ARTICLE IN PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Surface Science xxx (2006) xxx–xxx

SURFACE SCIENCE

www.elsevier.com/locate/susc

Electronic density of states in sequence dependent DNA molecules

B.P.W. de Oliveira [a], E.L. Albuquerque [a,*], M.S. Vasconcelos [b]

[a] *Departamento de Física, Universidade Federal do Rio Grande do Norte, Campus Universitario, 59072-970 Natal, RN, Brazil*
[b] *Departamento de Ciências Exatas, Centro Federal de Educação Tecnológica do Maranhão, 65025-001 São Luís, MA, Brazil*

Abstract

We report in this work a numerical study of the electronic density of states (DOS) in π-stacked arrays of DNA single-strand segments made up from the nucleotides guanine G, adenine A, cytosine C and thymine T, forming a Rudin–Shapiro (RS) as well as a Fibonacci (FB) polyGC quasiperiodic sequences. Both structures are constructed starting from a G nucleotide as seed and following their respective inflation rules. Our theoretical method uses Dyson's equation together with a transfer-matrix treatment, within an electronic tight-binding Hamiltonian model, suitable to describe the DNA segments modelled by the quasiperiodic chains. We compared the DOS spectra found for the quasiperiodic structure to those using a sequence of natural DNA, as part of the human chromosome Ch22, with a remarkable concordance, as far as the RS structure is concerned. The electronic spectrum shows several peaks, corresponding to localized states, as well as a striking self-similar aspect.

Keywords: Computer simulation; Electron density; Excitation spectra calculations; Green's function method; Electrical transport; Biological molecules–nucleic acids

1. Introduction

In the last decade, there have been dramatic advances toward the realization of devices and integrated computers at the molecular scale [1]. One reason for that lies on the striking advances in molecular biology and nanotechnology, which are open up the possibility to explore the interface between biology and electronics at the single-molecule level [2]. In fact, the use of molecules as electronic components is a powerful new direction in the science and technology of nanometer-scale systems, due to their scientific and engineering implications [3]. First pioneering experiments were performed demonstrating that individual molecules can operate as switches one thousand times smaller than those on conventional microchips [4]. The ultimate limit would be a device where electrons hop on to and off from a single atom between two contacts, which has been receiving considerable attention recently. Although the basic process that underlies the function of molecular electronic devices, i.e., electron transport through molecules and clusters, is still not well understood, several research groups are striving to fabricate novel molecular electronic structures and to develop fundamental insights about their behavior [5].

Researchers are working to join biology and nanotechnology, fusing useful biomolecules, such as chemically synthesize DNA, in arrangements that do everything from emitting light [6] to storing tiny bits of magnetic data [7]. One of their fundamental goals is the realization of nanoscale devices in which a few or a single biomolecule can be used to transfer and process an electronic signal. The biomolecules have particular functionality that can be exploited for the implementation of electronic devices [8]. The combination of molecular biology (for engineering proteins with the desired functional and/or self-assembling properties) and nanotechnology (for device fabrication) thus becomes the tool to realize a new class of nanoelectronic elements [9]. Different nanotechnological

* Corresponding author. Tel.: +55 842153793; fax: +55 842153791.
E-mail address: eudenilson@dfte.ufrn.br (E.L. Albuquerque).

0039-6028/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.susc.2006.01.081

strategies have been selected to implement the biomolecular devices, following a bottom-up or a top-down approach depending on the biomolecule and on its functionality [10].

The result is a merger of the biology's ability to assemble complex structures with the nanoscientists' capacity to build useful devices. One of the biggest drivers behind these tasks revolves around the nature's impressive ability to manufacture and assemble complex molecules with great accuracy and high efficiency using specific biological molecules such as DNA and proteins, with atomic precision [11]. The realization of complex DNA-based circuits will, however, require new concepts and additional biological machinery allowing, for example, feedback from the electronic functionality to direct the assembly process and adaptation mechanisms. Furthermore, although the use of DNA molecules in nano-electronic circuits is a very promising task due to their self-assembly and molecular recognition abilities, their conductivity properties are still under intense and controversial debate [12].

