J., Linial, M., Murphy, R.F., Schneider, K., Toma, C., eds.: *BIRD*. Volume 13 of *Communications in Computer and Information Science.*, Springer (2008) 114–127
- [22] Papandreou, G., Yuille, A.: Perturb-and-map random fields: Using discrete optimization to learn and sample from energy models. In: *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, Barcelona, Spain (November 2011) 193–200
- [23] Salari, R., Möhl, M., Will, S., Sahinalp, S., Backofen, R.: Time and space efficient RNA-RNA interaction prediction via sparse folding. In Berger, B., ed.: *Research in Computational Molecular Biology*. Volume 6044 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg (2010) 473–490
- [24] Sheikh, S., Backofen, R., Ponty, Y.: Impact of the energy model on the complexity of RNA folding with pseudoknots. In Karkkainen, J., Stoye, J., eds.: *CPM*. Volume 7354 of *Lecture Notes in Computer Science.*, Springer (2012) 321–333
- [25] Tarlow, D., Adams, R.P., Zemel, R.S.: Randomized optimum models for structured prediction. In: *Fifteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*. (2012)
- [26] Wainwright, M.J., Jaakkola, T., Willsky, A.S.: A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory* **51**(7) (2005) 2313–2335
- [27] Wang, F., Landau, D.P.: Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett* **86**(10) (March 2001) 2050–2053