

Revisiting the Hole Size in Double Helical DNA with Localized Orbital Scaling Corrections

Published as part of *The Journal of Physical Chemistry virtual special issue "David N. Beratan Festschrift"*.

Ye Jin,^{||} Xuyan Ru,^{||} Neil Qiang Su, Yuncai Mei, David N. Beratan,^{*} Peng Zhang,^{*} and Weitao Yang^{*}

 Cite This: *J. Phys. Chem. B* 2020, 124, 3428–3435

 Read Online

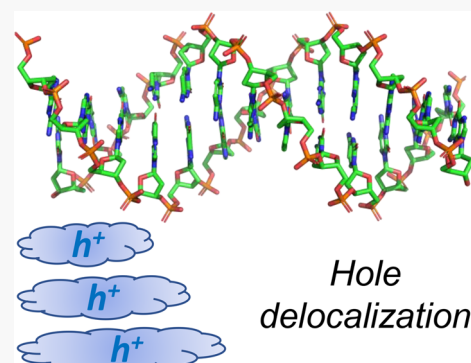
ACCESS |

 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: The extent of electronic wave function delocalization for the charge carrier (electron or hole) in double helical DNA plays an important role in determining the DNA charge transfer mechanism and kinetics. The size of the charge carrier's wave function delocalization is regulated by the solvation induced localization and the quantum delocalization among the π stacked base pairs at any instant of time. Using a newly developed localized orbital scaling correction (LOSC) density functional theory method, we accurately characterized the quantum delocalization of the hole wave function in double helical B-DNA. This approach can be used to diagnose the extent of delocalization in fluctuating DNA structures. Our studies indicate that the hole state tends to delocalize among 4 guanine–cytosine (GC) base pairs and among 3 adenine–thymine (AT) base pairs when these adjacent bases fluctuate into degeneracy. The relatively small delocalization in AT base pairs is caused by the weaker π – π interaction. This extent of delocalization has significant implications for assessing the role of coherent, incoherent, or flickering coherent carrier transport in DNA.



INTRODUCTION

The study of charge transport (CT) in DNA has been an area of active investigation and debate for more than two decades. Charge flow through DNA is associated with DNA damage and repair and also with applications in nanotechnology.^{1–8} Purines have lower oxidation potentials than pyrimidines, and guanine has the lowest oxidation potential. Thus, guanine has been used as a hole trap in studies of photoinduced electron transfer and as an intermediate in long-distance multistep hole transport in DNA. The hole transport rate through DNA has been shown to depend on the distance between the donor and the acceptor, and several candidate mechanisms were proposed.^{9–17} Charge transfer over a short distance (2–3 base pairs) is generally accepted to occur via a coherent tunneling (superexchange) or a flickering resonance mechanism, with an exponential rate dependence on distance,^{8,9,12,13,18} while longer transfer is generally believed to occur by incoherent hopping.¹¹ In the flickering resonance mechanism, the transient resonance among electronic states of adjacent base pairs is created by structural fluctuations, and the spatial extent of the resonances depends on the DNA sequence; the delocalization extends to 3–4 base pairs in DNA.^{8,9,19,20} This mechanism was also considered as an intermediate regime for DNA CT. In this regime, the hole delocalizes across a purine tract, and the delocalized hole contributes to CT, producing a mechanism that is distinct from either coherent tunneling or incoherent nearest neighbor

hopping.¹⁰ Conformational and structural influences on DNA CT were examined in recent theoretical studies.^{20–26} For long-range CT in DNA, a variable-range hopping mechanism was proposed. In this mechanism, charge relay stations in the hopping pathways may create delocalized islands, and incoherent hopping among delocalized islands determines the CT kinetics.²⁷ A common feature in those mechanisms is that the transient resonance (state delocalization) among bases occurs, and the extent of delocalization (the delocalization length) plays an important role in the DNA charge transfer, especially for homogeneous DNA sequences where strong temporal and spatial correlations in base energies and couplings are found among the neighboring base pairs.^{19,20,28}

The mechanistic transition between single-step tunneling and multistep hopping in DNA CT depends on several factors, including sequence, distance, conformation, solvation, and hole state energetics. Above all, the hole size, also referred to as the polaron size, plays an important role in understanding the long-range DNA CT rate.^{29,30} Charge delocalization over G

Received: April 7, 2020

Published: April 9, 2020



bases in DNA was first explored theoretically by Sugiyama and Saito,³¹ and recent studies focused on analyzing the charge delocalization from different perspectives, including wave function analysis,^{32,33} QM/MM calculation,³⁴ and semi-empirical calculation.³⁵ Conwell and Basko used a tight-binding model and found that the wave functions of holes trapped on guanines are extended over 6 base pairs.³⁶ Senthilkumar et al. examined the hole delocalization using the density functional theory derived charge distribution and found that the charge distribution is strongly dependent on the structure of the adjacent base pairs. In the gas phase, the hole is delocalized over neighboring guanine bases in repeated G runs (GG or GGG).³⁷ Voityuk included solvation effects and found that, for neighboring bases in G repeat sequences (GG or GGG), charge distributions are localized to a single guanine site.³⁸

The hole delocalization is determined by competition between the quantum delocalization of the hole electronic state and the solvation induced charge localization.³⁹ Also, structural disorder produces localization. Quantum delocalization of the hole wave function can be characterized by the dependence of the ionization potential (IP), or the hole state energy, on the length of the DNA sequence. IP values calculated using Hartree–Fock (HF) theory are of low accuracy. Post-HF wave function theories often give a better description of the molecular electronic structure but are computationally expensive. Density functional theory (DFT),^{40–44} on the other hand, has achieved enormous success. However, challenges remain for DFT, including the issue of delocalization error.^{45–47} Popular density functional approximations (DFAs) usually suffer from delocalization error and consequently underestimate the band gap, IP, reaction barriers, and charge transfer excitation energies.⁴⁷ In particular, traditional DFAs usually provide the wrong trend for the IP with respect to the length of a stacked DNA structure, and this error is caused by the delocalization error (see computed IPs in the Results and Discussion).

