Stacking Effects on Charge Transfer Dynamics in Fluctuating DNA

Zhong-wei Li and Ke-li Han*

State Key Laoratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian 116023 (P. R. China)

E-mail: klhan@dicp.ac.cn

Abstract

Base-stacked structure is an important feature of DNA molecules. But dynamics study on the influences of the stacking effects on charge transfer in DNA is yet rare. In this article, a general rule about the relationship of onsite energies of same bases in a stack is derived by Hückel theory. It is found that the base in the middle position of the stack has lower onsite energy than the bases at the terminals due to squeezing effect, which is different from previous studied neighboring base effect. The former is a long-range effect while the latter acts in a short range. Semiempirical MNDO calculations on $(A:T)_n$ (n=1∼10) systems verfied the Hückel analysis. From this perspective, the so-called incoherent hopping mechanism is actually somewhat coherent due to the squeezing effect. To understand these stacking effects on charge transfer in DNA, a cross-scale method which combines classical MD simulations, quantum mechanism calculations, Marcus electron transfer theory and kinetic Monte Carlo simulations is developed and applied on hole dynamics in $(A:T)_n(G:C)$ (n=1∼10) systems. Although no superexchange meachanism is explicitly involved in the studied systems, a crossover

---

*To whom correspondence should be addressed

1

from strong to weak distance-dependency of hole arrival rates, which is an experimentally observed property of hole dynamics in DNA and is thought an evidence of the conversion from superexchange to hopping mechanim, also appears. We attribute it to the stacking effects. Such a result provides a new idea on explaining the crossover of different distance-dependencies of charge transfer rates in DNA. In addition, the squeezing effect may be a new driving force for long-range charge transfer. At the same time, some technical methods developed in the dynamics, e.g. calculations of onsite energies and electronic couplings in a stack, and simulated hole dynamics, etc., can be generalized to other complex molecular systems with charge transfer behaviors.

INTRODUCTION

In cell, the proteins containing [4Fe-4S] clusters were found to be involved in DNA repair, replication, and transcription etc. via DNA-mediated charge transport (DNA CT, also means DNA charge transfer in the following text).[1] It motivates us to find some tools to explore the mysteries of DNA conductivity in the genome. Experimental scientists have built various sophisticated platforms to measure DNA CT and have provided us a lot of valuable information.[1–15] However, considering the reasearch cost and the massive amounts of DNA sequence needed to be analyzed, we hope to build a set of generalized and feasible methods to theoretically characterize the ability of a DNA segment to transport a charge.

To study DNA CT, it is imporant to be firstly clear that the DNA of interest is dry or wet. For dry DNA, the consideration of the effect of structural fluctuations is not required and the driving force of electric current is bias voltage.[1] But for wet DNA the situation becomes more complicated and charge transfer can occur when DNA (or its conjugates) fluctuates in solution even no exteral force exists.[16] Since we are ultimately interested in its biological functions that wet DNA is the focus of the present work.

To characterize the conductivity of a fluctuating DNA segment, it would be convenient if it is able to observe the CT along the segment in real time. Since one of the energy sources of

DNA CT is the motion of nuclei,[16] and the charge transfer would in turn affect the the DNA conformation by the interactions with the vibrational modes of bases as well as the polarity of surrounding environments,[17–19] so a nonadiabatic method is in principle necessary to describe the CT process. However, it has been never an easy task to directly perform electrons-nuclei coupled nonadiabatic dynamics on large molecular systems like DNA oligomers.[20] Considering that the main concern is the motion of the charge and the nonadiabatic effect on nuclei is far less than that on electrons, an approximate treatment is to perform classical molecular dynamics (MD) simulations for nuclei and then use quantum mechanism (QM) for electronic dynamics carried by the nuclear trajectories.[16]

Referring to previous theoretical studies, we note that there are generally two methodologies being able to perform DNA CT dynamics in the fluctuation-assisted framework.[16,18,21,22] One is to construct the time-dependent wave function (or density) as well as the Hamiltonian of the charge under the tight-binding approximation and then directly let it evolve in time.[16] In this way, Grozema et al. elucidated the importance of electrostaic interactions between the charge and the donor site on the superexchange CT during short distance and the role of conformational fluctuations in multistep incoherent hopping CT for longer range.[16] Kubar et al. introduced the electrostaic interactions with the environment so that the solvent effect could be studied in detail.[18] Renaud et al. included the interactions with molecular vibrational modes in the Hamiltonian and obtained CT rates quantitatively consistent with experimental data.[17] On the other hand, a cross-scale method combining the MD, QM, Marcus electron transfer theory and kinetic Monte Carlo (MC) simulations was also employed to study DNA CT.[21,23] In comparison with the former, the latter is less used and needs more exploration.

Either way, it is crucial to obtain a reasonable energy relationship for the bases in a stack, what is not yet an easy task due to the influences of neighboring bases and the stacking effects. Attempts to evaluate the onsite energy for a hole in special base sequences have been reported and mostly focused on the n-stacked system for $n \leq 3$ as the interactions

between non-adjacent bases are thought to be ignorable.[16,24–27] On the other hand, it is noticed that there is strong or weak correlation between the nucleobases along poly-G or poly-A strands for n>3,[3,18,19,28,29] implicating that it may be not enough to merely calculate the parameters from dimers or trimers for the consideration of stacking effects. Considering that the computational cost is bound to rapidly rise as n increases, it is debatable whether it is necessary to put the entire stack into calculation.

In the present work, we firstly analyze the bonding nature of π-stacking system with Hückel theory, which had been used on stacked nucleobases by Jakobsson et al.[30] but we here will have a different discussion. As a result, a curved onsite energy map for the bases of same type (e.g., for hole transfer, the base in the middle of the stack has lower onsite energy than that at the terminals) is derived by Hückel analysis and then supported by semiempirical MNDO calculations on the stacks consisted of continuous A:T base pairs, which usually act as a bridge for hole transfer in many DNA CT experiments. In fact, such a curved energy relationship had been exhibited by different calculations elsewhere[18,19,29] but a general conclusion about it has not yet derived. To distinguish it from other stacking effects discussed previously, e.g. neighboring base effect in trimers[24–26] and polarization effect in dimers,[27,31] we call it "squeezing" effect in the following text. Of course, they are essentially caused by the stacked structures.

To examine these stacking effects on DNA CT, a CT dynamics (detail presented in the next section) is developed based on the above mentioned cross-scale method and then applied on $(A:T)_n$ (n=1∼10) mediated hole transport, which have rich experimental data for reference.[2,6,7,32] Herein the hole is assumed to be initially generated at the first A:T pair and then transported via the $(A:T)_n$ bridge to a G:C acceptor. Surprisingly, the experimentally observed crossover from strong to weak distance-dependence of CT rates as the bridge length increases[6,7,32] can be predicted by the onsite energy map of the $(A:T)_n(G:C)$ systems if the stacking effects are considered, and then simulated hole dynamics indeed qualitatively reproduced it. About the crossover, it is thought to indicate a conversion between the

two mainstream mechanisms,[33,34] i.e. superexchange and incoherent hopping, which work in short distance and long range CT, respectively. Other theoretical explanations on it including the Coulomb interactions between donor and the transported charge,[16] and the effect of distance-dependent reorganizaiton energy.[35] In our model, no electrostatic potential and no distance-dependent reorganization energy is introduced, nor is there any presupposed mechanism must be obeyed. So the consideration of stacking effect provides a new suggestion about the mechanism of DNA CT.

In this work, some technical problems are also touched, e.g. the calculations of electronic coupling and onsite energy in a stacked system, and simulations of time-resolved spectroscopy. These parts can be generalized for the study of charge transfer behavior in complex molecular systems.

METHODOLOGY

A general scheme of the cross-scale method to perform DNA CT dynamics is outlined as follows. (i) Perform classic MD simulations to obtain fluctuating trajectories of the interested DNA segment in solution. (ii) Based on the fixed nuclei configurations along MD trajectories, use QM methods to calculate the electronic structures. (iii) Once the required parameters, e.g. electronic coupling and onsite energy, are obtained, Marcus electron transfer theory and kinetic MC method are employed to perform CT dynamics along the base stack. As the current implementation steps are described in detail below, similarities and differences from previous works are also briefly mentioned.

MD Simulations

For classical MD simulations, BSC1 DNA force field and TIP3P water force field in Amber 16 package[36] are used on the DNA molecule of interest. Sodium cations are added to neutralize negative charge of the DNA. To describe the dynamic disorder of the fluctuating DNA, 2

ns MD (300 K and 1 atm) with time step of 2 fs is performed. It is worth mentioning that the time interval of two outputted adjacent MD points is set at $\Delta t = 0.2$ ps referring to previous studies,[16,19] meaning that in later CT dynamics the electronic behavior within 0.2 ps will share the information with same nuclear configuration. Thus 10000 consecutive MD points are obtained. To reduce the computational cost of subsequent QM calculations, 500 snapshots are randomly picked up as a database for constructing many different kinds of MD trajectories, which denoted as $\{\mathbf{R}(t)\}$. Wherein the sugar-phosphate backbones are removed since CT is assumed occurs only via the stacked nucleobases.[11,37]

CT Parameters

Suppose now we have a constructed MD trajectory, based on each fixed nuclear configuration, the instantaneous rate constant k for a single-step CT between any two sites (e.g. i and j) at time t can be estimated by Marcus-Hush equation,[38–40]

$$k_{ij}(\mathbf{R}(t)) = \frac{J_{ij}(\mathbf{R}(t))^2}{\hbar} \left(\frac{\pi}{\lambda_{ij} k_{\mathrm{B}} T} \right)^{1/2} \exp \left(-\frac{(\Delta G_{ij}(\mathbf{R}(t)) + \lambda_{ij})^2}{4\lambda_{ij} k_{\mathrm{B}} T} \right),$$ (1)

wherein k_{B} is the Boltzmann constant, ΔG the Gibbs free energy and J the electronic coupling (also called charge transfer integral). Note that $k(\mathbf{R}(t))$, $J(\mathbf{R}(t))$ and $\Delta G(\mathbf{R}(t))$ are variables as nuclear configuration evolves with time while the reorganization energy λ is pre-prepared parameter in our model. Refer to previous theoretical study,[16,23,26,31,41,42] their definitions and calculations are as follows.

