Electronic Transport in Double-Strand DNA Segments

E.L. Albuquerque¹, P.W. Mauriz² and D.A. Moreira¹

¹Departamento de Física, Universidade Federal do Rio Grande do Norte
59072-970 Natal-RN, Brazil
²Departamento de Ciências Exatas, Centro Federal de Educação Tecnológica do Maranhão
65025-001 São Luís-MA, Brazil
E-mail: eudenilson@dfte.ufrn.br

Abstract. We report in this work a numerical study of the electronic density of states in π-stacked arrays of DNA double-strand segments made up from the nucleotides guanine G, adenine A, cytosine C and thymine T. In order to reveal the relevance of the underlying correlations in the nucleotides distribution, we compare the results for a genomic DNA sequence, considering a segment of the first sequenced human chromosome 22 (Ch 22), with those of two artificial sequences forming a Rudin-Shapiro (RS) as well as a Fibonacci (FB) polyGC quasiperiodic sequences. Our theoretical method uses an electronic tight-binding Hamiltonian suitable to describe the DNA segments modeled by the quasiperiodic chains.

A DNA chain can be considered as a symbolic sequence of four nucleotides, namely guanine (G), adenine (A), cytosine (C) and thymine (T), which define the structure of the amino acids to form proteins. Unlike proteins, however, 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 [1].

To address the above-mentioned issues, we report in this work the electronic conductivity of a DNA molecule. Our theoretical model is based on a tight-binding Hamiltonian, within a Dyson’s framework, together with a transfer matrix employed to simplify the algebra, which can be otherwise quite involved. The electronic density of states is thus calculated considering that the DNA molecule is sandwiched in a substrate, arranged as a Fibonacci (FB) and a Rudin-Shapiro (RS) quasiperiodic structures (for a review of the physical properties of these and other quasiperiodic structures see Ref. [2, 3]). The spectra are then compared with those found from a genomic DNA sequence, considering a segment of the first sequenced human chromosome 22 (Ch 22).

Consider the double-strand model for the DNA, that can represent the actual design of the DNA molecule in a more accurate approach. In order to simplify our model, we do not study the influence of the twist in the double helix. Instead, we assume that the double strand displays a two-dimensional layout [4]. The tight-binding Hamiltonian to be considered for this model is written as:

\[
H = \omega_1^n |n, 1\rangle\langle n, 1| + \omega_2^n |n, 2\rangle\langle n, 2| + V_{12}^{nn} [ |n, 1\rangle\langle n, 2| + |n, 2\rangle\langle n, 1| ]
\]
The ratio among the distances of consecutive generations tends to the golden mean, \( \omega = 1.618 \), with a limiting ratio of the elementary length to the nearest neighbors. Also, the term \( V_{j,j}^{n,n+\delta} \) (\( \delta = \pm 1 \)) is the hopping potential (we consider here hopping potentials limited to the nearest neighbors). Also, the term \( V_{12}^{\infty} = V_{21}^{\infty} \) corresponds to the hydrogen bonds between the bases \( A - T \) and \( G - C \) in antiparallel strands.

The Dyson equation is

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

where \( I \) is the identity matrix and \( H \) is the Hamiltonian given above.

First, let us model the DNA molecule structure by a Fibonacci (FB) sequence. Starting from a \( G \) (guanine) nucleotide as seed, the quasiperiodic FB sequence can be built through the inflation rules \( G \rightarrow GC \), \( C \rightarrow G \) for the first strand (grown in the direction \( 5' \rightarrow 3' \)). For the second strand, the inflation rules are: \( C \rightarrow CG \) and \( G \rightarrow C \), in such a way that we always have a \( GC \) or \( CG \) base pair, i.e., a poly(\( G \))-poly(\( C \)) double-stranded DNA. The quasiperiodic Rudin-Shapiro (RS) structure can be grown in a similar way (for more details see [5]).

We now turn our discussion to the determination of the electronic density of states (DOS), namely

\[
\rho(\omega) = -(1/\pi)\text{Im}[\text{Tr}(n|G(\omega)|n)],
\]

where Im means the imaginary part of the argument shown between brackets. The energies \( \omega_j \) are chosen from the ionization potential of the respective nucleotides, i.e.: \( \omega_A = 8.24 \text{ eV} \) (adenine), \( \omega_C = 8.87 \text{ eV} \) (cytosine), \( \omega_G = 7.75 \text{ eV} \) (guanine), and \( \omega_T = 9.14 \text{ eV} \) (thymine) [6, 7, 8]. All the hopping terms \( V_{nm} \) among the bases were taken equal to 0.4 eV, since theoretical calculations using \textit{ab initio} methods yield for this potential values in the range 0.4-1 eV [6, 7, 8]. 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 the HOMO’s (Highest Occupied Molecular Orbital) guanine state, giving us \( V_{GS} = 2.39 \text{ eV} \). The hopping term inside the electrode is 12 eV [9]. Further, the on-site energy for the substrate (platinum) is \( \omega_S = 5.36 \text{ eV} \), which is related with the work function of this metal [10], and the hopping potential due to the hydrogen bonds linking the two strands will be considered as 0.1 eV.

