Available online at www.sciencedirect.com
ScienceDirect
Physica A 370 (2006) 625–631
PHYSICA A
www.elsevier.com/locate/physa

Electronic transport in DNA sequences: The role of correlations and inter-strand coupling

E.L. Albuquerque[a], M.L. Lyra[b,*], F.A.B.F. de Moura[b]

[a]*Departamento de Física, Universidade Federal do Rio Grande do Norte, 59072-970 Natal-RN, Brazil*
[b]*Departamento de Física, Universidade Federal de Alagoas, 57072-970 Maceió-AL, Brazil*

Received 16 January 2006; received in revised form 17 February 2006
Available online 27 March 2006

Abstract

We investigate the electronic properties in sequences of single and double-strand DNA molecules made up from the nucleotides guanine *G*, adenine *A*, cytosine *C* and thymine *T*. Using a tight-binding formulation we solve the time-dependent Schrödinger equation to compute the spread of initially localized wave packets. We also compute the localization length in finite segments by employing a Green's function recursion method. We compare the results for the genomic DNA sequence with those of two artificial sequences, namely the quasiperiodic Rudin–Shapiro one, which has long-range correlations, and a intra-strand pair correlated DNA sequence. We found that the short-range character of the intra-strand correlations suffices for a quantitative description of the one-electron wave-packet dynamics in the double-strand real DNA sequences. Further, the inter-strand coupling promotes electronic transport over a longer segment.

Keywords: Electronic transport; Molecular nano-structures; Anderson localization

1. Introduction

The biological task to unveil the desoxyribo-nucleic-acid (DNA) molecule [1], the basic building block of living species, is one of the main challenging quest of nowadays science. Human DNA is around 6 mm long, has about 2×10^8 nucleotides and is tightly packed in a volume equal to $500\,\mu m^3$ [2]. If a set of three nucleotides can be assumed to be analogous to a byte, then these numbers represent either $1\,Kb\,\mu m^{-1}$ (linear density) or $1.2\,Mb\,\mu m^{-3}$ (volume density), an appreciation of how densely information can be stored in the DNA molecule.

On the other hand, the design of DNA-based devices for molecular nanoelectronics is crucially dependent upon elucidation of the mechanism and dynamics of electrons and hole transport in DNA [3]. Unlike proteins, DNA is not primarily an electron/hole-transfer problem, and its suitability as a potential building block for molecular devices may not depend only on long-distance transfer of electrons and holes through the molecule. The reason for that lies in the mechanism itself: it fails to explain the persistence of efficient charge transfer when the transfer rates do not decrease rapidly with the transfer distance [4]. However, its π-stacked array of

*Corresponding author. Fax: +55 82 214 1645.
E-mail address: marcelo@df.ufal.br (M.L. Lyra).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2006.02.043

base pairs, here considered as a symbolic sequence of a four-letter alphabet representing the four different nucleotides, namely guanine (G), adenine (A), cytosine (C) and thymine (T), does indeed provide appropriate pathway for long-range charge transport [5], although the mechanism for long-range transport and short-range transfer may differ entirely [6]. Strong stacking interactions result in the fastest electron-transfer kinetics, whereas dynamical motion of the base pairs and reactant energetics also modulate the distance dependence of DNA-mediated charge transport, reducing its efficiency [7].

Although accurate determination of the electronic conductivity in DNA molecule is important, the roles played by the long and short-range correlations on electronic delocalization still deserve further investigations [8]. Based on a simple ladder model of a double-strand DNA, it was shown that correlations and asymmetry of the sequence affect the electron localization [9,10]. Recently, we have proposed a single-strand DNA sequence modeling the long and short-range correlations by a quasiperiodic Rudin–Shapiro and a random sequences, respectively [11]. Although the long-range correlations were responsible for the effective electronic transport at specific resonant energies of finite DNA segments, much of the anomalous spread of an initially localized electron wave packet can be accounted by short-range pair correlations on DNA. This finding suggests that a systematic approach based on the inclusion of further short-range correlations on the nucleotide distribution can provide an adequate description of the electronic properties of DNA segments.