According to the tight-binding approximation, suppose that the wave function of the charge can be expressed as a linear combination of the frontier molecular orbitals on the individual base pairs[16,26,43,44] and the charge is always localized on a base pair[45,46] during the propagation, the electronic coupling can be evaluated by[47]

$$J_{ij} = \langle \psi_i^{\mathrm{HOMO(LUMO)}} | \hat{F} | \psi_j^{\mathrm{HOMO(LUMO)}} \rangle,$$ (2)

wherein $\hat{F}$ is the Fock operator of the charge in the whole molecular system and ψ_i the MO localized on i-th base pair. HOMO (the highest occupied molecular orbital) is used for hole transfer while LUMO (the lowest unoccupied molecular orbital) for excess electron transfer. Charge localizaiton on a single base in DNA has been an accepted model for many years.[27,28] On the other hand, Renaud et al. introduced the concept of partial localization (i.e., the charge is partial localized on a single base while others delocalized over several base pairs) to improve theoretical CT rates in comparison with experiments.[17] Here we use traditional localized model with a trick that one site is set as a base pair rather a single base.

About Gibbs free energy ΔG, Deng et al.[42] and Bag et al.[23] evaluated it by $\Delta G_{ij} = [E_i + E_{j+(-)}] - [E_{i+(-)} + E_j]$, while E is the electronic state energy based on corresponding optimized geometry. But what they studied were dry DNA so the fluctuating effects and the stacking effects were ignored. To consider these effects in wet DNA, ΔG is rewritten as[16]

$$\Delta G_{ij} = \epsilon_j - \epsilon_i, \tag{3}$$

wherein ϵ_i is the fluctuating onsite energy when a charge is localized on site i. Similar to Eq. (2), ϵ can be evaluated by

$$\epsilon_i^{\mathrm{HT}} = -\langle \psi_i^{\mathrm{HOMO}} | \hat{F} | \psi_i^{\mathrm{HOMO}} \rangle \tag{4}$$

for hole transfer (HT) and

$$\epsilon_i^{\mathrm{EET}} = \langle \psi_i^{\mathrm{LUMO}} | \hat{F} | \psi_i^{\mathrm{LUMO}} \rangle \tag{5}$$

for excess electron transfer (EET).[25,26,41]

For practical calculations of ϵ and J, Senthilkumar et al.[16,25,26,41] employed the charge transfer integrals module included in the Amsterdam Density Functional (ADF) theory program.[48] We here refer to the projective method proposed by Kirkpatrick.[49] First, for a n-stacking system, construct the coefficient matrix of C_{loc} which is consisted of the eigenvectors

of the localized MOs,

$$
C_{\text{loc}} =
\begin{bmatrix}
c^{\chi_1}_{m_1,\psi_1} & c^{\chi_2}_{m_1,\psi_1} & \cdots & & & & & & & \\
c^{\chi_1}_{m_1,\psi_2} & \cdots & \cdots & & & & & & & \\
\cdots & \cdots & \cdots & & & & & & & \\
& & & c^{\chi_1}_{m_2,\psi_1} & c^{\chi_2}_{m_2,\psi_1} & \cdots & & & & \\
& & & c^{\chi_1}_{m_2,\psi_2} & \cdots & \cdots & & & & \\
& & & \cdots & \cdots & \cdots & & & & \\
& & & & & & \cdots & & & \\
& & & & & & & c^{\chi_1}_{m_n,\psi_1} & c^{\chi_2}_{m_n,\psi_1} & \cdots \\
& & & & & & & c^{\chi_1}_{m_n,\psi_2} & \cdots & \cdots \\
& & & & & & & \cdots & \cdots & \cdots
\end{bmatrix},
\tag{6}
$$

wherein $c^{\chi_k}_{m_i,\psi_j}$ means the MO coefficient of the j-th MO on the k-th atomic orbital (AO) basis of the site i molecule. The blank is filled with zeros. Then, the MO coefficients of the whole stack in the basis of the localized MOs can be projected onto C_{loc} by

$$
C^{\text{loc}}_{\text{sys}} = C_{\text{loc}} S^{\text{AO}}_{\text{sys}} C^{\text{T}}_{\text{sys}},
\tag{7}
$$

wherein C_{sys} are the MO coefficients of the whole system in the same AO basis set as C_{loc} use. And $S^{\text{AO}}_{\text{sys}}$ is the overlap matrix. Note that Eq. (7) is different from the equation used by Kirkpatrick et al. in which orthogonal AO basis is assumed.[21,49] Finally, the Fock matrix of the stack in the basis of localized MOs can be formed by

$$
F^{\text{loc}}_{\text{sys}} = C^{\text{loc}}_{\text{sys}} \epsilon_{\text{sys}} C^{\text{loc}\,\text{T}}_{\text{sys}},
\tag{8}
$$

wherein ϵ_{sys} is the diagonal matrix of MO eigenvalues of the stack. J and ϵ defined in Eqs. (2), (4) and (5) can be directly picked up from $F^{\text{loc}}_{\text{sys}}$.

Unlike ϵ and J, reorganizatin energy λ is time-independent in our model. As usual,[16,23]

8

it is divided into two parts. For a single-step CT from site i to site j, internal reorganization energy deriving from the geometric relaxation of the base itself is evaluated by[40,50]

$$\lambda_{ij}^{\text{int}} = [E_i^* - E_i] + [E_{j+(-)}^* - E_{j+(-)}],\tag{9}$$

wherein E_i and $E_{j+(-)}$ is the electronic state energy based on corresponding optimized geometry while E_i^* and $E_{j+(-)}^*$ are the energies of the neutral and charged electronic states with the geometries of the charged and neutral species respectively. But the evaluation of outer reorganization energy coming from the environment is not an easy task. Referring to previous studies,[35,51] we artificially set it as $\lambda_{ij}^{\text{out}} = 0.2$ eV throughout this work.

The above QM calculation step is the most time-consuming in the overall approach. In order to achieve a balance between the computational cost and accuracy, we refer to previous studies in which DFTB,[18,19,52] ZINDO,[21] NDDO-G,[29] and SAOP[16] etc. were used, and selectively tested different methods covering semiempirical, *ab initio* and DFT with different basis. As expected, in terms of computational cost, it is difficult to employ the QM methods higher than HF/STO-3G (or the same level) to fluctuation-assisted long-distance DNA CT dynamics at ns time-scale. Grozema et al. used high level SAOP/TZP method to evaluate ϵ and J for hole dynamics on $(\text{A:T})_n$ stacks for $n \leq 6$ but the practical calculations were limited to the dimer of adenines.[16] They built a Gaussian distribution for the fluctuating ϵ from the calculations of two adenines and then used it to generate sample values for all adenines in DNA hairpins containing adenines more than two. Thus the squeezing effect was just neglected. Here we choose semiempirical MNDO method which allows us to track the evolution of the electronic structure of long-chain base-stacks up to ns time-scale without losing reasonability in predicting the parameters in Eq. (1) (see Figure S1 and Figure S2 in the supporting information (SI)). Wherein, $\lambda_{ij}^{\text{int}}$ is time-independent so it is evaluated by higher level method than MNDO. Concretely, we use B3LYP/6-31G predicted values (for HT) which are $\lambda_{\text{AA}}^{\text{int}} = 0.384$ eV, $\lambda_{\text{AG}}^{\text{int}} = 0.502$ eV, and $\lambda_{\text{GA}}^{\text{int}} = 0.457$ eV, respectively.

Electronic structure calculations are performed in GAMESS package.[53]

CT Dynamics

Hitherto, the parameters needed in Eq. (1) are prepared. Since $k(\mathbf{R}(t))$ obtained in this way is instantaneous so it would not be used directly to characterize the CT property of the DNA (see Figure 1). Instead, kinetic MC method is used to simulate the CT pathways based on obtained $k(\mathbf{R}(t))$.[23] Firstly, residence time of the charge at time-site (t,i) is estimated by

$$\tau(t,i) = -k_i(t)^{-1} \ln(r_1), \tag{10}$$

where

$$k_i(t) = \sum_{j \neq i}^{n} k_{ij}(\mathbf{R}(t)). \tag{11}$$

r_1 is a uniform random number between 0 and 1. n is the stacking number what means that all possible charge transfer reactions started from site i on the entire studied DNA segment are taken into account, including single-step hopping and superexchange. This is different from previously studied DNA CT dynamics[16,21,23] in which the electronic coupling interactions are considered only for neighboring nucleobases.

Next, recall that the time interval between two neighboring MD points was chosen as $\Delta t = 0.2$ ps, the following assumption is used (see Figure 2 for more intuitive description):

$$\begin{cases} \text{If } \tau(t,i) > \Delta t, \text{ charge transfer does not occur at } \mathbf{R}(t), \ t = t + \Delta t; \\ \text{If } \tau(t,i) \leq \Delta t, \text{ charge transfer occurs at } \mathbf{R}(t), \ t = t + \tau(t,i), \ i \to l. \end{cases} \tag{12}$$

And the accept site l is selected by the relationship satisfying

$$\sum_{j=0}^{l-1} k_{ij}(\mathbf{R}(t)) < r_2 k_i(t) \leq \sum_{j=0}^{l} k_{ij}(\mathbf{R}(t)) \tag{13}$$

with the definition of $k_{i0} = 0$. r_2 is another uniform random number between 0 and 1. Two

sets of $\tau(t, i)$ must be different even though they are obtained from the same k_{ij} data, but should be on an order of magnitude. Repeated use of the formulas from (10) to (13) gives a possible CT pathway along the fluctuating base stack (Figure 3). Similarly, it is not uniquely determined. Therefore, to characterize the CT properties of the DNA, statistical analysis of a large number of these pathways is required, as will be seen later. Incidentally, once a set of $\{k(\mathbf{R})\}$ is obtained, CT dynamics with different initial conditions can be started from Eq. (10) instead of Eq. (1) by selecting initial $k(\mathbf{R}(t_0))$ randomly.

