Electrical current through DNA containing mismatched base pairs

Neranjan Edirisinghe¹, Vadym Apalkov¹, Julia Berashevich² and Tapash Chakraborty²

¹ Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30303, USA
² Department of Physics and Astronomy, University of Manitoba, Winnipeg, R3T 2N2, Canada

Received 6 February 2010, in final form 19 April 2010
Published 20 May 2010
Online at stacks.iop.org/Nano/21/245101

Abstract

Mismatched base pairs, such as different conformations of the G·A mispair, cause only minor structural changes in the host DNA molecule, thereby making mispair recognition an arduous task. Electron transport in DNA that depends strongly on the hopping transfer integrals between the nearest base pairs, which in turn are affected by the presence of a mispair, might be an attractive approach in this regard. We report here on our investigations, via the I–V characteristics, of the effect of a mispair on the electrical properties of homogeneous and generic DNA molecules. The I–V characteristics of DNA were studied numerically within the double-stranded tight-binding model. The parameters of the tight-binding model, such as the transfer integrals and on-site energies, are determined from first-principles calculations. The changes in electrical current through the DNA chain due to the presence of a mispair depend on the conformation of the G·A mispair and are appreciable for DNA consisting of up to 90 base pairs. For homogeneous DNA sequences the current through DNA is suppressed and the strongest suppression is realized for the G(anti)·A(syn) conformation of the G·A mispair. For inhomogeneous (generic) DNA molecules, the mispair result can be either a suppression or an enhancement of the current, depending on the type of mispairs and actual DNA sequence.

1. Introduction

Self-assembly and self-recognition properties of double-stranded DNA make these molecules a very important nanoscale system with many unique properties. Although self-recognition of DNA means that the four nucleotides, i.e. guanine (G), adenine (A), cytosine (C) and thymine (T), are arranged only in C–G and A–T pairs through weak hydrogen bonds, mismatched base pairs can also be present in the DNA chain. These mispairs can result from chemical damage, ionizing radiation or many other sources from the environment. The appearance of mispairs within DNA can lead to genetic mutation, even leading up to cancer [1]. In this context, an efficient repair mechanism of mismatched base pairs remains crucial for preserving genome integrity [2–4]. A clear understanding of the mismatch repair mechanisms essentially requires knowledge of the intrinsic properties of the mispairs, which distinguish them from the canonical Watson–Crick base pairs. Such specific intrinsic properties of mispairs can perhaps be used for the development of sensors capable of detecting the mismatched base pairs [5]. The presence of mispairs in DNA alters the thermodynamic properties of the molecule [6, 7], and even influences the magnetic properties [8], and also changes locally the carrier transfer rates. Although a mispair causes only local geometric changes in DNA, for small molecules the integral characteristics of DNA are also modified by the mispair. For example, changes in magnetic susceptibility of DNA due to the presence of a mispair are supposed to occur for DNA containing up to 50 base pairs [8].

Below we address another physical property of DNA which is sensitive to the presence of mispairs. We study charge transport through a finite-length DNA molecule connected to electrical contacts [9–11]. The efficiency of charge transfer in DNA is governed by the geometry of the individual base pairs and their sequences. The electron transport is mainly determined by two energetic parameters: the on-site energy, i.e. the electron energy on a given base pair, and the hopping integral between the nearest-neighbor base pairs. Therefore, any structural change in the geometry of the base pairs would affect the transport properties of DNA. Incorporation
of mispairs into the DNA chain is a source of such geometric changes. Depending on the geometry of the mispairs and their interaction with the neighboring base pairs, the structural distortion of the DNA helix induced by mispairs can be global or local. A global change modifies the position of several base pairs nearest to the mispair. Such mispairs should be easier to recognize by the repair enzymes and would also significantly alter the transport properties of DNA. From this point of view, the recognition of a mispair generating only local changes in the structure of DNA represents a significant challenge. In our present investigation, we chose the type of mispairs that results in less distortion and, therefore, is poorly recognized by the repair enzymes in human cells [3, 12]. The G-A mispair belongs to such a class of mismatched pairs [7]. This mispair is known to have several conformations, but only the G(anti)-A(anti) and G(anti)-A(syn) conformations satisfy the required conditions, i.e. they induce only local changes in the DNA structure [7]. In what follows, we study these mispairs incorporated into a DNA molecule of finite length. Such a DNA molecule with mispairs is studied below within a double-stranded tight-binding model [14, 15]. The parameters of the model are calculated by first-principles methods.