A DNA chain is a sequence of four possible nucleotides which define the structure of the amino acids to form proteins. It can be considered as a symbolic sequence of a four letter alphabet, namely guanine (G), adenine (A), cytosine (C) and thymine (T). Unlike proteins, a stacked array of DNA base pairs derived from these nucleotides can provide the way to promote long-range charge migration, which in turn gives important clues to mechanisms and biological functions of transport [13]. Numerous algorithms have been introduced to characterize and graphically represent the genetic information stored in the DNA nucleotide sequence. The goal of these methods is to generate patterns for certain sequences or groups of sequences.

With this aim in mind, we report in this work the electronic density of states of a DNA molecule by using a tight-binding Hamiltonian, together with a transfer-matrix within a Dyson's framework, employed to simplify the algebra which can be otherwise quite involved. We consider a model in which the DNA molecule is sandwiched in a substrate, following a Fibonacci (FB) and a Rudin–Shapiro (RS) quasiperiodic structures, and compare them to the DNA sequence of the first sequenced human chromosome 22 (Ch22), entitled NT_{011520}, whose arrangement was retrieved from the internet page of the National Center of Biotechnology Information.

This paper is structured as follows: we present in Section 2 our theoretical model based on an electronic tight-binding Hamiltonian suitable to describe a single-strand of DNA segments with pure diagonal correlated disorder modelled by the quasiperiodic chain of FB and RS type. The numerical results and the conclusions of this work are presented in Section 3.

2. General theory

For a single-strand DNA chain, our tight-binding Hamiltonian is written in terms of a localized basis as

$$H = \sum_n \omega_n |n\rangle\langle n| + \sum_{n,m} V_{nm} |n\rangle\langle m|, \tag{1}$$

where ω_n represents the energy (in units of $\hbar$) of the site n, and V_{nm} is the hopping potential. The sum over m is limited to the nearest neighbors.

The Dyson equation is

$$G(\omega) = \omega^{-1}[I + HG(\omega)], \tag{2}$$

where I is the identity matrix and H is the Hamiltonian given by (1).

To setup a quasiperiodic chain of Fibonacci type, we start from a G (guanine) nucleotide as seed and the quasiperiodic FB sequence can be built through the inflation rules G → GC and C → G. The initial case that we investigate is for the first generation of the FB sequence, with only a guanine linked to the substrate. We determine its Green function by applying the tight-binding Hamiltonian (1) to the Dyson equation (2) to get

$$G_{nn}^{-1} = \omega - \omega_G + 2\gamma(1), \tag{3}$$

where

$$\gamma(1) = -\frac{V_{GS}^2}{\omega - \omega_S + V_{SS}T(\omega)}. \tag{4}$$

Here, $T(\omega)$ is the transfer function given by

$$T(\omega) = -(2V_{SS})^{-1}\left[(\omega - \omega_S) \pm \sqrt{(\omega - \omega_S)^2 - 4V_{SS}^2}\right]. \tag{5}$$

Repeating the procedure for any Fibonacci generation, we observe a difference between the G_{nn} terms for odd and even generations, namely,

$$G_{nn}^{-1} = \omega - \omega_G + \gamma(1) + \kappa(N) \tag{6}$$

for odd generations, while for even generations we have

$$G_{nn}^{-1} = \omega - \omega_G + \gamma(1) + \gamma(N), \tag{7}$$

where

$$\kappa(N) = -\frac{V_{ij}^2}{\omega - \omega_i + \kappa(N-1)}, \tag{8}$$

with a similar expression for $\gamma(N)$. Here N is the number of nucleotides in the strand. The hopping potential $V_{i,j}$, with $j = i - 1$, can assume two distinct values: V_{GG} and V_{GC}. Both V_{GS} and V_{CS} represent the interaction of the strand with the substrate. The symmetry $V_{i,i-1} = V_{i-1,i}$ holds.

The results for the single-stranded Fibonacci sequence are contrasted with the Rudin–Shapiro (RS) sequence (another substitutional sequence) which displays an absolutely continuous Fourier measure, a property which it shares with the random sequences [14]. Starting also from a G (guanine) nucleotide as seed, the quasiperiodic RS sequence is then built through the inflation rules G → GC, C → GA, A → TC, and T → TA.

