RNA matrix models with external interactions and their asymptotic behavior

I. Garg and N. Deo*
Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India
(Received 20 August 2008; revised manuscript received 28 January 2009; published 2 June 2009)

We study a matrix model of RNA in which an external perturbation on n nucleotides is introduced in the action of the partition function of the polymer chain. The effect of the perturbation appears in the exponential generating function of the partition function as a factor $\exp(1-n\alpha/L)$ (where α is the ratio of strengths of the original to the perturbed term and L is the length of the chain). The asymptotic behavior of the genus distribution functions as a function of length for the matrix model with interaction is analyzed numerically for all $n \leq L$. It is found that as $n\alpha/L$ is increased from 0 to 1, the term 3^L in the number of diagrams $a'_{L,g,\alpha}$ at a fixed length L, genus g and α, goes to 2^L $[(3-\frac{n\alpha}{L})^L$ for any $n\alpha/L]$ and the total number of diagrams $\mathcal{N}'_\alpha$ at a fixed length L and α but independent of genus g, undergoes a change in the factor $\exp(\sqrt{L})$ to 1 $(\exp[(1-n\alpha/L)\sqrt{L}]$ for any $n\alpha/L)$. However the exponent L of the dominant length dependent term in $a'_{L,g,\alpha}$ stays unchanged. Hence the universality is robust to changes in the interaction (α). The distribution functions also exhibit unusual behavior at small lengths.

DOI: 10.1103/PhysRevE.79.061903 PACS number(s): 87.14.gn, 02.10.Yn, 11.10.Jj, 87.15.−v

I. INTRODUCTION

Understanding of the process of RNA folding finds its ultimate use in the prediction of the fully folded, partially folded, and completely unfolded structures under physiological conditions [1]. Under these conditions, unfolding is a very slow process as compared to folding in the presence of a force. Application of a force increases the unfolding rate and we can therefore get the unfolded structures from the folded ones ([1] and references therein). Experimental techniques of force-induced measurements have proved successful in probing properties related to different aspects of RNA folding and unfolding, domain unfolding in proteins, in polysaccharides and nucleic acids ([2] and references therein). Over the years, optical tweezers and atomic force microscopy techniques have been employed to study the physical, elastic, and structural properties of the biomolecules by recording their force extension curves and studying the force-dependent dynamics and folding landscapes of the molecules ([3–9] and references therein).

Mechanical unfolding and refolding of single RNA has been studied using force-ramp, hopping, and force-jump methods ([10] and references therein). In mechanical unfolding experiments, it has been observed that at a critical value of the applied force, the hairpin structure toggles between the folded and the unfolded states [11–13]. In these experiments, the ionic concentrations play an important role. Ions (monovalent and divalent cations), specific proteins, and ligands are known to affect the stability of RNA structures and govern the transitions from the intermediate (secondary) to native folded (tertiary) states. The effects of ion-RNA interaction depend upon the environment which varies with the size of ions and their distance from the RNA molecule [14–16]. Experiments of Bustamante *et al.* [12,13] have shown that the denaturation of RNA by a constant force involves multiple trajectories (for RNA hairpins and Tetrahymena thermophila ribozyme) while undergoing a transition from the folded structure state to the unfolded state. These trajectories depend on the point at which the force is applied [1,17]. This diverseness in the folding-unfolding pathways is due to the rugged energy landscape of RNA (consisting of many minima). In these experiments, the RNA molecule is usually clamped at one end and the force is applied at the other end of the polymer. Extensions in the molecule are studied with respect to a varying force. The external force in experiments may be introduced into the random matrix models of RNA [18,19] by adding complicated potentials to the action of the partition function. We consider here (theoretically) the case where a simple external perturbation acts on a fixed number of nucleotides. The elongation of RNA in the model is done by adding a base and we calculate some of the extension properties with respect to the external perturbation. Important theoretical studies have been carried out for RNA under tension [20,21] in other statistical models.

We study here the effect of an external perturbation on the random matrix model of RNA folding in [18,19]. The model in [18,19] makes a correspondence between the Feynman diagrams and the graphical representation of real RNA structures (both secondary and tertiary, originally noticed in [22] for secondary structures only). Random matrix theory provides an analytic method to take into account the tertiary structures in addition to the planar structures in a natural way. We discuss very briefly a generalization of the random matrix model of RNA with external interaction (a topological model) proposed in [23], where the external perturbation acts on a single nucleotide and on n nucleotides, $n \leq L$, L being the length of polymer chain. We will refer to these models as 1-NP RNA model and n-NP RNA model, respectively (NP is nucleotide perturbation).

We outline in Sec. II the extended matrix model (with interaction) of [23] for completeness and understanding and introduce the 1-NP and n-NP RNA models. Further we present a detailed numerical analysis of the asymptotics of the genus distribution functions for the extended matrix model of RNA with perturbation [23] in Sec. III. The genus distribution functions: the total number of diagrams at a fixed length L but independent of genus g, $\mathcal{N}$ and the number of

*ndeo@physics.du.ac.in

diagrams at a fixed length L and genus g, $a_{L,g}$ of the matrix model of RNA in [19] are found to change in the presence of an external perturbation. We extend the numerical asymptotic analysis to the n-NP RNA model as well.

II. EXTENDED MATRIX MODELS OF RNA

We review here the matrix model with a perturbation in the action of the partition function studied in [23]. The partition function is

$$Z_{L,\alpha}(N) = \frac{1}{A_L(N)} \int \prod_{i=1}^{L} d\phi_i \times \exp\left[-\frac{N}{2} \sum_{i,j=1}^{L} (V^{-1})_{i,j} \mathrm{Tr}\ \phi_i \phi_j \right] \times \exp\left[-N \sum_{i=1}^{L} (W^{-1})_i \mathrm{Tr}\ \phi_i \right] \frac{1}{N} \mathrm{Tr} \prod_{i=1}^{L} (1+\phi_i), \tag{1}$$

