J Stat Phys (2008) 132: 551–560
DOI 10.1007/s10955-008-9540-0

Large Deviations for Random Trees

Yuri Bakhtin · Christine Heitsch

Received: 13 December 2007 / Accepted: 3 April 2008 / Published online: 30 April 2008

Abstract We consider large random trees under Gibbs distributions and prove a Large Deviation Principle (LDP) for the distribution of degrees of vertices of the tree. The LDP rate function is given explicitly. An immediate consequence is a Law of Large Numbers for the distribution of vertex degrees in a large random tree. Our motivation for this study comes from the analysis of RNA secondary structures.

Keywords Random trees · Gibbs distributions · Large deviations · RNA secondary structure

1 Introduction

In this note, we prove a Large Deviation Principle (LDP) for two models of equilibrium statistical mechanics. In both cases, we consider a set of trees on N vertices and we define the Gibbs distribution associated to a certain energy function on that set. The main goal of our work is to study some typical features of large random trees ($N \to \infty$) under these distributions.

Here, we provide rigorous proofs for the LDP results announced in [1]. As discussed there, our results are motivated by, and have applications to, the branching of RNA secondary structures. An RNA molecule is a linear biochemical chain which folds into a three dimensional structure via a set of 2D base pairings known as a nested secondary structure. For an introduction to the mathematics of RNA secondary structures, see review articles such as [2, 16] or Chap. 13 in [15], Chap. 10 in [3]. RNA secondary structures are frequently represented as combinatorial objects, for instance as in [6–8, 13, 14]. Here, the trees we consider are a useful abstraction of these biological structures, as well as relatively straightforward to analyze mathematically. In this simplified model of RNA folding, explained more fully in [1], we can address the interplay between entropy and energy in determining a "typical" branching configuration. We find that, due to the entropy factor, the typical configurations in our model differ from the arrangements which have minimal energy in interesting ways.

Y. Bakhtin (✉) · C. Heitsch
School of Mathematics, Georgia Tech., Atlanta, GA 30332-0160, USA
e-mail: bakhtin@math.gatech.edu

Our mathematical results support and extend recent developments in RNA secondary structure prediction (reviewed in [9, 10]) which broaden the focus beyond simply finding a structure with minimal free energy. In particular, we prove a Law of Large Numbers for the degree frequencies in our large random trees, and find that the most common trees are not the minimizers of the associated energies. This highlights the limitations of prediction methods focused solely on energy minimization and the significance of entropy considerations in computational structural biology.

2 Models and Results

In this section we describe our models and state the results. The proofs are given in the next section.

2.1 Labeled Trees

In our first model we fix a natural number $D \geq 2$ and for each $N \in \mathbb{N}$ consider the set $\mathbb{T}_N(D)$ of labeled trees on $N \in \mathbb{N}$ vertices such that the degree of each vertex does not exceed D. See Fig. 1 for an example. To define Gibbs distributions on $\mathbb{T}_N(D)$ we need a function $c : \{1, \ldots, D\} \to \mathbb{R}$ which plays the role of the energy associated with the degree of a vertex.

To each of the trees T in $\mathbb{T}_N(D)$ we associate the energy

$$H(T) = \sum_{j=1}^{N} c(d_j(T)) = \sum_{k=1}^{D} c(k)\chi_k(T), \tag{1}$$

where $d_j(T)$ denotes the degree of the j-th vertex, and $\chi_k(T)$ is the number of vertices of degree k in T. Now the Gibbs probability measure on $\mathbb{T}_N(D)$ associated with H is given by

$$P_N\{T\} = \frac{e^{-\beta H(T)}}{Z_N}, \quad T \in \mathbb{T}_N(D),$$

where $\beta > 0$ is the inverse temperature parameter and

$$Z_N = \sum_{T \in \mathbb{T}_N} e^{-\beta H(T)} \tag{2}$$

is the partition function.

Our first result is an LDP for the degree distribution of random labeled trees under measures P_N introduced above.

