Citation for published version:

DOI:
10.1016/j.cpc.2010.06.029

Publication date:
2011

Document Version
Peer reviewed version

Link to publication

NOTICE: this is the author's version of a work that was accepted for publication in Computer Physics Communications. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Computer Physics Communications, vol 182, issue 1, 2011, DOI 10.1016/j.cpc.2010.06.029

University of Bath

Alternative formats
If you require this document in an alternative format, please contact: openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Abstract

The electronic transmission properties of DNA molecules are believed to play a significant role in many physical phenomena taking place in living organisms [1]. Here we study the charge transport (CT) properties of cancer-related genes, including some of the most important tumor suppressors. We find that the changes in averaged CT around the sites of pathogenic and cancerous mutations are statistically smaller than those on sites where pathogenic mutations have not been observed. The results suggested that CT might be an indicator to discriminate between pathogenic and non-pathogenic mutations at an early stage. Mutations which cause little change in CT may be more likely to occur, or more likely to be missed by damage-repair enzymes which probe CT, and are therefore more likely to persist and cause disease.

Keywords: deoxyribonucleic acid, charge transport, cancer, mutation

We study the relation between the point mutations and CT properties of some of the most important tumor suppressors, together with other cancer-associated genes. These genes regulate cell proliferation by monitoring various molecular signaling pathways. Disfunction of the tumor suppressors will cause abnormal cell proliferation and the development of cancers [2]. CT through DNA is inhibited at the damaged sites of the sequence, owing to misalignments of base pair π-stacking [1]. The base excision repair (BER) enzymes such as Endonuclease III and MutY are believed to efficiently locate the DNA base lesions or mismatches by probing the inhibition of the DNA-mediated CT due to the damages [3].

We have previously studied [4] the CT properties of the genomic sequence of the p53 tumor suppressor gene, which is known as the “guardian of the genome”. The results show that on average the cancerous mutations of the gene yield smaller changes of the CT in contrast with non-cancerous mutations. Based on such a behavior of CT, we proposed a possible scenario of how cancerous mutations might circumvent the DNA damage-repair mechanism and survive to yield carcinogenesis [4]. Here we will add to this analysis by showing that the p53 gene with a randomly rearranged sequence does not show this effect at all. This test has also been applied to many other genetic sequences.

The native genetic sequences and mutations of cancer-related genes are retrieved from four different databases [5, 6]. The genomic sequence of a gene with length $N$ base pairs (bps) is defined as $(s_1, s_2, \cdots, s_N)$. Each point mutation of a given gene is characterized by the set $(k, s)$, where $k$ and $s$ are the position of the point mutation in the genomic sequence and the mutant nucleotide which replaces the nucleotide $s_k$ of normal DNA, respectively.

The simplest model of coherent hole transport in DNA is given by an effective one-dimensional
Figure 1: Schematic models for charge transport in DNA. The nucleobases are given as (dark grey) circles and ellipses. Electronic pathways are shown as solid lines of varying thickness to indicate variation in strength. The left model indicates the one-channel model where the sugar-phosphate backbone is ignored. In the right model, circles denote the smaller pyrimidines, ellipses are the large purines and (light grey) circles denote the sugar-phosphate backbone sites. Note that diagonal hopping between purines is favored, and between pyrimidines is disfavored, by the larger size of the purines.

Hückel-Hamiltonian for CT through nucleotide HOMO states [1], where each lattice point represents a nucleotide base (A,T,C,G) of the chain for \( n = 1, \ldots, N \). In this tight-binding formalism, the on-site potentials \( \epsilon_n \) are given by the ionization potentials \( \epsilon_G = 7.75 \text{ eV}, \epsilon_C = 8.87 \text{ eV}, \epsilon_A = 8.24 \text{ eV} \) and \( \epsilon_T = 9.14 \text{ eV} \), at the \( n \)th site, cp. Fig. 1; \( t_{n,n+1} \) is assumed to be nucleotide-independent with \( t_{n,n+1} = 0.4 \text{ eV} \) [1].

A model which is less coarse-grained is provided by the diagonal ladder model shown in Fig. 1. Here, both strands of DNA and the backbone are modelled explicitly and the different diagonal overlaps of the larger purines (A,G) and the smaller pyrimidines (C,T) are taken into account by suitable inter-strand couplings [7]. The intra-strand couplings are 0.35 eV between identical bases and 0.17 eV between different bases; the diagonal inter-strand couplings are 0.1 eV for purine-purine, 0.01 eV for purine-pyrimidine and 0.001 eV for pyrimidine-pyrimidine. Perpendicular couplings to the backbone sites are 0.7 eV, and perpendicular hopping across the hydrogen bond in a base pair is reduced to 0.005 eV.