In order to reduce the delocalization error, various corrections have been added to DFAs. Hybridization methods, including local hybrid,^{48,49} double hybrid,^{50,51} and range-separated functionals,^{52–54} are commonly used. IP tuned functionals, including Koopmans-compliant (KC) functionals⁵⁵ and tuned range-separated functionals,^{56–58} were also developed in recent years. However, these methods are either limited in their correction to the delocalization error or are not size-consistent with system-specific tuning. Recently, Yang and co-workers developed a localized orbital scaling correction (LOSC)⁵⁹ as a universal method to treat fractional electrons explicitly and to correct the delocalization error in diverse DFAs. This method is size-consistent and thus is particularly well suited to reduce the delocalization error in a system with an increasing size.

In this study, we investigate the transient hole state delocalization in double helical B-DNA with consecutive guanine–cytosine (GC) and adenine–thymine (AT) runs using the recently developed LOSC method. We examined the IPs of the DNA base pairs using LOSC, and we benchmarked this approach using high-level *ab initio* methods. We found that LOSC derived IPs agree well with those computed using high-level coupled-cluster methods. Combining molecular dynamics ensemble averaged IPs and solvation energies, we found that the transient hole state of GC base pairs can delocalize among up to 4 base pairs, while AT base pairs can delocalize among

up to 3 base pairs. The derived hole size of the GC base pairs is longer than the earlier HF-theory-based predictions³⁹ and is consistent with estimates obtained using classical molecular dynamics derived base pair energy fluctuations and nearest neighbor base pair electronic coupling strengths.¹⁹ Note that the lifetime of the transient resonance, and its quantitative contribution to DNA charge transfer kinetics, requires quantum dynamical simulations where the dephasing and decoherence need to be handled appropriately, and those further studies are beyond the scope of the present work.

METHODS

Previous studies showed that the interplay of the π stack interaction driven quantum delocalization and the solvation and disorder induced localization leads to hole states that are somewhat delocalized.³⁹ The hole size is defined as the number of nucleobases over which the hole state is delocalized in a purine tract. A critical parameter that characterizes the hole size in DNA is the base pair redox potential (see the Supporting Information for the correlation between redox potential change and charge delocalization). The two contributions to the base pair redox potential are shown from the four-state thermodynamic cycle in Figure 1: one

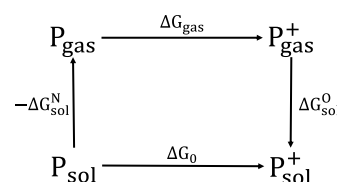


Figure 1. Thermodynamic cycle for the oxidation of DNA base pairs. P is the base pair run (AT or GC pairs). The oxidation free energy change of DNA base pairs in solution (ΔG_0) is calculated as a sum of both intrinsic electronic binding energy (vacuum IP ΔG_{gas}) and the difference in the solvation energies between DNA base pair oxidized state (ΔG_{sol}^0) and neutral state (ΔG_{sol}^N).

contribution arises from the intrinsic electronic binding energy (correlated with the vacuum IP), and the other contribution arises from the difference in the solvation energies of the neutral and oxidized states of the base pairs. Below, we address these two contributions to determine the upper bound of the transient hole size in homogeneous double helical B-DNA.

Ionization Potential from Localized Orbital Scaling Correction Analysis. The IPs of double helical DNA that contains a homogeneous GC sequence and AT sequence are calculated using the LOSC DFT method. In DFT, the connection between the IP and the highest occupied molecular orbital (HOMO) energy has been studied extensively.^{40,43,60} The exact Kohn–Sham HOMO orbital energy is $-\text{IP}$, based on asymptotic analysis of the electron density.⁴³ Perdew, Parr, Levy, and Balduz proved that, for the exact functional, the chemical potential for electron removal (addition) is equal to $-\text{IP}$ ($-\text{EA}$),⁶⁰ which is the derivative of the total energy with respect to the electron number. For a given density functional approximation, Cohen, Moris-Sanchez, and Yang proved that the chemical potential for electron removal (addition) is the Kohn–Sham HOMO (LUMO) orbital energy when the exchange–correlation energy is expressed as an explicit and differentiable functional of the electron density.⁶¹ However, when the exchange–correlation energy is expressed as an explicit and differentiable functional of the Kohn–Sham

density matrix, the chemical potential for electron removal (addition) is the generalized Kohn–Sham HOMO (LUMO) orbital energy.⁶¹ Thus, for a DFA with minimal delocalization error, the generalized Kohn–Sham HOMO (LUMO) should lead to the accurate prediction of the IP (EA) for molecules and bulk systems.⁴⁵ The LOSC method systematically corrects the delocalization error in common DFAs. The orbital energies are calculated with

$$\epsilon_m^{\text{LOSC}} = \epsilon_m^{\text{DFA}} + \Delta\epsilon_m^{\text{LOSC}} \quad (1)$$

by imposing a correction to the orbital energies of common DFAs.

In LOSC, the correction to the total energy is given by

$$\Delta E^{\text{LOSC}} = \sum_i \frac{1}{2} \kappa_{ii} (\lambda_{ii} - \lambda_{ii}^2) - \sum_{i \neq j} \frac{1}{2} \kappa_{ij} \lambda_{ij}^2 \quad (2)$$

which imposes the Perdew–Parr–Levy–Baldur (PPLB) condition⁶⁰ in each local region. In this correction, λ_{ij} is the occupation matrix of the localized orbitals (LOs), called orbitalets, $\{\phi_i(\mathbf{r})\}$, $\lambda_{ij} = \langle \phi_i | \rho_s | \phi_j \rangle$, and the curvature matrix elements, $\{\kappa_{ij}\}$, read

$$\begin{aligned} \frac{1}{2} \kappa_{ij} = \frac{1}{2} \iint \frac{\rho_i(\mathbf{r}) \rho_j(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \\ - \frac{\tau C_x}{3} \int [\rho_i(\mathbf{r})]^2 [\rho_j(\mathbf{r})]^2 d\mathbf{r} \end{aligned} \quad (3)$$