where ϕ_i are $i=1,\ldots,L$ independent $(N\times N)$ Hermitian matrices at each nucleotide position i in the chain, $\exp[-N\Sigma_{i=1}^{L}(W^{-1})_i \mathrm{Tr}\ \phi_i]$ is the perturbation term with W_i the strength of perturbation on nucleotide at position i and $V_{i,j}$ is the matrix element of a $(L\times L)$ symmetric matrix giving the interactions between the L nucleotides at positions i and j in the polymer chain (H bonds formed between any two nucleotides saturate). The observable $\Pi_i(1+\phi_i)$ is an ordered product over ϕ_i's which ensures that the diagonal elements do not appear in V. The normalization constant in the partition function is $A_L(N)=\int \Pi_{i=1}^{L} d\phi_i \exp[-N/2\Sigma_{i,j=1}^{L}(V^{-1})_{i,j}\mathrm{Tr}\ \phi_i\phi_j] \times \exp[-N\Sigma_{i=1}^{L}(W^{-1})_i \mathrm{Tr}\ \phi_i]$. We make the following simplifications in the model: (i) any type of base pairing is allowed with all the pairing probabilities the same and equal to $V_{i,j}=v$, i.e., no specificity toward Watson-Crick or Wobble pairings and (ii) the polymer chain is assumed to be infinitely flexible. In the matrix model of RNA, N is analogous to the role played by chemicals such as Mg^{2+} in a solution to separate secondary and tertiary structures of RNA [18]. The interaction energies between the pairs of bases in the nearest-neighbor models [24] can be taken into account in these RNA matrix models by using a more complicated interaction matrix V [18]. We consider $W_i=w$ where w gives the strength of the perturbation which is the same on all nucleotides in the chain. After carrying out a series of Hubbard Stratonovich transformations, the integral over L matrices ϕ_i in Eq. (1) reduces to an integral over a single $(N\times N)$ Hermitian matrix σ,

$$Z_{L,\alpha}(N) = \frac{1}{R_L(N)} \int d\sigma \exp[-N/2v\ \mathrm{Tr}(v/w+\sigma)^2] \times \frac{1}{N} \mathrm{Tr}(1+\sigma)^L, \tag{2}$$

where $R_L(N)=\int d\sigma \exp[-N/2v\ \mathrm{Tr}(v/w+\sigma)^2]$. Following the algebra in [23] [from Eq. (5) to Eq. (15)], the exponential generating function $G(t,N,\alpha)$ of the partition function $Z_{L,\alpha}(N)$ is

$$G(t,N,\alpha) \equiv \sum_{L=0}^{\infty} Z_{L,\alpha}(N) \frac{t^L}{L!} = \exp[vt^2/2N + t(1-\alpha)] \left[\frac{1}{N} \sum_{k=0}^{N-1} \binom{N}{k+1} \frac{(t^2 v)^k}{k!\, N^k} \right], \tag{3}$$

where $\alpha=\frac{v}{w}$ gives the ratio of strengths of the quadratic to the perturbed term in the action of the partition function [Eq. (1)].

For $\alpha=0$, i.e., no external perturbation we have the random matrix model in [19] where we get all the structures. However, for $\alpha=1$, i.e., when the strengths of base pairing interactions and applied perturbation are the same, it is observed that the partition function for odd lengths of the polymer chain vanishes completely. At this particular value of the applied external perturbation, the effect on the partition function is drastic. In the extended matrix model with interaction, each unpaired base of the polymer chain in the contact diagrams is weighted by a factor $(1-\alpha)$ which becomes zero when $\alpha=1$ thus removing structures with any unpaired bases at that length. We therefore have two regimes: (i) $0\leq\alpha<1$ comprising of both the unpaired and paired base structures and (ii) $\alpha=1$ where structures with only fully paired bases remain. After incorporating the simplifications, the effect of the external perturbation appears on the free bases of the polymer chain only. The genus distributions for the extended matrix model in [23] exhibit interesting features for different α's.

The RNA random matrix model with interaction requires a number of simplified assumptions in order to make the calculation tractable for studying the effect of external perturbations. This simplified model for RNA studies effects induced by the perturbation on the folding. The most important feature of this matrix model of RNA folding is the topological information that it contains which is important since the problem of RNA folding is topological in nature. Even in the absence of factors which constrain the formation of RNA structures, for example, finite flexibility of the chain, geometric, and steric constraints, we expect some topological properties found in these models to be true for real RNAs. In [25] it was found for the RNA matrix model and the real RNA sequences that the average genus of the possible structures for a given length L depends linearly on the length. Loop entropy and loop statistics have also been found to play an important role in determining the possible conformations and assessing RNA stability [26,27]. The inclusion of entropy (loop entropy) is nontrivial in these matrix models. An exciting future direction is to calculate the loop entropy in matrix models of RNA. A proposal which includes entropy in the matrix model of RNA is described in [28] which extends the models to multimatrix field theory models of RNA.

A. Extended matrix model of RNA with perturbation on a single base (1-NP) and n bases (n-NP)

We now consider a generalization of the matrix model with interaction proposed in [23] by adding a perturbation to

a single nucleotide in the polymer chain [Eq. (1)]. Thus we will consider $W_i = W_1$ only (1-NP). The motivation comes from the force-induced experiments in obtaining important characteristics of folding and unfolding of RNAs discussed in the introduction [1–13,17]. We keep all the assumptions the same as for the model in [23] discussed above. The interaction partition function $Z_{L,\alpha}(N)$ will be given by Eq. (1) with the perturbation term now being $\exp[-N(W^{-1})_1 \mathrm{Tr}\ \phi_1]$ and the normalization constant given by $A_L(N) = \int \Pi_{i=1}^{L} d\phi_i \exp[-N/2 \Sigma_{i,j=1}^{L} (V^{-1})_{i,j} \mathrm{Tr}\ \phi_i \phi_j] \times \exp[-N(W^{-1})_1 \mathrm{Tr}\ \phi_1]$. Carrying out a similar mathematical analysis employed in going from Eq. (1) to Eq. (3) above, we can write the exponential generating function of the partition function as in Eq. (3) with α/L in place of α. The partition functions $Z_{L,\alpha}(N)$ for different L can be found exactly from the exponential generating function by equating the coefficients of powers of t on both the sides of the equation. The structures in the extended matrix model for perturbation on a single nucleotide can be divided into the two regimes: (i) $0 \leq \frac{\alpha}{L} < 1$ having structures with a combination of paired and unpaired bases and (ii) $\frac{\alpha}{L} = 1$ comprising of only the completely paired base structures. In general, if the number of nucleotides with the perturbation is n where $n \leq L$, the sum in the interaction term goes from $i = 1, \ldots, n$, then α in Eq. (3) is replaced by $\frac{n\alpha}{L}$ (this gives the n-NP RNA matrix model).

