Sequence Dependence of Electronic Transport in DNA

Antonio Rodríguez¹, Rudolf A. Römer∗², and Matthew S. Turner²

¹ Dpto Matemática Aplicada y Estadística, E.U.I.T. Aeronáutics, U.P.M., PZA Cardenal Cisneros s/n, Madrid 28040, Spain
² Physics Department & Centre for Scientific Computing, University of Warwick, Coventry CV4 7AL, U.K.

Received 21 August 2005, revised 27 August 2005
Published online Revision : 1.12, compiled 2/2/2008

Key words DNA, electronic transport, localization
PACS 72.15.Rn, 87.15.Cc, 73.63.-b

We study electronic transport in long DNA chains using the tight-binding approach for a ladder-like model of DNA. We find insulating behavior with localization lengths ξ ≈ 25 in units of average base-pair separation. Furthermore, we observe small, but significant differences between λ-DNA, centromeric DNA, promoter sequences as well as random-ATGC DNA.

1 Introduction DNA is a macro-molecule consisting of repeated stacks of bases formed by either AT (TA) or GC (CG) pairs coupled via hydrogen bonds and held in the double-helix structure by a sugar-phosphate backbone. In most models of electronic transport [1, 2] it has been assumed — following earlier pioneering work [3, 4] — that the transmission channels are along the long axis of the DNA molecule and that the conduction path is due to π-orbital overlap between consecutive bases [5].

A simple quasi-1D model incorporating these aspects has been recently introduced [6], building on an earlier, even simpler 1D model [1]. For the model, electronic transport properties have been investigated in terms of localisation lengths [6, 7], crudely speaking the length over which electrons travel. Various types of disorder, including random potentials, had been employed to account for different real environments. It was found that random and λ-DNA have localisation lengths allowing for electron motion among a few dozen base pairs only. However, poly(dG)-poly(dC) and also telomeric-DNA have much larger electron localization lengths. In Ref. [6], a novel enhancement of localisation lengths has been observed at particular energies for an increasing binary backbone disorder.

2 The DNA tight-binding model A convenient tight binding model for DNA can be constructed as follows: it has two central conduction channels in which individual sites represent an individual base; these are interconnected and further linked to upper and lower sites, representing the backbone, but are not interconnected along the backbone. Every link between sites implies the presence of a hopping amplitude. The Hamiltonian $H_L$ for this ladder-like model is given by

$$H_L = \sum_{i=1}^{L} \sum_{\tau=1,2} (t_{i,\tau}|i,\tau\rangle\langle i+1,\tau| + \varepsilon_{i,\tau}|i,\tau\rangle\langle i,\tau|) + \sum_{q=1,1} (t_q^0|i,q\rangle\langle i,q| + \varepsilon_q^0|i,q\rangle\langle i,q|) + \sum_{i=1}^{L} t_{1,2}|i,1\rangle\langle i,2|$$

(1)

∗ Corresponding author: e-mail: r.roemer@warwick.ac.uk, Phone: +44 2476 574 328, Fax: +44 7876 858 246

* pss data will be provided by the publisher
where $t_{i,\tau}$ is the hopping amplitude between sites along each branch $\tau = 1, 2$ and $\varepsilon_{i,\tau}$ is the corresponding onsite potential energy. $t_{i}^{q}$ and $\varepsilon_{i}^{q}$ give hopping amplitudes and onsite energies at the backbone sites. Also, $q(\tau) = \mp 1$ for $\tau = 1, 2$, respectively. The parameter $t_{12}$ represents the hopping between the two central branches, i.e., perpendicular to the direction of conduction. Quantum chemical calculations with semi-empirical wave function bases using the SPARTAN package [8] results suggest that this value, dominated by the wave function overlap across the hydrogen bonds, is weak and so we choose $t_{12} = 1/10$. As we restrict our attention here to pure DNA, we also set $\varepsilon_{i,\tau} = 0$ for all $i$ and $\tau$.