where

$$\rho_i(\mathbf{r}) = |\phi_i(\mathbf{r})|^2, \quad C_x = \frac{3}{4} \left(\frac{6}{\pi} \right)^{1/3} \quad (4)$$

and τ is a nonempirical parameter, $\tau = 6(1 - 2^{-1/3})$. Here, the desired LOs, $\{\phi_i(\mathbf{r})\}$, are generated by a restrained Boys localization procedure from the mixture of canonical orbitals (COs) $\{\varphi_i(\mathbf{r})\}$. In particular, the LOs minimize the target function

$$\{\phi_i\} = \arg \min F[\{\phi_i\}, \{\varphi_i\}, \{\epsilon_i^{\text{CO}}\}] \quad (5)$$

where $\{\epsilon_i^{\text{CO}}\}$ is the set of CO energies. The target function, F , is given by

$$F = \sum_i [\langle \phi_i | \mathbf{r}^2 | \phi_i \rangle - \langle \phi_i | \mathbf{r} | \phi_i \rangle^2] + \sum_{ij} \omega_{ij} U_{ij}^2 \quad (6)$$

where U is the unitary overlap matrix $U_{ij} = \langle \phi_i | \phi_j \rangle$ and ω is a penalty function, $\omega_{im} = \omega(|\epsilon_i^{\text{LO}} - \epsilon_m^{\text{CO}}|)$. The role of this penalty function is to suppress the mixing between orbitals when they are far apart in energy. Reference 59 provides a detailed discussion of the penalty function. With a frozen orbital approximation, the LOSC correction to the orbital energy from eq 1 is

$$\Delta\epsilon_m^{\text{LOSC}} = \sum_i \kappa_{ii} \left(\frac{1}{2} - \lambda_{ii} \right) |U_{im}|^2 - \sum_{i \neq j} \kappa_{ij} \lambda_{ij} U_{im} U_{jm}^* \quad (7)$$

Solvation Energy. The solvation energy contribution to the redox potential can be estimated using a four-state thermodynamic cycle. The four states include a singly oxidized N base pair sequence in a vacuum and water (with their free energies denoted by E_{ov} and E_{ow}) as well as N neutral base pairs in a vacuum and water (with free energies denoted E_{nv} and E_{nw}). The solvation energy contribution to the redox

potential (ΔE_{sol}) is calculated as the difference between the oxidized and neutral stacked base pair energies³⁹

$$\Delta E_{\text{sol}} = -[(E_{\text{nw}} - E_{\text{nv}}) - (E_{\text{ow}} - E_{\text{ov}})]/e \quad (8)$$

The solvation energy is evaluated by solving the finite-difference Poisson–Boltzmann (FDPB) equations using Delphi.^{62,63} The FDPB method can produce solvation free energies of excellent accuracy compared to experiments,^{64,65} and the intrinsic error of FDPB may be nearly canceled out when computing the free energy difference.⁶⁶ In the analysis, the internal dielectric constant is 4 and the solvent dielectric constant is 80. A cubic lattice is generated for the double helical DNA with a grid size of 201 and 2.0 grids/Å. We used the PARSE atomic radii.⁶² The atomic charges of oxidized and neutral states of DNA bases are the Mulliken charges calculated using the B3LYP/6-31G* optimized single base pair structure. In the solvation energy calculations, the effect of structural fluctuations is included in the ensemble averaged solvation energy using molecular dynamics (MD) snapshots.

MD simulations are carried out on five double helical DNA structures for consecutive GC and AT pairs, $5' - A \dots G_N \dots A - 3'$ and $5' - G \dots A_N \dots G - 3'$, $N = 1-5$, with a total sequence length of 12. Starting structures of these DNA fragments were generated using the PyMOL builder module.⁶⁷ In all simulations, GROMACS 2019.2⁶⁸ was used with the Amber99bsc1 force field.⁶⁹ The DNA duplex was embedded in a rectangular box of TIP3P water molecules.⁷⁰ The DNA is placed in the center and at least 10 Å from the box edge. Twenty-two Na^+ ions were included to neutralize the system. 200 ps of equilibrium (100 ps NVT and 100 ps NPT) was carried out, and 1 ns of MD simulation was performed with conformations saved every 2 ps. For all MD simulations, 500 snapshots were extracted for each DNA duplex, and the last 400 snapshots were used to study the dependence of the IP on the molecular and solvation fluctuations.

RESULTS AND DISCUSSION

Ionization Potential. We benchmarked our LOSC-based IP calculations with experiments and with high-level coupled-cluster computations. All calculations are performed using a local QM4D program⁷¹ with the Def2-SVP basis set. The calculated IPs using LOSC with three different functionals appear in Table 1. Without LOSC, DFT determined IP values

Table 1. LOSC Determined Ionization Potentials (eV) Using Selected DFT Functionals^a

IP/eV	PBE	BLYP	B3LYP	Exp
guanine	7.66 (4.86)	7.51 (4.64)	7.65 (5.66)	7.8 ⁷²
GC-WC pair	6.97	6.75	7.10	
GC stack	7.81	7.60	7.97	

^aGC-WC refers to Watson–Crick pair. GC stack refers to a stacked guanine and cytosine in a single strand. The numbers in parentheses are DFT determined ionization potentials without LOSC. All geometries are optimized with B3LYP/6-31G*.

deviate from the experimental value by more than 2 eV. The LOSC computed IPs agree well with the measured IPs and are insensitive to the choice of functionals. The IPs derived using PBE and B3LYP are similar, and LOSC with BLYP produces slightly lower IPs. The LOSC/PBE calculated IP of a single guanine is very close to the experimental IP (7.8 eV from the

NIST WebBook⁷²). We use PBE in the following IP calculations.

