Macroscopic localization lengths of vibrational normal modes in a heuristic DNA model

C. J. Páez*
Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas (UNICAMP), Caixa Postal 6165, 13083-970 Campinas, SP, Brazil

R. Rey-González
Departamento de Física, Universidad Nacional de Colombia, Ciudad Universitaria, Bogotá, Colombia

P. A. Schulz
Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas (UNICAMP), Caixa Postal 6165, 13083-970 Campinas, SP, Brazil

(Received 5 October 2009; published 20 January 2010)

In this work we study the localization of vibrational modes in heuristic models for disordered DNA-like molecules. Within such approach, atomic groups are replaced by renormalized sites connected by effective springs. The oscillation amplitudes at each site are considered and the localization degree of the normal modes is analyzed by means of the participation ratio, as well as the relative fluctuation of an ensemble of disorder realizations for normal modes in different frequency ranges. The present results suggest that the dynamical properties at low frequencies are completely different for double-strand structures compared to single-strand ones. Irrespective to disorder, double-strand molecules show normal modes with macroscopic localization lengths at low frequencies, for a wide range of spring constants considered in the literature, in contrast to the strong localization in single strands.

DOI: 10.1103/PhysRevB.81.024203 PACS number(s): 63.22.−m, 63.50.−x, 63.20.Pw, 81.07.Nb

I. INTRODUCTION

DNA is probably the most fascinating structure in microscopic nature. Elucidating its structure as a double helix of nucleotides in sugar and phosphate backbones revealed the very mechanism of storing genetic information and how DNA plays a central role in inheritance mechanisms.[1]

Albeit the complexity of the environment, a useful framework is to consider a DNA molecule as a structure formed, on one hand, by an ordered backbone (sequence of groups of phosphates and sugars) embedding, on the other hand, a highly polymorphic ladder of base pairs (A-T or C-G). The genetic information is inscribed in the sequencing of base pairs, and there is no hindering in proposing design DNA sequencing for artificial purposes within the scope of nanoelectronics, leading to an extra motivation for investigating the electronic properties of these macromolecules.[2] Such studies, even when based on heuristic models, may lead to hints of important biological processes.[3,4]

In spite of the relevance of the electronic properties, several important biological processes are dependent on the elastic properties of DNA strands.[5] A throughout investigation of these properties from a microscopic point of view requires a full atomistic description, a framework showing the same drawbacks as in the electronic counterpart; the numerical cost hinders an investigation of macroscopic samples. A great deal of insight into such elastic properties may be gained by simple elastic rod models[6] but the intrinsic local polymorphism due to the base-pair sequencing is then completely lost. Therefore, intermediate heuristic models,[7–12] which take into account local inhomogeneity due to base pairing, but avoiding the complexity of fully microscopic descriptions are very useful, since long systems are feasible to handle numerically while retaining microscopic details.

In the present paper we address the localization of vibrational normal modes in ladder models that mimic DNA-like molecules described within an heuristic model in an intermediate framework mentioned above. As will become clear through the paper, the lack of a detailed atomistic description affects mainly higher-frequency modes which are out of the range of the relevant elastic properties.[13] We focus on the strandness (double or single) and sequencing effects, taking into account base pairing. The choice made here is a random sequencing, taken as an extreme case, corresponding to a disordered chain, where all normal modes are, in principle, localized.[14] On the other hand, correlation in the disorder may have additional effects on the degree of eigenstates localization.[15] These correlation-induced delocalization mechanisms have been investigated also for vibrational properties.[16] Moreover, recently have been considered in the context of electronic properties of DNA as discussed in Ref. 17.

The scenario of dynamic properties of DNA has evolved in different fronts in the past decades. The study of the molecular dynamics of DNA has gained increasing attention, some studies have developed detailed models[18] for the representation of exact classical solutions. Unfortunately, the computational cost is so high that only has been simulated DNA fragments,[19] which is far from characterize physical and biological processes such as replication, transcription, and denaturation.

