Current–voltage characteristics of double-strand DNA sequences

L.M. Bezerril\textsuperscript{a}, D.A. Moreira\textsuperscript{a}, E.L. Albuquerque\textsuperscript{a,}\textsuperscript{*,} U.L. Fulco\textsuperscript{b}, E.L. de Oliveira\textsuperscript{c}, J.S. de Sousa\textsuperscript{c}

\textsuperscript{a} Departamento de Física, Universidade Federal do Rio Grande do Norte, 59072-970, Natal-RN, Brazil
\textsuperscript{b} Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, 59072-970, Natal-RN, Brazil
\textsuperscript{c} Departamento de Física, Universidade Federal do Ceará, 60455-760, Fortaleza-CE, Brazil

We use a tight-binding formulation to investigate the transmissivity and the current–voltage (I–V) characteristics of sequences of double-strand DNA molecules. In order to reveal the relevance of the underlying correlations in the nucleotides distribution, we compare the results for the genomic DNA sequence with those of artificial sequences (the long-range correlated Fibonacci and Rudin–Shapiro one) and a random sequence, which is a kind of prototype of a short-range correlated system. The random sequence is presented here with the same first neighbors pair correlations of the human DNA sequence. We found that the long-range character of the correlations is important to the transmissivity spectra, although the I–V curves seem to be mostly influenced by the short-range correlations.

Due to their potential applications in nanoelectronics, there has been a growing interest in the synthesis, characterization, and the electronic properties of DNA-based molecules with periodic nucleotide sequences [1–4]. Using a full range of physical and biochemical methods, studies have now established that double helical DNA is a medium for the efficient transport of electrons, triggering a series of experimental and theoretical investigations [5,6]. The earliest studies involved physical measurements of current flow in DNA fibers, and led to a mixture of conclusions, some suggesting high electron mobility through DNA, others indicating no conductivity [7–10]. In fact, one experiment even pointed to DNA as a superconductor [11]. These physical studies have not yet been reconciled with one another. The variations probably depend heavily upon the connections between the DNA and the electrodes used, as well as upon the integrity of the DNA itself in the absence of water and exposed to very high voltages. Besides, these experiments have underscored not only that the DNA base pair stack can mediate hole and electron transport chemistry, but also the exquisite sensitivity of the charge transport through the DNA structure.

In addition to its own structural complexity, measuring charge transport in a DNA chain is strongly biased by the invasive role of contacts, interaction with some inorganic substrate, and temperature/atmosphere experimental conditions. As a result, a focus has been now shifted from asking whether DNA can mediate long-range charge transport to questioning how it works. Specifically, we are seeking answers to how do DNA structure and sequence affect this reaction, and how important is DNA-mediated charge transport.

To this end, several experimental groups have reported measurements of the current–voltage (I–V) characteristics obtained from electrical transport measurements throughout individual or small numbers of DNA-like molecules captured between two metal nanoelectrodes. The traditional molecular view of electron transfer between donor and acceptor species gives rise to a novel view of the molecule as a current-carrying conductor, and observables such as electron-transfer rates are replaced by the conductivities, or more generally by current–voltage relationships, in molecular junctions. Of primary importance is the need to understand the interrelationship between the molecular structure of such junctions and their function, i.e. their transmission and conduction.
properties. Such investigations, in which single molecules or small molecular assemblies operate as conductors connecting ‘traditional’ electrical components such as metal or semiconductor contacts, constitute a major part of what has become the active field of molecular electronics. Their diversity, versatility, and amenability to control and manipulation make molecules and molecular assemblies potentially important components in nanoelectronic devices. A standard electron transfer system thus containing a donor, an acceptor, and a molecular bridge connecting them. Our focus is the electron transfer between the two conducting electrodes (donor and acceptor) through a molecular medium, which bears strong similarity to the more conventional systems that involve at least one molecular species in the donor/acceptor pair. These developments have attracted much attention from the semiconductor industry and there is a great interest from an applied point of view to model and understand the capabilities of molecular conductors. At the same time, this is also a topic of great interest from the point of view of basic physics. Three possible mechanisms might be considered:

(a) Superexchange: the charge tunnels from the donor (DN) electrode to the acceptor (AC) electrode through the DNA segment in a nonadiabatic process. An exponential decrease in the rate of charge transport with increasing length of the DNA segment is predicted.

(b) Hopping: charge occupies the DNA segment in travelling from DN electrode to AC electrode by hopping between the DNA’s discrete molecular orbitals. If the rate of charge migration is faster than trapping, the charge should be able to migrate over long distances before getting trapped.