On general grounds, a mispair in a DNA molecule should be considered as a perturbation of the base pair stack or as a defect in the DNA molecule. Therefore, we should expect that replacement of a canonical base pair by a mispair suppresses the current through the molecule. Such suppression has been reported experimentally in [13], where it was observed that a single G-T or C-A mispair decreases the current through a DNA molecule by almost 300 times. The suppression of the current is large and is due to a strong and global perturbation of the DNA molecule by a G-T or C-A mispair. We would like to emphasize that in the present paper we consider a different type of mispair, i.e. G-A mispairs, which results in only local structural distortion of the molecule. Therefore, in our case the mispair can be considered as a weak defect in the DNA molecule. In this case it is not clear what the final effect of the mispair on the charge transport through the molecule will be. Indeed, below we show that, depending on the DNA sequence, the current through the molecule can be suppressed or enhanced by a single base pair. Thus, the G-A mispair in a DNA molecule can be an additional scattering center, which increases the resistance of the molecule, or can open a new charge transfer channel, increasing the transport through DNA.

Our paper is organized as follows. In section 2, we introduce a double-stranded tight-binding model of DNA and describe the method of calculating the electrical current through DNA under an applied bias voltage. In section 3, we describe the first-principles methods which are used to calculate the parameters of the tight-binding model of DNA. The corresponding transfer integrals and on-site energies for canonical Watson–Crick base pairs and for mispairs are presented. The results of our calculation of the $I–V$ characteristics of DNA with and without mispairs for homogeneous poly(G)–poly(C) DNA are presented in section 4. In section 5, we consider generic DNA molecules. They are described as DNA with a random sequence of base pairs.

2. Double-stranded tight-binding model of DNA

We study the current–voltage ($I–V$) characteristics of a long DNA chain within the double-stranded tight-binding model [14] of DNA connected to two electrodes, see figure 1. The model includes the interstrand and intrastand charge hoppings between the nearest base pairs and is described by the following Hamiltonian:

$$
\mathcal{H} = \mathcal{H}_t + \mathcal{H}_{\text{leads}} + \mathcal{H}_{\text{DNA–leads}},
$$

where $\mathcal{H}_t$ is the tight-binding double-stranded DNA Hamiltonian which describes the electron transfer within DNA, $\mathcal{H}_{\text{leads}}$ is the Hamiltonian of charge carriers in the two leads, i.e. right and left contacts, and $\mathcal{H}_{\text{DNA–leads}}$ describes the coupling between the DNA and the leads.

The two strands of DNA correspond to HOMO and LUMO orbitals of the base pairs, see figure 1. The corresponding tight-binding Hamiltonian, $\mathcal{H}_t$, includes both the intrastand and interstrand hopping and is given by the following expression:

$$
\mathcal{H}_t = \sum_{i,K,\sigma} \varepsilon_{K,i} n_{K,i,\sigma} + \sum_{i,K,\sigma} t_{K,i,\sigma} a_{K,i,\sigma}^\dagger a_{K,i+1,\sigma} + \sum_{i,\sigma} t_{L,i,\sigma} a_{L,i,\sigma}^\dagger + \text{h.c.},
$$

where $K = \text{H, L}$, $a_{K,i,\sigma}$ and $a_{L,i,\sigma}$ are the annihilation operators of an electron with spin $\sigma = \pm 1$ on site (base pair) $i$ at the HUMO and LUMO strands, respectively, $\varepsilon_{K,i} = \varepsilon_{\text{H},i} \sigma$ and $\varepsilon_{L,i}$ are the corresponding on-site energies, $n_{K,i,\sigma} = a_{K,i,\sigma}^\dagger a_{K,i,\sigma}$, and $t_{K,i,\sigma}$ and $t_{L,i}$ are the intrastand hopping integrals between the nearest base pairs, and $t_{L,i}$ is the interstand hopping integral within the same base pair. The index $i = 1, \ldots, N$ is the number of the base pair within the DNA. Here $N$ is the total number of base pairs in DNA. Each nucleotide of a base pair has one HOMO orbital and one LUMO orbital. Thus, each base pair has two HOMO orbitals and two LUMO orbitals, e.g. a G–C base pair has HOMO and LUMO orbitals for both guanine and cytosine. In the double-stranded model of the DNA molecule, for each base pair we take into account only one HOMO orbital (with the highest energy) and one LUMO orbital (with the lowest energy). For example, for a G–C base
pair we take into account only the HOMO orbital for guanine and LUMO orbital for cytosine.

