# Point-Mutation Effects on Charge-Transport Properties of the Tumor-Suppressor Gene *p*53

Chi-Tin Shih,[1] Stephan Roche,[2] and Rudolf A. Römer[3]

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

We report on a theoretical study of point mutations effects on charge transfer properties in the DNA sequence of the tumor-suppressor *p*53 gene. On the basis of effective tight-binding models which simulate hole propagation along the DNA, a statistical analysis of mutation-induced charge transfer modifications is performed. In contrast to noncancerous mutations, mutation hot spots tend to result in significantly weaker *changes of transmission properties*. This suggests that charge transport could play a significant role for DNA-repairing deficiency yielding carcinogenesis.

PACS numbers: 87.15.A−, 87.14.G−, 87.19.xj

The charge transfer properties and long-range oxidation mechanisms in DNA molecules play a critical role in living organisms [1,2]. For instance, it is believed that base excision repair (BER) enzymes locate the DNA base lesions or mismatches by probing the DNA-mediated charge-transport (CT) [3,4]. The *p*53 DNA is said to be the "guardian of the genome" since it encodes the *TP*53 protein that suppresses the tumor development by activating the DNA repair mechanisms or the cell apoptosis process if the damage of DNA is irreparable. More than 50% of human cancers are related to the mutations of the *p*53 gene which jeopardize the efficient functioning of *TP*53 [5]. Most of the cancerous mutations are point mutations—a base pair substituted by another—with distributions along the DNA sequence that are highly nonuniform [6]. The positions where the mutations occur most frequently are called the "hot spots" of mutations. Each point mutation can be characterized by two parameters $k$ and $s$, respectively, representing the position of the mutation on the sequence and the nucleotide substituting the original one. From the IARC database [6], one finds that most hot spots of *p*53 are located in the exons 5, 6, 7, and 8 in the interval from the 13 055th to the 14 588th nucleotide. The distribution of the point mutations in this range is reported in Fig. 1. In this Letter, by using single and double strands tight-binding models with parameters fitted from *ab initio* calculations [2,7], the charge transmission changes owing to cancerous and noncancerous point mutations are statistically studied for the *p*53 gene. We find that anomalously small changes of charge transfer efficiency modulations coincide with cancerous mutations. In contrast, noncancerous mutations result, on average, in much larger changes of the CT properties. We suggest a new scenario how cancerous mutations could shortcut the DNA damage or repair processes and hence yield carcinogenesis.

The electronic energetics of a DNA chain should take into account three different contributions coming from the nucleobases system, the backbone system and the environment [2]. We emphasize that in many of the models to be used here, simplifying assumptions about these energy scales have to be employed. Mostly, however, the ionization energies $\epsilon_G = 7.75$ eV, $\epsilon_C = 8.87$ eV, $\epsilon_A = 8.24$ eV, and $\epsilon_T = 9.14$ eV [8], are taken as suitable approximations for the onsite energetics at each base as well as 7.75 eV for the electrodes [8–12]. A simple but physically reasonable description of coherent hole transport in single strand DNA is then given by an effective tight-binding Hamiltonian [9]

$$H = \sum_n \epsilon_n c_n^\dagger c_n - \sum_n t_{n,n+1}(c_n^\dagger c_{n+1} + \text{H.c.}), \quad (1)$$

where each lattice point represents a nucleotide base (*A*, *T*, *C*, *G*) of the chain for $n = 1, \ldots, N$. This one-leg (1L) model is shown schematically in Fig. 2(a). In this tight-binding formalism, $c_n^\dagger$ ($c_n$) is the creation (destruction) operator of a hole at the $n$th site. The $t_{n,n+1}$ are the hopping integrals along the DNA. $\epsilon_n$ is chosen equal to the ionization potential at the $n$th site. Furthermore, in the 1L model $t_{n,n+1}$ is assumed to be nucleotide independent with
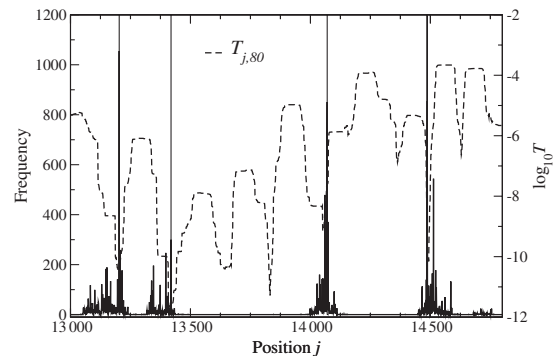


