Electric Transport Properties of the p53 Gene and the Effects of Point Mutations

Chi-Tin Shih¹*, Rudolf A. Römer², Stephan Roche³

¹ Department of Physics, Tunghai University, 40704 Taichung, Taiwan
² Department of Physics and Centre for Scientific Computing, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK
³ CEA/DSM/DRFMC/SPSMS/GT, 17 avenue des Martyrs, 38054 Grenoble, France

Received XXXX, revised XXXX, accepted XXXX
Published online XXXX

PACS 87.15.Aa, 87.14.Gg, 87.19.Xx

* Corresponding author: e-mail ctshih@thu.edu.edu, Phone +886-4-23590121 ext. 3474, Fax +886-4-23594643

In this work, charge transport (CT) properties of the p53 gene are numerically studied by the transfer matrix method, and using either single or double strand effective tight-binding models. A statistical analysis of the consequences of known p53 point mutations on CT features is performed.

1 Introduction The electronic transmission properties of DNA molecules are believed to play a critical role in many physical phenomena taking place in the living organisms [1,2,3,4]. For instance, it is believed that charge transfer (CT) through DNA is inhibited at the damaged sites of the sequence, owing to misalignments of base pair π-stacking. Similarly, base excision repair (BER) enzymes such as Endonuclease III and MutY are suggested to efficiently locate the DNA base lesions or mismatches by probing the DNA-mediated CT [5,6,7].

Besides, given that the development of cancers is closely related to the DNA damage/repair mechanism [8], the modifications of CT properties when mutations start to develop is therefore an important question to deepen. A most important gene in cancer research is p53 also known as the “guardian of the genome” [9]. Indeed, p53 encodes the tumor suppressor TP53 protein that suppresses the tumor development by activating the DNA repair mechanisms or the cell apoptosis process if DNA reparation is impossible. There are 20303 base pairs in the p53 sequence (NCBI accession number X54156). More than 50% of human cancers are related to the mutations of the p53 gene which usually jeopardize the efficient activity of TP53 [10]. Most of the cancerous mutations are point mutations — a base pair substituted by another — with distributions along the DNA sequence that are highly non-uniform [11]. Each point mutation can be described by two parameters (k, s), respectively giving the mutation position k on the sequence and the nucleotide type s (either A, C, G, or T) substituting the original one. The most frequent mutation locations found in the cancer cells are named mutation “hotspots”. From the International Agency for Research on Cancer (IARC) database [11], it is found that most hotspots of p53 are located in the exons 5 ~ 8 in the interval from the 13055th to the 14588th nucleotide. The 13203th base pair has the highest frequency of occurrence (1055 times) and more than 80% of the total 23544 cases in the database occur on 1% of the base pairs of the p53. The mutation (k, s) is said to be “cancerous” (“noncancerous”) if it is (not) found in the IARC database.

In this paper, the effects of all possible point mutations on CT are studied for the p53 gene using appropriate tight-binding models and energy parameters which are know to reproduce experimental results or first principle calculations [12,13]. We find that anomalously small changes of CT efficiency modulations coincide with cancerous mutations. In contrast, non-cancerous mutations result, on average, in much larger changes of the CT properties. From

Copyright line will be provided by the publisher
this analysis, we propose a new scenario for understanding the underlying origin of how cancerous mutations shortcut the DNA damage/repair processes.

2 Models for charge transport in DNA

The generic form of the simple but physically sounding 1-channel model of coherent hole transport of DNA is given by an effective tight-binding Hamiltonian (the “fishbone model” (FB)) \[12\]

\[
H_{FB} = \sum_{i=1}^{L} \sum_{q=\uparrow,\downarrow} \left( -t_{i}|i\rangle\langle i+1| - t_{q}^{q}|i,q\rangle\langle i| \right) + \varepsilon_{i}|i\rangle\langle i| + \varepsilon_{q}^{q}|q,i\rangle\langle i,q| + h.c.
\] (1)

where each lattice point stands for a nucleotide base pair of the chain for \(i = 1, \ldots, L\). \(t_{i}\) is the hopping amplitude between \(i\)th and \((i+1)\)th base pairs and \(\varepsilon_{i}\) is the on-site potential of the \(i\)th base pair. \(t_{q}^{q}\) with \(q = \uparrow, \downarrow\) is the hopping amplitude between the \(i\)th base pair and its neighboring (upper and lower) backbone sites \(|i,q\rangle\). The onsite energy at the sites \(|i,q\rangle\) is given by \(\varepsilon_{i}^{q}\). The model will be reduced to the simplest one-ladder (1L) model if the sugar-phosphate backbone sites \(|i,q\rangle\) of DNA are absent, that is, \(t_{q}^{q} = \varepsilon_{q}^{q} = 0\) \[14,15,16,17\]. This one-channel model is shown schematically in Fig. 1(a).