The model (1) clearly represents a dramatic simplification of DNA. Nevertheless, in Ref. [1] it had been shown that an even simpler model — in which base-pairs are combined into a single site — when applied to an artificial sequence of repeated GC base pairs, poly(dG)-poly(dC) DNA, reproduces experimental data current-voltage measurements when $t_{i} = 0.37\text{eV}$ and $t_{i}^{q} = 0.74\text{eV}$ are being used. This motivates the above parametrization of $t_{i}^{q} = 2t_{i}$ and $t_{i,\tau} \equiv 1$ for hopping between like (GC/GC, AT/AT) pairs. Assuming that the wave function overlap between consecutive bases along the DNA strand is weaker between unlike and non-matching bases (AT/GC, TA/GC, etc.) we thus choose $1/2$. Furthermore, since the energetic differences in the adiabatic electron affinities of the bases are small [9], we choose $\varepsilon_{i} = 0$ for all $i$. Due to the non-connectedness of the backbone sites along the DNA strands, the model (1) can be further simplified to yield a model in which the backbone sites are incorporated into the electronic structure of the DNA. The effective ladder model reads as

$$
\hat{H}_{L} = \sum_{i=1}^{L} t_{1,2}|i,1\rangle\langle i,2| + \sum_{\tau=1,2} t_{i,\tau}|i,\tau\rangle\langle i+1,\tau| + \left[ \varepsilon_{i,\tau} - \frac{\left(t_{i}^{q(\tau)}\right)^{2}}{\varepsilon_{i}^{q(\tau)} - E} \right] |i,\tau\rangle\langle i,\tau| + h.c. \quad (2)
$$

Thus the backbone has been incorporated into an energy-dependent onsite potential on the main DNA sites. This effect is at the heart of the enhancement of localization lengths due to increasing binary backbone disorder reported previously [6].

3 $\lambda$-DNA, centromers and promoters We shall use 2 naturally occurring DNA sequences (“strings”). (i) $\lambda$-DNA [28] is DNA from the bacteriophage virus. It has a sequence of 48502 base pairs and is biologically very well characterised. Its ratio $\alpha$ of like to unlike base-pairs is $\alpha_{\lambda} = 0.949$. (ii) centromeric DNA for chromosome 2 of yeast has 813138 base pairs [29] and $\alpha_{\text{centro.}} = 0.955$. This DNA is also rich in AT bases and has a high rate of repetitions which should be favourable for electronic transport.

Another class of naturally existing DNA strands is provided by so-called promoter sequences. We use a collection of 4986 is these which have been assembled from the TRANSFAC database and cover a range of organisms such as mouse, human, fly, and various viruses. Promoter sequences are biologically very interesting because they represent those places along a DNA string where polymerase enzymes bind and start the copying process that eventually leads to synthesis of proteins. On average, these promoters consist of approximately 17 base-pairs, much too short for a valid localization length analysis by TMM. Therefore, we concatenate them into a 86827 base-pair long super-promoter with $\alpha_{\text{super-p.}} = 0.921$. In order to obtain representative results, 100 such super-promoters have been constructed, representing different random arrangements of the promoters, and the results presented later will be averages$^1$.

Occasionally, we show results for “scrambled” DNA. This is DNA with the same number of A, T, C, G bases, but with their order randomised. Clearly, such sequences contain the same set of electronic potentials and hopping variations, but would perform quite differently in a biological context. A comparison of their transport properties with those from the original sequence thus allows to measure how important the exact fidelity of a sequence is. On average, we find for these sequences $\alpha_{\lambda/S} = 0.899$, $\alpha_{\text{centro.}/S} = 0.9951$ and $\alpha_{\text{super-p.}/S} = 0.901$.

$^1$ Averages of $\xi$ are computed by averaging $1/\xi$. 

© 2003 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
A convenient choice of artificial DNA strand is a simple, 100000 base-pair long random sequence of the four bases, random-ATGC DNA, which we construct with equal probability for all 4 bases ($\alpha_{\text{random}} = 0.901$). We shall also ‘promote’ these random DNA strings by inserting all 4086 promoter sequences at random positions in the random-ATGC DNA ($\alpha_{\text{random}}/P = 0.910$).

4 Results for localization lengths

For studying the transport properties of model (1), we use a variant of the iterative transfer-matrix method (TMM) [10–14]. The TMM allows us to determine the localisation length $\xi$ of electronic states in the present system with fixed cross sections $M = 2$ (ladder) and length $L \gg M$. Traditionally, a few million sites are needed for $L$ to achieve reasonable accuracy for $\xi$. However, in the present situation we are interested in finding $\xi$ also for much shorter DNA strands of typically only a few ten thousand base-pair long sequences. Thus in order to restore the required precision, we have modified the conventional TMM and can now perform TMM on a system of fixed length $L_0$ by repeating forward- and backward-TMM steps [6,15–17].