The Hamiltonian of the leads describes the electrons in the right and left contacts and have the following form:

$$\mathcal{H}_{\text{leads}} = \sum_{Q,k,\sigma} \varepsilon_{Q,k} c_{Q,k,\sigma}^\dagger c_{Q,k,\sigma},$$  \hspace{1cm} (3)

where $Q = L, R$ corresponds to left and right contacts and $c_{Q,k,\sigma}$ are the annihilation operators of an electron with spin $\sigma = \pm1$ and wavenumber $k$ in the contact $Q = L, R$. The electron system in each lead is characterized by the corresponding chemical potentials, $\mu_L$ and $\mu_R$. Then the bias voltage across the DNA molecule is introduced as $V_b = \mu_L - \mu_R$.

The coupling of a DNA chain to the leads is introduced through hopping between the HOMO and LUMO states of the first base pair of DNA and the left contact and between the HOMO and LUMO states of the $N$th base pair and the right contact. The corresponding Hamiltonian is

$$\mathcal{H}_{\text{DNA–leads}} = \sum_{K,k,\sigma} V_{1,k} c_{1,k,\sigma}^\dagger a_{K,1,\sigma} + \text{h.c.} + \sum_{K,k,\sigma} V_{N,k} c_{N,k,\sigma}^\dagger a_{K,N,\sigma} + \text{h.c.},$$  \hspace{1cm} (4)

where $V_{1,k}$ and $V_{N,k}$ are the hopping integrals, which characterize the coupling of a DNA chain to the leads. In the above expression we assumed that the coupling strength is the same for both HOMO and LUMO states of DNA. In the following we also assume that the hopping integrals $V_{1,k}$ and $V_{N,k}$ are the same for all states in the contacts, i.e. $V_{1,k} = V_{N,k} = V_0$.

Under an applied bias voltage, which is equal to the difference of the chemical potentials of the leads, the stationary current through the DNA wire can be calculated from the Landauer expression:

$$I = \frac{2e^2}{h} \int d\varepsilon \left( f_L(\varepsilon) - f_R(\varepsilon) \right) \text{Tr}[\mathbf{G}^L(\varepsilon)\mathbf{G}^R(\varepsilon)],$$  \hspace{1cm} (5)

where $f_L(\varepsilon)$ and $f_R(\varepsilon)$ are the Fermi–Dirac distribution functions for the left and right contacts, respectively; $\mathbf{G}^L$ and $\mathbf{G}^R$ are retarded and advanced Green functions in the DNA chain, and $\mathbf{R}$ and $\mathbf{\Gamma}$ are the level-width matrices, determining the coupling of the DNA states through the continuous states of the right and left contacts, respectively. The Green functions and the level-width matrices are $2N \times 2N$ matrices, $\mathbf{G}^L, \mathbf{G}^R, \mathbf{R}^L, \mathbf{R}^R$, where $K_1, K_2 = H, L$ and $i_1, i_2 = 1, \ldots, N$. The level-width matrices have the following form:

$$\mathbf{R}^L_{K_1,i_1;K_2,i_2} = 2\pi \rho_L V_0^2 \delta_{i_1,1} \delta_{i_2,1} \text{ and } \mathbf{R}^R_{K_1,i_1;K_2,i_2} = 2\pi \rho_R V_0^2 \delta_{i_1,N} \delta_{i_2,N}. \quad \rho_L, \rho_R$$

are the density of states in the left and right contacts.

The retarded and advanced Green functions can be found by means of the equation-of-motion approach. The Green functions have the following form:

$$\mathbf{G}^{L,R}(\varepsilon) = [\varepsilon - H_1 \pm \frac{i}{2}(\mathbf{R}^L + \mathbf{R}^R)]^{-1}. \quad \hspace{1cm} (6)$$

We numerically calculate the current through the DNA chain as a function of the applied bias voltage. We start from the tight-binding Hamiltonian of the DNA molecule, equation (2), and construct the Hamiltonian matrix. Then, with the known level-width matrices we calculate the Green functions from equation (6). Finally, for a given bias voltage, i.e. the chemical potentials of the leads, we calculate the current through DNA from equation (5).

The crucial parameters, which determine the current through DNA, are the parameters of the tight-binding Hamiltonian, $\mathcal{H}_c$, of DNA. These parameters are the on-site energies and the hopping integrals, which depend on the geometry of the base pairs and are calculated within first-principles methods for mispairs and canonical Watson–Crick base pairs.