The problem involving the spread of one electron wave-functions in low-dimensional disordered systems is a well-known issue with several connections with transport properties [12]. In general lines, the wave function of an electron moving in a perfectly periodic potential spreads linearly in time. In the presence of uncorrelated disorder, the scaling theory predicts the absence of extended eigenstates [13] in one-dimensional (1D) systems. Therefore, the width of the time-dependent wave-function saturates in the long-time limit, i.e., the electron wave function remains localized in a finite region around the initial position. The scaling theory prediction of exponential localization of all one-electron eigenfunctions in 1D systems can be violated when special short-range [14] or long-range [15] correlations are present in the disorder distribution. The influence of scale-free disorder in the 3D Anderson transition has also been recently addressed [16]. In particular, the presence of dimer-like correlations on a N-site binary chain produces $\sqrt{N}$ extended states. These states have random phase changes when crossing the dimer impurities which results in a finite coherence length. If the energy of the resonant extended state is within the band of allowed states of the underlying pure chain, the electron wave-packet experiences a super-diffusive spread [14]. The propagating properties in the random dimer-model was analyzed in detail in Ref. [17] where the action of a DC electric field was investigated. Other examples of failure of the scaling theory of localization are Fibonacci [18] and Thue–Morse [19] lattices, which show super-diffusive propagation.

In this paper, we report a numerical study of the one-electron wave-packet dynamics in arrays of DNA *single* and *double-strand* segments, made up from four nucleotides either following a Rudin–Shapiro (RS) quasiperiodic sequence, which has long-range correlations, or a random sequence, which is a kind of prototype of a short-range correlated system, presented here with the same first neighbors pair correlations of the human DNA sequence (we call it the PC sequence). The results are compared with those obtained using real DNA sequence, which is usually claimed to have long-range correlations in their base-pair sequences [5,8,9,20–22].

2. Model

Our Hamiltonian is an effective tight-binding model describing one electron moving in a ladder geometry and open boundaries. We consider a single orbital per site and nearest-neighbor interactions, whose time-dependent Schrödinger equation is given by ($\hbar = 1$) [23]

$$\mathrm{i}\,\mathrm{d}\psi_{j,m}/\mathrm{d}t = \sum_{x=-1}^{1} V_{j,j+x}^{m}\psi_{j+x,m} + t_j\psi_{j,n} + \varepsilon_{j,m}\psi_{j,m}, \tag{1}$$

where $n \neq m$. Here, $\varepsilon_{j,m}$ is the single energy at the orbital $|j,m\rangle$ (j is the index running along the segment and $m = 1, 2$ indexes each strand) and $\psi_{j,m}$ is the wave-packet coefficient in the single orbital basis

$$|\Psi\rangle = \sum_{j,m} \psi_{j,m}|j,m\rangle. \tag{2}$$

Also, $V^m_{j,j+x}$ is the longitudinal first-neighbor electronic hopping between sites j and $j+x$ at strand m. In the DNA molecule such hopping amplitude is effectively mediated by side chains of sugar phosphate. Further, t_j is the transversal hopping connecting the strands at point j. We consider a wave packet initially localized at site $|j_0, m_0\rangle$, i.e.,

$$\psi_{j_0,m_0}(t=0) = \delta_{j,j_0}\delta_{m,m_0}. \tag{3}$$

A fourth-order Runge–Kutta method is used to solve the above set of coupled differential equations. We will be particularly interested to calculate the square root of the mean-square displacement $\sigma_2(t)$ defined by

$$\sigma_2(t) = \sqrt{\sum_{j=1}^{N}\sum_{m=1}^{2}[(j-j_0)^2 + (m-m_0)^2]|\psi_{j,m}(t)|^2}, \tag{4}$$

where N is the number of nucleotides.