In analogy with the FB sequence, we start our study of the single-stranded DNA chain grown through a RS sequence by investigating the behavior of the organic

nanostructured circuit for a single guanine linked to the substrate. Its Green function will be the same as Eq. (3) because both FB and RS sequences are identical for this generation. The RS sequence starts to deviate from the FB sequence in the third generation, when we will have a sequence of four nucleotides GCGA connected to the substrate. Using a procedure similar to the quasiperiodic Fibonacci case, we can get for any even RS generation number the same expression as for the Fibonacci case (the strand ends with a cytosine nucleotide C), while for odd generation numbers (ending with a adenine nucleotide A) the expression is similar to Eq. (6) provided we replace $\kappa(N)$ by $\alpha(N)$ given by

$$\alpha(N) = -\frac{V_{i,i-1}^2}{\omega - \omega_i + \alpha(N-1)}. \tag{9}$$

Differently from the FB sequence, for the Rudin–Shapiro sequence $V_{i,i-1}$ represents four distinct values of hopping potentials: V_{CT}, V_{GC}, V_{GA}, and V_{TA} (while in FB we only have two hopping potentials: V_{GC} and V_{GG}), where we have assumed that $V_{i,i-1} = V_{i-1,i}$ in both cases.

3. Numerical results and conclusions

We will now discuss the main results found in this work. The density of state (DOS) for the systems is given by

$$\rho(\omega) = -(1/\pi)\mathrm{Im}[\mathrm{Tr}\langle n|G(\omega)|n\rangle], \tag{10}$$

where Im means the imaginary part of the argument shown between brackets. The energies ω_j are chosen from the ionization potential of the respective nucleotides, i.e., $\omega_A = 8.24$ eV (adenine), $\omega_C = 8.87$ eV (cytosine), $\omega_G = 7.75$ eV (guanine), and $\omega_T = 9.14$ eV (thymine) [15–17]. All the hopping terms V_{nm} among the bases were taken equal to 1 eV, considering that theoretical calculations using ab initio methods yield for this potential values in the range 0.4–1 eV [15–17]. The potential at the interface DNA-substrate (here considered as a platinum electrode) is considered to be the difference between the Fermi's level of the platinum and a HOMO's (highest occupied molecular orbital) isolate guanine state, giving us $V_{GS} = 2.39$ eV. We are aware that the HOMO state of the guanine may significantly change in the presence of the substrate, yielding a different potential at the interface DNA-substrate. Although we do not expect any relevant change in the DOS main features, the actual electron's localization length may be influenced, specially at the band edges. The hopping term inside the electrode is 12 eV [18]. Further, the on-site energy for the substrate (platinum) is $\omega_S = 5.36$ eV, which is related with the work function of this metal [19].

Fig. 1 depicts the density of states for a DNA quasiperiodic chain following an even (Fig. 1(a)) and odd (Fig. 1(b)) generation of a Fibonacci quasiperiodic sequence, respectively. Here N_{FB} means the sequence generation number, while n_{FB} corresponds to the number of nucleotides in a given sequence generation. From there we can infer the following main properties:

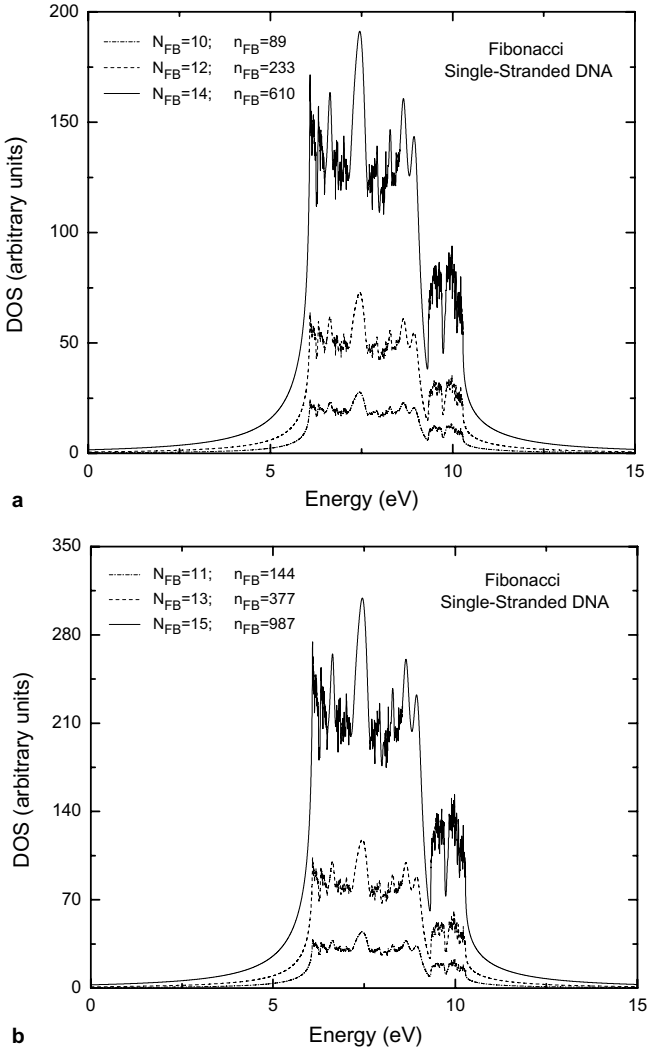

Fig. 1. Density of state spectra for the Fibonacci Poly GC DNA single-strand model: (a) even generation numbers corresponding to the 10th (chain-dotted line), 12th (dashed line), and 14th (full line) FB sequence generation and (b) odd generation numbers corresponding to the 11th (chain-dotted line), 13th (dashed line), and 15th (full line) FB sequence generation.

(a) The parity of the FB generation is not important for the DOS spectrum.
(b) Although the DOS for each generation as a whole does not show any symmetry, there are two very well defined and symmetrical regions, lying in the intervals (in units of eV) $5.75 < \omega < 9.30$ (we call it region I), and $9.30 < \omega < 10.30$ (region II).
(c) Region II, which appears as a sort of anomaly in the DOS spectrum, is due to the presence of the cytosine nucleotide in the quasiperiodic chain. We can also notice that this region represents a kind of the profile of the region I inverted and in a smaller scale.
(d) Each region defines a clearly auto-similar spectrum for different generations. The auto-similarity holds also for the whole spectrum (regions I + II).

(e) The central peak for region I is next to the guanine's ionization energy $\omega_G = 7.75$ eV, while the central valley in region II corresponds to $\omega_G + 2V_{nm} = 9.75$ eV.
(f) The ratio among the distances of consecutive generations tends to the gold mean, $\tau = (1 + \sqrt{5})/2$, a number intrinsically linked to the Fibonacci sequence.

The density of states for a DNA quasiperiodic chain following a Rudin–Shapiro quasiperiodic sequence are shown in Fig. 2(a), for the number of nucleotides $n_{RS} = 256$, corresponding to the 9th RS sequence generation, and Fig. 2(b), for $n_{RS} = 512$, corresponding to the 10th RS sequence generation. Although some similarities with the Fibonacci case persist (for instance, the asymmetry of the spectra and the fact that again the parity of the quasiperiodic generation is not important), they are completely different, indicating how important is the model considered to simulate the DNA structure. As their main features, their central peaks, which are sequence independent, lie around 6.8 eV (which is about $\omega_C - V_{nm}$), with the band-width approximately given by $\omega_G \pm 4V_{nm}$. More important, when one compares these spectra with those generated from a sequence of natural DNA, as part of the human chromosome Ch22, with the same number of nucleotides, a remarkable agreement is found, as they are depicted in Fig. 2(a) and (b).