3. Parameters of tight-binding model from first-principles calculations

The value of the charge transfer integral corresponding to the electron hopping between two neighboring base pairs depends on the geometry of the base pairs and their relative orientation [16]. Since the equilibrium geometry of a given base pair depends on the structure of the surrounding base pairs, to calculate the transfer integrals we need to consider the basic computational unit, consisting of at least three base pairs. We calculate the transfer parameters for three base pairs within the first-principles methods.

First, we generate the DNA sequences containing pure canonical base pairs using the structural parameters of the B-DNA molecule from [17]. Then incorporation of the G-A mispair inside a helix was done with the help of the CHARMM program [18]. The geometries of the mispairs have been earlier obtained by the DFT methods within the Jaguar [19] software package (see [7] for details). Optimization of the geometry of the generated DNA sequence with mispairs was performed with the empirical potential methods within the CHARMM [18] program applying the adopted basis Newton–Raphson minimization procedure for which the non-bonded interactions include both the electrostatic and the Lennard-Jones potentials. A cutoff radius of 14 Å was used for the non-bonded list of atomic pairs, while for the energy calculations the contribution of the atoms within a cutoff distance of 10 Å was taken into account.

The optimization of a DNA sequence containing a mispair was performed within the CHARMM program by the multistep optimization process. Under such optimization the positions of the terminated canonical base pairs were kept fixed. Because of the fixed positions of the canonical pairs a distortion of the sugar–phosphate backbone occurs in some places. To reconstruct the sugar–phosphate backbone, the sequences were optimized under the condition that the geometry and positions of all base pairs were fixed, while the atoms belonging to the backbone were free to move. The next step was to obtain the optimized position of the mispair relative to its neighboring pairs. This was done by applying the constraints to the position of the canonical base pairs,
while the backbone and mispair were relaxed. The surrounding environment was taken into account by placing the obtained DNA sequences into a water box. Then the optimization procedure with the same constraints as described above were performed. As a result, the energetically optimized location of the middle base pair in a given DNA sequence was obtained in the solvent environment.

The calculations of the charge transfer parameters were performed within the ADF program [20] with the TZ2p basis set. The HOMO and LUMO energies were calculated for the isolated nucleobases optimized in vacuum. The results thus obtained are presented in table 1.

The intrastrand charge transfer integrals \( t_{\text{H}} \) and \( t_{\text{L}} \) were calculated between the two nearest-neighbor base pairs for a given DNA sequence. The interstrand charge transfer integrals \( t_{\perp \text{H}} \) and \( t_{\perp \text{L}} \) were evaluated for each base pair. Within this procedure, each base pair was considered as a molecular fragment (see [16] for details). The obtained magnitudes of the intra- \( t_{\text{H}}, t_{\text{L}} \) and interstrand \( t_{\perp \text{H}}, t_{\perp \text{L}} \) hopping integrals are presented in table 2. The notation of the integrals is defined in figure 1.

We can see from table 2 that for some DNA sequences the replacing of the G–C base pair by a mispair increases the overlap of the \( \pi \) orbitals of the nearest-neighbor nucleobases, i.e. increases the intrastrand hopping integral. The magnitude of the intrastrand charge transfer integrals depends on the distance and the twist angle between nucleobases [16]. It was shown in [16] that even small fluctuations of the twist angle by 5° can change the charge transfer integral by ±0.1 eV. The G–A mispair induces only local changes in DNA structure and mainly changes the equilibrium twist angle. Thus the G–A mispair can result in both enhancement and suppression of the intrastrand inter-base coupling. For mispairs, which produce strong structural distortion of the DNA molecule, the only effect is strong suppression of the intrastrand hopping integral, which finally results in strong suppression of the charge transfer through the molecule [13]. For such a type of mispair the distortion of the DNA molecule is extended over many base pairs. In this case, to extract the hopping integrals for DNA with mispairs we need to perform \textit{ab initio} calculations for relatively large molecules. The mispairs, considered in the present paper, result in small structural changes of DNA, which are local, and the corresponding hopping integrals can be extracted from the considered small DNA sequence.