For the DNA sequence of the first sequenced human chromosome 22 (Ch22), entitled NT_{011520}, the number of letters of this sequence is about 3.4×10^6 nucleotides. This sequence was retrieved from the internet page of the National Center of Biotechnology Information. The on-site energies in one of the DNA strands $\varepsilon_{j,1}$ were chosen from the ionization potential of the respective nucleotides. The available theoretical values in the literature obtained from ab initio calculations are quite uncertain. Here we will use as a representative set of ionization potentials those provided in Ref. [24], i.e., $\varepsilon_A = 8.24$, $\varepsilon_T = 9.14$, $\varepsilon_C = 8.87$, and $\varepsilon_G = 7.75$, all units in eV, representing the adenine, thymine, cytosine and guanine molecules, respectively. The energies of the second strand $\varepsilon_{j,2}$ were generated by considering the inter-strand nucleotide correlations which imposes that only base pairs CG and AT can be allowed. In DNA molecules, the intra-strand hopping amplitude is smaller than the disorder width due to the variability of the on-site energies. The inter-strand coupling mediated by the hydrogen bonds between complementary sequences is weaker than the intra-strand coupling. In order to reproduce some specific features of real DNA molecules, it would be important to consider both inter- and intra-strand hopping variability. Here we will concentrate on a tight-binding ladder model with unitary intra- and inter-chain hopping amplitudes. Our aim is to explore the relative role played by intra-strand correlations and inter-strand coupling in the localized nature of the one-electron states as well as in the wave-packet spreading. Although specific details of the band structure and the precise spacial extension of the localized states can depend on the actual relative value of inter- and intra-strand couplings, the influence of correlations on the wave-packet propagation can be well described by the present model with a single value for both hopping amplitudes.

We also compute the Lyapunov exponents of double-strand segments using the following Green's function recursion method based on Dyson's equation (see Ref. [25] for details):

$$G^{n+1}_{n+1,n+1} = [E - H^0_{n+1,n+1} - V_{n+1,n}G^n_{n,n}V_{n,n+1}]^{-1}, \tag{5}$$

with

$$G^{n+1}_{1,n+1} = G^n_{1,n}V_{n,n+1}G^{n+1}_{n+1,n+1}, \tag{6}$$

where $G^{n+1}_{1,n+1}$ denotes the 2×2 Green's function operator between the first and the $(n+1)$th base pairs. Also, $G^{n+1}_{n+1,n+1}$ and $H^0_{n+1,n+1}$ are the Green's function operator and the free Hamiltonian for the isolated $(n+1)$ base pair, $V_{n,n+1}$ is the diagonal 2×2 matrix coupling the base pairs at position n and $n+1$, and E is the diagonal 2×2 matrix for the electron energy. We start this recursive equation considering as initial values $G^1_{1,1} = I$ (identity matrix) and $G^0_{0,0} = 0$.

The Lyapunov exponent for a double-strand DNA segment is given by

$$\gamma_2(E) = (1/2N)\ln[\mathrm{Tr}|G^{N+1}_{1,N+1}|^2]. \tag{7}$$

It is related to the localization length in the double-strand by $\lambda_2 = 1/\gamma_2$. For extended states, $\gamma_2(E)$ vanishes in the thermodynamic limit. In what follows, the quantities $\sigma_1(t)$ (wave-packet spread) and $\gamma_1(E)$ (Lyapunov coefficient), related to the single-strand segments, will also be reported allowing a quantitative analysis of the role played by the coupling between strands in the nature of one-electron states.

In order to elucidate the possible relevance of intra-strand nucleotide correlations, we will compare the one-electron behavior in segments of Ch22 (a natural DNA sequence) with the behavior in artificial sequences having long-range correlations, such as the quasi-periodic Rudin–Shapiro (RS) sequence, as well as in artificial sequences containing only first neighbors correlations. The quasiperiodic RS sequence can be built, starting from a G (guanine) nucleotide as seed, through the inflation rules $G \to GC$, $C \to GA$, $A \to TC$, and $T \to TA$. The RS sequence belongs to the family of the so-called substitutional sequences, which are characterized by the nature of their Fourier spectrum. It exhibits an absolutely continuous Fourier measure, a property it shares with the random sequence (for a review of the physical properties of these and others quasiperiodic structures see Ref. [26]). With the intention of comparing this sequence to the genomic one, we assume that the energies $\varepsilon_{j,m}$ take the four different values ε_G, ε_A, ε_C, and ε_T as in the DNA genomic sequence.