We also assessed the accuracy of the LOSC derived IPs of the stacked Watson–Crick GC pairs (GC-WC, abbreviated as (GC)_n, *n* = 1–4 below) by comparing with high-level coupled-cluster (CC) results. In the calculations, stacked GC pair geometries with a neighboring distance of 3.4 Å were taken from the starting structure of the MD simulation and the calculated IPs using LOSC and two CC methods are presented in Table 2. For (GC)₁ and (GC)₂, we compared LOSC results

Table 2. Ionization Potentials (eV) Derived Using HF and PBE with and without LOSC, Compared with Coupled-Cluster Methods: CCSD, CCSD(T), and CC2^a

IP/eV	HF	PBE	LOSC	CCSD	CCSD(T)	CC2
(GC) ₁	7.47	4.28	6.98	6.95	7.32	6.93
(GC) ₂	6.93	3.83	6.42	6.37	6.71	6.50
(GC) ₃	6.62	3.57	6.12			6.08
(GC) ₄	6.42	3.42	5.77			5.60

^aLOSC shows high accuracy compared to high-level CC methods.

with coupled-cluster singles and doubles (CCSD). The computed results showed that LOSC provides high-accuracy IPs compared to CCSD results. The differences in the computed IPs are smaller than 0.05 eV. For (GC)₃ and (GC)₄, we compared LOSC results with CC2 calculations because of the increased computational costs. LOSC derived IPs of 6.12 and 5.77 eV for (GC)₃ and (GC)₄ differ from the CC2 IPs by no more than 0.2 eV. For (GC)₁ and (GC)₂, we also computed their IPs using the domain-based local pair natural orbital (DLPNO) CCSD(T) theory⁷³ with a triple- ζ basis set (def2-TZVPP). LOSC IPs show reasonable agreements with the DLPNO-CCSD(T) results. Overall, LOSC IPs are very accurate compared with high-level CC methods. In contrast, PBE computed IPs are poor compared to the coupled-cluster results, and the error is more pronounced for the short GC pair lengths. IPs derived using the HF method show better performance than PBE calculations but systematically overestimate the IP by more than 0.5 eV. In particular, the error increases with the number of GC base pairs.

To evaluate the impact of structural fluctuations on the IP, we calculated the ensemble averaged IP for a single GC pair from the MD simulation with the sequence 5'-AAAAAAGA-AAAA-3'. The ensemble averaged LOSC IP is 6.77 eV, and the root mean squared error (RMSE) is 0.3 eV. The derived RMSE is higher than the reported value, on the order of 0.1 eV.^{28,74} However, when calculating the IP and RMSE using the LOSC corrected total energy difference between the neutral and ionized species, we obtained a standard deviation of 0.18 eV, in accordance with previously reported results. The computed ensemble averaged LOSC IP of 7.01 eV is consistent with the result derived using the LOSC corrected HOMO energy. Further calculations for dimer systems ((GC)₂) show that the energy difference between ensemble averaged IPs is similar to the results from the HOMO energies. In the following calculations, we use the ensemble averaged IP (LOSC corrected HOMO energy) derived from MD simulated trajectories to determine the hole state size.

Solvation Energy. We calculated the solvent contribution to the hole state redox potential following the strategies described earlier.³⁹ The DNA duplex backbone is not considered in the calculations of electrostatic energies in

both oxidized and neutral states. The contributions of IP and solvation to the redox potential are shown in Table 3 for

Table 3. IP and Solvation Energy Contribution to the Redox Potential for Different Lengths of GC Pairs (eV)

	IP	Δ IP	$e\Delta E_{\text{sol}}$	$e\Delta\Delta E_{\text{sol}}$	Δ IP + $e\Delta\Delta E_{\text{sol}}$
(GC) ₁	6.77	0	−1.93	0	0
(GC) ₂	6.32	−0.45	−1.54	0.39	−0.06
(GC) ₃	6.04	−0.73	−1.28	0.64	−0.09
(GC) ₄	5.91	−0.86	−1.11	0.81	−0.05
(GC) ₅	5.87	−0.90	−0.93	1.00	0.10

selected GC pairs. The computed solvation energies are nearly the same as the values reported in ref 39, indicating that the backbone contribution is small and negligible. This small backbone effect is caused by the energy cancellation between E_{xv} and E_{xw} (*x* corresponding to *n* and *o*) in eq 8.

Table 3 shows that the IP drops rapidly with the GC pair length increasing from 1 to 4 and converges to 5.87 eV at 5 GC pairs. Compared to the previous study of IPs calculated with the HF method,³⁹ the LOSC IP values decrease faster and converge more slowly (i.e., the states are delocalized), as shown in Figure 2. In the previous study,³⁹ the hole state

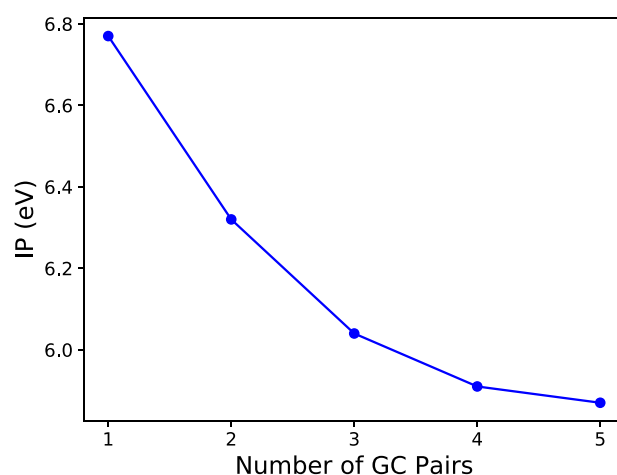


Figure 2. IP for different numbers of GC pairs. The IP drops when the base pair length increases from 1 to 4 and plateaus at 5.

delocalization is determined by the competition between the decrease of the IP and the increase of the solvation energy as the number of base pairs grows. To be consistent with the previous study, we used the ensemble averaged solvation energy reported in ref 39 for GC pairs. The combined IP and solvation energy contributions to the redox potentials are shown in Figure 3. At short lengths (1–4 base pairs), the solvent does not localize the hole since the quantum delocalization effect dominates. Although a redox potential increase is observed for 3–4 base pairs, the small energy change (0.04 eV or 1.5 $k_B T$) is insufficient to disrupt the hole delocalization. For 4–5 base pairs, since the IP converges and the solvation energy continues to increase, the hole is localized and stabilized around 4 base pairs. The hole size of GC pairs in our study (3–4 base pairs) is larger than the result found in the previous study³⁹ but is in accordance with recent studies in DNA CT, where the hole size was estimated using the hole