The current model clearly shows the fluctuation-assisted mechanism of DNA CT. Since no external force is introduced, it is easily deduced that the charge will stay at a certain site if the fluctuation stops, as experimentally proposed.[4] So far, since no assumption that the charge is positive or negative is made, it is generalized for HT and EET. As far as the following discussions concerned, it means HT if there is no special explanation.

Results and Discussions

Hückel Analysis on π-Stacked Systems

Since that all the HOMOs as well as the LUMOs of the four nucleobases (G, A, T, C) are π orbitals, it is easy to build a model of the π channel for a n-stacked nucleobases of same type by Hückel theory with the tight-binding approximation. Hückel bases, bond matrix, solutions and MO shapes are attached in the SI. As a result, the variation of energy levels of the Hückel MOs for n-stack (n=2~6) is plotted as in Figure 4. From the MO phases (see Figure S4) it can be known that what we should pay attention to is just ϕ_1. Because that of the n MOs only ϕ_1 actually reflects the nature of the stacking structure what is a linear combination of individual π orbitals via the weak interaction, while others are actually covalent bonds what would be hard to form in a stack. This is different from the discussion of Jakobsson et al. in which all the Hückel MOs were used for CT dynamics simulations.[30]

It can also be seen from Figure 4 that as n increases, the energy level of ϕ_1 becomes

11

higher and higher due to a stronger antibonding interaction. This is qualitatively consistent with the fact that GG doublets and GGG triplets have lower ionization potential than the isolated G bases.[27,54] Together with the shape variation of ϕ_1 (as shown in Figure 5) a more general conclusion can be inferred. Namely, as the basis in the middle position of the stack contributes the most to the energy increase, it is in turn implicating that the energy level of the localized π orbital on the middle site should raise higher than that at the ends. For HT along the nucleobases of same type, it means that the base in the middle of the stack should have lower onsite energy than the bases at the terminals, what can be seen from MNDO calculations in the next section.

Squeezing Effect

The averaged values of the fluctuating onsite energies for a hole in the $(A{:}T)_n$ systems predicted by MNDO are shown in Figure 6. For the sake of clarity, relative values of three stacks i.e. $(A{:}T)_3$, $(A{:}T)_6$, and $(A{:}T)_{10}$ are exhibited. The curved pattern of onsite energies appears clearly in all the examples. And the "well" becomes deeper as the stack grows. For $n{=}3$, the energy gap between terminals and the middle is around 0.1 eV. As n increases to 10, the gap enlarges to 0.3 eV. At the current computational level, it is even comparable to the ionization potential (IP) gaps between the four nucleobases (MNDO predicted relative $\langle \epsilon_i^{\mathrm{HT}} \rangle$ for the four isolated bases are respectively A 0.38 eV, C 0.76 eV and T 1.16 eV with respect to G, which are reasonable in comparison with the values predicted by SAOP/TZ2P method[25] and experimental results[55]). Besides, since reorganization energies are set to invariable in our model, the large fluctuations on k_{ij} are mainly attributed to the fluctuating onsite energies more than to electronic couplings (see Figure 1 and Figure S5). A bit change in onsite energy may make the instantaneous CT rates to fluctuate over several orders of magnitude and it is the motivation that prompts us to examine the squeezing effect.

In previous studies concerning the fluctuation-assisted mechanism of DNA CT, the driving force was found to act via the following ways: (i) Strong electronic couplings induced by

interbase motions.[56] (ii) Fluctuating onsite energy gaps generated by the polarization effect between neighboring bases.[16,29] (iii) Electrostatic potentials from the environment.[18,19,57] By considering one or more factors of them, the proposed models are ion-gated CT,[57] solvent fluctuations drived CT,[19] phonon-assisted polaron-like hopping CT,[28,58] and CT-active conformations gated CT,[3] et al. In the last two models, both have a well-like domain consists of approximately 3~5 base pairs. A polaron means a structural distortion induced by the charge delocalization over which an energy minimum is created, and the CT-active conformation means a domain generated by a transiently extended π-orbital over which charge can delocalize. The difference between them is that a charge is required in the former but not in another. So the wells in Figure 6, which is derived theoretically, is somewhat similar to the experimentally proposed concept of CT-active domain as both are intrinsic properties of π-stack. As the strand extends, it is difficult to maintain the stacked structure as a whole due to the flexibility. More likely, a well formed in a short range would provide a wave-like driving force to transmit the charge to a distance as the strand fluctuates (see Figure S6).

In A-mediated HT experiments, the hopping rate of a hole between two adjacent A bases is a concern.[2,6,7,59] However, Figure 6 implies that A bases of different positions in a stack may have different abilities on transmitting a hole. From this perspective, the so-called incoherent hopping mechanism is actually somewhat coherent due to the squeezing effect.

Simulated Hole Dynamics

The squeezing effect is one of the consequences of the stacked structure. Besides it, the influences from neighboring bases of different types also play an important role in DNA CT. These effects can be visible from the onsite energy map of $(A{:}T)_n(G{:}C)$ (n=1~10) systems as Figure 7 shows. Due to the existence of G:C pair, the onsite energies of A:T pairs decrease more or less as closing to it. This leads to a waterfall-like pattern of onsite energies for n=1~4. As n increases more, a flat energy relationship appears among the A:T pairs which are at a distance from the G:C terminal. At the same time, the energy gap at the distal

end generated by squeezing effect is retained. Such an energy pattern implies a variable distance-dependency of CT rates. To verify the causality between them, CT dynamics is performed.

Since what the dynamics experiments of DNA CT reflected are statistical results of the signals responded by a large number of single-molecule DNA (or its hairpin) in solutions, statistical analysis of many possible pathways (Figure 3) of different initial conditions provides a straight way to mimic the experimental approaches. In this way, an example of simulated hole dynamics on $(A:T)_6(G:C)$ system is shown in Figure 8. It clearly displays a consecutive hopping process that a hole transfers from the first A:T pair to the G:C acceptor via a $(A:T)_5$ bridge. The rate of hole arrival can be obtained by single exponential fits to the data of the hole population on the acceptor, e.g. $P(t)=\Phi(1 - e^{-k_a t})$, analoging to experimental treatment. But it is not recommended to compare it as the absolute rate to experimental results with the following reasons: (1) The hole injection step, which is generally included in the experiments, is not considered here. (2) The experiments are carried out in dissipative systems while what we use is a closed model, i.e., the hole will not escape into the environment. Ignoring these factors would have a simulated arrival rate faster than experimental values by several orders of magnitude.[16,17]

Instead, to understand the influence of stacking effects on the distance-dependency of CT rates, the calculated arrival rates of $(A:T)_n(G:C)$ ($n=1\sim10$) are collected and plotted in Figure 9. Wherein β is calculated by the fits of $k_a \propto 10^{-\beta N}$. In short distance, experiments obtained stronger dependency ($\beta 1^{\text{Exp}}=1.02$) than our simulations ($\beta 1^{\text{Theo}}=0.49$). This is due to the super-exchange CT in experiments, what is not explicitly involved in our $(A:T)_n(G:C)$ systems. However, from the simulated results, it can be seen that the stacking effects also gives a crossover from strong ($\beta 1^{\text{Theo}}=0.49$) to weak ($\beta 2^{\text{Theo}}=0.17$) distance-dependency. If no consideration of it, it is expected that CT rates would vary obeying weak dependency at all distance. In the region of weak dependency, the resulted $\beta 2$ values from simulations and experiments are consistent very well. It corresponds to the flat areas in Figure 7. Studies on

14

long-range DNA CT should focus on these areas.

CONCLUSIONS

Base-stacked structure is a feature of DNA molecules. Previous studies on the stacking effects usually focus on the effect from the neighboring bases. In this work, Hückel analysis on π-stacked structures predicts a curved energy relationship of n-mer ($n \geq 3$) stacked bases of same type, i.e. the base in the middle position of a stack should have lower hole onsite energy than the bases at the terminals. Semiempirical MNDO calculations on $(A:T)_n$ (n=1∼10) systems verified it. To the best of our knowledge, such an energy relationship is derived for the first time, and called "squeezing" effect here. The difference of it from the neighboring base effect is that the former is a long-range effect ($n \geq 3$) while the latter acts in a short range ($n \leq 3$). The onsite energy gap created by these effects reaches to 0.3∼0.4 eV, which is comparable to the IP gap between adenine and guanine, implying an unignorable role in DNA CT.

To understand the influences of the stacking effects on DNA CT, a CT dynamics based on the cross-scale method, which combines classical MD simulations, quantum chemistry calculations, Marcus electron transfer theory and kinetic Monte Carlo simulations, is developed. As its application on $(A:T)_n(G:C)$ (n=1∼10) systems, a crossover of different distance-dependencies of CT rates is resulted. In short distance ($n \leq 4$), the dependency is stronger than the experimentally observed hopping mechanism but weaker than super-exchange mechansim. We classify it as hopping mechanism still and the stonger dependency than normal hopping regions is attributed to the stacking effects. In longer distance ($n > 4$), simulated and experimentally measured dependencies are in a good agreement. It corresponds to the normal regions of hopping mechanism, which dominates the long-range DNA CT.

The current model is not yet accurate in predicting the absolute rates. To improve it,

the interactions with the environments should be considered. On the other hand, some technical methods developed in this work, e.g. the calculations of onsite energies and electronic couplings in a stack, and simulated hole dyanmics etc., can be generalized to other complex molecular systems in which the charge transfer behavior is concerned.