Let us recall that a sequence of probability measures $(\mu_N)_{N\in\mathbb{N}}$ on a compact metric space (E, ρ) satisfies an LDP with a lower-semicontinuous nonnegative rate function $I : E \to \mathbb{R}$

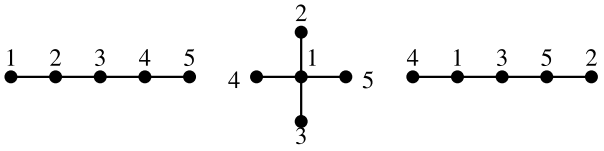

Fig. 1 Three distinct labeled trees. The trees are neither rooted nor ordered

if

$$\limsup_{N\to\infty} \frac{1}{N} \ln \mu_N(C) \le -I(C), \quad \text{for any closed set } C \subset E,$$

and

$$\liminf_{N\to\infty} \frac{1}{N} \ln \mu_N(O) \ge -I(O), \quad \text{for any open set } O \subset E,$$

where for $U \subset E$,

$$I(U) = \inf_{p\in U} I(p).$$

See [5, Sect. II.3] or [4, Sect. 1.2] for further details.

Informally, an LDP means that if we consider random variables X_N with distribution μ_N, then for all p and large N we have

$$\mu_N\{X_N \approx p\} \approx e^{-NI(p)}.$$

In particular, if the minimal value 0 is attained by I at a unique point p^* then for any neighborhood O of p^*, $\mu_N(O^c)$ decays exponentially in N. This can be restated as a Law of Large Numbers with exponential convergence in probability to the limit point p^*.

We can view $(\chi_1, \dots, \chi_D)$ as a random vector defined on the probability space $\mathbb{T}_N(D)$ equipped with the Gibbs measure P_N. We would like to study the frequencies of vertex degrees, so for each N we introduce a probability measure ν_N on $[0,1]^D$ defined as the distribution of the random vector $\frac{1}{N}(\chi_1, \dots, \chi_D)$ under P_N. It is natural to formulate an LDP for ν_N on the set

$$\mathcal{M} = \left\{ p \in [0,1]^D : \sum_{k=1}^{D} p_k = 1, \ \sum_{k=1}^{D} k p_k = 2 \right\}$$

equipped with Euclidean distance. Notice that $\mathcal{M}$ is nonempty if $D \ge 2$. It is a one-point set for $D = 2$, and, as an elementary calculation shows, it is a $(D-2)$-dimensional simplex that can be parametrized by $(p_3, p_4, \dots, p_D)$:

$$p_1 = \sum_{k=3}^{D} (k-2) p_k,$$

$$p_2 = 1 - \sum_{k=3}^{D} (k-1) p_k,$$

$$p_3, p_4, \dots, p_D \ge 0,$$

$$\sum_{k=3}^{D} (k-1) p_k \le 1,$$

so that it is natural to consider $\mathcal{M}$ as a subset of $(D-2)$-dimensional Euclidean space spanned by $p_3, p_4, \dots, p_D$.

Though the random vector $\frac{1}{N}(\chi_1, \dots, \chi_D)$ does not belong to $\mathcal{M}$, it is asymptotically close to $\mathcal{M}$:

$$\sum_{k=1}^{D} \frac{\chi_k}{N} = 1, \qquad \sum_{k=1}^{D} k \frac{\chi_k}{N} = 2 - \frac{2}{N}.$$

So instead of formulating an LDP for the sequence of random vectors $\frac{1}{N}(\chi_1, \ldots, \chi_D)$, we shall formulate and prove an LDP for a sequence of random vectors that is close to it and belongs to $\mathcal{M}$.

To define the rate function, we introduce $J : \mathcal{M} \to \mathbb{R}$ via

$$J(p) = -h(p) + \beta E(p) + G(p),$$

where

$$h(p) = -\sum_{k=1}^{D} p_k \ln p_k$$

is the entropy of the probability vector $p = (p_1, \ldots, p_D)$,

$$E(p) = \sum_{k=1}^{D} p_k c(k)$$

is the energy associated with p, and $G(p)$ is defined by

$$G(p) = \sum_{k=1}^{D} p_k \ln((k-1)!). \tag{3}$$

In Sect. 3, we shall see that the function G appears naturally in the analysis of random trees.

The function J is strictly convex down and continuous on $\mathcal{M}$. Therefore, it attains its minimal value at a uniquely defined point $p^* \in \mathcal{M}$. Since G and E are linear in p, and $h(p)$ has infinite normal derivative at the boundary of $\mathcal{M}$, considered as a subset of $(D-2)$-dimensional Euclidean space, p^* is an interior point of $\mathcal{M}$. Consider now

$$I(p) = J(p) - J(p^*). \tag{4}$$

It is easy to see that I is bounded, convex and continuous on $\mathcal{M}$.