We also constructed a random sequence with the same inter-chain pair correlation of the Ch22 chromosome sequence [11]. To this end we firstly measured the fraction of pairs $p(I,J)$ ($I,J = G,A,C$ and T) in one of the strands of the Ch22 sequence considered. After that, we generate an artificial sequence starting with a cytosine C site. During the sequence construction, the nucleotide following a type I nucleotide is chosen to be of type J with probability $P(J,I) = p(I,J)/\sum_K p(I,K)$. The resulting pair correlated (PC) sequence has only short-range correlations which, by construction, are the same as in the original Ch22 sequence. For completeness, we will further report the results obtained for uncorrelated (fully random) sequences containing the same fraction of each nucleotide found in the Ch22 sequence considered.

3. Results and discussion

We start by reporting the localization length as computed using the Green's function technique for single and double-strand sequences consisting of $N = 512$ bases. We used a typical Ch22 sequence starting at the first cytosine after the 10^5th nucleotide. The segment size and initial position were chosen arbitrarily to represent typical DNA sequences. Although some specific features may change for different choices, we expect that the main picture concerning the interplay between coupling and correlations will be fairly represented. All artificial sequences have the same length. In Fig. 1a we compare the results for all single-strand sequences studied (Ch22, RS, PC and random). Notice that the localization length for the Ch22 sequence reaches a maximum value around the band center with an average value of the order of 40 nucleotides. A similar picture is found for a completely uncorrelated sequence but the average maximum value of the localization length is

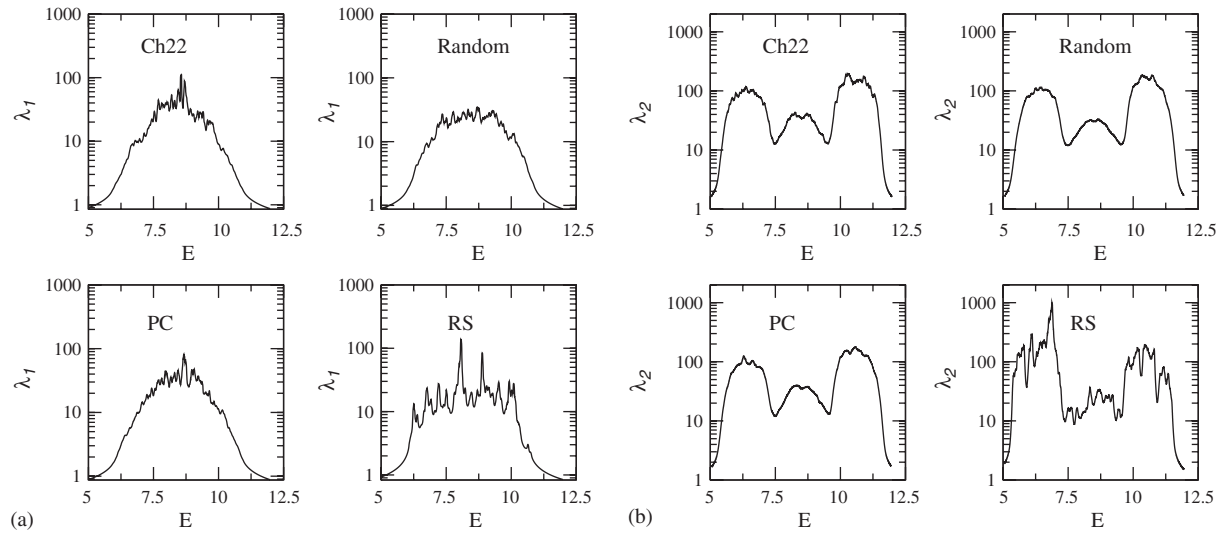

Fig. 1. Localization length as a function of the energy (E) for (a) single and (b) double-strand DNA segments. The coupling between strands promotes the emergence of sidebands. Although Ch22, PC and the random sequences display roughly the same behavior, the long-range character of the correlations in the RS sequence is signaled by a strong asymmetry, larger localization lengths and a complex peak structure.