For the homogeneous chains, such as \((G–C)_3\) and \((A–T)_3\) sequences, the interstrand hopping integrals \( t_{\perp \text{L}} \) are identical for all base pairs, while for sequences containing a mispair in the middle of a chain, the interstrand hopping integral, \( t_{\perp \text{L}(2)} \), corresponding to the mispair differs from \( t_{\perp \text{L}(1)} \) and \( t_{\perp \text{L}(3)} \). A similar tendency was observed for the intrastrand integrals \( t_{\text{H}} \) and \( t_{\text{L}} \), where the presence of a mispair changes the hopping integrals. Therefore, incorporation of a mispair into the \((G–C)_3\) or \((A–T)_3\) DNA sequence leads to a change of sign and the magnitude of the interstrand and intrastrand transfer integrals, which should affect the transport characteristics of the DNA. However, in addition to alteration of the charge transfer integrals in the presence of a mispair, modification of the energetic profile within the DNA chain containing a mispair occurs, which would also significantly influence the charge transfer within DNA.

With the known parameters of the double-stranded tight-binding model of a DNA molecule we calculate its energy spectrum and then find the corresponding retarded and advanced Green functions. Substituting the calculated Green functions into expression (5) we obtain the current through the DNA at a given bias voltage. Since DNA is a quasi-one-dimensional molecule the current through DNA is sensitive to the changes in the parameters of a base pair within the molecule. Therefore if a canonical Watson–Crick base pair is replaced by a mispair then the current through the system is changed.

### Table 1. The HOMO and LUMO energies \( (\varepsilon_{\text{H}} \text{ and } \varepsilon_{\text{L}}) \) of the single nucleobases (in eV).

<table>
<thead>
<tr>
<th></th>
<th>Guanine</th>
<th>Adenine</th>
<th>Thymine</th>
<th>Cytosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOMO</td>
<td>-9.40</td>
<td>-9.79</td>
<td>-10.46</td>
<td>-10.27</td>
</tr>
<tr>
<td>LUMO</td>
<td>-5.37</td>
<td>-5.85</td>
<td>-6.57</td>
<td>-5.90</td>
</tr>
</tbody>
</table>

### Table 2. The values of the intra- and interstrand charge transfer integrals (in eV) for different DNA sequences, where the middle pair \((X_1–X_2)\) has been replaced by G–C, A–T or G(anti)-A(anti), G(anti)-A(anti) mispers.

<table>
<thead>
<tr>
<th>((X_1–X_2))</th>
<th>( t_{\text{H(1)}} )</th>
<th>( t_{\text{H(2)}} )</th>
<th>( t_{\perp \text{L}} )</th>
<th>( t_{\text{L(1)}} )</th>
<th>( t_{\text{L(2)}} )</th>
<th>( t_{\perp \text{L}(3)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>((G–C))</td>
<td>-0.133</td>
<td>-0.133</td>
<td>-0.041</td>
<td>-0.041</td>
<td>-0.050</td>
<td>-0.050</td>
</tr>
<tr>
<td>((A–T))</td>
<td>-0.218</td>
<td>-0.102</td>
<td>0.066</td>
<td>0.067</td>
<td>-0.001</td>
<td>-0.063</td>
</tr>
<tr>
<td>G(anti)–A(\text{anti})</td>
<td>-0.213</td>
<td>0.146</td>
<td>-0.072</td>
<td>-0.071</td>
<td>0.024</td>
<td>0.020</td>
</tr>
<tr>
<td>G(anti)–A(\text{syn})</td>
<td>0.254</td>
<td>-0.136</td>
<td>-0.153</td>
<td>-0.073</td>
<td>0.024</td>
<td>0.091</td>
</tr>
</tbody>
</table>

| \((A–T)\)              | 0.011                  | 0.011                  | -0.075                 | -0.075                 | 0.054                  | 0.054                  |
| A\(\text{anti}\)–G      | -0.002                 | -0.055                 | -0.079                 | -0.117                 | 0.054                  | 0.131                  |
| A\(\text{syn}\)–G       | 0.027                  | 0.207                  | -0.092                 | -0.011                 | 0.054                  | -0.161                 |
the values of the hopping integrals between the mispair and the nearest-neighbor canonical base pairs. For both types of conformations, the on-site energies are the same. Figure 2 clearly illustrates that for both types of mispairs the current through the DNA is suppressed with the introduction of a mispair. The stronger suppression of the current is caused by the G(anti)-A(syn) mispair.