These introductory remarks can be summarized by stating the aim of the present work. Motivated by recent investigations on the localization of electronic states in ladder models for DNA molecules, we address the analog problem for vibrational models. The results suggest that in spite of a complete random sequencing, double-strand DNA-like models reveal a low-frequency window of normal modes that, although localized, reveal macroscopic localization lengths.

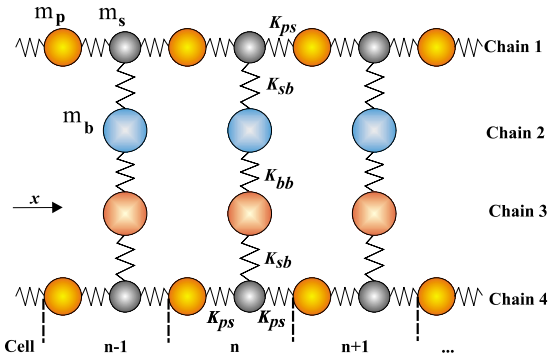

FIG. 1. (Color online). Heuristic model of DNA. m_p represents the PO_4 groups, m_s represents the sugar rings, and m_b represents one of the four bases that are distributed randomly; these bases are always paired.

II. MODEL

Empiric dynamical models,[20,21] taking into account the full symmetry of DNA, i.e., the helical structure and all individual atomic site belong to the early approaches to the problem. However they handle the main hindrance of a large number of parameters which do not allow an easy evaluation of the contribution of different factors on the dynamic properties of these macromolecules. Since mechanical deformations of DNA are necessary in fulfilling their biological processes, we will be mainly interested in the low-frequency range of the vibrational modes, namely, the range of macroscopic bending, stretching, and torsion motions, extensively discussed within homogeneous rod models. Our heuristic model fits into an intermediate category approach, substituting the full atomic description mostly relevant at higher frequencies but taking into account relevant features as base pairing and sugar-phosphate backbones. The present model is illustrated in Fig. 1, revealing the mass and spring constants necessary in the harmonic approximation considering only nearest-neighbor interactions.

In Fig. 1 we consider six sites per unit cell where each site represents the center of mass of one of the following molecular subunits: a phosphate that is represented by m_p, a sugar ring that is represented by m_s, and the four bases that represents the nucleotides, A, T, C, and G with specific binding properties; only A-T and C-G pairs are possible. Along the chain, we consider the nucleotide sequencing as completely random.

We also consider only small displacements parallel to the helix, in the x direction, such approximation is good only for the low-frequency modes.[11] Considering other displacements would increase further the number of empiric parameters, leading to a useless scenario of arbitrary complexity. It should be mentioned, however, that the robustness of our qualitative findings has been checked against the introduction of extra force constants. We neglect the influence of interhelical forces in oriented and aligned DNA chains. The forces between atomic groups are van der Waals forces that in the approximation of small displacements are harmonic forces.

Hence, in the model we use a set of three force constants: k_{ps}, k_{sb}, and k_{bb} that represent the coupling between the phosphate-sugar, sugar-base, and base-base, respectively. The force constant k_{ps} is chosen by interpolating the data from single DNA molecule force-extension experimental curve[22] in the range of intermediate forces ($5 < f < 60$ pN), which is due to the purely elastic contribution associated with the stretching of the double helix along its axis. The force constants K_{sb} and k_{bb} are taken from the approximation of potentials that describe the hydrogen bonds between the two bases in a given pair.[23–26] Thus, we have two types of constant for K_{bb}: k_{AT} and k_{CG} for the base pairs A-T and C-G, respectively. The values are shown in the Table I. Once more it should be mentioned that a search in the literature shows a wide range of parameter sets and therefore our study has to focus on physical properties that are robust against these parameter choices.

TABLE I. Parameters of the model calculation.