Acknowledgement

Supporting Information Available

This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Grodick, M. A.; Muren, N. B.; Barton, J. K. *Biochemistry* **2015**, *54*, 962–973.

(2) Takada, T.; Kawai, K.; Cai, X.; Sugimoto, A.; Fujitsuka, M.; Majima, T. *J. Am. Chem. Soc.* **2004**, *126*, 1125–1129.

(3) O'Neill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 11471–11483.

(4) ONeill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2004**, *126*, 13234–13235.

(5) Peter, K.; Elke, M.-E.; Anton, T.; Torsten, F.; Hans-Achim, W. *Angew. Chem. Int. Ed. 44*, 1636–1639.

(6) D., L. F.; Huihe, Z.; Pierre, D.; Boiko, C.; R., W. M. *Angewandte Chemie International Edition 45*, 7982–7985.

(7) Lewis, F. D.; Zhu, H.; Daublain, P.; Fiebig, T.; Raytchev, M.; Wang, Q.; Shafirovich, V. *J. Am. Chem. Soc.* **2006**, *128*, 791–800.

(8) Valis, L.; Wang, Q.; Raytchev, M.; Buchvarov, I.; Wagenknecht, H.-A.; Fiebig, T. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10192–10195.

(9) Lewis, F.; Daublain, P.; Cohen, B.; VuraWeis, J.; Wasielewski, M. *Angew. Chem. Int. Ed.* **2008**, *47*, 3798–3800.

(10) Daublain, P.; Thazhathveetil, A. K.; Wang, Q.; Trifonov, A.; Fiebig, T.; Lewis, F. D. *J. Am. Chem. Soc.* **2009**, *131*, 16790–16797.

(11) Genereux, J. C.; Barton, J. K. *Chem. Rev.* **2010**, *110*, 1642–1662.

(12) Conron, S. M. M.; Thazhathveetil, A. K.; Wasielewski, M. R.; Burin, A. L.; Lewis, F. D. *J. Am. Chem. Soc.* **2010**, *132*, 14388–14390.

(13) Park, M. J.; Fujitsuka, M.; Kawai, K.; Majima, T. *J. Am. Chem. Soc.* **2011**, *133*, 15320–15323.

(14) Genereux, J. C.; Wuerth, S. M.; Barton, J. K. *J. Am. Chem. Soc.* **2011**, *133*, 3863–3868.

(15) Renaud, N.; Harris, M. A.; Singh, A. P. N.; Berlin, Y. A.; Ratner, M. A.; Wasielewski, M. R.; Lewis, F. D.; Grozema, F. C. *Nat. Chem.* **2016**, *8*, 1015–1021.

(16) Grozema, F. C.; Tonzani, S.; Berlin, Y. A.; Schatz, G. C.; Siebbeles, L. D. A.; Ratner, M. A. *J. Am. Chem. Soc.* **2008**, *130*, 5157–5166.

(17) Renaud, N.; Berlin, Y. A.; Lewis, F. D.; Ratner, M. A. *J. Am. Chem. Soc.* **2013**, *135*, 3953–3963.

(18) Kubar, T.; Elstner, M. *The Journal of Physical Chemistry B* **2008**, *112*, 8788–8798.

(19) Kubar, T.; Kleinekathofer, U.; Elstner, M. *J. Phys. Chem. B* **2009**, *113*, 13107–13117.

(20) Yonehara, T.; Hanasaki, K.; Takatsuka, K. *Chem. Rev.* **2012**, *112*, 499–542.

(21) Reha, D.; Barford, W.; Harris, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5436–5444.

(22) Gutiérrez, R.; Caetano, R. A.; Woiczikowski, B. P.; Kubar, T.; Elstner, M.; Cuniberti, G. *Phys. Rev. Lett.* **2009**, *102*, 208102.

(23) Bag, S.; Mogurampelly, S.; Goddard III, W. A.; Maiti, P. K. *Nanoscale* **2016**, *8*, 16044–16052.

(24) Voityuk, A. A.; Jortner, J.; Bixon, M.; Rsch, N. *Chem. Phys. Lett.* **2000**, *324*, 430–434.

(25) Senthilkumar, K.; Grozema, F. C.; Guerra, C. F.; Bickelhaupt, F. M.; Siebbeles, L. D. A. *J. Am. Chem. Soc.* **2003**, *125*, 13658–13659.

(26) Senthilkumar, K.; Grozema, F. C.; Guerra, C. F.; Bickelhaupt, F. M.; Lewis, F. D.; Berlin, Y. A.; Ratner, M. A.; Siebbeles, L. D. A. *J. Am. Chem. Soc.* **2005**, *127*, 14894–14903.

(27) Sugiyama, H.; Saito, I. *J. Am. Chem. Soc.* **1996**, *118*, 7063–7068.

(28) Henderson, P. T.; Jones, D.; Hampikian, G.; Kan, Y.; Schuster, G. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 8353–8358.

(29) Voityuk, A. A.; Siriwong, K.; Rsch, N. *Angew. Chem. Int. Ed. 43*, 624–627.

(30) Jakobsson, M.; Stafstrm, S. *J. Chem. Phys.* **2008**, *129*, 125102.

(31) Valeev, E. F.; Coropceanu, V.; da Silva Filho, D. A.; Salman, S.; Brdas, J.-L. *J. Am. Chem. Soc.* **2006**, *128*, 9882–9886.

(32) Giese, B.; Amaudrut, J.; Khler, A.; Spormann, M.; Wessely, S. *Nature* **2001**, *412*, 318–320.

(33) Berlin, Y. A.; Burin, A. L.; Ratner, M. A. *Chem. Phys.* **2002**, *275*, 61–74.

(34) Bixon, M.; Jortner, J. *Chem. Phys.* **2002**, *281*, 393–408.

(35) LeBard, D. N.; Lilichenko, M.; Matyushov, D. V.; Berlin, Y. A.; Ratner, M. A. *J. Phys. Chem. B* **2003**, *107*, 14509–14520.

(36) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1996**, *118*, 2309–2309.

(37) Gorodetsky, A. A.; Green, O.; Yavin, E.; Barton, J. K. *Bioconjugate Chem.* **2007**, *18*, 1434–1441.

(38) Marcus, R. A. *Rev. Mod. Phys.* **1993**, *65*, 599–610.

(39) Barbara, P. F.; Meyer, T. J.; Ratner, M. A. *J. Phys. Chem.* **1996**, *100*, 13148–13168.

(40) Brdas, J.-L.; Beljonne, D.; Coropceanu, V.; Cornil, J. *Chem. Rev.* **2004**, *104*, 4971–5004.

(41) Senthilkumar, K.; Grozema, F. C.; Bickelhaupt, F. M.; Siebbeles, L. D. A. *J. Chem. Phys.* **2003**, *119*, 9809–9817.

(42) Deng, W.-Q.; Sun, L.; Huang, J.-D.; Chai, S.; Wen, S.-H.; Han, K.-L. *Nat. Protocols* **2015**, *10*, 632–642.

(43) Siebbeles, L. D.; Berlin, Y. A. *Chem. Phys.* **1998**, *238*, 97–107.

(44) Grozema, F. C.; Berlin, Y. A.; Siebbeles, L. D. A. *Int. J. Quantum Chem.* **1999**, *75*, 1009–1016.

(45) Hatcher, E.; Balaeff, A.; Keinan, S.; Venkatramani, R.; Beratan, D. N. *J. Am. Chem. Soc.* **2008**, *130*, 11752–11761.

(46) Gu, J.; Leszczynski, J.; Schaefer, H. F. *Chem. Rev.* **2012**, *112*, 5603–5640.

(47) Newton, M. D. *Chem. Rev.* **1991**, *91*, 767–792.

(48) te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931–967.

(49) Kirkpatrick, J. *Int. J. Quantum Chem.* *108*, 51–56.

(50) Olofsson, J.; Larsson, S. *J. Phys. Chem. B* **2001**, *105*, 10398–10406.

(51) Siriwong, K.; Voityuk, A. A.; Newton, M. D.; Rsch, N. *J. Phys. Chem. B* **2003**, *107*, 2595–2601.

(52) Gutirrez, R.; Caetano, R.; Woiczikowski, P. B.; Kubar, T.; Elstner, M.; Cuniberti, G. *New Journal of Physics* **2010**, *12*, 023022.

(53) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(54) Saito, I.; Nakamura, T.; Nakatani, K.; Yoshioka, Y.; Yamaguchi, K.; Sugiyama, H. *J. Am. Chem. Soc.* **1998**, *120*, 12686–12687.

(55) Hush, N.; Cheung, A. S. *Chem. Phys. Lett.* **1975**, *34*, 11–13.

(56) Voityuk, A. A. *J. Chem. Phys.* **2008**, *128*, 045104.

(57) Barnett, R. N.; Cleveland, C. L.; Joy, A.; Landman, U.; Schuster, G. B. *Science* **2001**, *294*, 567–571.

(58) Ly, D.; Sanii, L.; Schuster, G. B. *J. Am. Chem. Soc.* **1999**, *121*, 9400–9410.