For a measure Q on $[0,1]^D \times \mathcal{M}$ we define $Q^{(1)}$ and $Q^{(2)}$ as the marginal distributions of Q on $[0,1]^D$ and $\mathcal{M}$ respectively.

Theorem 1 *There is a sequence of probability measures $(Q_N)_{N\in\mathbb{N}}$ defined on $[0,1]^D \times \mathcal{M}$ with the following properties.*

1. *For each N, we have $Q_N^{(1)} = \nu_N$.*
2. *For each N,*

$$Q_N\left\{(x,y) \in [0,1]^D \times \mathcal{M} : \sum_{k=1}^{D} |x_k - y_k| > \frac{2}{N}\right\} = 0.$$

3. *The sequence $(Q_N^{(2)})_{N\in\mathbb{N}}$ satisfies an LDP on $\mathcal{M}$ with the rate function I defined in (4).*

Remark 1 This theorem says that although the random vector χ/N does not belong to $\mathcal{M}$, one can find another random vector that is, on the one hand, very close to χ/N and on the other hand belongs to $\mathcal{M}$ and satisfies the LDP.

Theorem 1 immediately implies the following Law of Large Numbers:

Corollary 1 *As* $N \to \infty$,

$$\left(\frac{\chi_1}{N}, \ldots, \frac{\chi_D}{N}\right) \to p^*$$

in probability.

Remark 2 The statements above show that with high probability the degree frequencies are close to p^*. Note that in most cases the minimum of the energy E on $\mathcal{M}$ is not attained at p^*.

2.2 Plane Trees

We now consider a similar model for plane trees (sometimes also called ordered trees). These are rooted trees such that subtrees at any vertex are linearly ordered, see e.g. [12]. See Fig. 2 for an example.

We choose the notation of this section to be parallel to that of the previous one, but we have to redefine it completely to adapt it to the plane tree model.

We fix a number $D \in \mathbb{N}$ and for each $N \in \mathbb{N}$ let $\mathbb{T}_N(D)$ denote the set of ordered trees on $N \in \mathbb{N}$ vertices such that the branching (i.e. the number of children) at each vertex does not exceed D. Notice that for a nonroot vertex, its branching number and its degree differ by 1.

The energy of each vertex depends only on its branching and is given by a function $c : \{0, 1, \ldots, D\} \to \mathbb{R}$. With each tree $T \in \mathbb{T}_N(D)$ we associate the energy

$$H(T) = \sum_{k=0}^{D} c(k)\chi_k(T), \tag{5}$$

where $\chi_k(T)$ is now the number of vertices with k children in T. The Gibbs probability measure on $\mathbb{T}_N(D)$ associated with H is given by

$$P_N\{T\} = \frac{e^{-\beta H(T)}}{Z_N}, \quad T \in \mathbb{T}_N(D),$$

where $\beta > 0$ is the inverse temperature and Z_N is a normalizing constant.

For each N, we introduce a probability measure ν_N on $[0, 1]^{D+1}$ defined as the distribution of the random vector $\frac{1}{N}(\chi_0, \chi_1, \ldots, \chi_D)$ under P_N.

We redefine $\mathcal{M}$ to be

$$\mathcal{M} = \left\{ p \in [0, 1]^{D+1} : \sum_{k=0}^{D} p_k = 1, \ \sum_{k=0}^{D} k p_k = 1 \right\}.$$

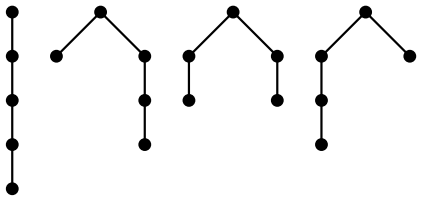

Fig. 2 Four different plane trees. The trees are rooted at the top and have linearly ordered subtrees, but are unlabeled

In this case, $\mathcal{M}$ can be viewed as a $(D-1)$-dimensional simplex in $(D-1)$-dimensional Euclidean space parametrized by $p_2, \dots, p_d$:

$$p_0 = \sum_{k=2}^{D} (k-1) p_k,$$

$$p_1 = 1 - \sum_{k=2}^{D} k p_k,$$

$$p_2, \dots, p_D \geq 0,$$

$$\sum_{k=2}^{D} k p_k \leq 1.$$

To formulate an LDP for this model we define $J : \mathcal{M} \to \mathbb{R}$ via

$$J(p) = -h(p) + \beta E(p),$$

where

$$h(p) = -\sum_{k=0}^{D} p_k \ln p_k$$

is the entropy of the probability vector $p = (p_0, p_1 \dots, p_D)$, and

$$E(p) = \sum_{k=0}^{D} p_k c(k)$$

is the energy associated with $p \in \mathcal{M}$.