(c) Domain hopping: charge occupies the DNA segment by delocalizing over several bases, or a domain. This domain hops along the DNA segment to travel from DN electrode to AC electrode. As in a pure hopping mechanism, the charge should be able to travel long distances before getting trapped.

We are all aware that the DNA found in cells is a double helix consisting of two antiparallel strands held together by specific hydrogen-bonded base pairs: adenine (A) always pairs with thymine (T), and guanine (G) always pairs with cytosine (C). The specificity of this base pairing and the ability to ensure that it occurs in this fashion (and not some other) is key to the use of DNA in materials applications. The double helical arrangement of the two molecules leads to a linear helix axis, linear not in the geometrical sense of being a straight line, but in the topological sense of being unbranched. Such model (the two-leg ladder model) was also used by other groups to discuss the transport properties of DNA sequences [12–14]. 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. So far, 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 [15].

With this aim in mind, we report in this work an analytical as well as numerical investigation of the one-electron states in double-strand binary DNA segments. Our theoretical model is based on a tight-binding Hamiltonian, together with a transfer matrix technique employed to simplify the algebra which can be otherwise quite involved. We consider a model in which the DNA molecule is sandwiched by two electrodes (donor-DN and acceptor-AC, respectively), following a Fibonacci (FB) and a Rudin–Shapiro (RS) quasiperiodic structures [16–18], and compare them to the DNA sequence of the first sequenced human chromosome 22 (Ch22), whose arrangement was retrieved from the internet page of the National Center of Biotechnology Information. We investigate the conductivity of the DNA molecule models through their electron transmission coefficient. Furthermore, by solving numerically a time-independent Schrödinger equation, we compute also some basic properties of the I–V characteristics, for all DNA models considered here.

The Hamiltonian for a double-strand DNA chain is an effective tight-binding model describing one electron moving in a geometry, composed by two interconnected chains of sites, side by side, with a single orbital per site and nearest-neighbor interactions [19], i.e.,

\[ \mathcal{H} = \sum_n (\epsilon_n^{\alpha} |\psi_n^{\alpha}\rangle \langle \psi_n^{\alpha}| + \epsilon_n^{\beta} |\psi_n^{\beta}\rangle \langle \psi_n^{\beta}|) + \sum_n w(\langle \psi_n^{\alpha}|\psi_{n+1}^{\beta}\rangle + |\psi_n^{\beta}| \langle \psi_n^{\alpha}|) + \sum_{n,\delta=\pm 1}(\langle \psi_n^{\alpha}|\psi_{n+\delta}^{\alpha}\rangle + |\psi_n^{\beta}| \langle \psi_n^{\beta}|\psi_{n+\delta}^{\alpha}\rangle). \]

(1)

where \( \epsilon_n^{\alpha} \) is the single energy at the orbital \( \psi_n^{\alpha} \) (the upper index refers to the chain, while the lower index refers to the site position in each chain). Also \( t \) and \( w \) are the intra-chain and the inter-chain first-neighbor electronic overlaps (hopping amplitude), respectively (for simplicity, we are considering that the charge carriers are travelling from DN to AC electrodes by hopping between the DNA’s discrete molecular orbitals). The corresponding time-independent Schrödinger equation is given by:

\[ t(\psi_{n+1}^{\alpha} + \psi_{n-1}^{\alpha}) + w\psi_n^{\beta} = (E - \epsilon_n^{\alpha})\psi_n^{\alpha}, \]

(2)

\[ t(\psi_{n+1}^{\beta} + \psi_{n-1}^{\beta}) + w\psi_n^{\alpha} = (E - \epsilon_n^{\beta})\psi_n^{\beta}. \]

(3)

Within this framework, the Schrödinger equation can be written as:

\[ \begin{pmatrix} \psi_{n+1}^{\alpha} \\ \psi_{n+1}^{\beta} \\ \psi_n^{\alpha} \\ \psi_n^{\beta} \end{pmatrix} = M(n) \begin{pmatrix} \psi_n^{\alpha} \\ \psi_n^{\beta} \\ \psi_{n-1}^{\alpha} \\ \psi_{n-1}^{\beta} \end{pmatrix}, \]

(4)

where \( M(n) \) is the transfer matrix

\[ M(n) = \begin{pmatrix} (E - \epsilon_n^{\alpha})/t & -w/t & -1 & 0 \\ -w/t & (E - \epsilon_n^{\beta})/t & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \]

(5)

After successive applications of the transfer matrix \( M(n) \), we have

\[ \begin{pmatrix} \psi_{n+1}^{\alpha} \\ \psi_{n+1}^{\beta} \\ \psi_n^{\alpha} \\ \psi_n^{\beta} \end{pmatrix} = M(n)M(n-1) \cdots M(2)M(1) \begin{pmatrix} \psi_0^{\alpha} \\ \psi_0^{\beta} \\ \psi_{n-1}^{\alpha} \\ \psi_{n-1}^{\beta} \end{pmatrix}. \]

(6)

Calculation of the products of these transfer matrices is completely equivalent to solving the Schrödinger equation for the system. The criterion for allowed energy is when \( (1/2)\text{Tr}[\rho] < 1 \), with \( \text{Tr}[\rho] \) meaning the trace of the matrix \( \rho \), and \( \rho = M(N)M(N-1) \cdots M(2)M(1) \) [20].