somewhat smaller. This feature points out to a relevant role played by the nucleotide correlations on the nature of the one-electron states. When first neighbors correlations are introduced, the localization length assumes roughly the same quantitative behavior found in the Ch22 sequence. On the other hand, for the long-range correlated RS sequence, the localization length within the energy band depicts new features. Although, in average, it is not larger than in the Ch22 sequence, the localization length in RS segments exhibits a sequence of peaks over a wider energy window which reflects the scale invariance of the spectrum.

For double-strand sequences (see Fig. 1b) the coupling between strands induces the emergence of sidebands. The localization length for either Ch22, PC and random sequences are quite similar, which points out to the irrelevance of correlations to the nature of one-electron states. Although the localization length at the center of the band is of the same magnitude as in the single-strand sequence, at the sidebands it reaches a maximum average value which is roughly three times larger. Therefore, the double-strand nature has a strong impact on the spacial extent of the electronic states. The coupling between strands, thus favors electron transport over longer distances. However, the localization length is finite indicating an effective absence of long-range electron transport. The localization length in the double-strand RS sequence is qualitatively and quantitatively distinct. First of all, it exhibits a strong asymmetry due to the fact that long segments of a single nucleotide pair are generated by the RS inflation rule. The quasi-periodicity also leads to longer localization lengths, which for the double-strand RS sequence has a small energy window with $\lambda_2 > N$ at which electron transmission is allowed. For this quasiperiodic sequence the coupling between strands leads to a substantial increase of the average localization length.

To study the spread of one-electron wave function, we start from a wave packet localized at the guanine G site closer to the center of the single- and double-strand segments. In order to avoid finite-size effects, we used larger segments with $N = 1500$. For the Ch22, PC and random sequences an average over 20 distinct segments was employed, to account for configurational variability. Typical results are depicted in Fig. 2. For the wave-packet spread over a single-strand sequence (see Fig. 2a), the long-range correlations in the RS sequence results in a wave-packet spread over a segment which overpass the one achieved in the Ch22 strand by a factor of the order of 1.5. On the other hand, the spread in a completely uncorrelated sequence is just half of that in Ch22, pointing to the importance of the nucleotide correlations. The fact that the spread in the PC sequence is already $\frac{3}{4}$ of the spread in the natural Ch22 sequence led to the conjecture that the systematic inclusion of further short-range correlations might be enough to capture the correct one-electron dynamics in DNA molecules [11]. Prior to the saturation the mean-square displacement presents in all cases a diffusive-like spread after a short initial transient.

The above trend concerning the role played by short and long-range correlations is further strengthened when we analyze the wave-packet spread in double-strand sequences as shown in Fig. 2b. In this case, the