FIG. 1. Mutation frequency of each site (thin lines) and averaged transmission coefficient $\bar{T}_{j,80}$ (dashed line). Vertical dotted lines denote known regions of frequent mutations (hot spots).
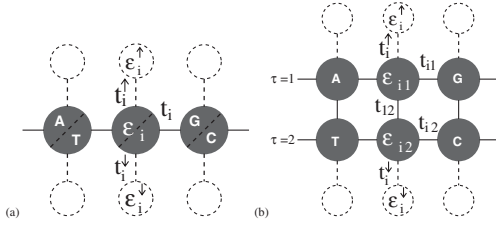
FIG. 2. Schematic models for hole transport in DNA. The nucleobases are given as (gray) circles. Electronic pathways are shown as lines, and dashed lines and circles denote the sugar-phosphate backbone. Graph (a) shows effective models 1L and FB (with dashed backbone) for transport along a single channel, whereas graph (b) depicts possible two-channel transport models 2L and LM (with dashed backbone).

$t_{n,n+1} = 0.4$ eV following prior modelling in agreement with *ab initio* calculations [9,13].

A straightforward generalization of model (1) includes a two-leg ladder model (2L) as shown in Fig. 2(b). The hopping between like base pairs (AT-AT, GC-GC, etc.) is chosen as 0.35 eV, between unlike base pairs it is 0.17 eV; the interchain hopping $t_\perp = 0.1$ eV. Other models [2] include the presence of sites which represent the sugar-phosphate backbone of DNA but along which no electron transport is allowed (cf. Fig. 2). In the following we call the one-channel variety a *fishbone* (FB) and the two-channel version *ladder model* (LM). The additional hopping onto the backbone is 0.7 eV and the backbone onsite energy is taken to be 8.5 eV, roughly equal to the mean of all onsite energies for the base pairs.

The most convenient method for studying the transport properties of these four quasi-one-dimensional tight-binding models is the transfer-matrix method [14], which allows to determine the transmission coefficient $T(E)$ of hole states in systems with varying cross section $M$ and length $L \gg M$. Briefly, we can solve for the eigenstates $|\Psi\rangle = \sum_n \psi_n |n\rangle$ of the Hamiltonian, where $|n\rangle$ represents the state that the hole is located in the $n$th site, as $(\psi_L, \psi_{L-1})^T = \tau_L (\psi_1, \psi_0)^T$ where $\tau_L(E)$ is the global transfer matrix [14]. $E$ is the energy of the injected carrier. The transmission $T(E)$ is given in terms of $\tau_L(E)$ by a simple analytic formula [10] for the 1L and FB models and can be computed from the localization lengths for the 2L and LM models [14].

Let us define $S = (s_1, s_2, \cdots, s_{20\,303})$ as the sequence of the $p53$ gene (NCBI access number $X54\,156$, $20\,303$ base pairs) [15], whereas $S_{j,L}$ is a segment of $S$ with length $L$ starting at the $j$th base pair, i.e. $S_{j,L}(n) = S(j-1+n)$ with $n = 1, 2, \ldots, L$. Next, we denote by $T_{j,L}(E)$ the transmission coefficient corresponding to $S_{j,L}$. We then characterize the energy-averaged CT for the $j$th site with segment length $L$ as the value $\bar{T}_{j,L}$ obtained by integrating $T_{j,L}(E)$ for all incident energies and all possible $L$ subsequences of all $p53$ segments of length $L$ containing the $j$th site such that

$$\bar{T}_{j,L} = \frac{1}{L} \sum_{n=j-L+1}^{j} \frac{1}{E_1 - E_0} \int_{E_0}^{E_1} T_{n,L}(E)dE. \quad (2)$$

where $n$ is further restricted to $1 \le n \le 20\,304 - L$ 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. In Fig. 1 we show $\bar{T}_{j,80}$ for model 1L and base pair range $13\,000 < j < 14\,800$ where the most cancerous mutations occur. The positions of four groups of hot spots, i.e., peaks of the mutation frequency, corresponding to the four exons (5th–8th) coincide with local minima of $\bar{T}_{j,80}$.