To account for the full double-strand nature of DNA, an alternative two-channel ladder model (LM) shown in Fig. 1(b) is also used. The corresponding Hamiltonian is given as \[18\]

\[
H_{LM} = \sum_{i=1}^{L} \sum_{\tau=1,2} \left( t_{i,\tau}|i,\tau\rangle\langle i+1,\tau| + \varepsilon_{i,\tau}|i,\tau\rangle\langle i,\tau| \right) + \sum_{q=\uparrow,\downarrow} \left( \sum_{\tau=1,2} t_{q}^{q}|i,\tau\rangle\langle i,q(\tau)| + \varepsilon_{q}^{q}|i,q\rangle\langle i,q| \right) + t_{1,2}|i,1\rangle\langle i,2| + h.c.
\] (2)

where \(t_{i,\tau}\) is the hopping amplitude between the sites along each branch \(\tau = 1, 2\) and \(\varepsilon_{i,\tau}\) is the corresponding onsite energy. \(t_{1,2}\) represents the hopping between the nucleotides of each base pair. Again, the model will be reduced to a two-leg (2L) model if the backbone sites are not taken into account \[19,20,21\].

The onsite energies for each base are chosen according to the ionization energies, \(\varepsilon_{A} = 8.24\text{eV}, \varepsilon_{G} = 8.87\text{eV}, \varepsilon_{T} = 7.75\text{eV}\) and \(\varepsilon_{C} = 9.14\text{eV}\) \[22,23,24,25,26\] for each model. For model 1L, the hopping term between pairs base are all set as \(t_{n} = 0.4\text{eV}\). Other values ranging from 0.1 to 1 eV are also used to investigate the robustness of our conclusion. For model FB, \(t_{n} = 0.4\text{eV}\) as in 1L. The additional hopping terms linking to the backbone are taken as 0.75 eV, whereas all backbone onsite energies are assumed to be 8.5 eV, roughly equal to the average value of all onsite energies for the base pairs. The hopping terms in model 2L between the same kind of base pairs \((AT/AT, GC/GC, etc.)\) are chosen as 0.35 eV, and 0.17 eV otherwise \[12\]. Interchain coupling constant is fixed to \(t_{\perp} = 0.1\text{eV}\). Last, in model LM, intrachain and interchain hopping strengths are taken as in the two-leg model 2L. Additionally, the backbone energetics is treated as in the fishbone model case.

3 Method

The most convenient method to evaluate the transport properties of these quasi-one-dimensional tight-binding models is known as the transfer matrix method (TMM) \[27,28,29,30,31\]. This approach allows to determine the hole transmission coefficient \(T(E)\) in systems varying cross section \(M\) and length \(L\gg M\). In brief, the eigenstates \(|\Psi\rangle = \sum_{n} \psi_{n} |n\rangle\) (here \(|n\rangle\) denotes the \(n\)th site position of the hole) of the Hamiltonian are computed from \((\psi_{L}, \psi_{L-1})^T = \tau_{L} \cdot (\psi_{1}, \psi_{0})^T\) where \(\tau_{L}(E)\) is the global transfer matrix \[30\]. \(E\) is the energy of the injected carrier. The localization lengths are deduced from the scaling analysis of \(T(E)\), whatever the used effective model \[27,28,29,30\]. Besides, when assuming that the DNA sequences are connected to the semi-infinite metallic electrodes \[2\], \(T(E)\) takes the following analytical form \[16,22,33,34,35\]

\[
T(E) = \frac{4 - E^2}{P + 2 - E^2 \tau_{11} \tau_{22} + E (\tau_{11} - \tau_{22})(\tau_{12} - \tau_{21})}
\] (3)

with \(E = (E - \varepsilon_{m})/t_{0}\) and \(P = \sum_{i,j=1,2} \tau_{ij}^{2}\). \(\varepsilon_{m}\) and \(t_{0}\) are the onsite energies and the hopping integral of the electrode energetics, respectively. It is readily shown that

\[
\tau_{L} = \begin{pmatrix}
\tau_{11} & \tau_{12} \\
\tau_{21} & \tau_{22}
\end{pmatrix} = M_{L} M_{L-1} \ldots M_{2} M_{1}
\] (4)

with

\[
M_{n} = \begin{pmatrix}
\frac{E-\varepsilon_{m}}{t_{n}} & \frac{t_{n-1}}{t_{n}} \\
1 & 0
\end{pmatrix}
\] (5)