The first sequenced human chromosome 22 (Ch22) contains about \( 3.49 \times 10^8 \) nucleotides, its largest segment (NT01520) about \( 2.33 \times 10^5 \) nucleotides, and the second largest one about \( 4.25 \times 10^5 \) nucleotides. Here, we consider a very short sequence (the maximum number of nucleotides is 512) selected from the largest segment NT01520 (for a statistical study of this sequence see Ref. [21]), starting from the G (guanine) nucleotide. The energies \( \epsilon_n \) are chosen from the ionization potential of the respective
The quasiperiodic Fibonacci sequence is constructed starting from a 9 nucleotide as seed, the quasiperiodic RS sequence can be built through the inflation rules

\[ \alpha = \begin{pmatrix} a & 0 \\ e^{ika} & 0 \end{pmatrix}, \]

\[ \beta = \begin{pmatrix} 0 & e^{ika} \\ 1 & 0 \end{pmatrix}, \]

where the matrix \( T = 69^{-1} S^{-1} P S \). We consider here \( B_\alpha = B_\beta = 1 \). Also,

\[ A_\alpha = (-T_{41} T_{34} - T_{42} T_{34} + T_{44} T_{31} + T_{44} T_{32})/(T_{43} T_{34} - T_{33} T_{44}), \]

\[ A_\beta = (-T_{43} T_{31} - T_{43} T_{32} + T_{43} T_{41} + T_{33} T_{42})/(T_{43} T_{34} - T_{33} T_{44}). \]

Furthermore,

\[ S = \begin{pmatrix} e^{-ika} & 0 & e^{ika} & 0 \\ 0 & e^{-ika} & 0 & e^{ika} \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \]

\[ \Theta = \begin{pmatrix} e^{-ika} & 0 & 0 & 0 \\ 0 & e^{-ika} & 0 & 0 \\ 0 & 0 & e^{ika} & 0 \\ 0 & 0 & 0 & e^{ika} \end{pmatrix}, \]

where \( k \) is given by

\[ k = \cos^{-1}\left(1 - \epsilon_s/(2\epsilon_S)\right). \]

We use the energy of the platinum electrode \( \epsilon_S = 5.36 \text{ eV} \), which is related to the work function of this metal [31]. Also the potential inside the electrode \( \epsilon_S \) is considered to be 12 eV.

The transmission coefficient is a useful quantity to describe the transport efficiency in quantum systems. Nonetheless, \( T_N(E) \) is usually difficult to be directly measured experimentally. Access to transmission properties can be performed by measuring their \( I-V \) characteristics. However, applying a voltage bias in the conducting leads (DN and AC electrodes) contacting the DNA segment has also some influence on the scattering properties inside the molecule, and direct information on intrinsic effects of sequences on transmission should thus be considered with care. With the effective Hamiltonian given above, one can evaluate the \( I-V \) characteristics by applying the Landauer–Büttiker formulation [32]

\[ I(V) = \frac{2e}{h} \int_{-\infty}^{+\infty} T_N(E) \left( f_{\text{DN}}(E) - f_{\text{AC}}(E) \right) dE. \]

where the Fermi–Dirac distribution \( f_{\text{DN(AC)}} = [\exp((E - \mu_{\text{DN(AC)}})/k_B T) + 1]^{-1} \). Here, \( \mu_{\text{DN(AC)}} \) is the electrochemical potential of the two leads (donor-DN and acceptor-AC) fixed by the applied bias voltage \( V \) as \( |\mu_{\text{DN}} - \mu_{\text{AC}}| = eV \) [33]. We are assuming the Fermi level energy equal to zero [34]. The current onset is crucially dependent on the electrochemical potentials of the leads that can be altered by the coupling to molecules, which is another important issue to be separately considered. For simplicity, before bias voltage is applied, the electrochemical potential of the whole system is taken to be zero. It is important to emphasize that the transmission \( T_N(E) \) should be calculated also for negative values of energy.