(59) Kawai, K.; Takada, T.; Tojo, S.; Majima, T. *J. Am. Chem. Soc.* **2003**, *125*, 6842–6843.
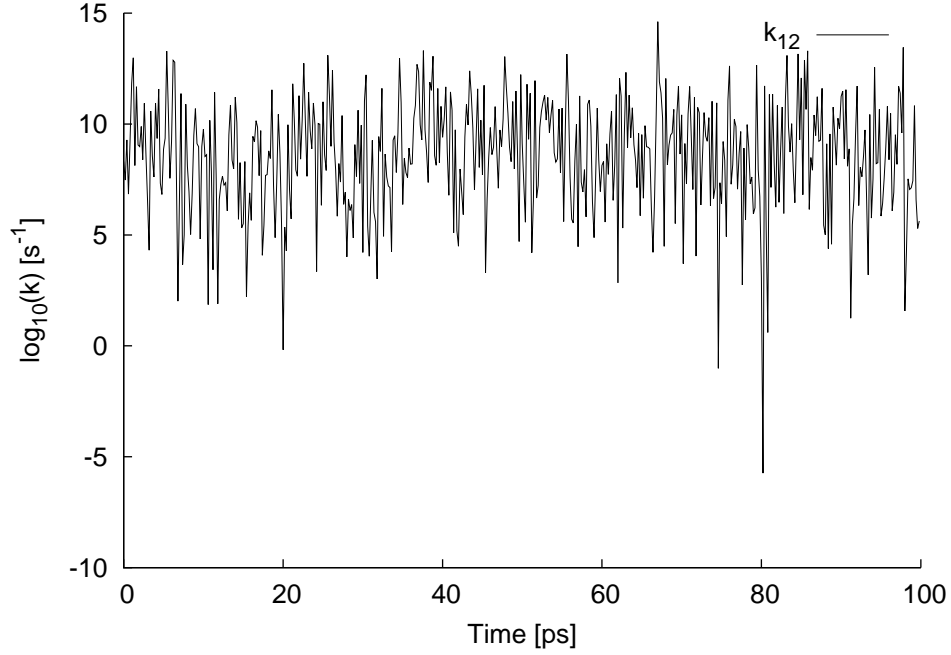
Figure 1: Fluctuating instantaneous rate constants $k(\mathbf{R}(t))$ defined in Eq. (1). The example is k_{12} for hole transfer in $(A{:}T)_2$.
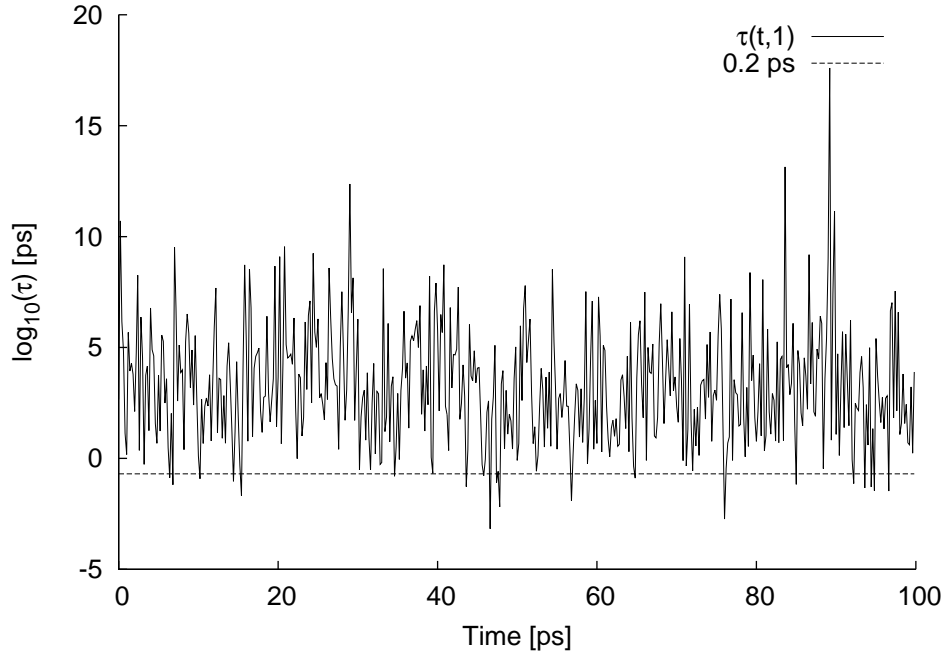
Figure 2: Fluctuating residence time τ defined in Eq. (10). The example is $\tau(t,1)$ for hole transfer in $(A{:}T)_2$. The occurrence threshold of CT, which is set to 0.2 ps is also plotted for reference. According to the assumption in Eq. (12), CT only occurs at the time points below the threshold line.
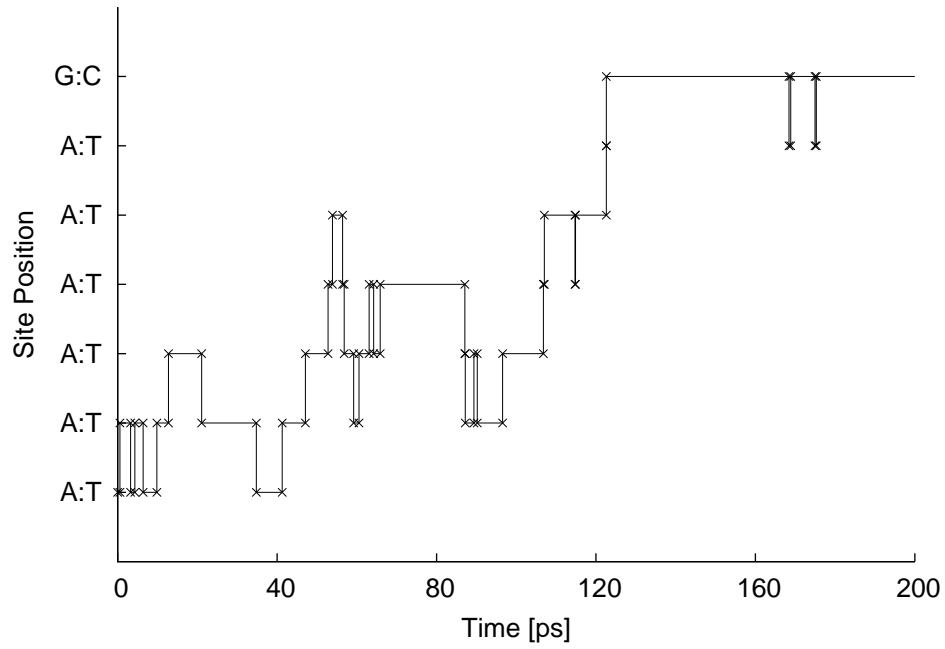
Figure 3: A possible CT pathway given by repeated use of the formulas from (10) to (13). The example is hole transfer in $(A:T)_6(G:C)$ system.
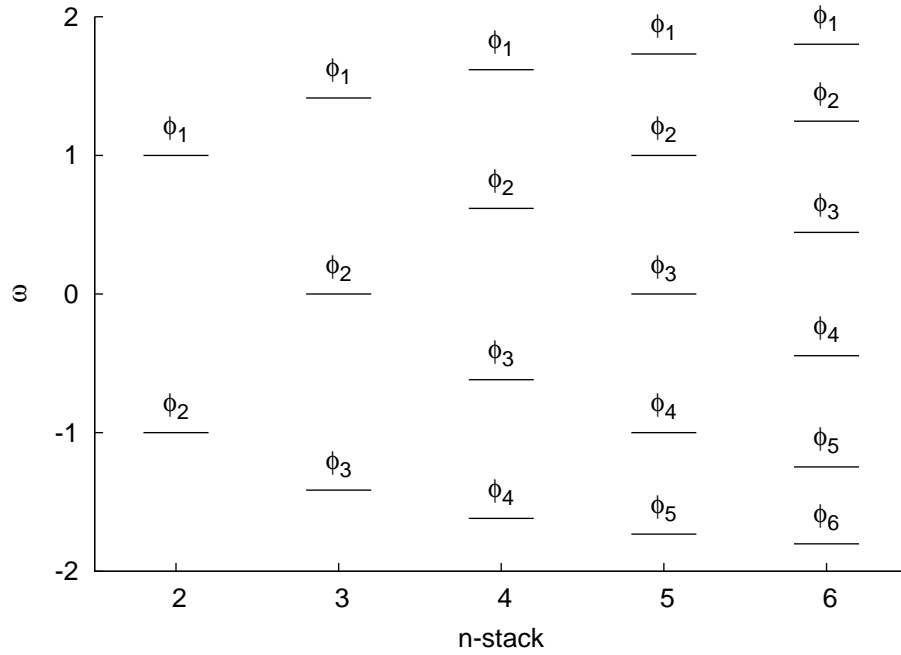
Figure 4: MO energy levels (indicated by ω) of n-mer π-stacked systems predicted by Hückel theory. MO shapes see Figure S4.
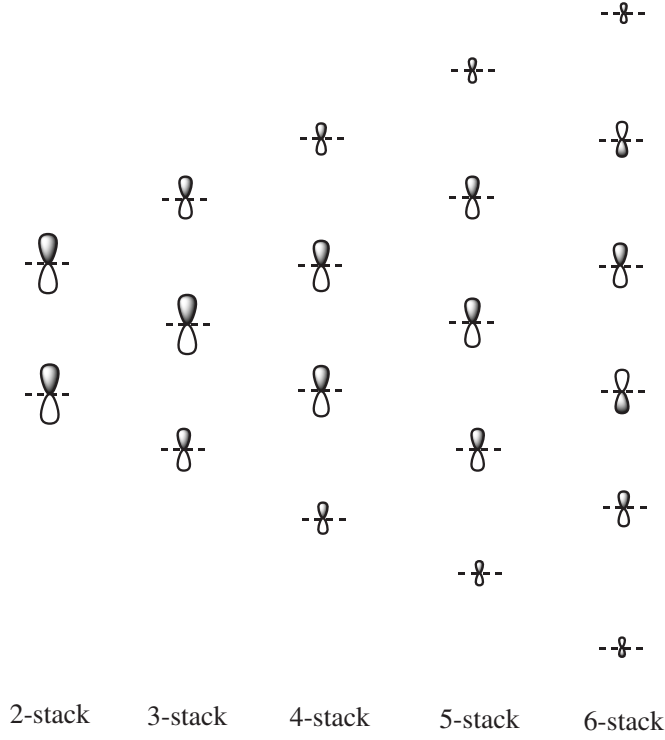
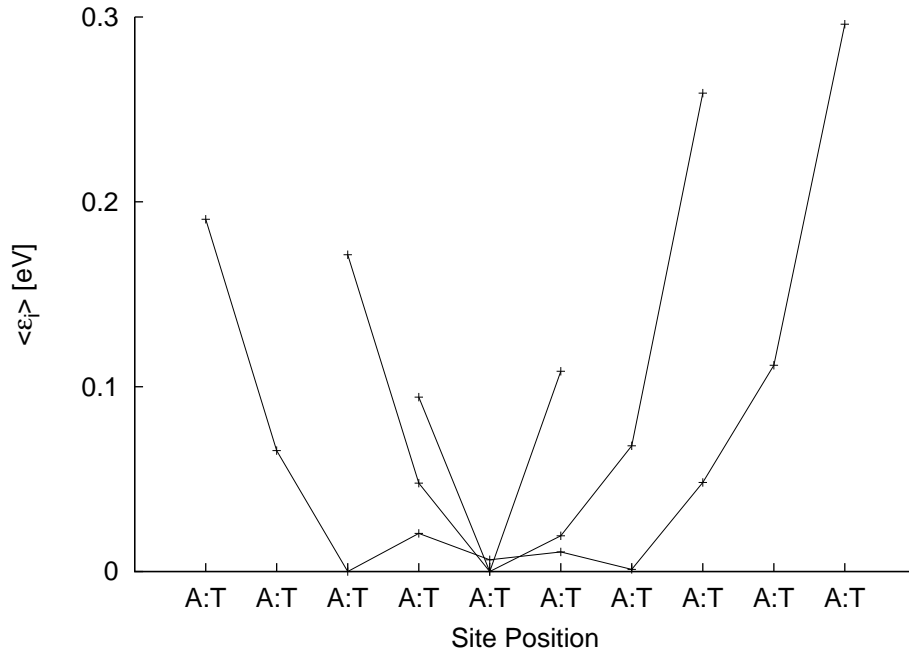Figure 5: Diagrams of the highest orbitals (ϕ_1) in a n-mer π-stacked systems predicted by Hückel theory.