As in the first model, the function J attains its minimum on $\mathcal{M}$ at a unique point inside $\mathcal{M}$ that we denote by p^*. Let

$$I(p) = J(p) - J(p^*). \tag{6}$$

This function will play the role of the rate function. Notice that in the case of plane trees it does not involve the function $G(p)$ that appeared in the construction of the rate function for the case of labeled trees.

For a measure Q on $[0,1]^{D+1} \times \mathcal{M}$ we define $Q^{(1)}$ and $Q^{(2)}$ as the marginal distributions of Q on $[0,1]^{D+1}$ and $\mathcal{M}$ respectively.

Theorem 2 *There is a sequence of probability measures $(Q_N)_{N\in\mathbb{N}}$ defined on $[0,1]^{D+1} \times \mathcal{M}$ with the following properties.*

1. *For each N, we have $Q_N^{(1)} = \nu_N$.*
2. *For each N,*

$$Q_N \left\{ (x, y) \in [0,1]^{D+1} \times \mathcal{M} : \sum_{k=0}^{D} |x_k - y_k| > \frac{1}{N} \right\} = 0.$$

3. *The sequence $(Q_N^{(2)})_{N\in\mathbb{N}}$ satisfies an LDP on $\mathcal{M}$ with the rate function I defined in (6).*

An immediate consequence is the following Law of Large Numbers:

Corollary 2 *As* $N \to \infty$,

$$\left(\frac{\chi_0}{N}, \frac{\chi_1}{N}, \ldots, \frac{\chi_D}{N}\right) \to p^*$$

in probability.

3 Proofs

We start with the proof of Theorem 1, adopting the notation and setting for labeled trees from Sect. 2.1.

The crucial fact for our analysis is the following formula for the number of trees on N vertices with degrees $d_1, \ldots, d_N$:

$$\binom{N-2}{d_1-1,\ d_2-1, \ldots, d_N-1}$$

if $d_1 + \ldots + d_N = 2N-2$, and 0 otherwise, see [11, Formula (2.1)]. Therefore, the total number of N-trees T with $\chi(T) = (n_1, \ldots, n_D)$ is given by

$$\begin{aligned}&\binom{N-2}{\underbrace{0,\ldots,0}_{n_1}, \underbrace{1,\ldots,1}_{n_2}, \ldots, \underbrace{D-1,\ldots,D-1}_{n_D}}\binom{N}{n_1,\ldots,n_D}\\ &\quad = \frac{(N-2)!}{(2!)^{n_3}\ldots((D-1)!)^{n_D}} C(N,n),\end{aligned}$$

where $C(N,n) = \binom{N}{n_1,\ldots,n_D}$. All these trees T have the same energy $H(T)$, so that

$$P_N\left\{\frac{\chi(T)}{N} = \frac{n}{N}\right\} = \frac{e^{-NF(\frac{n}{N})}C(N,n)}{Z_N}, \tag{7}$$

where Z_N is defined in (2), and we notice that

$$Z_N = \sum_{\substack{n_1+\ldots+n_D=N\\ n_1+\ldots+Dn_D=2N-2}} e^{-NF(\frac{n}{N})}C(N,n),$$

and

$$F(p) = \beta E(p) + G(p) = \beta\sum_{k=1}^{D} c(k)p_k + \sum_{k=1}^{D}\ln((k-1)!)p_k, \quad p \in [0,1]^D,$$

with $G(p)$ defined in (3).

Our plan is to use the LDP for multinomial distribution that manifests itself in coefficients $C(N,n)$ in the r.h.s. of (7), and then apply a version of Varadhan's lemma for Gibbs transformation via the exponential factor $e^{-NF(\frac{n}{N})}$.