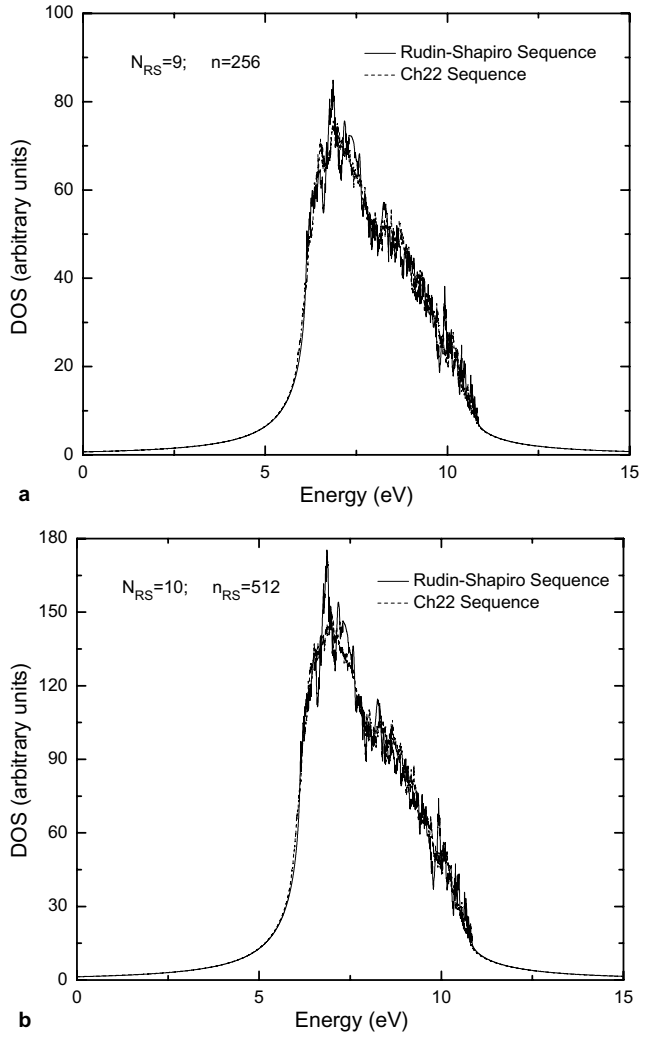

Fig. 2. Density of state spectra for the Rudin–Shapiro (full line) and part of the human chromosome Ch22 (dashed line) models: (a) $n_{RS} = 256$, corresponding to the 9th RS sequence generation and (b) $n_{RS} = 512$, corresponding to the 10th RS sequence generation.

To summarize, we have studied the electronic density of states in one-dimensional DNA single-strand structure modelled by the quasiperiodic Fibonacci and Rudin–Shapiro sequences, aiming to further contribute to the present understanding of the role played by correlations on the electronic properties of DNA segments. In order to unreveal the actual relevance of long-range correlations, which is a kind of signature of the quasiperiodic sequences, we compared the DOS spectra considering segments of the Ch22 human chromosome with those resulting from the quasiperiodic Rudin–Shapiro sequence, with a remarkable agreement. Furthermore, the long-range correlations present in Ch22 and RS sequences are responsible for the slow vanishing of their DOS spectra, which may promote an effective electronic transport at specific resonant energies of finite DNA segments. Also, in order to model specific transport properties of DNA molecules it would be important not only to consider their double-strand character [20], but the different values assumed by the coupling constant between distinct pairs of nucleotides [21] as well, which in turn have important consequences in the on-site (ionization) energies ω_j.

Acknowledgements

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

References

[1] C. Joachim, J.K. Gimzewski, A. Aviram, Nature 408 (2000) 541.
[2] A. Nitzan, M.A. Ratner, Science 300 (2003) 1384.
[3] S. Luryi, J.M. Xu, A Zaslavsky (Eds.), Future Trends in Microelectronics: The Road Ahead, Wiley, New York, 1999.
[4] M.A. Reed et al., Appl. Phys. Lett. 78 (2001) 3735.
[5] J. Park et al., Nature 417 (2002) 722;
W. Liang et al., Nature 417 (2002) 725.
[6] H. Wadati et al., Appl. Phys. Lett. 86 (2005) 23901.
[7] M. Zwolak, M. Di Ventra, Appl. Phys. Lett. 81 (2002) 925.
[8] C. Nicoloni, Molecular Bioelectronics, World Scientific, Singapore, 1996.
[9] S.-T. Yau, G. Qian, Appl. Phys. Lett. 86 (2005) 103508.
[10] E. Braun, K. Keren, Adv. Phys. 53 (2004) 441.
[11] R.G. Endres, D.L. Cox, R.R.P. Singh, Rev. Mod. Phys. 76 (2004) 195.
[12] E.L. Albuquerque, M.S. Vasconcelos, M.L. Lyra, F.A.B.F. de Moura, Phys. Rev. E 71 (2005) 021910.
[13] 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).

[14] E.L. Albuquerque, M.G. Cottam, Polaritons in Periodic and Quasiperiodic Structures, Elsevier, Amsterdam, 2005.
[15] H. Sugiyama, I. Saito, J. Am. Chem. Soc. 118 (1996) 7063.
[16] A.A. Voityuk et al., J. Chem. Phys. 114 (2001) 5614.
[17] H. Zhang et al., J. Chem. Phys. 117 (2002) 4578.
[18] E. Maciá, F. Triozon, S. Roche, Phys. Rev. B 71 (2005) 113106.
[19] Y.A. Berlin, A.L. Burin, M.A. Ratner, Chem. Phys. 275 (2002) 61.
[20] H. Yamada, Phys. Lett. A 332 (2004) 65.
[21] Y.J. Yan, H. Zhang, J. Thermoplast. Compos. Chem. 1 (2002) 225.