The lesson learnt for this simple, do-able, interaction is that the interaction distinguishes in the Feynman diagrams (structures of RNA) the paired and free bases by assigning a weight $(1 - \frac{n\alpha}{L})$ to the free bases and 1 to the paired bases. The important point to note is the dependence of the weight on L.

III. ASYMPTOTICS OF THE EXTENDED MATRIX MODELS FROM NUMERICS

The asymptotic behavior (large L) of the genus distribution functions for the matrix model of RNA in [19] showed universal characteristics. We investigate here numerically the changes that the genus distribution functions: (i) the total number of diagrams at a fixed length L but independent of genus g, $\mathcal{N}$ [defined as $\mathcal{N} = Z_L(N=1)$] and (ii) the number of diagrams at a fixed length L and genus g, $a_{L,g}$ [defined through $Z_L(N) = \Sigma_{g=0}^{\infty} a_{L,g} \frac{1}{N^{2g}}$], of the model in [19] undergo when a perturbation is added. The asymptotics of the genus distribution functions are computed for the extended matrix model (i) when the perturbation is on all the bases, $n = L$ [23] and (ii) when perturbation is on n bases, n-NP. We will represent the genus distribution functions for the different matrix models as follows: (i) $\mathcal{N}$ and $a_{L,g}$ will represent the asymptotic formulas for the model in [19], (ii) $\mathcal{N}'_\alpha$ and $a'_{L,g,\alpha}$ will represent the new asymptotic formulas for the matrix model of RNA with interaction [23], and (iii) $\mathcal{N}_\alpha$ and $a_{L,g,\alpha}$ will represent the numerical values of the genus distribution functions for different α's for the extended matrix model. We start with the asymptotic expressions of [19], (i) $\mathcal{N} = L^{L/2} \exp[-(L/2) + \sqrt{L} - (1/4)]/\sqrt{2}$ and (ii) $a_{L,g} = k_g 3^L L^{(3g-3/2)}$, where $v = 1$ and $k_g = \frac{1}{3^{(4g-3/2)} 2^{2g+1} g! \sqrt{\Pi}}$, and find the asymptotic formulas $\mathcal{N}'_\alpha$ and $a'_{L,g,\alpha}$ for the matrix model with interaction for lengths up to $L = 40$, $\alpha = 0, 0.25, 0.5, 0.75, 1$, and $n = L$.

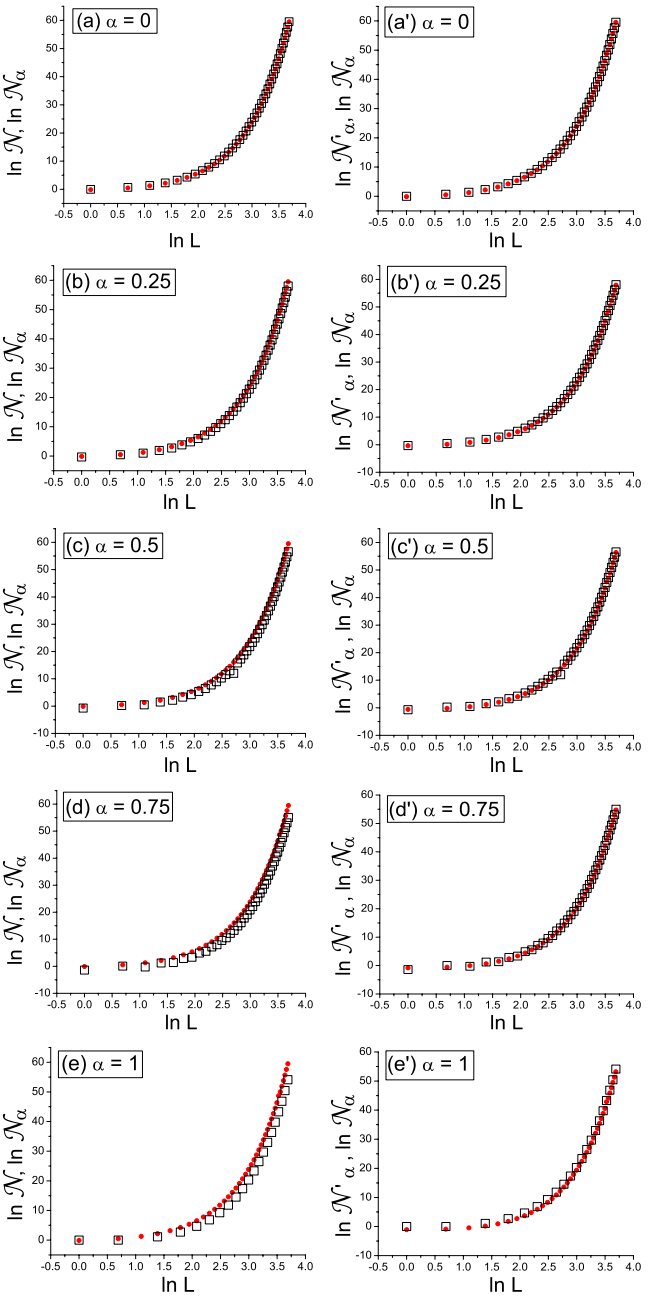

FIG. 1. (Color online) (a)–(e) Plots of the natural logarithm of the asymptotic formula $\ln(\mathcal{N})$ in [19] (red dotted curves) and the numerically calculated $\ln(\mathcal{N}_\alpha)$ values (black boxed curves) for $\alpha = 0$, 0.25, 0.5, 0.75, and 1 for different lengths L. (a′)–(e′) The natural logarithm of the new asymptotic formula $\ln(\mathcal{N}'_\alpha)$ (red dotted curves) for the extended matrix model of RNA [23] is plotted along with the numerical $\ln(\mathcal{N}_\alpha)$ values (black boxed curves) for $\alpha = 0$, 0.25, 0.5, 0.75, and 1 for different lengths L.