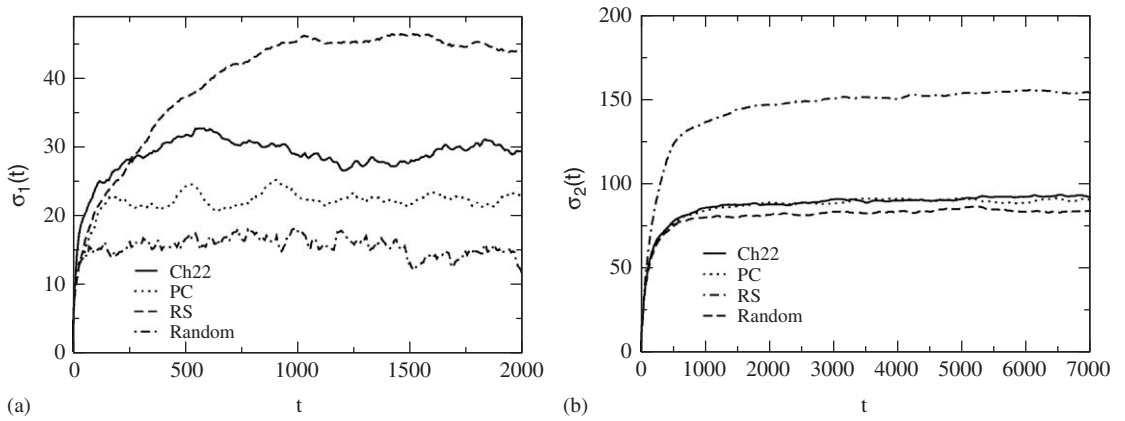

Fig. 2. (a) Spread of the wave function, defined by the time-dependent square root of the mean-square displacement, as a function of time (t) for several kinds of single-strand sequences. (b) The wave-packet spread in double-strand sequences. The long-range correlations in the RS sequence induces a large wave-function spread as compared to the Ch22 sequence. Short-range correlation of the PS sequence matches nicely those of the Ch22 human chromosome. Error bars at the long time-regime are of the order of 3 base pairs.

spread in the long-range correlated RS sequences becomes much larger when compared to the single-strand case. This fact is associated to the larger values achieved by the localization length of stationary states, as previously discussed. When compared with the spread in double-strand Ch22 sequences, the RS one allows for the wave-packet spread over a segment which is twice as large due to the excess of correlations. The coupling between strands in Ch22 favored the electron spread which now reaches a segment almost three times larger than in the single-strand sequences. It is instructive to notice that the spread in an uncorrelated double-strand sequence is already close to that in the Ch22, and the inclusion of first neighbors correlations suffices to achieve the same wave-packet spread in Ch22. The above results indicate that long-range correlations are not relevant for the one-electron dynamics in DNA and that the inclusion of just first-neighbors correlations may be enough to have a quantitative description of the wave-packet spread in double-strand sequences.

4. Final remarks and conclusion

In summary, aiming to unreveal the actual relevance of short and long-range correlations on the electronic properties of DNA-like sequences, we numerically study the one-electron wave-packet dynamics in arrays of single- and double-strand segments. We have modeled DNA sequences considering its four nucleotides following two artificial sequences, namely the RS quasiperiodic sequence, with a long-range pair correlation, and a pair-correlated random sequence (PC), with the same first neighbors pair correlations of the human Ch22 DNA sequence, which shows short-range pair correlation behavior. The anomalous spread of an initially localized electron wave packet is accounted by the short-range pair correlations, suggesting that the inclusion of just first-neighbors intra-strand pair correlations on the nucleotide distribution provides an adequate description of the DNA's electronic properties. Further, we have shown that the coupling between strands allows for the electron transport over a longer segment. However, such enhancement shall depend strongly on the intra-chain coupling and further studies considering more realistic model parameters would be needed in order to infer about the actual relevance of this spreading enhancement in DNA molecules.

We would like to stress that, although in all simulations we have chosen to start the wave-packet spread with the electron initially localized at a guanine nucleotide, the main trend observed shall be mainly independent of this choice. Also, the double-strand model used here does not account for variability of the hopping amplitudes and their dependence on the electron energy. Such features may be included by explicitly taking into account the transport along the sugar phosphate side chains. Although, we do not expect any relevant change in the main features related to the one-electron dynamics, the energy band structure and the actual localization length may be influenced, specially at the band edges. A few other points deserve further investigation on the light of the present tight-binding modeling. Transport in coding and non-coding sequences may have a signature of the distinct nucleotide correlations present on them. Also, the choice of the set of ionization potentials of the DNA bases and the inter- and intra-strand hopping amplitudes determines the main features of the density of states. The possible emergence of sub-bands has to be considered once energy gaps usually play a relevant role in transport properties [27]. Finally, the backbone of sugar phosphate that connects the DNA bases may be included explicitly in the tight-binding Hamiltonian. This will allow the use of more appropriate parameters provided by ab initio calculations necessary to reproduce closer some details of the electronic density of states. We expect to address these points in a future communication.

