Charge Transport in Cancer-Related Genes and Early Carcinogenesis

Chi-Tin Shih[a], Yun-Yin Cheng[a], Stephen A. Wells[b], Ching-Ling Hsu[c], Rudolf A. Römer[b]

[a] *Department of Physics, Tunghai University, 40704 Taichung, Taiwan and The National Center for Theoretical Sciences, 30013 Hsinchu, Taiwan*

[b] *Department of Physics and Centre for Scientific Computing, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK*

[c] *Department of Physics, Chung-Yuan Christian University, Chung-Li, Taiwan*

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: deoxyribonucleicacid, charge transport, cancer, mutation *Revision* : 1.28

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 $\mathcal{N}$ base pairs (bps) is defined as $(s_1, s_2, \cdots, s_{\mathcal{N}})$. 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

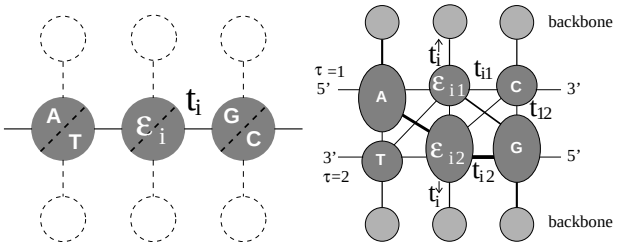

Figure 1: Schematic models for charge transport in DNA. The nucleobases are given as (dark grey) circles and ellipses. Electronic pathways are shown as solid lines of varying thickness to indicate variation in strength. The left model indicates the one-channel model where the sugar-phosphate backbone 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 ϵ_n are given by the ionization potentials $\epsilon_{\rm G} = 7.75e$V, $\epsilon_{\rm C} = 8.87e$V, $\epsilon_{\rm A} = 8.24e$V and $\epsilon_{\rm T} = 9.14e$V, at the nth site, cp. Fig. 1; $t_{n,n+1}$ is assumed to be nucleotide-independent with $t_{n,n+1} = 0.4e$V [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 $0.35e$V between identical bases and $0.17e$V between different bases; the diagonal inter-strand couplings are $0.1e$V for purine-purine, $0.01e$V for purine-pyrimidine and $0.001e$V for pyrimidine-pyrimidine. Perpendicular couplings to the backbone sites are $0.7e$V, and perpendicular hopping across the hydrogen bond in a base pair is reduced to $0.005e$V.

The transmission coefficient $T(E)$ for a DNA sequence with length Nbps 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

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

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 $9.75e$V, respectively; for the diagonal model the bounds are 7 and $11e$V. Then we examine the difference between $T(E)$ of the normal and mutated genomic sequence of a point mutation [4]

$$\bar{\Delta}^{k,s}_{j,L} = \int_{E_0}^{E_1} \frac{|T_{j,L}(E) - T^{k,s}_{j,L}(E)|^2}{E_1 - E_0} dE \qquad (2)$$

where $T^{k,s}_{j,L}(E)$ 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 $3\mathcal{N}$ possible point mutations of a gene with $\mathcal{N}$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 Γ 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 thsoe for the other two potential mutations at the same position, obtaining a ranking $\gamma_{\rm LR}(k, s; L)$ for the mutation (k, s) of 1,2 or 3 for lowest, middle and highest CT change.