Figure 6: MNDO predicted relative onsite energies for a hole in the individual sites of n-stacked systems. The examples are from fluctuating $(A:T)_3$, $(A:T)_6$ and $(A:T)_{10}$. The values are averaged on 500 samples randomly picked up from 2 ns MD simulations.
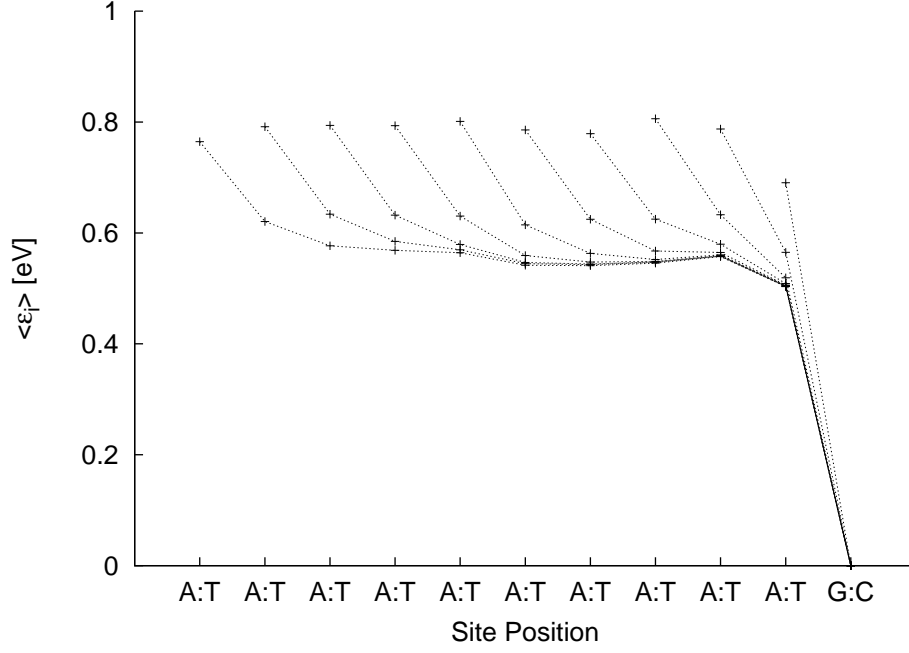
Figure 7: MNDO predicted relative onsite energies for a hole in the individual sites of fluctuating $(A:T)_n(G:C)$ ($n=1\sim10$) systems. The values are averaged on 500 samples randomly picked up from 2 ns MD simulations.
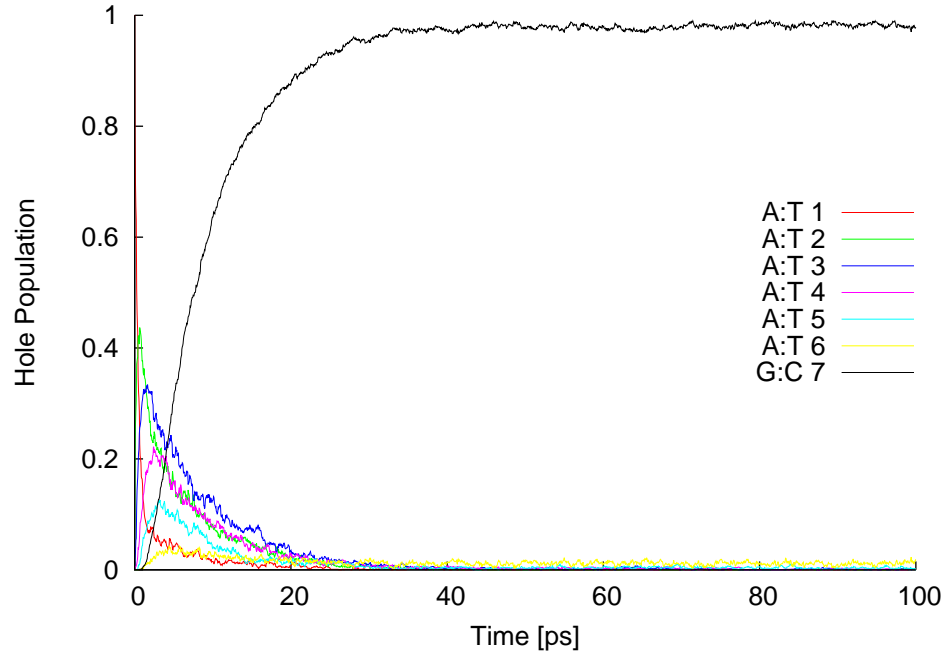

Figure 8: Simulated hole dynamics on DNA. The example is on $(A:T)_6(G:C)$.
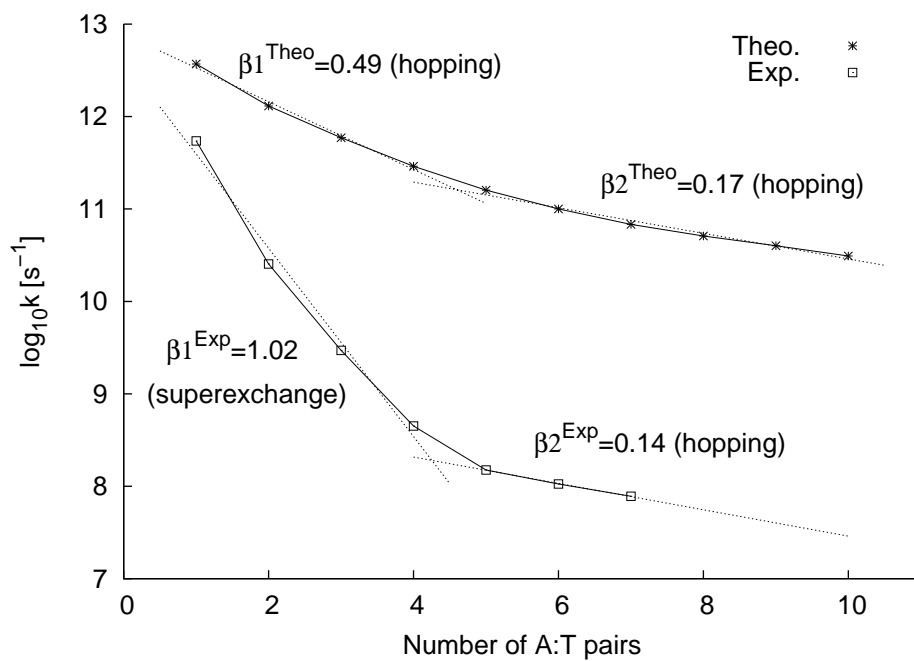
Figure 9: Distance-dependency of CT rates. Experimental values are from ref.[6]

Supporting information for: Stacking Effects on Charge Transfer Dynamics in Fluctuating DNA

Zhong-wei Li and Ke-li Han[*]

State Key Laoratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Science, Dalian 116023 (P. R. China)

E-mail: klhan@dicp.ac.cn

MNDO Predicted CT Parameters

Take fluctuating $(A{:}T)_2(G{:}C)$ for example, the probability distribution of MNDO predicted onsite energies as well as the electronic couplings for hole transfer are plotted in Figure S1 and Figure S2. They fit the Gaussian distributions with averaged values and mean square deviations listed as in Table S1

Table S1: Averaged values μ and mean square deviations σ in Figure S1 and Figure S2.