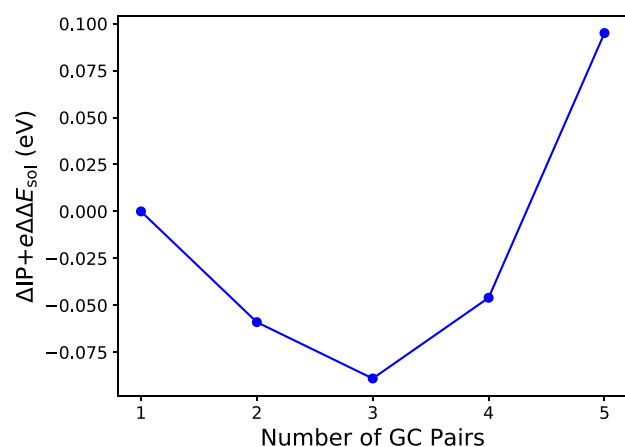


Figure 3. Combined contribution of IP and solvation energy for different lengths of GC pairs. There is a significant increase between 4 and 5 base pairs, indicating that the size of the hole is around 3–4 base pairs.

energy fluctuations and the nearest neighbor GC pair electronic coupling strength.^{14,19}

We also calculated the contributions of the IP and the solvation energy to the redox potential for AT base pairs, and the results are shown in Table 4 and Figure 4. Compared to

Table 4. IP and Solvation Energy Contribution to the Redox Potential for Different Lengths of AT Pairs (eV)

	IP	ΔIP	$e\Delta E_{sol}$	$e\Delta\Delta E_{sol}$	$\Delta IP + e\Delta\Delta E_{sol}$
(AT) ₁	7.76	0	−1.65	0	0
(AT) ₂	7.36	−0.40	−1.30	0.35	−0.05
(AT) ₃	7.14	−0.62	−1.10	0.55	−0.07
(AT) ₄	7.00	−0.76	−0.76	0.90	0.14

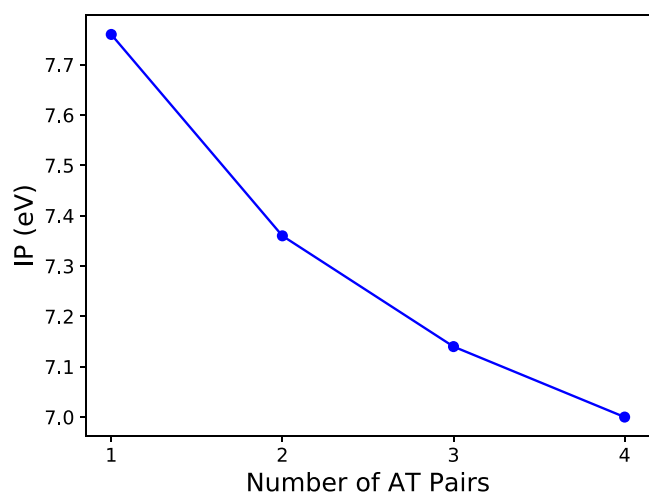


Figure 4. IP for different numbers of AT pairs. The IP drops when the base pair length increases from 1 to 4, and the difference is smaller compared to that of GC pairs.

GC pairs, the IP of AT base pairs drops slowly as the number of AT base pairs grows. The IP change from 1 to 4 base pairs is 0.76 eV, smaller than the 0.86 eV change for GC pairs. In contrast, the solvation energy changes from 1 to 4 base pairs are comparable to those for the GC pairs. The combined contribution to the redox potential, $\Delta IP + e\Delta\Delta E_{sol}$, stabilizes

the hole state when the length of AT pairs grows from 1 to 3. Further increasing the AT pair length from 3 to 4 raises the redox potential by about 0.21 eV, shown in Figure 5. This

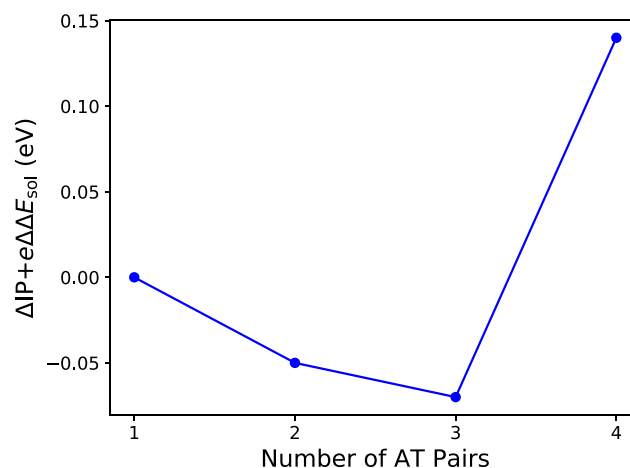


Figure 5. Combined contribution of IP and solvation energy for different lengths of AT pairs. The redox potential of AT pairs decreases with increasing base length from 1 to 3 and increases from 3 to 4, indicating that the hole size is 3 AT base pairs.

indicates that the hole can delocalize among 3 AT base pairs. The smaller delocalization in AT pairs may be caused by the weaker electronic interaction between intrastrand nearest neighbor adenines (24 meV) as compared to that of intrastrand nearest neighbor guanines (37 meV).⁷⁵ The small electronic interaction among AT pairs impedes the wave function delocalization and thus produces a small hole size compared to GC pairs. A previous study by Elstner et al.³⁴ indicated that the average electronic coupling between AT pairs in poly(A) (A_8) is larger than the coupling between GC pairs in poly(G) (G_8), leading to strong hole delocalization for AT base pairs. The CT parameters were computed using the semiempirical self-consistent charge density functional tight binding (SCC-DFTB) method. Our results contradict this SCC-DFTB study, and the disparity may be caused by the poor descriptions of electronic delocalization with the semiempirical SCC-DFTB method.