For a homogeneous DNA sequence a mispair can be considered as the defect which breaks the periodicity of the tight-binding model. Such a defect results in additional scattering and finally suppresses the current through DNA. There are two modifications the mispair introduces into the homogeneous system: a mispair changes the on-site energy and a mispair changes locally the hopping integrals. Both of these effects result in suppression of the transport through the chain, although the transport is more sensitive to the changes in the intrastrand hopping integrals. Comparing the hopping integrals for G(anti)-A(syn) and G(anti)-A(anti) mispairs (see table 2), we can conclude that the changes in the hopping integrals for the G(anti)-A(syn) mispair is larger than for the G(anti)-A(anti) mispair. As a result suppression of current through a DNA chain is larger for a G(anti)-A(syn) mispair, which can be seen in figure 2. A similar suppression of the charge transport through homogeneous DNA should be expected if the G-C base pair is replaced by a canonical A-T base pair. In this case we also see the suppression of the current, which is shown in figure 2. The saturated values of the current through DNA molecules with a A-T base pair and a G(anti)-A(anti) mispair are close, which can be explained by similar values of the hopping integrals for a A-T base pair and a G(anti)-A(anti) mispair (see table 2).

To describe the effect of a mispair on the electrical current through DNA, we introduce two characteristics. The first one is the change of the saturated value of the current through DNA at a large bias voltage, \(V_b = -4\) eV. This change is calculated as the difference between the current through the DNA without a mispair and that when a mispair is present in the DNA:

\[
\Delta I = I_{\text{without mispair}} - I_{\text{with mispair}}.
\]  

Another characteristic is the relative change of the current, which is defined by the following expression:

\[
\delta I = \frac{\Delta I}{I_{\text{without mispair}}}. \tag{8}
\]

The relative change of current is a more appropriate characteristic of the effect of a mispair, especially when the dependence on the length of the DNA is considered.

Since the mispair modifies the structure of DNA only locally, which means that the parameters of the tight-binding model are changed only near the mispairs, the changes in the current due to the presence of the mispair should be suppressed with increasing length of the DNA wire. In figure 3 the change of the current through DNA due to the introduction of a G(anti)-A(syn) or a G(anti)-A(anti) mispair is shown as a function of the length of the DNA. For all values of the DNA length the suppression of the current is stronger for the G(anti)-A(syn) mispair. The results shown in figure 3 illustrate

4. Homogeneous DNA sequences

For homogeneous poly(G)–poly(C) DNA the coupling between the nearest-neighbor base pairs results in the formation of HOMO and LUMO minibands. The widths of these minibands are determined by the intrastrand hopping integrals. Introduction of a mispair in such a homogeneous DNA results in scattering of otherwise ‘freely’ propagating charges. This scattering suppresses the current through the DNA. In figure 2, the current through DNA is shown as a function of the applied bias voltage for a poly(G)–poly(C) DNA without and with a mispair. The typical \(I–V\) characteristic has a step-like behavior, where transitions between the steps occur when the chemical potentials of the contacts cross the HOMO or LUMO energy bands. The steps are clearly visible in figure 2 for DNA without and with a mispair. The actual positions of the steps depend on the gate voltage, while the widths of the steps are determined by the HOMO–LUMO energy gaps. The width of the transition region between the steps is determined by the width of the HOMO and LUMO minibands, i.e. by the values of the hopping integrals, and the temperature. In what follows, we study the saturated value of the current through the DNA. In figure 2, this saturated value corresponds, for example, to the bias voltage of \(-4\) eV. At this bias voltage both the HOMO and LUMO states contribute to the current through the molecule. The saturated value of the current also has a weak dependence on the temperature of the system.

As already mentioned above, we consider below only two types of mispairs: the G·A mispair in two of its conformations: G(anti)-A(syn) and G(anti)-A(anti). Within the tight-binding model the differences between these two conformations are

\[
\begin{align*}
\text{G(anti)-A(syn)} & = V - \text{G(anti)-A(anti)} \\
\text{syn} & = \text{anti}.
\end{align*}
\]

\[
\text{G(anti)-A(syn)} = V - \text{G(anti)-A(anti)} = \text{A·T}.
\]

\[
\text{G(anti)-A(anti)} = \text{G·A}.
\]

\[
\text{G(anti)-A(syn)} = \text{A·T}.
\]

\[
\text{G(anti)-A(anti)} = \text{G·A}.
\]