	μ (eV)	σ (eV)
ϵ_1	8.80	0.21
ϵ_2	8.58	0.23
ϵ_3	8.01	0.22
J_{12}	-0.00025	0.047
J_{23}	-0.0021	0.038

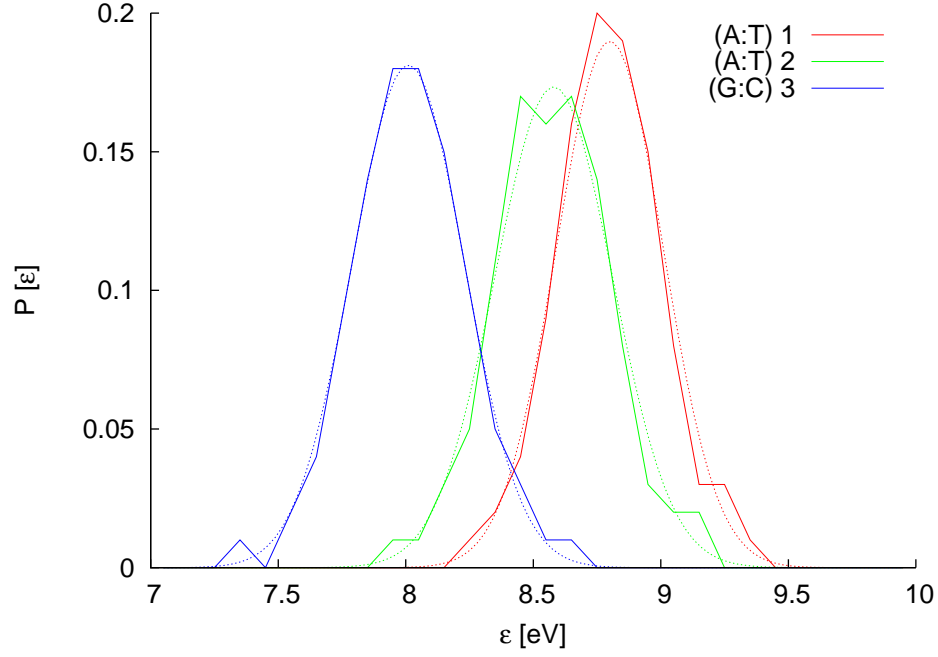
[*]To whom correspondence should be addressed

Figure S1: Probability distribution of the onsite energies of fluctuating $(A:T)_2(G:C)$ system for hole transfer.
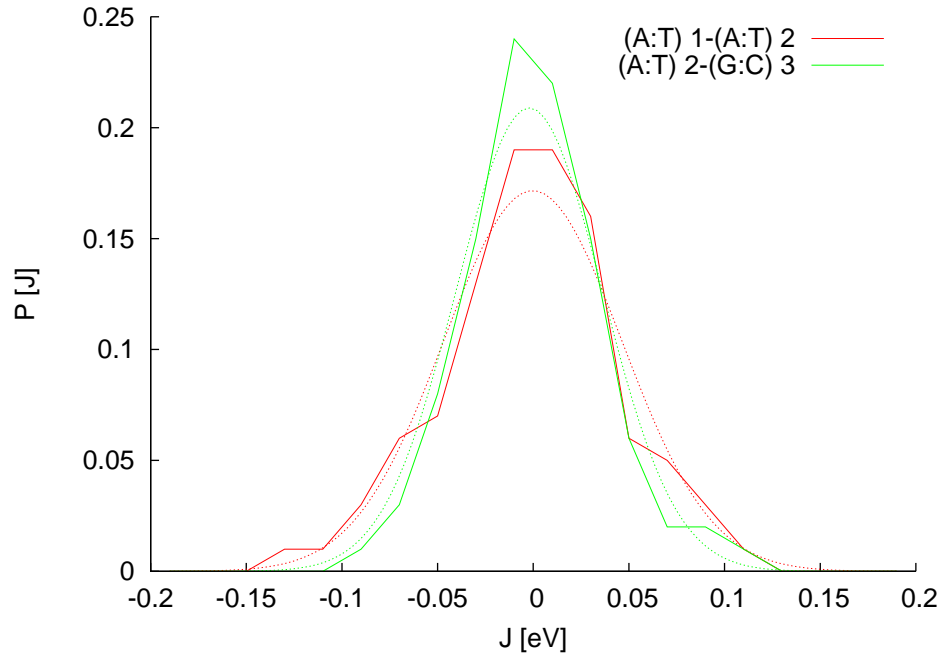

Figure S2: Probability distribution of the electronic couplings of fluctuating $(A:T)_2(G:C)$ system for hole transfer.

Hückel Analysis on π-Stacked Structure

According to Hückel theory, the basis set used in the π-stack is schematically shown as in Figure S3.
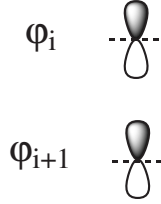

$$\varphi_i$$

$$\varphi_{i+1}$$

Figure S3: Schematic diagram of π-stack bases used in Hückel theory.

Corresponding bond matrix is

$$
\begin{bmatrix}
-\omega & 1 & & & & \\
1 & -\omega & 1 & & & \\
& 1 & -\omega & 1 & & \\
& & \cdots & \cdots & \cdots & \\
& & & 1 & -\omega & 1 \\
& & & & 1 & -\omega
\end{bmatrix}_n ,
\tag{S1}
$$

wherein only the interactions between neighboring sites are considered. Thus the Hückel MO energy of the n-stack can be indicated by the value of ω which has a generalized solution of

$$
\omega_s = 2\cos\left(\frac{\pi}{n+1}s\right), \quad (s = 1, 2, \cdots, n),
\tag{S2}
$$

with corresponding MO coefficients of

$$
\xi_{rs} = \left(\frac{2}{n+1}\right)^{1/2} \sin\left(\left(\frac{\pi s}{n+1}\right)r\right), \quad (r, s = 1, 2, \cdots, n).
\tag{S3}
$$

ω and ξ for $n = 2 \sim 6$ are listed in Table S2.

3

Table S2: Predicted MO eigenvalues and eigenvectors of π-stack systems by Hückel method.

n-stack	MO	ω	ξ_1	ξ_2	ξ_3	ξ_4	ξ_5	ξ_6
2	ϕ_1	1.000	0.707	0.707				
	ϕ_2	-1.000	0.707	-0.707				
3	ϕ_1	1.414	0.500	0.707	0.500			
	ϕ_2	0.000	0.707	0.000	-0.707			
	ϕ_3	-1.414	0.500	-0.707	0.500			
4	ϕ_1	1.618	0.372	0.602	0.602	0.372		
	ϕ_2	0.618	0.602	0.372	-0.372	-0.602		
	ϕ_3	-0.618	0.602	-0.372	-0.372	0.602		
	ϕ_4	-1.618	0.372	-0.602	0.602	-0.372		
5	ϕ_1	1.732	0.289	0.500	0.578	0.500	0.289	
	ϕ_2	1.000	0.500	0.500	0.000	-0.500	-0.500	
	ϕ_3	0.000	0.577	0.000	-0.577	0.000	0.577	
	ϕ_4	-1.000	0.500	-0.500	0.000	0.500	-0.500	
	ϕ_5	-1.732	0.289	-0.500	0.577	-0.500	0.289	
6	ϕ_1	1.802	0.232	0.418	0.521	0.521	0.418	0.232
	ϕ_2	1.247	0.418	0.521	0.232	-0.232	-0.521	-0.418
	ϕ_3	0.445	0.521	0.232	-0.418	-0.418	0.232	0.521
	ϕ_4	-0.445	0.521	-0.232	-0.418	0.418	0.232	-0.521
	ϕ_5	-1.247	0.418	-0.521	0.232	0.232	-0.521	0.418
	ϕ_6	-1.802	0.232	-0.418	0.521	-0.521	0.418	-0.232

Corresponding MO shapes are drawn schematically as in Figure S4.