where \(\varepsilon_{m} = \varepsilon_{n}\) for 1L and \(\varepsilon_{m} - \sum_{q} \frac{\varepsilon_{q}^{q}}{E^{q}}\) for FB, respectively \[18\]. In the following \(\varepsilon_{m} = \varepsilon_{G} = 7.75\text{eV}\) and \(t_{0} = 1\text{eV}\).
To analyze the position-dependent transport properties of p53 gene and the effect of point mutations, let us define \( S = (s_1, s_2, \ldots, s_{20303}) \) as a finite-length sequence of the p53 gene. \( S_{jL} \) is a segment of \( S \) starting from the \( j \)th base pair with length \( L \), that satisfies \( S_{jL}(n) = S(n + j - 1) \) with \( n = 1, 2, \ldots, L \). The transmission coefficient as a function of energy is denoted as \( T_{jL}(E) \). CT for the \( j \)th site with propagation length \( L \) is defined as the averaged value of the integrated \( T_{jL}(E) \) (for all incident energies) of all \( L \) possible subsequences of p53 containing the \( j \)th site and with length \( L \)

\[
T_{jL} = \frac{1}{L} \sum_{n=j-L+1}^{j} \frac{1}{E_1 - E_0} \int_{E_0}^{E_1} T_{n,L}(E)dE.
\] (6)

where \( n \) is further restricted to \( 1 \leq n \leq 20304 - L \) close to the boundaries; \( E_0 \) and \( E_1 \) denote a suitable energy window which we shall normally choose to equal the extrema of the energy spectrum for each model, i.e. [6.5, 10.5] for model 1L, [7.5, 10.5] for 2L, [8, 9.5] for FB and [5, 15] for LM.

If the \( k \)th base on the p53 sequence is mutated from \( s_k \) to \( s \) and \( j \leq k \leq j + L - 1 \) (i.e., the mutated site belongs to the segment \( S_{jL} \)), the mutated sequence will be denoted as \( S'_{jL} \). \( S'_{jL}(k - j + 1) = s \) and \( S'_{jL}(i \neq k - j + 1) = S_{jL}(i) \). The transmission coefficients of the original and mutated sequences are denoted as \( T_{jL}(E) \) and \( T'_{jL}(E) \), respectively. The squared difference of the transmission coefficients between the wild and mutated sequences is defined as

\[
\Delta_{jL}^{ks} = [T_{jL}(E) - T'_{jL}(E)]^2.
\] (7)

And \( \Delta_{jL}^{ks}(E) \) is then summed for all incident energy \( E \) as

\[
\bar{\Delta}_{jL}^{ks} = \frac{1}{E_1 - E_0} \int_{E_0}^{E_1} dE \Delta_{jL}^{ks}(E).
\] (8)

Finally, \( \bar{\Delta}_{jL}^{ks} \) is averaged over all segments with length \( L \) containing the mutation site \( (k, s) \), to give the average effect of the mutation \( (k, s) \) on the change of CT for p53

\[
\Gamma(k, s, L) = \frac{1}{L} \sum_{j} \bar{\Delta}_{jL}^{ks}.
\] (9)

4 Results and discussion The 14585th base (exon 8, codon 306) of the p53 sequence is found 133 times in the IARC database that mutates from \( C \) to \( T \) and causes various types of cancer [16]. On the other hand, the mutations \( C \rightarrow G \) and \( C \rightarrow A \) are said to be noncancerous since they are never found in cancer cells. The effects of the cancerous \( (C \rightarrow T) \) and noncancerous mutations \( T_{14570,31}(E) \) and \( \Delta_{14570,31}(E) \) for the models FB and 2L are shown in Figure 2 respectively. The overall effect of these three mutations \( \Gamma(14585, s, L = 20, \ldots, 100) \) for all the 4 models is given in Table 1.

It is clear from Table 1 that for many cases the CT change due to cancerous mutation is much smaller than noncancerous mutations. These results are stable over a wide range of \( L \) and model parameters. This suggests a scenario to understand how specific mutation hotspots could be robust against repair mechanism, and trigger carcinogenesis. Experimentally, the BER enzymes can locate the damaged sites on DNA by probing the CT of the segment.
bound by the enzymes [6,7]. If a mutation only weakly changes the CT, the enzymes will not be able to find it and the repair mechanism will not be activated. Such mutations will survive DNA repair mechanisms and yield cancers. In contrast, those mutations that strongly affect CT could be more easily detected by the CT probing mechanism of enzymes and therefore repaired.