CONCLUSIONS

Wave function delocalization of a hole in double helical DNA plays an important role in determining CT mechanisms and kinetics. Previous theoretical studies of hole states in stacked DNA produced diverse predictions concerning the hole size. The disparity was caused largely by the poor description of the electronic delocalization associated with commonly used density functional approximations. The localized orbital scaling correction (LOSC) method, which reduces the delocalization error in most density functional approximations, allows us to investigate the ionization potentials and the hole electronic state delocalization of DNA base pairs systematically. In the gas phase, the LOSC derived IP of a single guanine, 7.51–7.66 eV, agrees closely with the experimental value of 7.8 eV, and the values are virtually independent of the choice of functional. For stacked DNA base pairs, the LOSC derived gas phase IPs of (GC)_{1–4} differ by no more than 0.2 eV from IPs computed using coupled-cluster methods (CCSD and CC2), indicating that LOSC accurately captures the electronic delocalization in stacked DNA base pairs.

We further assessed the hole state delocalization of the GC and AT pairs in poly(G) and poly(A) DNA duplexes using the computed redox potentials. Combined with Poisson–Boltzmann solvation energy calculations and molecular dynamics ensemble averaging, we found that the hole state tends to delocalize among 4 GC base pairs and 3 AT base pairs. The hole size estimates for GC sequences are larger than the earlier HF-based predictions and other predictions, that delocalize the hole over 3 base pairs. The current analysis is consistent with previous estimates that were obtained using base pair energy fluctuations and nearest neighbor base pair electronic coupling strength. For AT base pairs, the hole state delocalization is smaller than the prior SCC-DFTB predictions.

Our assessment on the hole size in poly(A) DNA duplexes is consistent with previous experimental findings that the dependence of the charge transport mechanisms (tunneling and hopping) on the DNA duplex length indicates a delocalization length of up to 3 AT base pairs.^{76,77} Homogeneous DNA sequences have the capacity to create structures with “delocalized islands” that will impact the charge transport kinetics considerably by altering the reorganization energy and the mutual electronic interactions between delocalized islands. This delocalization effect on the charge transport kinetics was examined theoretically in helical aromatic residues,⁷⁸ where theoretical studies show that hopping among delocalized islands produces transport kinetics that differs by orders of magnitude from the sequential nearest neighbor hopping kinetics. Charge hopping among delocalized islands in homogeneous DNA sequences is expected and should be considered in the associated modeling.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.0c03112>.

Details of the correlation between the redox potential change and the charge delocalization (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

David N. Beratan – Department of Chemistry, Department of Biochemistry, and Department of Physics, Duke University, Durham, North Carolina 27708, United States; orcid.org/0000-0003-4758-8676; Email: david.beratan@duke.edu

Peng Zhang – Department of Chemistry, Duke University, Durham, North Carolina 27708, United States; Email: peng.zhang@duke.edu

Weitao Yang – Department of Chemistry, Duke University, Durham, North Carolina 27708, United States; orcid.org/0000-0001-5576-2828; Email: weitao.yang@duke.edu

Authors

Ye Jin – Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Xuyan Ru – Department of Chemistry, Duke University, Durham, North Carolina 27708, United States; orcid.org/0000-0001-8395-0666

Neil Qiang Su – Department of Chemistry, Duke University, Durham, North Carolina 27708, United States; orcid.org/0000-0001-7133-2502

Yuncai Mei – Department of Chemistry, Duke University, Durham, North Carolina 27708, United States

Complete contact information is available at:

<https://pubs.acs.org/doi/10.1021/acs.jpcb.0c03112>

■ AUTHOR CONTRIBUTIONS

Y.J., X.R.: Contributed equally to this work.

■ NOTES

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

W.Y. acknowledges support from the National Institutes of Health (Grant No. R01 GM061870-17). Y.J. acknowledges support from the Peter W. Jeffs Fellowship from Duke University and from the Center for Complex Materials from First-Principles, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0012575. X.R., D.N.B., and P.Z. acknowledge support from the National Institute of Health (Grant No. GM-48043). N.Q.S. and Y.M. acknowledge the support from the National Science Foundation (CHE-1900338).

■ REFERENCES

- (1) Beratan, D. N. Why are protein and DNA electron transfer so different? *Annu. Rev. Phys. Chem.* **2019**, *70*, 71–97.
- (2) Beratan, D. N.; Liu, C.; Migliore, A.; Polizzi, N. F.; Skourtis, S. S.; Zhang, P.; Zhang, Y. Charge transfer in dynamical biosystems, or the treachery of (static) images. *Acc. Chem. Res.* **2015**, *48*, 474–481.
- (3) Boon, E. M.; Livingston, A. L.; Chmiel, N. H.; David, S. S.; Barton, J. K. DNA-mediated charge transport for DNA repair. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12543–12547.
- (4) Holman, M. R.; Ito, T.; Rokita, S. E. Self-repair of thymine dimer in duplex DNA. *J. Am. Chem. Soc.* **2007**, *129*, 6–7.
- (5) Lewis, F. D.; Young, R. M.; Wasielewski, M. R. Tracking Photoinduced Charge Separation in DNA: From Start to Finish. *Acc. Chem. Res.* **2018**, *51*, 1746–1754.
- (6) Teo, R. D.; Rousseau, B. J. G.; Smithwick, E. R.; Felice, R. D.; Beratan, D. N.; Migliore, A. Charge transfer between [4Fe4S] proteins and DNA is unidirectional. Implications for biomolecular signaling. *Chem* **2019**, *5*, 122–137.
- (7) Teo, R. D.; Smithwick, E. R.; Migliore, A.; Beratan, D. N. A single AT-GC exchange can modulate charge transfer-induced p53-DNA dissociation. *Chem. Commun.* **2019**, *55*, 206–209.
- (8) Blumberger, J. Recent advances in the theory and molecular simulation of biological electron transfer reactions. *Chem. Rev.* **2015**, *115*, 11191–11238.
- (9) Zhang, Y.; Liu, C.; Balaeff, A.; Skourtis, S. S.; Beratan, D. N. Biological charge transfer via flickering resonance. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 10049–10054.
- (10) Renaud, N.; Berlin, Y. A.; Lewis, F. D.; Ratner, M. A. Between superexchange and hopping: an intermediate charge-transfer mechanism in poly (A)-poly (T) DNA hairpins. *J. Am. Chem. Soc.* **2013**, *135*, 3953–3963.
- (11) Giese, B. Long-distance charge transport in DNA: the hopping mechanism. *Acc. Chem. Res.* **2000**, *33*, 631–636.
- (12) Blaustein, G. S.; Lewis, F. D.; Burin, A. L. Kinetics of charge separation in poly (A)- poly (T) DNA hairpins. *J. Phys. Chem. B* **2010**, *114*, 6732–6739.
- (13) Lewis, F. D.; Liu, X.; Liu, J.; Miller, S. E.; Hayes, R. T.; Wasielewski, M. R. Direct measurement of hole transport dynamics in DNA. *Nature* **2000**, *406*, 51.
- (14) Xiang, L.; Palma, J. L.; Bruot, C.; Mujica, V.; Ratner, M. A.; Tao, N. Intermediate tunnelling–hopping regime in DNA charge transport. *Nat. Chem.* **2015**, *7*, 221.
- (15) Xu, B.; Zhang, P.; Li, X.; Tao, N. Direct conductance measurement of single DNA molecules in aqueous solution. *Nano Lett.* **2004**, *4*, 1105–1108.