We start by reporting the transmission coefficients \( T_N(E) \), as given by Eq. (7), which are depicted in Fig. 1 as a function of the energy, in units of eV. We have considered the Fibonacci sequence (Fig. 1a, with generation number \( N_F = 12 \), corresponding to \( n_F = 233 \) nucleotides), the Rudin–Shapiro one (Fig. 1b, with generation number \( N_R = 7 \), corresponding to \( n_S = 64 \) nucleotides), the random case (Fig. 1c, with \( n_{\text{RD}} = 64 \) nucleotides), and the human chromosome Ch22 (Fig. 1d, with \( n_{\text{Ch22}} = 64 \) nucleotides), respectively. Observe that the transmission bands in all cases are fragmented, which is related to the localized nature of the electron’s eigenstates in disordered chains, and reflects the number of passbands in each structure (when the localization factor is zero, the corresponding frequency intervals are known as passbands). It is relevant to stress that the presence of long-range correlations in the disorder distribution is a possible mechanism to induce delocalization in low dimensional systems [35]. However, the actual correlations in our model (hopping mechanism) 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.
Transmittance coefficient $T_N(E)$ as a function of the energy $E$, in units of eV, for: (a) the Fibonacci sequence, with generation number $N_F = 12$, corresponding to $n_F = 233$ nucleotides; (b) the Rudin–Shapiro structure, with generation number $N_{RS} = 7$, corresponding to $n_{RS} = 64$ nucleotides; the random case, with $n_{RD} = 64$ nucleotides; (d) the human chromosome Ch22, with $n_{Ch22} = 64$ nucleotides. Notice that the presence of correlations contributes to the survival of resonant transmission peaks for sequences up to hundreds of nucleotides.

and, therefore, transmission resonances survive in larger segments as compared with a non-correlated random sequence (see, for instance, the Fibonacci case). 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.

Current–voltage characteristics of double-strand DNA sequences are plotted in Fig. 2 for Fibonacci (Fig. 2a), Rudin–Shapiro (Fig. 2b), the random case (Fig. 2c) and the human chromosome Ch22 (Fig. 2d), respectively. We are assuming a linear voltage drop across the DNA molecules by means of the usual expression, numerically computed near zero temperature, as given by Eq. (15). To reproduce the potential mismatch at zero bias, the energy difference between the guanine HOMO energy level and the metallic Fermi level of the electrode is set to 1.2 eV [36]. As the voltage drop is switched on, the transmission coefficient $T_N(E)$ becomes voltage-dependent, resulting in transmission band shifts (shown in Fig. 2 for all cases studied here), which in turn lead to a voltage threshold modulation.

To extract the main features of tunneling currents in DNA chains, let us compare the behavior of the genomic Ch22 (Fig. 2d) with those characterizing the quasiperiodic and random structures (Figs. 2a, b, c) under the resonance condition given by the hopping term choice $t = 1$ eV. In this case, if the potential barrier between the metallic contacts and the DNA is set to zero, a staircase in the plot $I$–$V$ is found [34,37].

As soon as a potential barrier between the DNA and the metals is introduced (1.2 eV), the I–V characteristic curves show the profiles depicted in Fig. 2. The current threshold at a given voltage scale is not sensitive concerning the different structures considered here, mainly due to the electronic correlations presented by the structures. However, such correlations 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 threshold enhancement in DNA molecules. Observe the striking agreement between the I–V characteristic curves for the random and genomic case. Such agreement can be accounted by the short-range pair correlations shared by them, suggesting that the inclusion of just first-neighbors intra-strand pair correlations on the nucleotide distribution provides an adequate description of the DNA electronic properties.

In summary, we numerically study the one-electron conductivity dynamics in arrays of double-strand DNA segments. We have modelled the DNA sequences considering its four nucleotides following three artificial sequences, two of them characterizing the quasiperiodic structures of Fibonacci and Rudin–Shapiro types, both presenting a long-range pair-correlation. The other artificial sequence is a pair-correlated random sequence (RD), with the same first neighbors pair correlations of the human Ch22 DNA sequence, which shows short-range pair correlation behavior. The I–V characteristic curve seems to be 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. However, as the electron transmissivity depends strongly on the intra-chain coupling, further studies considering more realistic model parameters would be needed in order to infer about the actual relevance of this behavior in DNA molecules.

Finally, we would like to stress that 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. We expect to address this point in a future communication.
Fig. 2. Current–voltage characteristics of double-strand DNA sequences for (a) Fibonacci; (b) Rudin–Shapiro; (c) the random case; (d) the human chromosome Ch22, respectively.

Acknowledgements

This work was partially financed by the Brazilian Research Agencies CAPES (PROCAD and Rede NanoBioTec), CNPq (Rede NanoBioestruturas Project No. 555183/2005-0 and INCT-Nano(Bio)-Simes, Project No. 573925/2008-9) and FAPERN/CNPq (Pronex).

References