The transmission coefficient \( T(E) \) for a DNA sequence with length \( N \) bps for different injection energy can be calculated for both models by using the transfer matrix method (TMM) [8, 9]. The position-dependent averaged transmission coefficient at the \( j \)-th base pair for transmission length \( L \) bps is defined as

\[
\bar{T}_{j,L} = \frac{1}{L} \sum_{n=j-L+1}^{j} \int_{E_0}^{E_1} \frac{T_{n,L}(E)}{E_1 - E_0} dE \tag{1}
\]

where \( T_{n,L}(E) \) is the transmission coefficient of a subsequence of the gene with length \( L \) and starting from the position \( n \). \( n \) ranges from \( j - L + 1 \) to \( j \) such that the subsequence contains the \( j \)-th base pair. \( E_0 \) and \( E_1 \) are the lower and upper bounds of the incident energy of the carriers, e.g. for the 1D model used here, the values are 5.75 and 9.75 eV, respectively; for the diagonal model the bounds are 7 and 11 eV. Then we examine the difference between \( T(E) \) of the normal and mutated genomic sequence of a point mutation [4]

\[
\Delta_{j,L}^{k,s} = \int_{E_0}^{E_1} \left| T_{j,L}(E) - T_{j,L}^{k,s}(E) \right|^2 \frac{1}{E_1 - E_0} dE \tag{2}
\]

where \( T_{j,L}^{k,s}(E) \) is the transmission coefficient of the same segment of DNA but with the point mutation \((k, s)\). \( \Gamma(k, s; L) \) is the averaged effect of the point mutation \((k, s)\) on CT properties for all subsequences containing the mutation and with length \( L \). We make use of values of \( L \) between 10 and 60 bps.

There are a total of \( 3^N \) possible point mutations of a gene with \( N \) bps, a set which we denote as \( M_a \). The subset of pathogenic mutations is \( M_p \). We calculate the averaged CT effect, \( \Gamma(k, s; L) \) for the full set \( M_a \) and compare the distributions for \( M_a \) and \( M_p \). Fig. 2 shows the distributions of \( \Gamma \) for the TP53 sequence of the p53 gene. For the 1-D model the full distribution is close to log-normal. For both the 1-D and diagonal models the distribution for pathogenic mutations is visibly shifted towards lower values compared to the full distribution. This difference disappears when the sequence is randomised.

We obtain a local ranking (LR) for each pathogenic mutation \((k, s)\) by comparing its CT change \( \Gamma(k, s; L) \) to those for the other two potential mutations at the same position, obtaining a ranking \( \gamma_{LR}(k, s; L) \) for the mutation \((k, s)\) of 1, 2 or 3 for lowest, middle and highest CT change.
Those $k$ with more than one pathogenic mutations are excluded in the LR analysis. Similarly, we define a global ranking (GR) by ranking the normalized CT changes for all possible $3N$ mutations. We assign global ranks $\gamma_{GR}(k, s; L)$ of 1 for the lowest third of the global distribution of CT changes, 2 for the middle third and 3 for the top third. In Fig. 3 we show the incidences of the three possible local and global ranks for the cancerous mutations of p53. In local rankings, the incidence of pathogenic mutations with lowest CT change, $\gamma_{LR} = 1$, is significantly higher than for $\gamma_{LR} = 2, 3$. However, this result is barely distinguishable from the distributions of local rankings for shuffled sequences. In the global ranking, however, the incidence of pathogenic mutations with $\gamma_{GR} = 1$ is higher than for $\gamma_{GR} = 2, 3$ and the distribution for the native sequence is distinguishable from the distributions for shuffled sequences.

In total, we have performed these statistical tests for 35 cancer DNA sequences [5, 6]. The results [10] show that in about 95% of the DNA strands the pathogenic mutations are biased towards smaller changes in CT properties. Hence it seems very likely the scenario proposed here for early pathogenesis can be applied for an analysis of a wide range of genetic diseases.

This work was supported by the National Science Council in Taiwan (97-2112-M-029-002-MY3), the UK Leverhulme Trust (F/00215/AH) and the National Center for High-Performance Computing in Taiwan.

References