2-stack

4-stack

φ1 φ2 φ3 φ4 φ5

φ1 φ2 φ3

3-stack

5-stack

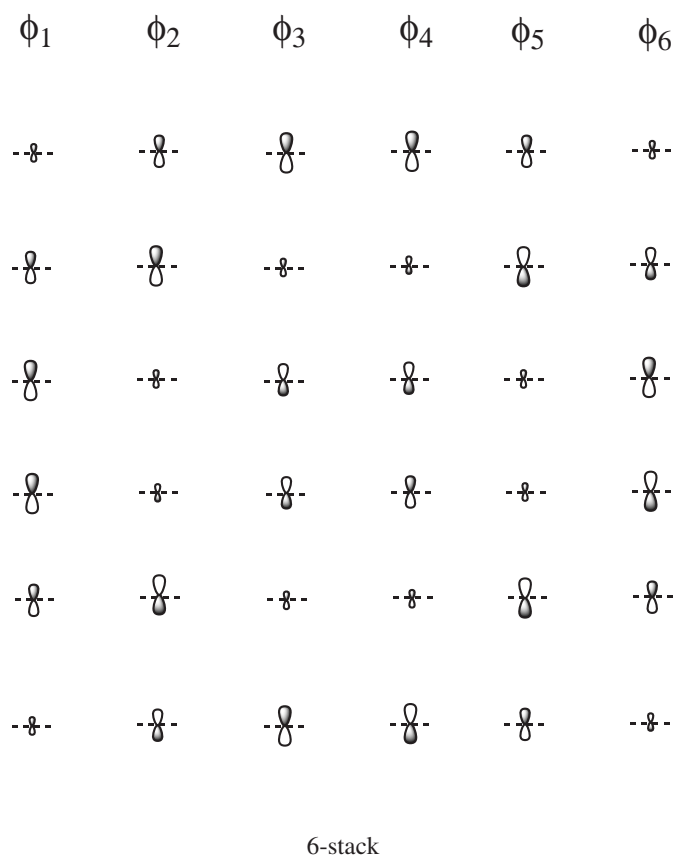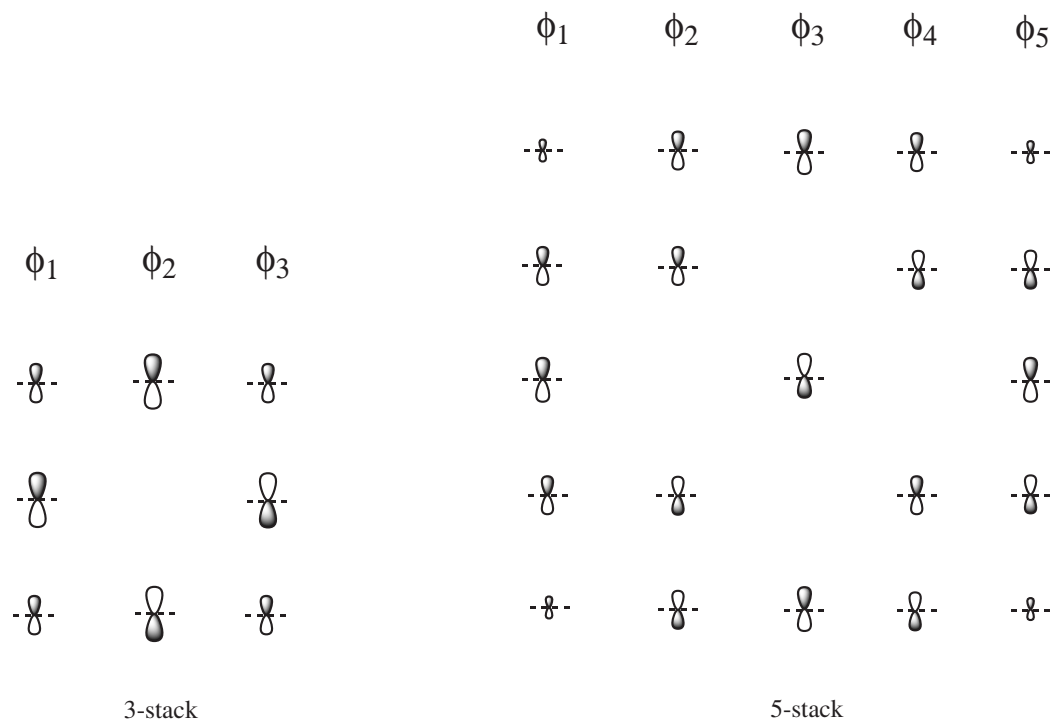φ1 φ2 φ3 φ4 φ5 φ6
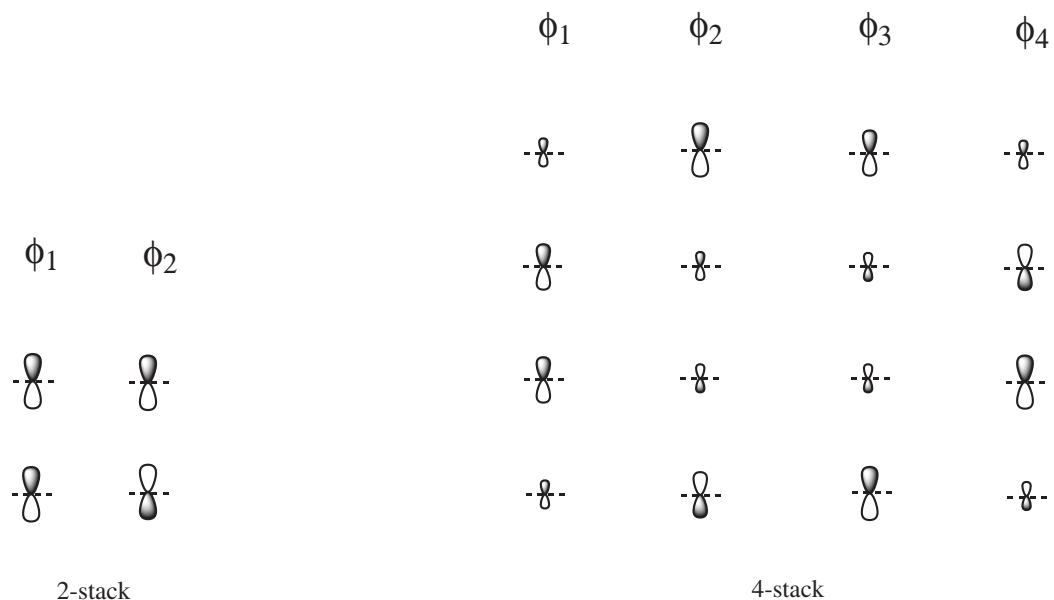
6-stack

Figure S4: MO shapes of n-mer (n=2∼6) π-stack systems predicted by Hückel theory.

The Dependence of k on CT Parameters

The Marcus-Hush equation of Eq. (1) in the main context can be rewritten as

$$k = \frac{J^2}{\hbar} \left(\frac{\pi}{k_B T} \right)^{1/2} \exp \left(-\frac{A}{4 k_B T} \right) \tag{S4}$$

wherein A here called the effective Gibbs free energy is

$$A = \frac{(\Delta G + \lambda)^2}{\lambda} - 4 k_B T \ln \left(\lambda^{-1/2} \right) \tag{S5}$$

Thus the Dependence of k on J and A can be seen as in Figure S5.
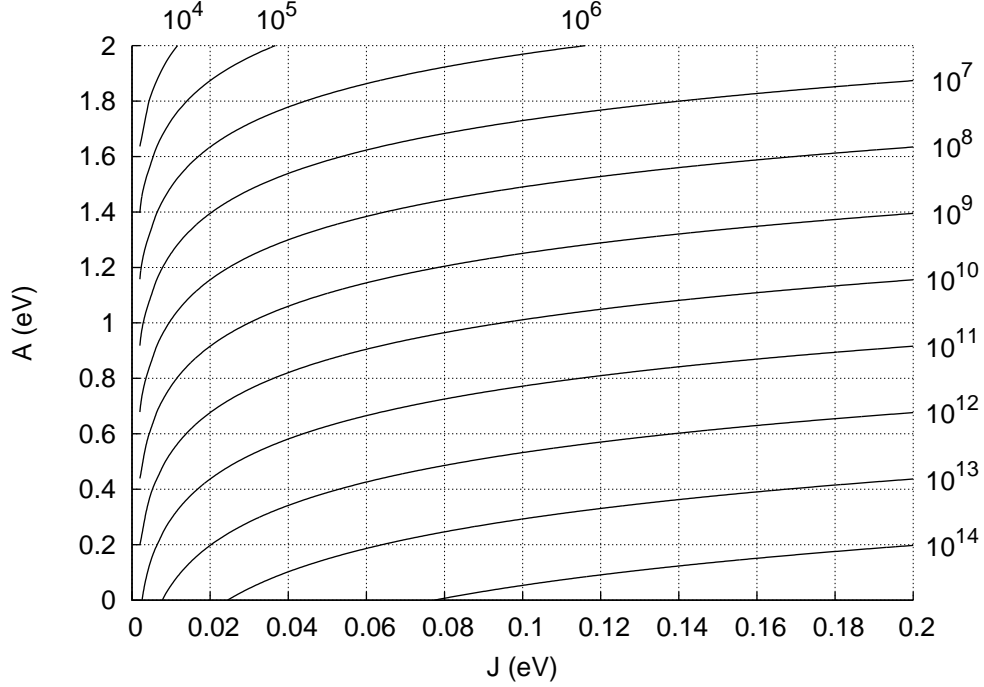
fig

Figure S5: Contour lines of rate constant k (in power series) as a function of electronic coupling J and effective Gibbs free energy A.

As Figure S5 shows, A has a greater impact on k than J does. Since λ is set to invariable in our model, the large fluctuations of k are mainly attributed to the fluctuating onsite energies more than to electronic couplings.

Squeezing Effect in Short Time

Figure 6 in the maintext shows the averaged onsite energies of individual sites in 2 ns. In a short time, the stacking structure is hard to maintain in a long distance due to the flexibility. Instead, a "well" formed in a short range would provide a wave-like driving force to transmit a charge to a distance as the strand fluctuates (see Figure S6).
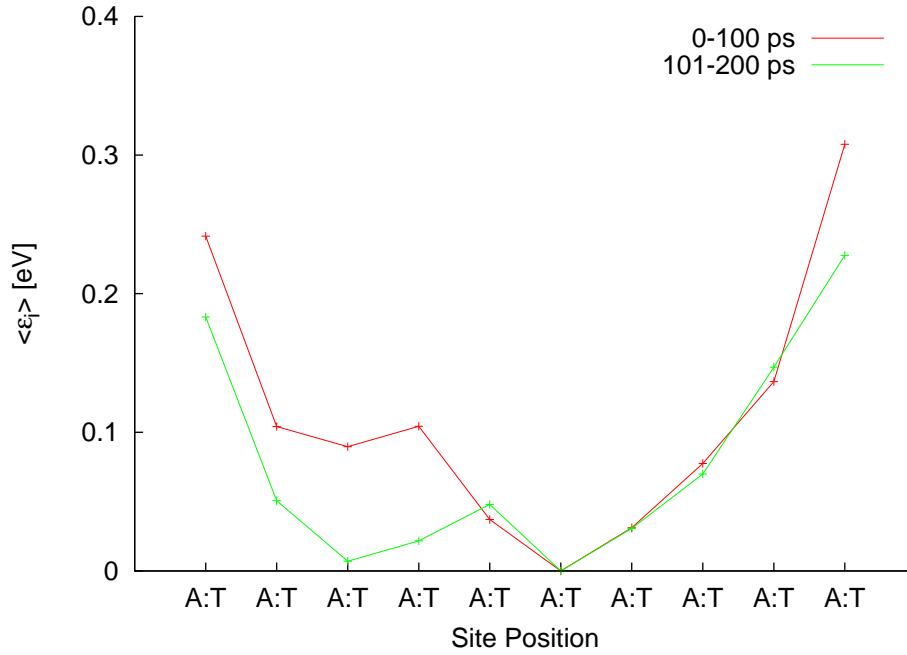

Figure S6: MNDO predicted relative onsite energies for a hole in the individual sites of $(A:T)_{10}$ system. The values are averaged on 500 samples from the first (red) and the second (green) 100 ps MD simulations, respectively.