Masses	amu	Force constants	N/m
m_p	95.98	K_{ps}	195
m_s	83.11	K_{sb}	70
m_A	134.12	K_{AT}	19.5
m_T	125.11	K_{CG}	56.3
m_C	110.10		
m_G	150.12		

III. RESULTS AND DISCUSSION

For the sake of completeness, we briefly discuss the textbook steps toward calculating the vibrational normal modes. The displacement of the ith mass in the unit cell n from its equilibrium position is denoted as $x_{n,i}$. The equations of motion of the masses of the first and second chains are

$$m_p \ddot{x}_{n,p} = k_{ps}(x_{n,s} - x_{n,p}) - k_{ps}(x_{n,p} - x_{n-1,s}), \tag{1}$$

$$m_s \ddot{x}_{n,s} = k_{ps}(x_{n+1,p} - x_{n,s}) - k_{ps}(x_{n,s} - x_{n,p}) + k_{sb}(x_{n,b} - x_{n,s}), \tag{2}$$

$$m_b \ddot{x}_{n,b} = k_{bb}(x_{n,b'} - x_{n,b}) - k_{sb}(x_{n,b} - x_{n,s}). \tag{3}$$

Here $x_{n,b'}$ is the displacement of the base corresponding to the pair in the unit cell n. The equations of motion of the third and fourth chains are completely analogous. We assumed a finite system with rigid boundary conditions.

Using $s_i = m_i^{1/2} x_i$, we can take the set of Eqs. (1)–(3) to form

$$\ddot{\vec{s}} + \mathbf{W}\vec{s} = 0. \tag{4}$$

$\mathbf{W}$ is the dynamical matrix. The squared normal frequencies of vibration of the system are given by the eigenvalues of $\mathbf{W}$ and the normal modes by the corresponding eigenvectors. The element a_i of an eigenvector represents the maximum amplitude of displacement of the ith mass in the associated normal mode.

The localization degree of the vibrational normal modes is investigated by means of the participation ratio (PR),[27,28] defined by

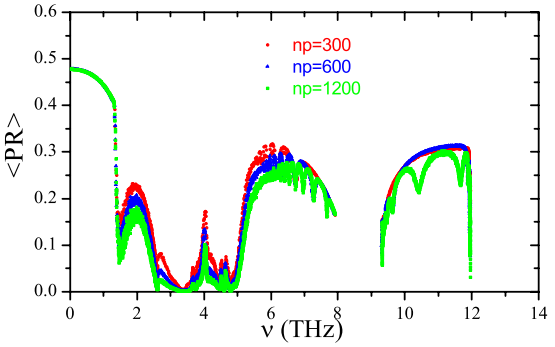

FIG. 2. (Color online). Average participation ratio as a function of frequency for a double-strand DNA for different chain lengths (see text).

$$\mathrm{PR} = \frac{1}{N\sum_{i=1}^{N} |a_i|^4} \tag{5}$$

with N is the total number of masses. In general the PR is a good estimate of the number of sites that participate in the vibrational normal modes. In the limit $N \to \infty$, the PR $\to 0$ for localized vibrational normal modes and for extended PR is independent of the size of the system, and reaches the maximum value 2/3 in one-dimensional ordered systems.[29] It is worth mentioning that PR has been normally employed in scrutinizing localization of electronic states. Nevertheless a pioneering use in the context of localized modes in disordered chains goes back to a work in 1976,[30] that confirms that disordered harmonic chains show only localized states, with an important exception, namely, the zero-frequency mode, which is necessarily delocalized in order to ensure the momentum conservation and the translational symmetry of the Hamiltonian.[14,16] In the context of the present work, Dominguez-Adame and Macia transposed to the vibrational properties scene, the effect of short-range correlation in disorder leading to delocalization of particular normal modes.[16]

As a first glimpse in our results, we present the average PR of the vibrational normal modes as a function of frequency for a double-strand DNA with backbone chains in Fig. 2, considering different chain lengths.