\[
\begin{align*}
\text{G(anti)-A(syn)} & = V - \text{G(anti)-A(anti)} \\
\text{syn} & = \text{anti}.
\end{align*}
\]
that the suppression of current through DNA with a mispair is discernible even for a long DNA with up to 90 base pairs. Although the decrease of \( \Delta I \) with increasing DNA length is clearly seen in figure 3, the current, \( I_{\text{without mispair}} \) itself is also suppressed. As a result, the relative changes \( \delta_I \) of the current, shown in figure 4, remain almost the same. For the G(anti)-A(anti) mispair the relative change is \( \delta_I \approx 0.22 \), while for the G(anti)-A(syn) mispair it is \( \delta \approx 0.55 \). In both cases, \( \delta_I \) has a weak dependence on the length of DNA. The reason for such a weak dependence is the homogeneous sequence of DNA without a mispair. For such a sequence the corresponding wavefunctions have the form of propagating waves, \( \exp(-ikx) \), extended over the entire DNA. As a result, each of the electron wavefunctions becomes sensitive to the presence of a mispair and the effect of the mispair on the wavefunction is appreciable even at large distances.

The effect of a mispair on the current through DNA depends also on the actual position of the mispair within the molecule. In the above analysis we assumed that the mispair is placed exactly in the middle of the DNA chain. In figure 5 the relative change, \( \delta_I \), of the current through DNA is shown as a function of the position of the mispair. For both types of mispairs the maximum changes in the current occur when the mispair is close to the middle of the molecule, although the dependence on the position of the mispair is weak. There is also some irregularity in the dependence of \( \delta_I \) on the position of the mispair. This irregularity is due to the extended nature of the wavefunctions of charge carriers in the homogeneous DNA.

5. Inhomogeneous DNA sequences

For a homogeneous DNA sequence, all electron wavefunctions corresponding to the tight-binding model are extended and occupy the whole DNA molecule. Therefore, all electron states are sensitive to the presence of a mispair which results in strong changes, \( \delta_I \), in the current through the DNA and also in the dependence of these changes on the position of the mispair. For an inhomogeneous DNA sequence the situation is different. The electron states for such a DNA become more localized and the mispair modifies only a few of them. In order to study the properties of an inhomogeneous DNA sequence, we consider...
a suppression of the current through DNA, i.e. for such DNA molecules $\delta_I > 0$. There is also a very broad region with negative values of $\delta_I$, which corresponds to an enhancement of the current through DNA. For a short DNA (10 base pairs in figure 6(d)) there is also a local peak at $\delta_I = 0$, which disappears for longer DNA (figure 6(b)). Here $\delta_I = 0$ means that the current through DNA does not change with introduction of a mispair. A small peak at $\delta \approx -6$ is due to the finite size effect and disappears at larger lengths of DNA.

For the G(anti)-A(syn) mispair we have a different behavior (see figures 6(a) and (c)). Similar to the case of G(anti)-A(anti), there is a large number of different DNA sequences for which the current through DNA is suppressed with the introduction of a G(anti)-A(syn) mispair ($\delta_I > 0$), but this number depends on the length of DNA. For example, for a DNA with 10 base pairs there are 78% of generic DNA sequences for which the current through DNA is suppressed. For a DNA with 15 base pairs it is 66%. Hence with increasing length of DNA the number of different DNA sequences with $\delta_I > 0$ decreases. For a G(anti)-A(syn) mispair there is also a pronounced peak near $\delta_I = 0$. This peak disappears when the length of the DNA is increased. For negative values of $\delta_I$ we have an enhancement of the current through the DNA.

Although a mispair can be considered as a defect in a DNA chain, the effect of a mispair on the charge transport through the molecule can be different for different DNA sequences. If the current through DNA without a mispair is large, i.e. there are well-connected transport paths through the DNA molecule, then the mispair in such a system becomes a real defect which suppresses the transport. In this case we observe suppression of the current through DNA. If the current through DNA without a mispair is originally small, then introduction of a mispair in such a system can open additional transport channels, which can increase the current through the molecule. Although the enhancement of the current through DNA exists for both types of mispair, the more pronounced enhancement is visible for a G(anti)-A(syn) mispair. This can be related to stronger modifications of the intrastrand hopping integrals by a G(anti)-A(syn) mispair.