If the $k$th base on the $p53$ sequence is mutated from $s_k$ to $s$ and $j \le k \le j + L - 1$, we will denote the mutated segment containing this mutation as $S_{j,L}^{k,s}$ such that $S_{j,L}^{k,s}(k - j + 1) = s$ and $S_{j,L}^{k,s}(i) = S_{j,L}(i)$ for all $i \ne k - j + 1$. The corresponding transmission coefficients of the original and mutated sequence are denoted as $T_{j,L}(E)$ and $T_{j,L}^{k,s}(E)$, respectively. Similarly, we define the energy-averaged squared differences in transmission coefficient between original and mutated sequence as

$$\bar{\Delta}_{j,L}^{k,s} = \frac{1}{E_1 - E_0} \int_{E_0}^{E_1} |T_{j,L}(E) - T_{j,L}^{k,s}(E)|^2 dE. \quad (3)$$

The $14\,585$th base pair of the $p53$ sequence is a particularly active hot spot with 133 entries in the IARC database [6]. It exhibits mutations from C to T [16] and causing various types of cancer. However, the mutations $C \to G$ and $C \to A$ at the same position are not cancerous. The effects of the cancerous $C \to T$ and the noncancerous $C \to A$, $C \to G$ mutations on the CT properties are shown in Fig. 3. The transmission coefficients $T_{14\,575,20}(E)$ and $T_{14\,575,20}^{14\,585,s}(E)$ with $s = T$, A and G are given in Fig. 3. We
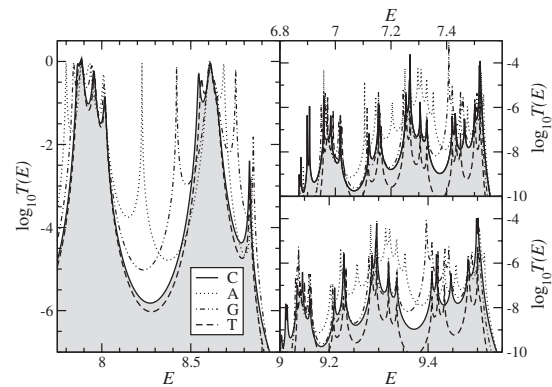


FIG. 3. Energy-dependence of logarithmic transmission coefficients $T_{14575,20}^{14585,s}(E)$ of the original sequence (C shaded solid line) and mutated (A dotted, G dotted-dashed, T dashed) sequences with length $L = 20$ (from $14\,575$th to $14\,594$th nucleotide) of $p53$. The left panel shows results for model 1L, the right two panels denote the two transport windows for the fishbone model [19].

find that for most energies the mutation $C \rightarrow T$ results in the weakest change in $T(E)$.

To evaluate the change of CT for *all* mutations in $p53$ quantitatively, $\bar{\Delta}_{14\,575,20}^{14\,585,s}$ are computed for all the four models. The results are shown in Table I. We see that the cancerous mutation $C \rightarrow T$ shows the smallest relative change in CT for nearly all models. The only differences occur for small $L = 20$ in models FB and LM but vanish quickly for larger $L$. Hence for a damage-repair process which uses a CT-based criterion as a detection mechanism, this mutation will be the hardest to identify. These results seem to suggest a scenario in which certain mutations might avoid the CT-driven DNA-damage-repair mechanism and survive to develop cancerous tumors. We have checked that this trend is independent of the specific model and hot spot chosen by analyzing also the hot spots 13 117, 13 203, 13 334, 13 419, 14 060, 14 069, 14 070, 14 074, 14 076, 14 486, 14 487, 14 501, 14 513, and 17 602 of the IARC $TP53$ data base [6] for DNA segment lengths $L = 10, 20, \ldots, 160$. We find that the number of cases in which a cancerous mutation corresponds to a segment of low transmission change is within 5%–15% the same for models L1, FB, L2 and LM, with results for L1 and FB very similar to each other. The models L2 and LM are within 15% of each other and have only a slightly smaller occurrence of these cases of low transmission change and high cancerous probability than L1 and FB. Thus in the following, we shall restrict our analysis to the simple case of the strictly 1D model 1L given by (1).