We averaged $\langle \mathrm{PR}(\nu) \rangle$ over 200 disorder configurations where the masses A, T, C, and G are randomly assigned with equal probability, leading to an average concentration of 25% for each one. The $\langle \mathrm{PR}(\nu) \rangle$ as a function of normal frequency are for systems 300, 600, and 1200 base pairs long. In opposition to electronic states, normal modes may be extended but the vibration amplitude of the sites may be dramatically different, as one can remember from the textbook results for optical modes of a simple diatomic chain in which every second atomic site is at rest. Therefore a completely extended mode will show a PR of 1/3 instead of 2/3. Therefore, a simple inspection of $\langle \mathrm{PR}(\nu) \rangle$ value does not allow to suggest a degree of localization. Indeed, Fig. 2 would indicate a low-frequency range of extended modes as well as a higher-frequency (optical-like branch) band of delocalized modes. However, comparing chains with increasing lengths reveal that the optical branch show a systematic decrease in PR, a hint for actually disorder-induced localized vibrations.[31] On the contrary, the lower-frequency range shows a collapse of PR curves for different system lengths, indicating the necessity of a closer look on the real degree of localization of these modes, such as an inspection of the atomic displacements.[8,25] The intermediate frequency range present rather noise average PR, a signature of truly localized modes. Recalling first the higher-frequency window, a direct inspection of one of these modes in Fig. 3(b), where one eigenmode among the average at $\nu = 2.589$ THz (600 base-pairs' case) is plotted, shows a strong localization. It should be noticed that the mode extension may be different at the backbone and base-pair chains. The chosen example in Fig. 3(b) indicates that the contribution from the backbone sites to the mode is clearly less significant than from the base pairs. On the other hand, we systematically find that normal modes look effectively delocalized in a low-frequency interval, $\sim 0-1.3$ THz, where $\langle \mathrm{PR}(\nu) \rangle$ curves collapse.

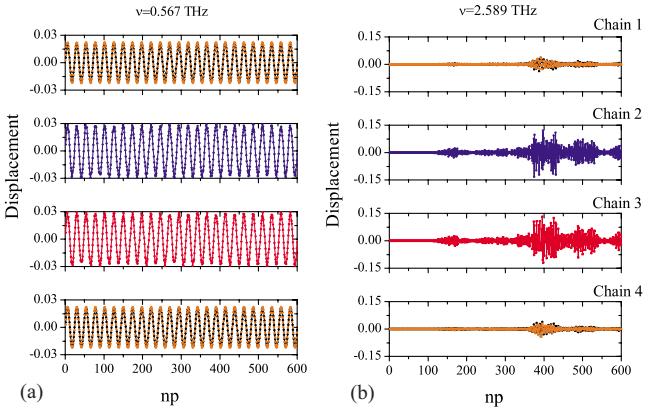

FIG. 3. (Color online). Displacement of the masses for two vibrational normal modes of the double-strand DNA-like chain with 600 base pairs. (a) Vibrational normal mode corresponding to $\nu = 0.567$ THz and $\langle \mathrm{PR} \rangle = 0.47$ and (b) vibrational normal mode corresponding to $\nu = 2.589$ THz and $\langle \mathrm{PR} \rangle = 0.038$.

In Fig. 3(a), a vibrational normal mode corresponding to the frequency 0.567 THz with 600 base pairs is plotted; it can be observed that it is homogeneously distributed over the whole system. This preliminary analysis suggests the possibility of a window of effective delocalized vibration in the low-frequency range which correspond to the range of interest concerning the important elastic properties of DNA that constitute the motivating background of our study.[6] Nevertheless, the PR clearly does not suffice to ensure that low-frequency vibrational normal modes are truly delocalized. Although the base pairing represents a correlation in the disorder, this mechanism may lead to only effectively delocalized states in the electronic case.[17] In the context of this methodological limitation, we introduce a tool originally proposed for electronic systems; the relative fluctuation of PR $\eta(\nu)$, defined initially by Moura *et al.* for an electronic system[32] and given by

$$\eta(\nu) = \frac{\sqrt{\langle \mathrm{PR}(\nu)^2 \rangle - \langle \mathrm{PR}(\nu) \rangle^2}}{\langle \mathrm{PR}(\nu) \rangle}. \tag{6}$$