Acknowledgments

We would like to thank partial financial support from CNPq, CNPq-CTEnerg, CNPq-BioNanoStructures network, FINEP-CTInfra, and FAPEAL (Brazilian Research Agencies).

References

[1] J.D.H. Watson, F.H.C. Crick, Nature 171 (1953) 737.
[2] H. Lodish, D. Baltimore, A. Berk, S.L. Zipursky, P. Matsudaira, J. Darnell, Molecular Cell Biology, Scientific American Books Inc., 1995.
[3] F.D. Lewis, X.Y. Liu, J.Q. Liu, S.E. Miller, R.T. Hayes, M.R. Wasielewski, Nature 406 (2000) 51.

[4] B. Giese, J. Amaudrut, A.K. Kohler, M. Spormann, S. Wessely, Nature 412 (2001) 318.
[5] S. Roche, D. Bicout, E. Maciá, E. Kats, Phys. Rev. Lett. 91 (2003) 228101;
S. Roche, D. Bicout, E. Maciá, E. Kats, Phys. Rev. Lett. 92 (2004) 109901 (E).
[6] D. Porath, A. Bezryadin, S. de Vries, C. Dekker, Nature 403 (2000) 635;
H.W. Fink, C. Schonenberger, Nature 398 (1999) 407.
[7] S.O. Kelley, J.K. Barton, Science 283 (1999) 375.
[8] F.A.B.F. de Moura, M.L. Lyra, Phys. Rev. Lett. 81 (1998) 3735.
[9] H. Yamada, Phys. Lett. A 332 (2004) 65.
[10] K.D. Klotsa, R.A. Römer, M.S. Turner, Biophys. J. 89 (2005) 2187.
[11] E.L. Albuquerque, M.S. Vasconcelos, M.L. Lyra, F.A.B.F. de Moura, Phys. Rev. E 71 (2005) 021910.
[12] B. Kramer, A. MacKinnon, Rep. Prog. Phys. 56 (1993) 1469;
F. Domínguez-Adame, V.A. Malyshev, F.A.B.F. de Moura, M.L. Lyra, Phys. Rev. Lett. 91 (2003) 197402;
R.A. Romer, H. Schulz-Baldes, Europhys. Lett. 68 (2004) 247.
[13] E. Abrahams, P.W. Anderson, D.C. Licciardello, T.V. Ramakrishnan, Phys. Rev. Lett. 42 (1979) 673.
[14] D.H. Dunlap, H.L. Wu, P.W. Phillips, Phys. Rev. Lett. 65 (1990) 88;
H.-L. Wu, P. Phillips, Phys. Rev. Lett. 66 (1991) 1366.
[15] F.M. Izrailev, A.A. Krokhin, Phys. Rev. Lett. 82 (1999) 4062.
[16] M.L. Ndawana, R.A. Romer, M. Schreiber, Europhys. Lett. 68 (2004) 678.
[17] C.A.A. da Silva, P.E. de Brito, H.N. Nazareno, Phys. Rev. B 52 (1995) 7775.
[18] G. Gumbs, M.K. Ali, Phys. Rev. Lett. 60 (1988) 1081.
[19] Y. Avishai, D. Berend, Phys. Rev. B 45 (1992) 2717.
[20] J. Maddox, Nature 358 (1992) 103.
[21] V.R. Chechetkin, A.Y. Turygin, Phys. Lett. A 179 (1995) 75.
[22] C. Vaillant, B. Audit, A. Arneodo, Phys. Rev. Lett. 95 (2005) 068101.
[23] M. Kohmoto, L.P. Kadanoff, C. Tang, Phys. Rev. Lett. 50 (1983) 1870.
[24] H. Sugiyama, I. Saito, J. Am. Chem. Soc. 118 (1996) 7063.
[25] A. MacKinnon, B. Kramer, Phys. Rev. Lett. 47 (1981) 1546.
[26] E.L. Albuquerque, M.G. Cottam, Polaritons in Periodic and Quasiperiodic Structures, Elsevier, Amsterdam, 2004.
[27] O.R. Davies, J.E. Inglesfield, Phys. Rev. B 69 (2004) 195110.