Experimentally, the BER enzymes can locate the damaged sites at a distance of 19 base pairs on the DNA strand by probing the CT of the segment bound by the enzymes [4]. If a mutation changes the CT only slightly, the enzymes might thus not be able to find it and the repair mechanism will not be activated. On the other hand, the fact that mutations $C \rightarrow G$ and $C \rightarrow A$ are not found in cancer cells

does not mean that these mutations do not occur. Rather, the changes in CT induced by them are more significant which could allow an easier detection by CT-probing enzymes—these types of mutations will be repaired and cancer will not develop.

In order to challenge such a scenario, the change of CT for all $20\,303 \times 3 = 60\,909$ types of possible point mutations are examined. The average effect of a mutation $(k, s)$ of a subsequence with length $L$ on the CT of $p53$ is defined as

$$\Gamma(k, s; L) = \frac{1}{L} \sum_{j=k-L+1}^{k} \bar{\Delta}_{j,L}^{k,s} \tag{4}$$

where $j$ also satisfies $1 \leq j \leq 20\,304 - L$ close to boundaries. Figure 4 shows the scatter plots of $\Gamma(k, s; L)$ versus frequency of all cancerous mutations for (a) $L = 20$ and (b) 80. The sharp peaks at small $\Gamma$ agree with the scenario that the most cancerous mutations—namely those with high frequency—change the CT only slightly and thus have smaller $\Gamma$.

Let us now compare the CT change (i) for the set $\mathcal{M}$ of all 60 909 possible point mutations of $p53$ (ii) for the set $\mathcal{M}_c$ of the 1953 cancerous point mutations in the IARC database [6] and (iii) for the set $\mathcal{M}_{c,10}$ of the 366 mutations which are found more than 10 times in the cancer tissues. For given $L$, we sort the CT results for $\mathcal{M}$ according to the computed magnitude of $\Gamma(k, s; L)$ and determine the rank $r(k, s; L) \in [1, 60\,909]$ of the CT change for each mutation $(k, s)$. A smaller rank means less CT change for the mutation. $\gamma(k, s; L) = 100\% \times r(k, s; L)/60\,909$ is then the relative rank in percentage.

The histograms of the distribution of $\gamma(k, s; L)$ are shown in Fig. 5(a) and 5(b) for the mutations $(k, s)$ of $\mathcal{M}_c$ and $\mathcal{M}_{c,10}$. The vertical axis is the percentage of mutations in $\mathcal{M}_c$ (gray bars) and $\mathcal{M}_{c,10}$ (black bars) whose $\gamma(k, s; L)$ belong to the corresponding bin range with each width set to 5%. For $\mathcal{M}$, the result is the dashed line at a value of 5%. The distributions for $\mathcal{M}_c$ and $\mathcal{M}_{c,10}$ are clearly biased to smaller values of $\gamma$, especially for the $L = 80$ case. E.g. there are about 9% for $L = 20$ and 27% of

TABLE I. Renormalized values of the energy-averaged changes $\bar{\Delta}_{14\,575,20}^{14\,585,s}$ in transmission properties for the 4 tight-binding models. All data are shown with at most 3 significant figures. Common multiplication factors for each group of data for given $L$ and mutations with $C \rightarrow A$, G and T are suppressed. Bold entries denote minima for the CT change.