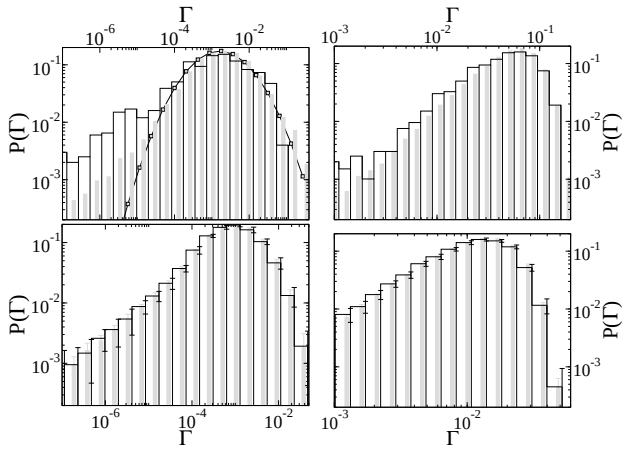

Figure 2: Distribution of the change in charge transport Γ for pathogenic (white bars) and all possible (grey bars) mutations for the $p53$ DNA strand with 20304 base pairs. The left/right columns show results for the 1D/diagonal models while the top/bottom rows indicate original/shuffled DNA sequences. For the randomly shuffled cases, error bars are calculated using 20 different sequence realizations. All results are for $L = 40$. The line for the 1-D case is a fit to a log-normal distribution.

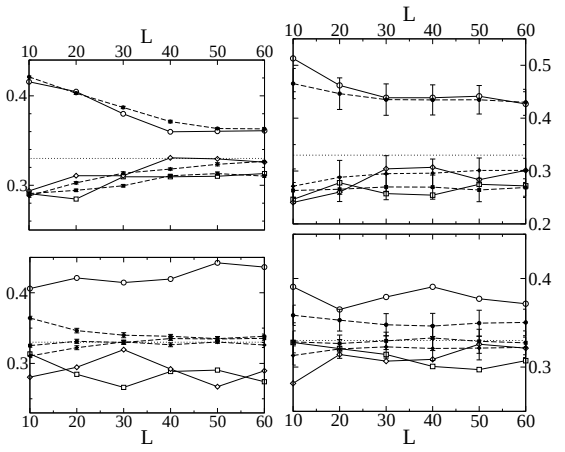

Figure 3: Distribution of the *local*, LR (top row) and *global*, GR (bottom row) ranking results of pathogenic mutations of $p53$ (open symbols) as a function of window lengths L. Circles indicate the proportion of pathogenic mutations with $\gamma = 1$; squares, $\gamma = 2$; diamonds, $\gamma = 3$. The closed symbols indicate averaged results for 20 randomly shuffled sequences. The left/right columns distinguish results for the 1D/diagonal models. The dotted horizontal lines show the 33% mark expected for a uniform random distribution.

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* $3\mathcal{N}$ mutations. We assign global ranks $\gamma_{\rm GR}(k, s; L)$ 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, $\gamma_{\rm LR} = 1$, is significantly higher than for $\gamma_{\rm LR} = 2, 3$. 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 $\gamma_{\rm GR} = 1$ is higher than for $\gamma_{\rm GR} = 2, 3$ 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

[1] Chaborty, T., editor, *Charge Migration in DNA: Perspectives from Physics, Chemistry and Biology*, Springer Verlag, Berlin, 2007.
[2] Sherr, C. J., Cell **116** (2004) 235.
[3] Rajski, S. R., Jackson, B. A., and Barton, J. K., Mutat. Res. **447** (2000) 49.
[4] Shih, C.-T., Roche, S., and Römer, R. A., Phys. Rev. Lett. **100** (2008) 018105.
[5] Steson, P. D. et al., Hum. Mutat. **21** (2003) 577, http://www.hgmd.cf.ac.uk/ac/index.php.
[6] Petitjean, A. et al., Hum. Mutat. **28** (2007) 622, http://www-p53.iarc.fr/index.html, R11.
[7] Wells, S. A., Shih, C.-T., and Römer, R. A., Int. J. Mod. Phys. B (2009) 4138.
[8] Roche, S., Phys. Rev. Lett. **91** (2003) 108101.
[9] Ndawana, M. L., Römer, R. A., and Schreiber, M., Europhys. Lett. **68** (2004) 678.
[10] Shih, C.-T., Cheng, Y.-Y., Wells, S. A., Hsu, C.-L., and Römer, R. A., (2010), unpublished.