We start with the family of distributions μ_N on $\mathcal{M}$ defined by

$$\mu_N\left\{\left(\frac{n_1}{N},\dots,\frac{n_D}{N}\right)\right\}=\begin{cases}\frac{C(N,n)}{Z'_N}, & \text{if } \left(\frac{n_1}{N},\dots,\frac{n_D}{N}\right)\in\mathcal{M},\\ 0, & \text{otherwise,}\end{cases}$$

where

$$Z'_N=\sum_{n/N\in\mathcal{M}} C(N,n).$$

Lemma 1 *The sequence of measures $(\mu_N)_{N\in\mathbb{N}}$ satisfies an LDP on $\mathcal{M}$ with rate function I_1 defined by*

$$I_1(p)=h^*-h(p),$$

where

$$h^*=\sup_{p\in\mathcal{M}} h(p).$$

Proof The proof of this lemma literally repeats that of Sanov's theorem (an LDP for the multinomial distribution, see [4, Theorem 2.1.10]). It is based on the formula:

$$\frac{1}{N}\ln C(N,n)=-\sum_{k=1}^{D}\frac{n_k}{N}\ln\frac{n_k}{N}+O\left(\frac{\ln N}{N}\right),\quad \text{as } N\to\infty,$$

which holds true uniformly in n, see e.g. [5, Lemma I.4.4].

Let us now introduce the Gibbsian weight

$$q_N\left(\frac{n}{N}\right)=e^{-NF\left(\frac{n}{N}\right)},$$

and a new family of measures λ_N on $\mathcal{M}$:

$$\lambda_N\left\{\frac{n}{N}\right\}=\frac{q_N\left(\frac{n}{N}\right)\mu_N\left\{\frac{n}{N}\right\}}{Z''_N},\quad \text{for } \frac{n}{N}\in\mathcal{M},$$

where

$$Z''_N=\sum_{\frac{n}{N}\in\mathcal{M}} q_N\left(\frac{n}{N}\right)\mu_N\left\{\frac{n}{N}\right\}=\int_{\mathcal{M}} e^{-NF(p)}\mu_N(dp).$$

In other words,

$$\lambda_N(dp)=\frac{e^{-NF(p)}\mu_N(dp)}{\int_{\mathcal{M}} e^{-NF(p)}\mu_N(dp)}.$$

Let us also denote $J_1(p)=F(p)+I_1(p)$ and $J_{1,*}=\inf_{p\in\mathcal{M}} J_1(p)$. □

Lemma 2 *The sequence of measures $(\lambda_N)_{N\in\mathbb{N}}$ satisfies an LDP on $\mathcal{M}$ with rate function I_2 given by $I_2(p)=J_1(p)-J_{1,*}$.*

Proof This lemma follows directly from a variant of Varadhan's lemma for Gibbs transformations (Theorem II.7.2 in [5]). □

Remark 3 Notice that $I_2(p) = I(p)$ for all $p \in \mathcal{M}$. So we have proven the desired LDP on $\mathcal{M}$ for $(\lambda_N)_{N\in\mathbb{N}}$, and in order to prove Theorem 1 we shall have to compare λ_N to ν_N.

Proof of Theorem 1 We consider the distribution P_N on $\mathbb{T}_N(D)$, so that $\frac{\chi}{N}$ is distributed according to ν_N. For each x that belongs to the support of ν_N we introduce the set

$$R(x) = \left\{ y \in \mathcal{M} : \ y_k = \frac{m_k}{N}, \ m_k \in \mathbb{Z}, k = 1, \ldots, D, \text{ and } \sum_{k=1}^{D} |x_k - y_k| = \frac{2}{N} \right\}.$$

It is easy to see that $1 \le |R(x)| \le D^2$ for all x, where $|R|$ denotes the number of elements in R. □

Let us now define the measure Q_N. We start with random variables χ/N, and define a random vector Y so that, given χ/N, the conditional distribution of Y is uniform on $R(\chi/N)$. Now Q_N denotes the joint distribution of χ/N and Y. Clearly, the first two desired properties of Q hold true by the definition of Q_N. The third one follows from Lemma 2 and the following statement claiming that measures $Q_N^{(2)}$ and λ_N differ by a subexponential factor, thus obeying an LDP with the same rate function:

Lemma 3 *There is a constant $C > 0$ such that for all N and all sets $U \subset \mathcal{M}$,*

$$\frac{1}{CN^4} \le \frac{Q_N^{(2)}(U)}{\lambda_N(U)} \le CN^4.$$

This lemma is a straightforward consequence of the following fact: there is a constant K such that if $|n_1 - n_1'| + \cdots + |n_D - n_D'| = 2$ then

$$\frac{1}{KN^2} \le \frac{e^{-NF(\frac{n}{N})} C(N,n)}{e^{-NF(\frac{n'}{N})} C(N,n')} \le KN^2.$$

The proof of Theorem 2 is essentially the same. It is based on the following expression for the number of ordered trees of order N with n_k nodes having k children:

$$\frac{1}{N} \binom{N}{n_0, \ n_1, \ n_2 \ldots}$$

if $n_1 + 2n_2 + \cdots = N - 1$, and 0 otherwise (see e.g. Theorem 5.3.10 in [12]).

Acknowledgements The authors thank the anonymous reviewers whose comments significantly improved the paper, and the ABC Math Program in the School of Mathematics at Georgia Tech for fostering interdisciplinary research linking mathematics with the biological sciences.

This research of Yuri Bakhtin is supported in part by NSF CAREER DMS-0742424.

This research of Christine E. Heitsch is supported in part by a Career Award at the Scientific Interface (CASI) from the Burroughs Wellcome Fund (BWF) and by NIH NIGMS 1R01GM083621-01.

References

1. Bakhtin, Y., Heitsch, C.: Large deviations for random trees and the branching of RNA secondary structures. Bull. Math. Biol., submitted online version available at http://arxiv.org/abs/0803.3990

2. Condon, A.: Problems on RNA secondary structure prediction and design. In: Automata, Languages and Programming. Lecture Notes in Comput. Sci., vol. 2719, pp. 22–32. Springer, Berlin (2003)
3. Durbin, R., Eddy, S.R., Krogh, A., Mitchison, G.: Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press, Cambridge (1998)
4. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, 2nd edn. Applications of Mathematics, New York, vol. 38. Springer, New York (1998)
5. Ellis, R.S.: Entropy, Large Deviations, and Statistical Mechanics. Classics in Mathematics. Springer, Berlin (2006). Reprint in the 1985 original
6. Gan, H.H., Pasquali, S., Schlick, T.: Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. Nucl. Acids Res. **31**(11), 2926–2943 (2003)
7. Heitsch, C.E., Condon, A., Hoos, H.H.: From RNA secondary structure to coding theory: A combinatorial approach. In: Ohuchi, A., Hagiya, M. (eds.) DNA8: Revised Papers from the 8th International Workshop on DNA Based Computers. Lecture Notes in Computer Science, vol. 2568, pp. 215–228. Springer, London (2003)
8. Le, S.-Y., Nussinov, R., Maizel, J.V.: Tree graphs of RNA secondary structures and their comparisons. Comput. Biomed. Res. **22**(5), 461–473 (1989)
9. Mathews, D.H.: Revolutions in RNA secondary structure prediction. J. Mol. Biol. **359**(3), 526–32 (2006)
10. Mathews, D.H., Turner, D.H.: Prediction of RNA secondary structure by free energy minimization. Curr. Opin. Struct. Biol. **16**(3), 270–278 (2006)
11. Moon, J.W.: Counting Labelled Trees. From Lectures Delivered to the Twelfth Biennial Seminar of the Canadian Mathematical Congress, Vancouver, vol. 1969. Canadian Mathematical Congress, Montreal (1970)
12. Stanley, R.P.: Enumerative Combinatorics, Vol. 2. Cambridge Studies in Advanced Mathematics, vol. 62. Cambridge University Press, Cambridge (1999). With a foreword by Gian-Carlo Rota and Appendix 1 by S. Fomin
13. Schmitt, W.R., Waterman, M.S.: Linear trees and RNA secondary structure. Discrete Appl. Math. **51**(3), 317–323 (1994)
14. Shapiro, B.A., Zhang, K.: Comparing multiple RNA secondary structures using tree comparisons. Comput. Appl. Biosci. **6**(4), 309–18 (1990)
15. Waterman, M.S.: Introduction to Computational Biology: Maps, Sequences and Genomes. Chapman & Hall/CRC, London/Boca Raton (1995)
16. Zuker, M.: RNA folding prediction: the continued need for interaction between biologists and mathematicians. In: Some Mathematical Questions in Biology—DNA Sequence Analysis, New York, 1984. Lectures Math. Life Sci., vol. 17, pp. 87–124. Amer. Math. Soc., Providence (1986)