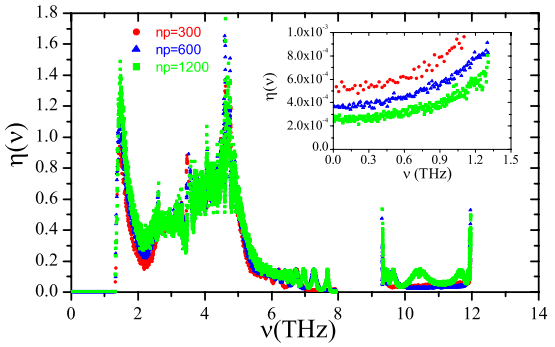

FIG. 4. (Color online). The relative fluctuation of the participation function $\eta(\nu)$ with $np=300$,600, and 1200 base pairs.

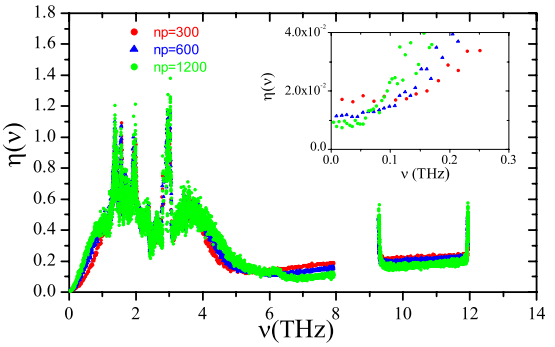

FIG. 5. (Color online). The relative fluctuation of the participation function $\eta(\nu)$ with $np=300$, 600, and 1200 base pairs, where the masses m_A, m_T, m_C, and m_G are in a proportion of 1, 2, 3, and 4, respectively, with respect to a mass average $m=129.86$ amu.

For extended vibrational normal modes, the relative fluctuation vanishes continuously with increasing system size. In the opposite regime of localized normal modes, the relative fluctuation grows with increasing system size, converging to a finite value. Figure 4 shows the $\eta(\nu)$ for the systems of 300, 600, and 1200 base pairs, already addressed in Fig. 2. Indeed, the fluctuation of the PR seems to be vanishingly small along a rather wide range in the low-frequency limit in the scale of Fig. 4. Nevertheless, the inset reveals low but finite values of the fluctuations, that nevertheless continuously diminish with increasing system length, a further suggestion of potentially truly delocalized modes. The high-frequency branch, on the other hand shows finite value fluctuations already in the scale of Fig. 4. This result is robust up to 10 000 base pairs long systems (corresponding to 3.4 μm), for a wide range and for an arbitrary change in spring constants.

Having in mind that base pairing is a correlation in disordered sequencing that constitute a mechanism for enhancing the localization length, that may lead to effective delocalization but not to truly extended electronic systems, the present behavior for low-frequency modes is rather intriguing. The key to the result is that the mass difference among the nucleotides is quite small and in DNA-like systems one is considering actually a low disorder limit. If we take artificially very different mass ratios for the nucleotides, the truly localized nature of these modes appears, as can be seen in Fig. 5.

Indeed, as shown in Table I, the masses that represent the bases present a maximum difference around 25% from each other. In a heuristic system, where the masses have a greater difference, the range of low frequencies where the normal modes are apparently extended diminishes continuously, even in the relatively short chain length range shown in Fig. 5; the $\eta(\nu)$ as a function of frequency for systems 300, 600, and 1200 base pairs long but now the masses m_A, m_T, m_C, and m_G are respectively, one, two, three, and four times the average mass $m=129.86$ amu.