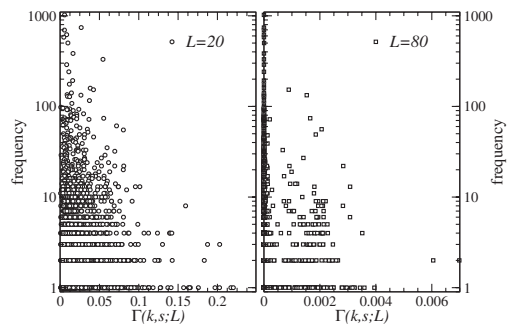| $s$ | $L$ | 1L | FB | 2L | LM |
|---|---|---|---|---|---|
| $C \rightarrow A$ | 20 | 23.1 | 8.46 | 2.24 | **0.43** |
| $C \rightarrow G$ | 20 | 37.6 | **0.73** | 0.83 | 0.57 |
| $C \rightarrow T$ | 20 | **5.63** | 1.08 | **0.34** | 0.66 |
| $C \rightarrow A$ | 30 | 15.7 | 54.8 | 96.2 | 1.76 |
| $C \rightarrow G$ | 30 | 21.4 | 0.55 | 2.75 | 0.40 |
| $C \rightarrow T$ | 30 | **9.14** | **0.0006** | **0.39** | **0.15** |
| $C \rightarrow A$ | 40 | 1.16 | 30.7 | 31.6 | 17.7 |
| $C \rightarrow G$ | 40 | 2.21 | 0.72 | 0.41 | 0.16 |
| $C \rightarrow T$ | 40 | **0.40** | **0.009** | **0.26** | **0.04** |



FIG. 4. Scatter plots of $\Gamma(k, s; w)$ versus occurrence frequency of all cancerous mutations $s$ corresponding to the hot spots $\{k\}$ for (a) $L = 20$ and (b) 80.
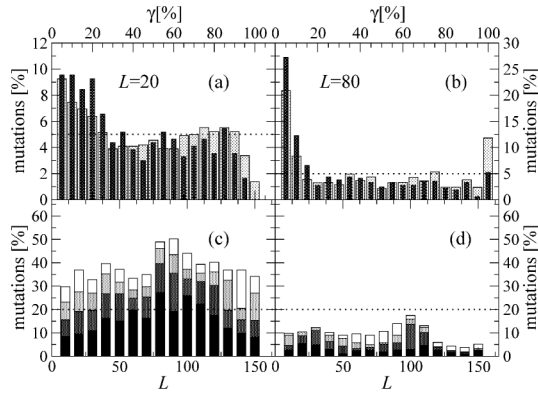
FIG. 5.    Histogram of the distribution of $\gamma(k, s; L)$ in $\mathcal{M}_c$ (light wide bars) and $\mathcal{M}_{c,10}$ (dark thin bars) which changes the $k$th nucleotide to $s$ for (a) $L = 20$ and (b) 80. For $\mathcal{M}$, all values are equal to 5% in the 20 intervals as indicated by the horizontal dashed lines. (c) shows the percentage of $\Gamma(k, s; L)$ values in $\mathcal{M}_{c,10}$ for small CT change as a function of DNA lengths in the range 0%–5% (black), 5%–10% (dark gray), 10%–15% (light gray), and 15%–20% (white). Similarly, (d) indicates large CT change for $\mathcal{M}_{c,10}$ in the ranges 80%–85% (black), 85%–90% (dark gray), 90%–95% (light gray), and 95%–100% (white). The horizontal dashed lines in (c) and (d) indicates the distributions for $\mathcal{M}$.

mutations for $L = 80$ in the $\mathcal{M}_{c,10}$ set whose $\gamma(k, s; L)$ values are smaller than 5%. This indicates that the cancerous mutations in $\mathcal{M}_c$ and $\mathcal{M}_{c,10}$ result in smaller CT changes than noncancerous ones. The distribution bias is more apparent in $\mathcal{M}_{c,10}$ than that in $\mathcal{M}_c$ in agreement with the choice of mutations.

Let us also evaluate the dependence of the CT change on different $L$. Figures 5(c) and 5(d) show the accumulated percentage of mutations in $\mathcal{M}_{c,10}$ whose $\gamma(k, s; L)$ values are smaller than 20% and larger than 80%, respectively. We see that around $L = 90$, more than 50% of mutations in $\mathcal{M}_{c,10}$ change the CT less than 20%, and the number of cancerous mutations with an 80% or more change in CT is much less than average for all $L$.