We have computed the energy dependence of the localization lengths for all sequences of section 3. In addition, $\lambda$-DNA, centromeric DNA and the super-promoter DNA where also scrambled 100 times and the localization length of each resulting sequence measured and the appropriate average constructed. Also, we constructed 100 promoted random-ATGC DNA sequences. As shown previously [6], the energy dependence of $\xi$ reflects the backbone-induced two-band structure. The obtained $\xi(E)$ values for the lower band are shown in Fig. 1. In the absence of any onsite-disorder, we find two prominent peaks separated by $t_{1,2}$ and $\xi(E) = \xi(-E)$. We also see that $\lambda$-DNA has roughly the same $\xi(E)$ dependence.
as random-ATGC-DNA. Promoting a given DNA sequence leads to small increases in localization length $\xi$, whereas scrambling can lead to increase as well as decrease. The super-promoter has larger $\xi$ values compared to random-atgc- and $\lambda$-DNA. Most surprisingly, centromeric DNA — the longest investigated DNA sequence — has a much larger localization length than all other DNA sequences and this even increases after scrambling.

5 Conclusions We have shown that the ladder model (1) is a simple, yet non-trivial representation of DNA within the tight-binding formalism. While keeping the number of parameters small, we manage to reproduce the wide-gap structure observed in much more accurate quantum chemical calculations of short DNA strands [1, 18–21]. In order to study the transport properties, we employ a variant of the TMM which provides useful information about the spatial extent $\xi$ of electronic states along a DNA strand in the quantum regime at $T = 0$. We note that the values of $\xi$ which we find are around 25 in the band which is surprisingly close to studies of range dependence of electron transfer [5, 22–27].

From our results, we find clear differences in localization lengths which are not simple related to a difference in DNA composition, but also reflect the order of base-pairs. Still the differences are within $10–20\%$ and it remains unclear how relevant these findings are biologically, i.e., whether electronic transport plays a role in the biological mechanism of DNA repair and protein generation.

Acknowledgements This work has been supported in part by the Royal Society. We thank A. Croy and C. Sohrmann for useful discussions.

References
[1] G. Cuniberti, L. Craco, D. Porath, and C. Dekker, Phys. Rev. B 65, 241314 (2002).
[2] J. Zhong, in Proceedings of the 2003 Nanotechnology Conference, Computational Publications, edited by M. Laudon and B. Romamowicz (PUBLISHERVERS, ADDRESS, 2003), Vol. 2, pp. 105–108, (Molecular and Nano Electronics).
[3] J. Ladik, M. Seel, P. Otto, and A. Bakhshi, Chem. Phys. 108, 203 (1986).
[4] A. Bakhshi, P. Otto, L. J., and M. Seel, Chem. Phys. 108, 215 (1986).
[5] C. R. Treadway, M. G. Hill, and J. K. Barton, Chemical Physics 281, 409 (2002).
[6] D. K. Klotza, R. A. Römer, and M. S. Turner, Biophys. J. 89, (2005).
[7] H. Yamada, Phys. Lett. A (2004).
[8] SPARTAN version 5.0, User’s Guide, Wavefunction Inc., 18401 Von Karman Ave., Suite 370 Irvine, CA 92612.
[9] S. S. Wesolowsky, M. L. Leininger, P. N. Pentchev, and H. F. Schaefer III, J. Am. Chem. Soc. 123, 4023 (2001).
[10] J.-L. Pichard and G. Sarma, J. Phys. C 14, L127 (1981).
[11] J.-L. Pichard and G. Sarma, J. Phys. C 14, L617 (1981).
[12] A. MacKinnon and B. Kramer, Z. Phys. B 53, 1 (1983).
[13] B. Kramer and A. MacKinnon, Rep. Prog. Phys. 56, 1469 (1993).
[14] A. MacKinnon, J. Phys.: Condens. Matter 6, 2511 (1994).
[15] K. Frahm, A. Müller-Groeling, J. L. Pichard, and D. Weinmann, Europhys. Lett. 31, 169 (1995).
[16] R. A. Römer and M. Schreiber, Phys. Rev. Lett. 78, 4890 (1997).
[17] M. L. Ndawana, R. A. Römer, and M. Schreiber, Europhys. Lett. 68, 678 (2004).
[18] O. R. Davies and J. E. Inglesfield, Phys. Rev. B 69, 195110 (2004).
[19] V. Bhalla, R. P. Bajpai, and L. M. Bharadwaj, European Molecular Biology reports 4, 442 (2003).
[20] I. L. Garzon et al., Nanotechnology 12, 126 (2001).
[21] A. Rakitin et al., Phys. Rev. Lett. 86, 3670 (2001).
[22] C. Wan et al., Proc. Natl. Acad. Sci. 97, 14052 (2000).
[23] E. Boon et al., Proc. Natl. Acad. Sci. 100, 12543 (2003).
[24] S. O. Kelley and J. K. Barton, Science 283, 375 (1999).
[25] C. J. Murphy et al., Science 262, 1025 (1993).
[26] M. A. O’Neil, C. Dohno, and J. K. Barton, Journal of the American Chemical Society Communications 126, 1316 (2004).
[27] S. Delaney and J. K. Barton, J. Org. Chem. 68, 6475 (2003).
[28] Bacteriophage lambda, complete genome [gi|9626243|ref|NC_001416.1|[9626243]], Genbank Accession number NC_001416, http://www.ncbi.nlm.nih.gov/entrez/
[29] CEN2, Chromosome II centromere, http://www.yeastgenome.org/
Greek symbols – w-greek.sty