Two further disorder sources can be introduced in the present model, which mimic actual physical processes. The first one is the hydration of a dry DNA (Refs. 11 and 33) as considered here. Hydration would be modeled by attaching waterlike sites randomly to the backbone chains, increasing the degree of disorder. A second additional disorder source occurs in DNA denaturation. An extreme case of denaturation would be to consider randomly positioned holes,[22,34,35] i.e., breaking base pairs, in the double strand. In our model the holes are simulated by setting the value of the spring constant, K_{AT} or K_{CG} to zero for a randomly chosen pair. The effect of the holes on the dynamical properties is shown in Fig. 6 where $\eta(\nu)$ as a function of frequency is plotted for the 600 base-pair long system with increasing hole concentration: $h=0\%$, $h=25\%$, and $h=50\%$. It is observed that increasing h, the range of effectively extended vibrational normal modes decreases. In the limit of $h=100\%$, i.e., the double strand is divided into a pair of single strands, only strongly localized modes are present.

IV. FINAL REMARKS

From the point of view of the dynamical properties, a DNA-like double chain shows strictly localized normal modes in the limit of random (disordered) sequencing. The specific characteristic that the A, T, C, and G masses are only slightly different (within a range of 25%) leads to a set of low-frequency modes that seem to be extended over macroscopic length. On the other hand, additional disorder mechanisms, such as partial hydration and denaturation produce stronger localization. In particular, the limit of completely breaking of the base pairs recovers single strands that reveal only strongly localized states. This limit, compared to the core of results shown in the present paper renders a picture in which a double strand shows effectively delocalized modes

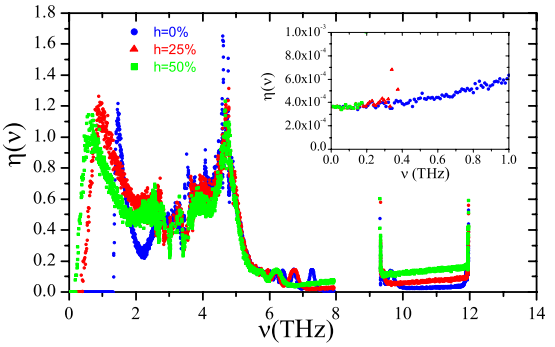

FIG. 6. (Color online). The relative fluctuation of the participation function with $h=0\%$, $h=25\%$, and $h=50\%$ holes.

over a finite range of frequencies while single strands are always strongly localized. On the other hand, a heuristic change in nucleotide mass relations reveals that the effectively extended modes up to macroscopic lengths that could occur in nature is proven to be localized in the thermodynamic limit, as shown by an model artifact.

Roles of further mechanisms as hydration and denaturation have been also introduced and should be focus of following investigations. A final comment should be addressed to recall the long difficult and inconclusive road of trying to parameterize dynamic properties within empirical models in the past decades. Nevertheless, the findings here are qualitative sensitive to a chemical signature of DNA molecules, namely, the similarity of the nucleotide masses, and not on a bona fide physical emulation of the chemical bonds, here emulated by the spring constants.

ACKNOWLEDGMENTS

C.J.P. acknowledges support from CNPp (at the beginning of this work) and FAPESP. P.A.S. receives partial support from CNPq.