A. Asymptotics for $\mathcal{N}_\alpha$

Figures 1(a)–1(e) show the combined plots of the natural logarithm of the asymptotic expression $\mathcal{N}$ (red dotted curve)

TABLE I. Table lists, for different values of α, slopes of the linearly fitted plots before and after the multiplication of $(1-\alpha)$ with the $\sqrt{L}$ term for (i) $[\ln(\mathcal{N}_\alpha)-\sqrt{L}+\frac{L}{2}]$ versus $L\ln(L)$ (slope 1), (ii) $[\ln(\mathcal{N}_\alpha)-(1-\alpha)\sqrt{L}+\frac{L}{2}]$ versus $L\ln(L)$ [slope 1(a)], (iii) $[\ln(\mathcal{N}_\alpha)-\sqrt{L}-\frac{1}{2}L\ln(L)]$ versus L (slope 2) and (iv) $[\ln(\mathcal{N}_\alpha)-(1-\alpha)\sqrt{L}-\frac{1}{2}L\ln(L)]$ versus L [slope 2(a)].

α	Slope 1	Slope 1(a)	Slope 2	Slope 2(a)
0	0.499	0.499	−0.5026	−0.5026
0.25	0.4885	0.499	−0.5353	−0.5022
0.5	0.4767	0.4987	−0.5683	−0.5027
0.75	0.4624	0.4981	−0.6025	−0.5060
1	0.4556	0.5003	−0.6331	−0.4992

with the natural log of the numerically computed $\mathcal{N}_\alpha$ values for $\alpha=0,0.25,0.5,0.75,1$ (black boxed curves) in the extended matrix model. It is observed that as α is increased from 0 to 1, the black boxed curves shift downward continuously indicating an α dependence in $\mathcal{N}_\alpha$ for the extended matrix model of RNA. We investigate this dependence in the following numerical analysis.

Taking natural log of $\mathcal{N}$ we get $\ln(\mathcal{N})\sim\frac{L}{2}\ln(L)-\frac{L}{2}+\sqrt{L}-\frac{1}{4}-\ln(\sqrt{2})$. We are interested in the large length (L) behavior and we see that the dependence of $\ln\mathcal{N}$ on L is strongest in $L\ln(L)$. We linearly fit the plots (i) $[\ln(\mathcal{N}_\alpha)-\sqrt{L}+\frac{L}{2}]$ versus $L\ln(L)$ (slope 1, Table I), (ii) $[\ln(\mathcal{N}_\alpha)-\sqrt{L}-\frac{1}{2}L\ln(L)]$ versus L (slope 2, Table I), and (iii) $[\ln(\mathcal{N}_\alpha)+\frac{L}{2}-\frac{1}{2}L\ln(L)]$ versus $\sqrt{L}$ (Fig. 2) for different α and find their slopes. In the linearly fitted plots of (i) and (ii), we find that there is a continuous decrease in the slope as α goes from 0 to 1 strongly suggesting a dependence of $\mathcal{N}_\alpha$ on α. In the fitted plots of (iii) we observe a remarkable behavior for $\alpha=0.75$ and $\alpha=1$ plots. In the $\alpha=0.75$ plot [Fig. 2(d)], the points for odd and even lengths separate out into two very distinct curves at small lengths and for the $\alpha=1$ plot [Fig. 2(e)], the odd length points vanish completely leaving only the even length points in the figure. This indicates that $[\ln(\mathcal{N}_\alpha)+\frac{L}{2}-\frac{1}{2}L\ln(L)]$ versus $\sqrt{L}$ is very sensitive to changes in α.

We put a factor $(1-\alpha)$ with the $\sqrt{L}$ term in the exponent of the $\mathcal{N}$ expression and then fit the plots: (i) $[\ln(\mathcal{N}_\alpha)-(1-\alpha)\sqrt{L}+\frac{L}{2}]$ versus $L\ln(L)$ [slope 1(a), Table I] and (ii) $[\ln(\mathcal{N}_\alpha)-(1-\alpha)\sqrt{L}-\frac{1}{2}L\ln(L)]$ versus L [slope 2(a), Table I] for different values of α. We observe that now all the slopes are nearly the same and equal to $+\frac{1}{2}$ and $-\frac{1}{2}$ for (i) and (ii), respectively. This proves that the factor $(1-\alpha)$ with the $\sqrt{L}$ term in the exponent of $\mathcal{N}$ is the correct choice. We can therefore write the new asymptotic expression of the total number of diagrams at a fixed length L and α but independent of genus g, $\mathcal{N}'_\alpha$ for the matrix model with perturbation for $n=L$ as

$$\mathcal{N}'_\alpha = L^{L/2}\exp[-L/2+(1-\alpha)\sqrt{L}-1/4]/\sqrt{2}. \quad (4)$$

We see from Eq. (4) that the total number of structures for the extended matrix model changes considerably, for example, when $\alpha=1$ the $\sqrt{L}$ term vanishes from the exponent. We repeat the exercise as before and plot the natural logarithm of the new asymptotic formula $\mathcal{N}'_\alpha$ (red dotted curves) for the matrix model of RNA with interaction given by Eq. (4) together with the natural log of numerically obtained $\mathcal{N}_\alpha$ values (black boxed curves) for different α's, Fig. 1(a′)–1(e′). The plot for the new asymptotic formula coincides with the numerical data $\mathcal{N}_\alpha$ confirming the new formula.