- (16) Hines, T.; Diez-Perez, I.; Hihath, J.; Liu, H.; Wang, Z.-S.; Zhao, J.; Zhou, G.; Müllen, K.; Tao, N. Transition from tunneling to hopping in single molecular junctions by measuring length and temperature dependence. *J. Am. Chem. Soc.* **2010**, *132*, 11658–11664.
- (17) Kawai, K.; Majima, T. Hole transfer kinetics of DNA. *Acc. Chem. Res.* **2013**, *46*, 2616–2625.
- (18) Risser, S. M.; Beratan, D. N.; Meade, T. J. Electron transfer in DNA: predictions of exponential growth and decay of coupling with donor-acceptor distance. *J. Am. Chem. Soc.* **1993**, *115*, 2508–2510.
- (19) Liu, C.; Xiang, L.; Zhang, Y.; Zhang, P.; Beratan, D. N.; Li, Y.; Tao, N. Engineering nanometre-scale coherence in soft matter. *Nat. Chem.* **2016**, *8*, 941.
- (20) Liu, C.; Zhang, P.; Beratan, D. N. Coarse-grained theory of biological charge transfer with spatially and temporally correlated noise. *J. Phys. Chem. B* **2016**, *120*, 3624–3633.
- (21) Zhang, Y.; Young, R. M.; Thazhathveetil, A. K.; Singh, A. P.; Liu, C.; Berlin, Y. A.; Grozema, F. C.; Lewis, F. D.; Ratner, M. A.; Renaud, N.; et al. Conformationally gated charge transfer in DNA three-way junctions. *J. Phys. Chem. Lett.* **2015**, *6*, 2434–2438.
- (22) Grozema, F. C.; Tonzani, S.; Berlin, Y. A.; Schatz, G. C.; Siebbeles, L. D.; Ratner, M. A. Effect of structural dynamics on charge transfer in DNA hairpins. *J. Am. Chem. Soc.* **2008**, *130*, 5157–5166.
- (23) Cramer, T.; Krapf, S.; Koslowski, T. DNA charge transfer in an external field: an atomistic approach. *J. Phys. Chem. C* **2007**, *111*, 8105–8110.
- (24) Gutierrez, R.; Caetano, R.; Woiczikowski, B.; Kubar, T.; Elstner, M.; Cuniberti, G. Charge transport through biomolecular wires in a solvent: bridging molecular dynamics and model Hamiltonian approaches. *Phys. Rev. Lett.* **2009**, *102*, 208102.
- (25) Troisi, A.; Orlandi, G. Hole migration in DNA: a theoretical analysis of the role of structural fluctuations. *J. Phys. Chem. B* **2002**, *106*, 2093–2101.
- (26) Young, R. M.; Singh, A. P. N.; Thazhathveetil, A. K.; Cho, V. Y.; Zhang, Y.; Renaud, N.; Grozema, F. C.; Beratan, D. N.; Ratner, M. A.; Schatz, G. C.; et al. Charge Transport Across DNA-based Three-Way Junctions. *J. Am. Chem. Soc.* **2015**, *137*, 5113–5122.
- (27) Renger, T.; Marcus, R. Variable-range hopping electron transfer through disordered bridge states: Application to DNA. *J. Phys. Chem. A* **2003**, *107*, 8404–8419.
- (28) Kubar, T.; Elstner, M. What governs the charge transfer in DNA? The role of DNA conformation and environment. *J. Phys. Chem. B* **2008**, *112*, 8788–8798.
- (29) Henderson, P. T.; Jones, D.; Hampikian, G.; Kan, Y.; Schuster, G. B. Long-distance charge transport in duplex DNA: the phonon-assisted polaron-like hopping mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8353–8358.
- (30) Conwell, E. M.; Rakhmanova, S. V. Polarons in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 4556–4560.
- (31) Sugiyama, H.; Saito, I. Theoretical studies of GG-specific photocleavage of DNA via electron transfer: significant lowering of ionization potential and S'-localization of HOMO of stacked GG bases in B-Form DNA. *J. Am. Chem. Soc.* **1996**, *118*, 7063–7068.
- (32) Mantz, Y. A.; Gervasio, F. L.; Laino, T.; Parrinello, M. Solvent effects on charge spatial extent in DNA and implications for transfer. *Phys. Rev. Lett.* **2007**, *99*, 058104.
- (33) Kravec, S.; Kinz-Thompson, C.; Conwell, E. Localization of a Hole on an Adenine-Thymine Radical Cation in B-Form DNA in Water. *J. Phys. Chem. B* **2011**, *115*, 6166–6171.
- (34) Kubar, T.; Elstner, M. Coarse-grained time-dependent density functional simulation of charge transfer in complex systems: application to hole transfer in DNA. *J. Phys. Chem. B* **2010**, *114*, 11221–11240.
- (35) Renaud, N.; Harris, M. A.; Singh, A. P.; Berlin, Y. A.; Ratner, M. A.; Wasielewski, M. R.; Lewis, F. D.; Grozema, F. C. Deep-hole transfer leads to ultrafast charge migration in DNA hairpins. *Nat. Chem.* **2016**, *8*, 1015–1021.
- (36) Conwell, E. M.; Basko, D. M. Hole traps in DNA. *J. Am. Chem. Soc.* **2001**, *123*, 11441–11445.
- (37) Senthilkumar, K.; Grozema, F. C.; Guerra, C. F.; Bickelhaupt, F. M.; Siebbeles, L. D. A. Mapping the sites for selective oxidation of guanines in DNA. *J. Am. Chem. Soc.* **2003**, *125*, 13658–13659.
- (38) Voityuk, A. A. Are radical cation states delocalized over GG and GGG hole traps in DNA? *J. Phys. Chem. B* **2005**, *109*, 10793–10796.
- (39) Kurnikov, I.; Tong, G.; Madrid, M.; Beratan, D. Hole size and energetics in double helical DNA: competition between quantum delocalization and solvation localization. *J. Phys. Chem. B* **2002**, *106*, 7–10.
- (40) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964**, *136*, B864.
- (41) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965**, *140*, A1133.
- (42) Levy, M. Universal variational functionals of electron densities, first-order density matrices, and natural spin-orbitals and solution of the v-representability problem. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 6062–6065.
- (43) Parr, R. G.; Yang, W. *Density-functional theory of atoms and molecules*; Oxford University Press: 1989.
- (44) Gross, E. K.; Dreizler, R. M. *Density functional theory*; Springer Science & Business Media: 2013.
- (45) Mori-Sánchez, P.; Cohen, A. J.; Yang, W. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.* **2008**, *100*, 146401.
- (46) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into current limitations of density functional theory. *Science* **2008**, *321*, 792–794.
- (47) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for density functional theory. *Chem. Rev.* **2012**, *112*, 289–320.
- (48) Jaramillo, J.; Scuseria, G. E.; Ernzerhof, M. Local hybrid functionals. *J. Chem. Phys.* **2003**, *118*, 1068–1073.
- (49) Arbuznikov, A. V.; Kaupp, M. Local hybrid exchange-correlation functionals based on the dimensionless density gradient. *Chem. Phys. Lett.* **2007**, *440*, 160–168.
- (50) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
- (51) Neese, F.; Schwabe, T.; Grimme, S. Analytic derivatives for perturbatively corrected double hybrid density functionals: theory, implementation, and applications. *J. Chem. Phys.* **2007**, *126*, 124115.
- (52) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.
- (53) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51–57.
- (54) Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*, 234109.
- (55) Borghi, G.; Ferretti, A.; Nguyen, N. L.; Dabo, I.; Marzari, N. Koopmans-compliant functionals and their performance against reference molecular data. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 075135.
- (56) Baer, R.; Livshits, E.; Salzner, U. Tuned range-separated hybrids in density functional theory. *Annu. Rev. Phys. Chem.* **2010**, *61*, 85–109.
- (57) Kronik, L.; Stein, T.; Refaely-Abramson, S.; Baer, R. Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals. *J. Chem. Theory Comput.* **2012**, *8*, 1515–1531.
- (58) Autschbach, J.; Srebro, M. Delocalization error and functional tuning in Kohn–Sham calculations of molecular properties. *Acc. Chem. Res.* **2014**, *47*, 2592–2602.
- (59) Li, C.; Zheng, X.; Su, N. Q.; Yang, W. Localized orbital scaling correction for systematic elimination of delocalization error in density functional approximations. *Natl. Sci. Rev.* **2018**, *5*, 203–215.
- (60) Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L., Jr. Density-functional theory for fractional particle number: derivative discontinuities of the energy. *Phys. Rev. Lett.* **1982**, *49*, 1691.
- (61) Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Fractional charge perspective on the band gap in density-functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 115123.