Fig. 1a depicts the density of states for a DNA quasi-periodic chain following a Fibonacci quasiperiodic sequence. 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: (a) 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) \( 5.0 < \omega < 9.30 \) (we call it region I), and \( 9.30 < \omega < 12.50 \) (region II); (b) Region II, which appears as a sort of anomaly in the DOS spectrum, is due to the presence of the cytosine nucleotide in the quasi-periodic chain. (c) Each region defines a clearly a strike auto-similar spectrum for different generations. The auto-similarity holds also for the whole spectrum (regions I+II); (d) The central peak for region I is next to the guanine’s ionization energy \( \omega_G = 7.75 \text{ eV} \), while the central peak in region II corresponds to \( \omega_G + 3V_{12}^{\infty} = 10.75 \text{ eV} \); (e) 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 quasi-periodic chain following a Rudin-Shapiro quasiperiodic sequence is shown in Fig. 1b, for the number of nucleotides \( n_{RS} = 256 \) and 512, corresponding to the 9th and 10th RS sequence generation (full line). 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
Figure 1. (a) Density of state spectra for the Fibonacci Poly-GC DNA double-strand model corresponding to the 10th (chain-dotted line), 12th (dashed line), and 14th (full line); (b) Same as in (a), but for the Rudin-Shapiro (full line) and part of the human chromosome Ch22 (dashed line) models, considering \( n_{RS} = 256 \) and 512, corresponding to the 9th and 10th RS sequence generation.

how important is the model considered to simulate the DNA structure. As their main features, their central peaks, which are sequence independent, lie next to the guanine’s ionization energy \( \omega_G = 7.75 \) eV, with the band-width approximately given by \( \omega_G \pm 2V_{nm} \). 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 (dashed line in Fig. 2b), a remarkable agreement is found.

We also report in this work the conductivity of the single strand DNA molecule models through their electron transmittance coefficient. For this case, the (discrete) Schrödinger equation can be written as

\[
\begin{pmatrix}
\psi_{j+1} \\
\psi_j
\end{pmatrix} = M(j) \begin{pmatrix}
\psi_j \\
\psi_{j+1}
\end{pmatrix},
\]

where \( M(j) \) is the transfer matrix. The transmission coefficient \( T_N(E) \) is defined by [11]

\[
T_N(E) = \left[ X^2(E) - 4 \right] \left[ X^2(E)(P_{12}P_{21} + 1) - X(E)(P_{11} + P_{22})(P_{12} - P_{21}) - \sum_{i,j} P_{ij}^2 - 2 \right]^{-1}
\]

where \( X(E) = (E - \omega_m)/t_m \), and \( P_{ij} \) are elements of the transfer matrix \( P = M(N)M(N - 1)...M(2)M(1) \). For a given energy \( E \), \( T_N(E) \) measures the level of backscattering events in the electron (or hole) transport through the chain.

In Fig. 2 we plot the transmission coefficient \( T_N(E) \), as given by Eq. (5), as a function of the energy, in units of eV, considering the Fibonacci (Fig. 2a), the Rudin-Shapiro (Fig. 2b), and the human chromosome Ch22 (Fig. 2c) sequences, respectively. Here, the number of nucleotides is equal to 233. Observe that the transmission bands in Fig. 2a, corresponding to the Fibonacci case, are fragmented, which is related to the localized nature of the one-electron eigenstates in disordered chains. It is relevant to stress that the presence of long-range correlations in the disorder distribution was recently shown to be a possible mechanism to induce delocalization in low dimensional systems. However, the actual correlations in DNA sequences are not strong enough to produce this correlation-induced transition and the stationary states remain all localized. Moreover, the presence of long-range correlations enhances the localization
length and, therefore, transmission resonances survive in larger segments as compared with a non-correlated random sequence. Observe also that the transmission coefficient for long-range correlated Rudin-Shapiro sequences, depicts a trend similar to the one produced by the genomic Ch22 sequence.

To summarize, we have proposed in this paper an analytical tight binding simple model, within Dyson’s framework, to study the electronic density of states (DOS) of a double-strand DNA molecule. For completeness, we have presented also the conductivity of the single-strand DNA through their electron transmittance coefficient. In order to unreveal the actual relevance of long range correlations, we compared the transmission spectra of the quasiperiodic strutures (Fibonacci ans Rudin-Shapiro) with those segments of the Ch22 human chromosome. We obtained that the long-range correlations, present in Ch22 and RS sequences, are responsible for the slow vanishing of some transmission peaks, which may promote an effective electronic transport at specific resonant energies of finite DNA segments.

The authors would like to acknowledge the financial support received from the Brazilian Research Agencies CAPES, CNPq (projects CT-ENERG 554889/2006-4 and CNPq-Rede NanoBioestruturas 555183/2005-0), FINEP-CTINFRA and FAPEMA.

References

[1] Albuquerque E L, Vasconcelos M S, Lyra M L and de Moura F A B F 2005 Phys. Rev. E 71 021910.
[2] Albuquerque E L and Cottam M G 2004 Polaritons in Periodic and Quasiperiodic Structures (Amsterdam: Elsevier).
[3] Albuquerque E L and Cottam M G 2003 Phys. Rep. 376 225.
[4] Yamada H 2004 Phys. Lett. A 332 65.
[5] de Oliveira B P W, Albuquerque E L and Vasconcelos M S 2006 Surf. Sci. 600 3770.
[6] Sugiyama H and Saito I 1996 J. Am. Chem. Soc. 118 7063.
[7] Voityuk A A et al. 2001 J. Chem. Phys. 114 5614.
[8] Zhang H et al. 2002 J. Chem. Phys. 117 4578.
[9] Maciá E Triozon F and Roche S 2005 Phys. Rev. B 71 113106.
[10] Berlin Y A, Burin A L and Ratner M A 2002 Chem. Phys. 275 61.
[11] Roche S, Bicout D, Maciá E and Kats E 2004 Phys. Rev. Lett. 92 109901.