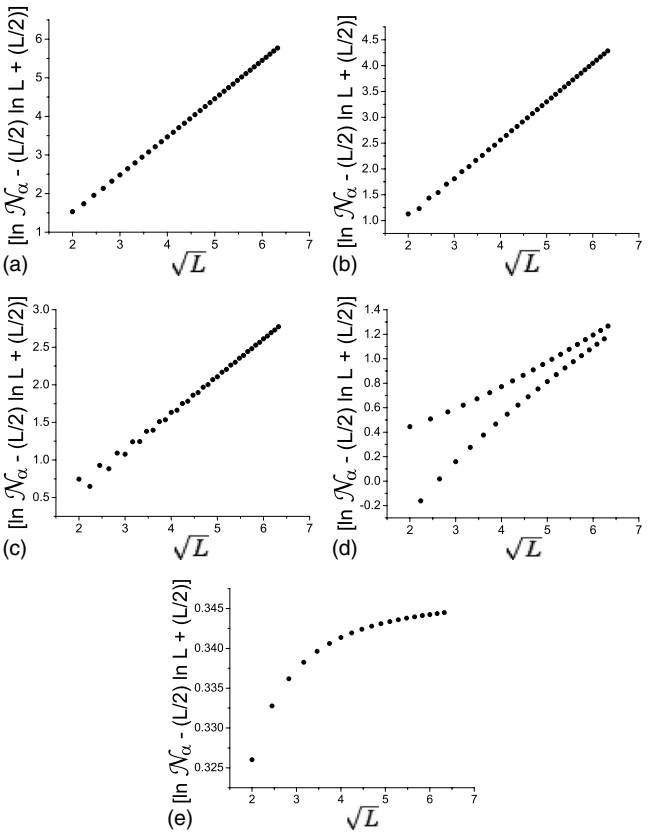

FIG. 2. $[\ln(\mathcal{N}_\alpha)+\frac{L}{2}-\frac{1}{2}L\ln(L)]$ versus $\sqrt{L}$ plots for different values of α along with their linearly fitted slope values: (a) $\alpha=0$ (slope=0.9818), (b) $\alpha=0.25$ (slope=0.7359), (c) $\alpha=0.5$ (slope =0.4926), (d) $\alpha=0.75$ (linear fit to the two curves gives slope =0.3595), and (e) $\alpha=1$ (the plot is not linear).

B. Asymptotics for $a_{L,g,\alpha}$

The natural logarithm of the asymptotic formula $a_{L,g}$ (black dotted curve in Fig. 3) plotted together with the natural log of the numerically calculated $a_{L,g,\alpha}$ values (green boxed curve) for different α's (shown here for $\alpha=0.75$, for $\alpha=0$, 0.25, 0.5, and 1 we get similar figures) in the extended matrix model clearly indicates that the asymptotic formula of the model in [19] needs to be changed to give the asymptotic behavior of the matrix model of RNA with external interaction [23]. The green boxed data curves move further and further away from the black dotted $a_{L,g}$ curve [19] as α goes from 0 to 1. This behavior is studied and the correct asymptotic expression $a'_{L,g,\alpha}$ for the extended matrix model is found.

We start with the asymptotic expression of $a_{L,g}=k_g 3^L L^{(3g-3/2)}$. Taking ln on both the sides and fixing $g=1$ we

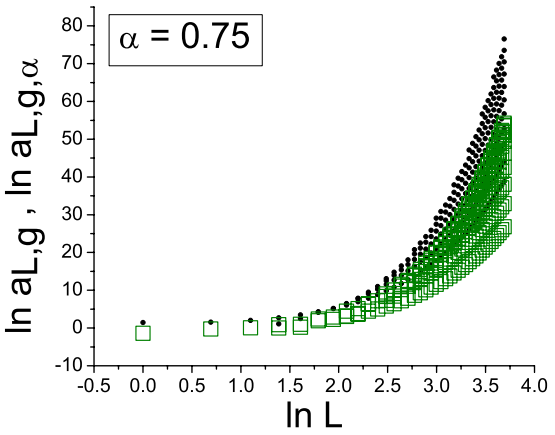

FIG. 3. (Color online) The natural logarithm of the asymptotic formula $\ln(a_{L,g})$ [19] (black dotted curves) is plotted together with the numerical $\ln(a_{L,g,\alpha})$ values (green boxed curves) for α=0.75 for different lengths L. Note: The figure plots $\ln(a_{L,g,\alpha})$'s for all genii corresponding to a particular length L of the polymer chain. The lowest curve (black dotted or green boxed) corresponds to genus g=0 for all the lengths (0 to 40) and the successive curves in the upward direction correspond to next higher genii with the maximum genus given by $g_{\max}=L/4$.

get, $\ln(a_{L,g=1}) \sim \ln[\frac{1}{3^{(5/2)}2^{(3)}\sqrt{\Pi}}]+L\ln(3)+\frac{3}{2}\ln(L)$. In $\ln(a_{L,g=1})$, L dependence is present in the form of L and $\ln(L)$. We are interested in the large L behavior so we first look for the dominant L dependence. The linear fits to the plots of $\ln(a_{L,g=1,\alpha})$ versus L in Table II (slope 1) show that the slopes of the numerical $a_{L,g,\alpha}$ curves for different α's are not the same and not equal to the slope of the $a_{L,g}$ asymptotic curve (slope should be ln(3) according to [19], marked Analytical in Table II). This indicates an α dependence in the factor 3 of the 3^L universal part of $a_{L,g}$. We represent this dependence by $x(\alpha)$ where $x(\alpha)$=exp(slope 1) (Table II). We write the asymptotic formula by replacing 3 with $x(\alpha)$. The expression for $a'_{L,g,\alpha}$ after taking ln on both of the sides becomes $\ln(a'_{L,g,\alpha}) \sim \ln(k_g)+L\ln[x(\alpha)]+(3g-\frac{3}{2})\ln(L)$. To determine the form of $x(\alpha)$, we plot $x(\alpha)$ versus α which is a straight line with slope=−1.133 and intercept=3.466. In the same way as the asymptotic expression for $a_{L,g}$ in [19] had the universal term 3^L, we find $x(\alpha)^L$ to be $(3-\alpha)^L$ for all α. We therefore have $\ln(a'_{L,g,\alpha}) \sim \ln(k_g)+L\ln(3-\alpha)+(3g-\frac{3}{2})\ln(L)$.

TABLE II. Table lists, for different values of α, the measures of slopes obtained from the linear fits to the plots between L and $\ln(a_{L,g=1,\alpha})$ (slope 1), the $x(\alpha)$ values for each α and slopes from the linear fit of plots between $\ln(L)$ and $[\ln(a_{L,g=1,\alpha})-L\ln(3-\alpha)]$ for each α (slope 2).