In table 3 a few random DNA sequences corresponding to the peaks in figure 6 are shown with the corresponding values of $\delta_I$ for different conformations of the G-A mispair. There is some correlation between the values of $\delta_I$ for the G(anti)-A(syn) and G(anti)-A(anti) mispairs, that is, $\delta_I$ is either positive or negative for both types of mispairs. This confirms the general tendency, that enhancement or suppression of the current is determined by the DNA sequence. That is, if the current through DNA without a mispair is small then we should expect enhancement of the transport, and if the initial current is large then the mispair suppresses the transport. Such a tendency is valid for both conformations of G-A mispair.

6. Conclusion

Replacing a canonical Watson–Crick base pair by a mispair results in local changes of the DNA geometry. Such changes modify the local electronic properties of DNA, e.g. the electron
energy at a base pair and the charge transfer matrix elements between the nearest base pairs. Within the tight-binding model of DNA these changes can be described as local changes of the on-site energy and the hopping integrals. Although these changes are local they influence the long-range transport through DNA due to the quasi-one-dimensional nature of electron dynamics in DNA. Therefore a mispair in DNA can change the long-range electron transport through the molecule, which is reflected in the electrical current through the DNA under an applied bias voltage.

For a homogeneous DNA sequence, a mispair can be considered as a point defect which changes both the on-site energy and the transport integrals, and as a result introduces additional scattering of otherwise freely propagating electrons. Such scattering results in additional resistance and suppresses the current through DNA. The strength of such suppression depends on the type of mispair. For the G(anti)-A(syn) and G(anti)-A(anti) mispairs, the change in current through the DNA chain is larger for the G(anti)-A(syn) mispair. This is because the changes of the intrastand hopping integrals near the mispair are larger for the G(anti)-A(syn) mispair compared to G(anti)-A(anti). Thus the G(anti)-A(syn) mispair defect is stronger than the G(anti)-A(anti) one. The canonical base pair, A-T, introduced in a homogeneous DNA molecule, also modifies locally the hopping integrals and can also be considered as a point defect. In this case we also observe the suppression of the current through DNA. The strength of such suppression is comparable to the suppression due to a G(anti)-A(anti) mispair.

The strength of suppression of the current through a homogeneous DNA molecule also depends on the position of the mispair within DNA. The strongest suppression occurs when the mispair is in the middle of the molecule.

The generic DNA sequence is inhomogeneous and the effect of a mispair can be either suppression or enhancement of the current through the molecule. The actual behavior of the current depends on the type of the mispair and the DNA sequence. If the current through DNA without a mispair is large then the corresponding electronic states of the molecule are well extended and provide good connectivity between two contacts. The mispair in such a DNA molecule can be considered as a point defect, which introduces an additional scattering and makes the extended states more localized. In this case the mispair suppresses the current through the DNA molecule. This behavior is similar to the behavior of a homogeneous DNA sequence.

If the current through a DNA molecule is relatively small, then the electronic states within such a molecule are almost localized and the charge transport through the molecule is suppressed. In this case introduction of a mispair can result in additional coupling between the quasilocalized states, making these states more extended. Therefore in this case the mispair increases the charge transport through DNA.

Thus for an inhomogeneous DNA sequence a mispair can be either (i) a defect, which makes the extended electronic states more localized and suppresses the current through the molecule, or (ii) a coupling element between the quasilocalized states of the molecule, which makes the quasilocalized states more extended and increases the current through the molecule. In the present paper we considered only a G-A mispair, which results in small local distortion of the DNA helix. In this case we observe both features of mispair, (i) and (ii). The mispair, for which the distortion of the DNA molecule is strong, can be only a defect, resulting in suppression of the current through the molecule, see [13].

For both types of mispairs considered here, the main effect is a suppression of the current through DNA. Namely, about 78% and 66% of different random sequences of a DNA molecule show suppression of the current through the molecule under introduction of G(anti)-A(anti) and G(anti)-A(syn) mispairs, respectively. With an increase of the length of DNA the number of such sequences remains almost the same for G(anti)-A(anti) but decreases for the G(anti)-A(syn) mispair. The difference in the properties of G(anti)-A(anti) and G(anti)-A(syn) mispairs is related to the fact that G(anti)-A(syn) produces larger changes in the intrastand hopping integrals compared to the G(anti)-A(anti) mispair. As a result, depending on the DNA sequence, the G(anti)-A(syn) mispair becomes a stronger defect or a stronger coupling element compared to G(anti)-A(anti). This also results in a broader distribution function of the changes of the current through DNA for the G(anti)-A(syn) mispair.

Acknowledgments

The work was supported in part by the Canada Research Chairs Program and the NSERC Discovery Grant (both awarded to TC).

References