\begin{table}
\begin{tabular}{cccc}
\alpha & \theta & \omicron & \tau \\
\beta & \vartheta & \pi & \upsilon \\
\gamma & \iota & \varpi & \phi \\
\delta & \kappa & \rho & \varphi \\
\epsilon & \lambda & \varrho & \chi \\
\varepsilon & \mu & \sigma & \psi \\
\zeta & \nu & \varsigma & \omega \\
\eta & \zeta & \xi & \\
\end{tabular}
\caption{Slanted greek letters}
\end{table}

\begin{table}
\begin{tabular}{cccc}
\Gamma & \Lambda & \Sigma & \Psi \\
\Delta & \Xi & \Upsilon & \Omega \\
\Theta & \Pi & \Phi & \\
\end{tabular}
\caption{Upright greek letters}
\end{table}
Table 3: Boldface variants of slanted greek letters

| Lower Case | Uppercase |
|------------|-----------|
| \(\alpha\) \text{\textbackslash \alpha}\) | \(\Theta\) \text{\textbackslash \Theta}\) |
| \(\beta\) \text{\textbackslash \beta}\) | \(\Pi\) \text{\textbackslash \Pi}\) |
| \(\gamma\) \text{\textbackslash \gamma}\) | \(\Sigma\) \text{\textbackslash \Sigma}\) |
| \(\delta\) \text{\textbackslash \delta}\) | \(\Upsilon\) \text{\textbackslash \Upsilon}\) |
| \(\epsilon\) \text{\textbackslash \epsilon}\) | \(\Phi\) \text{\textbackslash \Phi}\) |
| \(\zeta\) \text{\textbackslash \zeta}\) | \(\Psi\) \text{\textbackslash \Psi}\) |
| \(\eta\) \text{\textbackslash \eta}\) | \(\Omega\) \text{\textbackslash \Omega}\) |

Table 4: Boldface variants of upright greek letters

| Lower Case | Uppercase |
|------------|-----------|
| \(\alpha\) \text{\textbackslash \alpha}\) | \(\Theta\) \text{\textbackslash \Theta}\) |
| \(\beta\) \text{\textbackslash \beta}\) | \(\Pi\) \text{\textbackslash \Pi}\) |
| \(\gamma\) \text{\textbackslash \gamma}\) | \(\Sigma\) \text{\textbackslash \Sigma}\) |
| \(\delta\) \text{\textbackslash \delta}\) | \(\Upsilon\) \text{\textbackslash \Upsilon}\) |
| \(\epsilon\) \text{\textbackslash \epsilon}\) | \(\Phi\) \text{\textbackslash \Phi}\) |
| \(\zeta\) \text{\textbackslash \zeta}\) | \(\Psi\) \text{\textbackslash \Psi}\) |
| \(\eta\) \text{\textbackslash \eta}\) | \(\Omega\) \text{\textbackslash \Omega}\) |