α	Slope 1	$x(\alpha)$	Slope 2
0	1.198	3.313	1.646
0.25	1.109	3.03	1.639
0.5	1.012	2.75	1.633
0.75	0.9065	2.476	1.623
1	0.7891	2.2	1.655
Analytical	1.24	3.4556	1.495

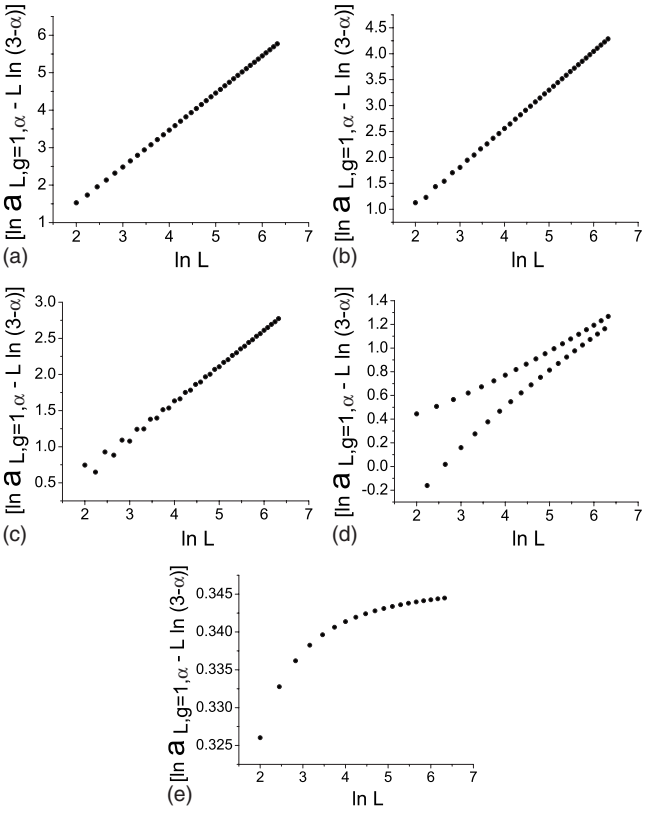

FIG. 4. $[\ln(a_{L,g=1,\alpha})$-$L\ln(3-\alpha)]$ versus ln(L) plots for (a) α=0, (b) α=0.25, (c) α=0.5, (d) α=0.75, and (e) α=1. The slopes for these values of α are listed in Table II (slope 2).

The dominant term in the asymptotic form of $a_{L,g}$, 3^L, [19] changes to $(3-\alpha)^L$ for the matrix model with interaction [23]. The asymptotic formula thus gets modified to $a'_{L,g,\alpha} \sim k_g(3-\alpha)^L L^{(3g-3/2)}$.

Analyzing the $\ln(L)$ dependence now, we assume that there exists an α dependence in the exponent of L which we represent by $f(\alpha)$. We can therefore write from $a'_{L,g,\alpha}$ after taking ln on both the sides and substituting g=1, $\ln(a'_{L,g=1,\alpha}) \sim \ln(\frac{1}{3^{(5/2)}2^{(3)}\sqrt{\Pi}})+L\ln(3-\alpha)+\frac{3}{2}[f(\alpha)]\ln(L)$. Linear fitted plots of $[\ln(a_{L,g=1,\alpha})-L\ln(3-\alpha)]$ versus $\ln(L)$ for different α values are shown in Fig. 4. The figure shows a continuous separation of data points belonging to the even and odd lengths as α is increased from 0 to 1. There are two distinct lines of data points at small lengths L which merge into a single line at higher lengths L. For α=1 the odd length points vanish completely from the plot. The slopes (Table II, slope 2) show that the difference between analytical and numerical values for different α is ~0.01, which is small. The $\ln(L)$ term therefore shows no significant α dependence. So we fix $f(\alpha)$=1. This gives the asymptotic formula of the number of diagrams at a fixed length L, genus g and α, $a'_{L,g,\alpha}$ for the matrix model of RNA with perturbation for $n=L$ to be

$$a'_{L,g,\alpha} \sim k_g(3-\alpha)^L L^{(3g-3/2)}. \quad (5)$$

The natural logarithm of the asymptotic formula (black dotted curve), Eq. (5) thus obtained is plotted together with the natural log of numerical $a_{L,g,\alpha}$ values (green boxed curve)

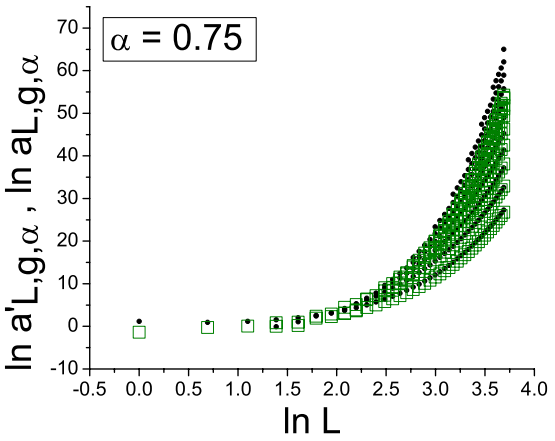

FIG. 5. (Color online) The plot for the natural logarithm of the new asymptotic formula, $\ln(a'_{L,g,\alpha})$ (black dotted curve) for the extended matrix model of RNA is shown together with the numerically obtained $\ln(a_{L,g,\alpha})$ for $\alpha=0.75$ (green boxed curve) as function of $\ln L$.

for different α's (Fig. 5, shown here for only $\alpha=0.75$, similar figures are found for $\alpha=0,0.25,0.5,1$) and it is seen that the formula matches with the numerical results for large L. To verify the final form of the formula, we substitute different α's and $g=1$ in Eq. (5) and plot $[\ln(a'_{L,g=1,\alpha})-L\ln(3-\alpha)]$ versus $\ln(L)$. The slopes are found to be 1.495 in all the cases. This result will hold for any genus g, though we have shown here the result for only $g=1$. It is interesting to note here that in $a'_{L,g,\alpha}$ for the extended matrix model, the universal term changes from 3^L in [19] to 2^L when α goes from 0 to 1, the completely paired base region. Other statistical models of pseudoknots in RNA [29–31] have studied combinatorial and statistical properties such as the number and fractions of pseudoknot structures. In particular a graph theoretic approach toward the combinatorial problem of RNA structures with pseudoknots shows that the number of bisecondary structures (secondary structures with non-nested pseudoknots) grows asymptotically as $\sim(\beta)^L$ where β is a combinatorial factor and L is the length of the chain [29]. In the matrix model with interaction discussed here we find that the dominant length dependence in the asymptotic behavior of the number of structures $a'_{L,g,\alpha}$ goes as $(3-\alpha)^L$.