*cjpaezg@ifi.unicamp.br

[1] J. D. Watson and F. H. C. Crick, Nature (London) **171**, 737 (1953).
[2] R. G. Endres, D. L. Cox, and R. R. P. Singh, Rev. Mod. Phys. **76**, 195 (2004).
[3] C. T. Shih, S. Roche, and R. A. Romer, Phys. Rev. Lett. **100**, 018105 (2008).
[4] A. A. Krokhin, V. M. K. Bagci, F. M. Izrailev, O. V. Usatenko, and V. A. Yampol'skii, Phys. Rev. B **80**, 085420 (2009).
[5] M. Barbi, S. Cocco, M. Peyrard, and S. Ruffo, J. Biol. Phys. **24**, 97 (1999).
[6] A. Matsumoto and W. K. Olson, Biophys. J. **83**, 22 (2002).
[7] S. Cuesta-López, J. Errami, F. Falo, and M. Peyrard, J. Biol. Phys. **31**, 273 (2005).
[8] K. Forinash, A. R. Bishop, and P. S. Lomdahl, Phys. Rev. B **43**, 10743 (1991).
[9] J. A. D. Wattis, S. A. Harris, C. R. Grindon, and C. A. Laughton, Phys. Rev. E **63**, 061903 (2001).
[10] S. C. Gómez-Carrillo, K. M. Fonseca, and R. R. Rey-González, Braz. J. Phys. **36**, 3b (2006).
[11] A. Wittlin, L. Genzel, F. Kremer, S. Haseler, A. Poglitsch, and A. Rupprecht, Phys. Rev. A **34**, 493 (1986).
[12] F. Zhang and M. A. Collins, Phys. Rev. E **52**, 4217 (1995).
[13] S. Srivastava, S. Srivastava, S. Singh, V. D. Gupta, and V. P. Gupta, Eur. Polym. J. **38**, 1423 (2002).
[14] P. Dean, Proc. Phys. Soc. Jpn. **84**, 727 (1964).
[15] D. H. Dunlap, H. L. Wu, and P. W. Phillips, Phys. Rev. Lett. **65**, 88 (1990).
[16] F. Domínguez-Adame, E. Macia, and A. Sanchez, Phys. Rev. B **48**, 6054 (1993).
[17] H. Carrillo-Nuñez and P. A. Schulz, Phys. Rev. B **78**, 235404 (2008), and references therein.
[18] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[19] S. Falsafi and N. O. Reich, Biopolymers **33**, 459 (1993).
[20] Y. Kim, K. V. Devi-Prasad, and E. W. Prohofsky, Phys. Rev. B **32**, 5185 (1985).
[21] Y. Z. Chen and E. W. Prohofsky, Phys. Rev. E **47**, 2100 (1993).
[22] M. C. Williams, I. Rouzina, and V. A. Bloomfield, Acc. Chem. Res. **35**, 159 (2002).
[23] K. Drukker, G. Wu, and G. C. Schatz, J. Chem. Phys. **114**, 579 (2001).
[24] V. V. Prabhu, L. Young, and E. W. Prohofsky, Phys. Rev. B **39**, 5436 (1989).
[25] K. Forinash and J. Keeney, J. Phys. A **25**, 6087 (1992).
[26] Y. Z. Chen, Yong-li Zhang, and E. W. Prohofsky, Phys. Rev. E **55**, 5843 (1997).
[27] J. T. Edwards and D. J. Thouless, J. Phys. C **5**, 807 (1972).
[28] R. J. Bell and P. Dean, Discuss. Faraday Soc. **50**, 55 (1970).
[29] D. J. Thouless, Phys. Rep. **13**, 93 (1974).
[30] R. D. Painter and W. M. Hartmann, Phys. Rev. B **13**, 479 (1976).
[31] X. Yang, Y. Zhuo, and X. Wu, Phys. Lett. A **234**, 152 (1997).
[32] I. F. dos Santos, F. A. B. F. de Moura, M. L. Lyra, and M. D. Coutinho-Filho, J. Phys.: Condens. Matter **19**, 476213 (2007).
[33] K. N. Woods, S. A. Lee, H.-Y. N. Holman, and H. Wiedemann, J. Chem. Phys. **124**, 224706 (2006).
[34] T. S. van Erp, S. Cuesta-Lopez, and M. Peyrard, Eur. Phys. J. E **20**, 421 (2006).
[35] D. J. Bicout and E. Kats, Phys. Rev. E **70**, 010902(R) (2004).