The results presented thus far are for a particular hotspot. To further challenge this scenario, many more hotspots of the p53 gene have been analyzed. We have thus calculated $\Delta(k, s, L)$ for 14 hotspots with the highest mutation frequencies and for $L$ up to 160 in all 4 models. The results show that the qualitative behavior of each hotspot for all models is similar. Thus the following analysis is performed on all hotspots for the 1L model. Fig. 4(b) shows the correlation between frequency found in the cancer cells and the CT change $\Gamma(k, s, L = 80)$ with $t_n = 0.4$ eV. It is clear that the hotspots with highest frequencies correspond to smaller $\Gamma$. Thus the correlation observed in Fig. 2 for the 14585th site is common for most of the hotspots. Fig. 4(a) and (c) show similar behaviors for $t_n = 0.1$ and 1 eV, respectively. The scenario is thus found to be robust for a wide range of $t_n$.

5 Conclusion The CT modifications due to all possible point mutations of the p53 tumor suppressor gene have been analyzed by TMM together with statistical methods. The results show that on average the cancerous mutations of the gene yield smaller changes of the CT in contrast with non-cancerous mutations. The tendency is valid for the 4 studied tight-binding models (1L, FB, 2L, and LM) and is robust for a wide range of the hopping integral $t_n$ ($0.1 \sim 1.0$ eV).

These results suggest a possible scenario of how cancerous mutations might circumvent the DNA damage-repair mechanism and survive to yield carcinogenesis. However, our analysis is only valid in a statistical sense and we do observe occasional non-cancerous mutations with weak changes of CT. For these, other DNA repair processes should exist and we therefore do not intend to claim that the DNA-damage repair solely uses a CT-based criterion. Still, our results exhibit an intriguing and new correlation between the electronic structure of DNA hotspots and the DNA damage-repair process.

One notes that to further support the abovementioned scenario, additional complexities of the DNA energetics

![Figure 4](image-url)
should also be considered. This includes to investigate the role of electron-phonon coupling, polaronic transport, more detailed sequence-dependent energetics such as two-strand couplings, electronic correlations, as well as metal/DNA contact interactions. Ultimately, experimental studies of short strands of wild and mutated subsequences of the p53 gene should be performed to challenge our theory. The lengths scales of DNA required to unveil our mechanism are already within the scope of experimental measurements.

Acknowledgements We thank J. Cadet, D. Gasparutto, Th. Douki, M. F. Yang, and M. Turner for useful discussions. This work is partially supported by the National Science Council and National Center for Theoretical Sciences in Taiwan (CTS, grant no. 95-2112-M-029-003-) and the UK Leverhulme Trust (RAR, grant no. F/00 215/AH). Part of the calculations were performed in the National Center for High-Performance Computing in Taiwan.