The asymptotic behavior of $a_{L,g,\alpha,n}$ and $\mathcal{N}_{\alpha,n}$ for the model with perturbation on n bases is the same as for the model with $n=L$ except that α is replaced by $\frac{n\alpha}{L}$ [as is evident from the discussion in Sec. II A, the expression of the exponential generating function $G(t,N,\alpha)$ given by Eq. (3) with $\frac{n\alpha}{L}$ in place of α]. Thus we can write the asymptotic expressions of the genus distribution functions for a perturbation acting on n bases as

$$a'_{L,g,\alpha,n} \sim k_g\left(3-\frac{n\alpha}{L}\right)^L L^{(3g-3/2)} \tag{6}$$

and

$$\mathcal{N}'_{\alpha,n} = L^{L/2}\exp[-L/2+(1-n\alpha/L)\sqrt{L}-1/4]/\sqrt{2}. \tag{7}$$

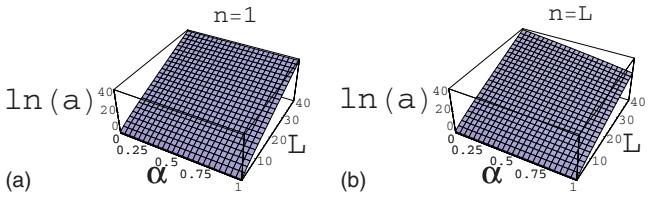

FIG. 6. (Color online) The figure plots natural logarithm of the genus distribution function $a_{L,g,\alpha}$ [represented by (a) in the figure] at a fixed genus, $g=1$ for (a) $n=1$ and (b) $n=L$ by varying α and L. It is seen that for $n=L$, the slope changes gradually from ln(3) to ln(2) as α is increased from 0 to 1 while the linear dependence on L stays unchanged.

The asymptotics of the genus distribution functions [Eqs. (4)–(7)] for the matrix model with external interaction therefore show marked changes in the presence of the perturbation from $a_{L,g}$ and $\mathcal{N}$ of the model in [19].

An important lesson learnt from this exercise is that the dominant term in the asymptotic, large L behavior of the number of diagrams $a'_{L,g,\alpha}$ at a fixed L, g and α is $(3-\frac{n\alpha}{L})^L$ with the same exponent L. Thus the universality is preserved on introduction of interactions (see Fig. 6). A change may occur for other interactions.

IV. CONCLUSIONS

In this work, we develop on the footsteps of the matrix model of RNA with interaction proposed in [23], the effect of an external perturbation on a single nucleotide. We argue that α in the exponential generating function of the partition function, Eq. (3) will be replaced by $\frac{\alpha}{L}$. Further, we generalize this result to a finite number $n\leq L$ of perturbations on the nucleotides where α in the exponential generating function of the partition function gets replaced by $n\alpha/L$ [in Eq. (3)]. The parameter space (α and n) of the model can thus be split into two regimes as: (i) $0\leq\alpha\leq 1$, $n<L$ and $0\leq\alpha<1$, $n=L$ consisting of structures with a combination of paired and unpaired bases and (ii) $\alpha=1$, $n=L$ comprising of only the fully paired base structures. We find numerically for the matrix model with interaction that the average genus varies linearly with the length of the polymer chain as $L/4$. The effect of introduction of this external perturbation changes the weight of the free bases [by $(1-\frac{n\alpha}{L})$] in the chain while the paired bases remain unaffected compared to [19]. Though the model is limited by the exclusion of factors such as stacking energies, it takes into account effectively the perturbation-induced changes on both the planar and secondary structures with pseudoknots (i.e., the tertiary structures). In the matrix models of RNA, we find the number of pseudoknots and their fraction out of the total possible conformations for a given length of the polymer chain from the partition function of the model.

We find numerically the asymptotic behavior of the genus distribution functions for the matrix model of RNA with interaction in [23] and the n-NP model. The numerical analysis shows that the term 3^L in $a_{L,g}$, found in [19], changes to $(3-\alpha)^L$ when $n=L$ [which becomes $(3-\frac{n\alpha}{L})^L$ when the perturbation is on n bases]. The power law term $L^{(3g-3/2)}$ in $a_{L,g}$

[19] remains the same for the asymptotic formula $a'_{L,g,\alpha}$ in the matrix model with perturbation for $n=L$ [23] and for any $n<L$. The total number of diagrams $\mathcal{N}$ also changes to $\mathcal{N}'_\alpha = L^{L/2} \exp[-L/2+(1-\alpha)\sqrt{L}-1/4]/\sqrt{2}$ with the term $\exp(\sqrt{L})$ in $\mathcal{N}$ [19] going to $\exp[(1-\alpha)\sqrt{L}]$ for $n=L$ (which becomes $\exp[(1-n\alpha/L)\sqrt{L}]$ when $n \leq L$). The most striking change found in $a'_{L,g,\alpha}$ is when α takes the value 1 (and $n=L$) as the slope changes from ln(3) to ln(2) (see Fig. 6) and in the $(1-\alpha)\sqrt{L}$ term in the exponent of $\mathcal{N}'_\alpha$ which goes to zero. It is shown in Figs. 2 and 4 that as α is increased from 0 to 1 in steps of 0.25, the points corresponding to even and odd lengths of the chain start splitting up into two different curves at small lengths, but converge into a single linear curve as the length is increased. At small lengths, this difference is most pronounced for $\alpha=0.75$, for both $\mathcal{N}_\alpha$ and $a_{L,g,\alpha}$. The $\alpha=1$ plots of $\mathcal{N}_\alpha$ and $a_{L,g,\alpha}$ [Fig. 2(e) and Fig. 4(e), respectively] show the absence of odd length data points. It is interesting to note that the genus distributions show different behavior at small and large lengths. This result is novel to matrix models of RNA with interactions. The large L (asymptotic) behavior of the distribution functions [Eqs. (4)–(7)] found for the RNA matrix model with external perturbation shows interesting changes.