To summarize, we find that (i) the conductance of hot spots of cancerous mutations is smaller than that of other sites, (ii) on average the cancerous mutations yield smaller changes of the CT in contrast with noncancerous mutations, (iii) the tendency in (ii) is stronger in the set of highly cancerous mutations with occurrence frequency $>10$. These results suggest a possible scenario of how cancerous mutations could circumvent the DNA-damage-repair mechanism and survive to yield carcinogenesis. Our analysis is however only valid in a statistical sense since occasional noncancerous mutations are found to yield weak change of CT. For these, other DNA repair processes should exist so that the DNA-damage repair might not solely uses a CT-based criterion. Still, our results exhibit an intriguing and new correlation between the electronic structure of DNA hot spots and the damage-repair process.

Provided the ideal structure of DNA is preserved by a careful sample preparation [17], mesoscopic transport measurements of DNA sequences of several tens of base pairs can be performed [18]. Therefore, our results could be experimentally challenged by investigating CT in wild and mutated short synthesized sequences of the $p$53 gene.

Further theoretical studies should also investigate how robust our conclusions are with regards to electron-phonon coupling effects, or other phenomena [2]. One finally notes that long-range correlations have been shown to critically influence long-range CT features [10]. The degree of long-range correlation disruption due to point-mutation defects in $p$53 would thus also deserve further exploration.

[1]  R. G. Endres, D. L. Cox, and R. P. Singh, Rev. Mod. Phys. **76**, 195 (2004).
[2]  *Charge Migration in DNA: Perspectives from Physics, Chemistry and Biology*, edited by T. Chakraborty (Springer-Verlag, Berlin, 2007).
[3]  S. R. Rajski, B. A. Jackson, and J. K. Barton, Mutat. Res. **447**, 49 (2000).
[4]  E. Yavin *et al.*, Proc. Natl. Acad. Sci. U.S.A. **102**, 3546 (2005).
[5]  C. J. Sherr, Cell **116**, 235 (2004).
[6]  A. Petitjean *et al.*, Human Mutation **28**, 622 (2007); http://www-p53.iarc.fr/index.html, R11.
[7]  G. Cuniberti, L. Craco, D. Porath, and C. Dekker, Phys. Rev. B **65**, 241314(R) (2002).
[8]  A. A. Voityuk, J. Jortner, M. Boxin, and N. Rösch, J. Chem. Phys. **114**, 5614 (2001); H. Sugiyama and I. Saito, J. Am. Chem. Soc. **118**, 7063 (1996).
[9]  Y. A. Berlin, A. L. Burin, and M. A. Ratner, Chem. Phys. **275**, 61 (2002).
[10]  S. Roche, D. Bicout, E. Macià, and E. Kats, Phys. Rev. Lett. **91**, 228101 (2003).
[11]  C. T. Shih, Phys. Rev. E **74**, 010903(R) (2006).
[12]  E. Díaz, A. Sedrakyan, D. Sedrakyan, and F. Domínguez-Adame, Phys. Rev. B **75**, 014201 (2007).
[13]  We emphasize that the results presented in the following also hold for $t_{n,n+1} = 0.1$ eV and 1 eV.
[14]  B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993).
[15]  P. A. Futreal *et al.*, Nucleic Acids Res. **19**, 6977 (1991).
[16]  Point mutations at a hot spot change a complete base *pair*, e.g., C → T corresponds to C/G → T/A.
[17]  A. Yu. Kasumov *et al.*, Appl. Phys. Lett. **84**, 1007 (2004).
[18]  D. Porath, A. Bezryadin, S. Vries, and C. Dekker, Nature (London) **403**, 635 (2000); B. Xu, P. Zhang, X. Li, and N. Tao, Nano Lett. **4**, 1105 (2004).
[19]  D. K. Klotsa, R. A. Römer, and M. S. Turner, Biophys. J. **89**, 2187 (2005).