References

[1] R. G. Endres, D. L. Cox, and R. P. Singh, Rev. Mod. Phys. 76, 195–214 (2004).
[2] T. Chakraborty (ed.), Charge Migration in DNA: Perspectives from Physics, Chemistry and Biology (Springer-Verlag, Berlin, 2007).
[3] R. Gutierrez and G. Cuniberti, NanoBioTechnology: Bioinspired device and materials of the future, (Humana Press, Totowa, 2007), chap. Hamiltonian approaches to charge transport in DNA molecular wires.
[4] Y. A. Berlin, I. V. Kurnikov, D. N. Beratan, M. A. Ratner, and A. L. Burin, Topics in Current Chemistry, (Springer, Berlin, 2004), pp. 1–36.
[5] S. R. Rajski, B. A. Jackson, and J. K. Barton, Mutat. Res. 447, 49–72 (2000).
[6] E. Yavin, A. K. Boal, E. D. A. Stemp, E. M. Boon, and A. L. Livingston, V. O. L. O’Shea, S. S. David, and J. K. Barton, Proc. Natl. Acad. Sci. 102, 3546 (2005).
[7] E. Yavin, E. D. A. Stemp, V. L. O’Shea, S. S. David, and J. K. Barton, Proc. Natl. Acad. Sci. 103, 3610 (2007).
[8] S. Loft and H. E. Poulsen, J. Mol. Med. 74, 297–312 (1996).
[9] D. P. Lane, Nature 358, 15 (1992).
[10] C. J. Sherr, Cell 234, 116 (2004).
[11] A. Petitjean, E. Mathe, S. Kato, C. Ishioke, S. Tavtigian, P. Hainaut, and M. Olivier, Hum. Mutat. 28(6), 622–29 (2007), http://www-p53.iarc.fr/index.html, R11.
[12] G. Cuniberti, L. Craco, D. Porath, and C. Dekker, Phys. Rev. B 65, 241314–4 (2002).
[13] E. B. Starikov, Phil. Mag. 85(29), 3435–3462 (2005).
[14] Y. A. Berlin, A. L. Burin, and M. A. Ratner, Superlattices and Microstructures 28(4), 241–252 (2001).
[15] Y. A. Berlin, A. L. Burin, and M. A. Ratner, Chem. Phys. 275, 61–74 (2002).
[16] S. Roche, Phys. Rev. Lett. 91, 108101–4 (2003).
[17] S. Roche, D. Bicout, E. Maciá, and E. Kats, Phys. Rev. Lett. 91, 228101–4 (2003).
[18] D. K. Klotz, R. A. Römer, and M. S. Turner, Biophys. J. 89, 2187–2198 (2005).
[19] A. V. Malyshev, Phys. Rev. Lett. 98(9), 096801 (2007).
[20] X. F. Wang and T. Chakraborty, Phys. Rev. Lett. 97, 106602 (2006).
[21] R. Gutierrez, S. Mohapatra, H. Cohen, D. Porath, and G. Cuniberti, Phys. Rev. B 74, 235105 (2006).
[22] H. Sugiyama and I. Saito, J. Am. Chem. Soc. 118(30), 7063–7068 (1996).
[23] A. A. Voityuk, J. Jortner, M. Boixin, and N. Rösch, J. Chem. Phys. 114(13), 5614–5620 (2001).
[24] H. Zhang, X. Q. Li, P. Han, X. Y. Yu, and Y. Yan, Journal of Chemical Physics 117(9), 4578–4584 (2002).
[25] X. Yang, X. B. Wang, E. R. Vorpagel, and L. S. Wang, Proc. Nat. Acad. Sci. 101(51), 17588–17592 (2004).
[26] E. Cauet, D. Dehareng, and J. Lievin, J. Phys. Chem. A 110(29), 9200–9211 (2006).
[27] J. L. Pichard and G. Sarma, J. Phys. C 14, L127–L132 (1981).
[28] J. L. Pichard and G. Sarma, J. Phys. C 14, L617–L625 (1981).
[29] A. MacKinnon and B. Kramer, Z. Phys. B 53, 1–13 (1983).
[30] B. Kramer and A. MacKinnon, Rep. Prog. Phys. 56, 1469–1564 (1993).
[31] A. MacKinnon, J. Phys.: Condens. Matter 6, 2511–2518 (1994).
[32] A. Chakrabarti, R. A. Römer, and M. Schreiber, Phys. Rev. B 68(19), 195417 (2003).
[33] E. Maciá, Phys. Rev. B 60(14), 10032–10036 (1999).
[34] C. T. Shih, phys. stat. sol. (b) 243(2), 378–381 (2006).
[35] C. T. Shih, Phys. Rev. E 74(1), 010903 (2006).
[36] All three types of mutations on the positions of the eight most frequent mutations are cancerous. So we choose this 9–th hotspot because it has two non-cancerous mutations for comparison.
[37] E. Díaz, A. Sedrakyan, D. Sedrakyan, and F. Domínguez-Adame, Phys. Rev. B 75, 014201 (2007).
[38] K. Senthilkumar, F. C. Grozema, C. F. Guerra, F. M. Bickelhaupt, and L. D. A. Siebbeles, J. Am. Chem. Soc. 125, 13658–13659 (2003).
[39] E. M. Conwell and S. V. Rahkmanova, Proc. Nat. Acad. Sci. 97(9), 4556–4560 (2000).
[40] E. M. Conwell, Proc. Natl. Acad. Sci. 102, 8795–8799 (2005).
[41] J. H. Wei, L. X. Wang, K. S. Chan, and Y. Yan, Phys. Rev. B 72(6), 064304 (2005).
[42] E. B. Starikov, Phil. Mag. Lett. 83, 699 (2003).
[43] E. Maciá, F. Triozon, and S. Roche, Phys. Rev. B 71, 113106 (2005).
[44] D. Porath, A. Bezryadin, S. Vries, and C. Dekker, Nature 403, 635–638 (2000).
[45] B. Xu, P. Zhang, X. Li, and N. Tao, Nano Lett. 4, 1105 (2004).
[46] H. Cohen, C. Nogues, R. Naaman, and D. Porath, Proc. Nat. Acad. Sci. 102, 11589 (2005).
[47] A. Y. Kasumov, M. Kociak, S. Guérin, B. Reulet, V. T. Volkov, D. V. Klinov, and H. Bouchiat, Science 291, 280–282 (2001).