The main outcome of this exercise is that the introduction of interactions in this simple random matrix model of RNA allows us to study the effect of interactions on the genus distribution functions as a function of the length. We find a result for the exponential generating function Eq. (3) which changes the singularity structure of the model (details will be reported elsewhere). This produces a change in the asymptotic form for the genus distribution functions which has been derived here using very simple graphical analysis. However the exponent L in $(3-\frac{n\alpha}{L})^L$ of the genus distribution function $a'_{L,g,\alpha}$ remains unchanged. Hence the universality is robust to the change in interaction (α values, Fig. 6). This is an exciting theoretical prediction. It would be very interesting to look for RNAs with interactions (acting on a single base or n bases) considered in this work and explore the challenging question whether the universal results reported here for the genus distributions (even though the results are true for simple homopolymers with interactions) can be reproduced in the experiments.

In order to make contact between the results of the matrix model with the pulling experiments we note that as the distribution functions are given as functions of the length (Figs. 1–6 and [23]), information about which structures, a structure at a certain length and genus can go to, under extension can be made. In other words, the fraction of planar and pseudoknot structures at different lengths can be found. This is very useful for the stretching experiments. Further more complicated interactions may need to be considered in the potential for explicit comparison with the experiments and this is a goal for the future.

ACKNOWLEDGMENTS

We would like to thank Professor H. Orland and Professor G. Vernizzi for very valuable and encouraging discussions. The work was financially supported by CSIR Project No. 03(1019)/05/EMR-II.

[1] B. Onoa and I. Tinoco, Jr., Curr. Opin. Struct. Biol. **14**, 374 (2004).
[2] T. E. Fisher, P. E. Marszalak, and J. M. Fernandez, Nat. Struct. Biol. **7**, 719 (2000).
[3] J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).
[4] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997); Phys. Rev. E **58**, 2386 (1998).
[5] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11935 (1997).
[6] A. D. Mehta, M. Rief, J. A. Spudich, D. A. Smith, and R. M. Simmons, Science **283**, 1689 (1999).
[7] U. Gerland, R. Bundschuh, and T. Hwa, Biophys. J. **81**, 1324 (2001).
[8] U. Gerland, R. Bundschuh, and T. Hwa, Phys. Biol. **1**, 19 (2004).
[9] C. Hyeon, G. Morrison, and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **105**, 9604 (2008).
[10] P. T. X. Li, D. Collin, S. B. Smith, C. Bustamante, and I. Tinoco, Jr., Biophys. J. **90**, 250 (2006).
[11] C. Hyeon and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **102**, 6789 (2005).
[12] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante, Science **292**, 733 (2001).
[13] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, Science **296**, 1832 (2002).
[14] A. Stein and D. M. Crothers, Biochemistry **15**, 160 (1976); Z. J. Tan and S. J. Chen, Biophys. J. **90**, 1175 (2006).
[15] V. K. Misra and D. E. Draper, Proc. Natl. Acad. Sci. U.S.A. **98**, 12456 (2001); V. K. Misra and D. E. Draper, J. Mol. Biol. **317**, 507 (2002); V. K. Misra, R. Shiman, and D. E. Draper, Biopolymers **69**, 118 (2003); D. E. Draper, D. Grilley, and A. M. Soto, Annu. Rev. Biophys. Biomol. Struct. **34**, 221 (2005); D. Grilley, V. Misra, G. Caliskan, and D. E. Draper, Biochemistry **46**, 10266 (2007); D. E. Draper, Biophys. J. **95**, 5489 (2008).
[16] V. K. Misra and D. E. Draper, J. Mol. Biol. **299**, 813 (2000); V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, and S. Doniach, Curr. Opin. Chem. Biol. **12**, 619 (2008).
[17] B. Onoa, S. Dumont, J. Liphardt, S. B. Smith, and I. Tinoco, Jr., Science **299**, 1892 (2003).
[18] H. Orland and A. Zee, Nucl. Phys. B **620**, 456 (2002).
[19] G. Vernizzi, H. Orland, and A. Zee, Phys. Rev. Lett. **94**, 168103 (2005).
[20] M. Muller, Phys. Rev. E **67**, 021914 (2003); M. Mueller, F. Krzakala, and M. Mezard, Eur. Phys. J. E **9**, 67 (2002).
[21] F. David, C. Hagendorf, and K.-J. Wiese, Europhys. Lett. **78**, 68003 (2007).
[22] E. Rivas and S. R. Eddy, J. Mol. Biol. **285**, 2053 (1999).
[23] I. Garg and N. Deo, e-print arXiv:0802.2440.

[24] M. J. Serra and D. H. Turner, Methods Enzymol. **259**, 242 (1995).
[25] M. Bon, G. Vernizzi, H. Orland, and A. Zee, J. Mol. Biol. **379**, 900 (2008).
[26] S. Cao and S. J. Chen, RNA **11**, 1884 (2005).
[27] T. R. Einert, P. Nager, H. Orland, and R. R. Netz, Phys. Rev. Lett. **101**, 048103 (2008).
[28] A. Zee, Acta Phys. Pol. B **36**, 2829 (2005).
[29] C. Haslinger and P. F. Stadler, Bull. Math. Biol. **61**, 437 (1999).
[30] A. Xayaphoummine, T. Bucher, F. Thalmann, and H. Isambert, Proc. Natl. Acad. Sci. U.S.A. **100**, 15310 (2003).
[31] F. W. D. Huang, L. Y. M. Li, and C. M. Reidys, BMC Bioinformatics **10**, S39 (2009).