- (62) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
- (63) Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* **2012**, *5*, 9.
- (64) Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science* **1995**, *268*, 1144–1149.
- (65) Friedrichs, M.; Zhou, R.; Edinger, S. R.; Friesner, R. A. Poisson-Boltzmann analytical gradients for molecular modeling calculations. *J. Phys. Chem. B* **1999**, *103*, 3057–3061.
- (66) Shen, J.; Wendoloski, J. Electrostatic binding energy calculation using the finite difference solution to the linearized Poisson-Boltzmann equation: Assessment of its accuracy. *J. Comput. Chem.* **1996**, *17*, 350–357.
- (67) Schrödinger, LLC. *The PyMOL Molecular Graphics System*, version 2.3; 2019.
- (68) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- (69) Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A.; et al. Parmbsc1: a refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55.
- (70) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
- (71) An in-house program for QM/MM simulations. Available at www.qm4d.info. Accessed July 31, 2018.
- (72) Lias, S. G.; Levin, R. D.; Kafafi, S. A. *NIST Chemistry WebBook: Ion Energetics Data*; National Institute of Standards and Technology: 2019.
- (73) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. Communication: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method [DLPNO-CCSD (T)]. *J. Chem. Phys.* **2018**, *148*, 011101.
- (74) Voityuk, A. A.; Siri Wong, K.; Rösch, N. Environmental Fluctuations Facilitate Electron-Hole Transfer from Guanine to Adenine in DNA π Stacks. *Angew. Chem., Int. Ed.* **2004**, *43*, 624–627.
- (75) Voityuk, A. A. Electronic couplings and on-site energies for hole transfer in DNA: Systematic quantum mechanical/molecular dynamic study. *J. Chem. Phys.* **2008**, *128*, 115101.
- (76) Giese, B.; Spichy, M. Long distance charge transport through DNA: Quantification and extension of the hopping model. *ChemPhysChem* **2000**, *1*, 195–198.
- (77) Lewis, F. D.; Zhu, H.; Daublain, P.; Cohen, B.; Wasielewski, M. R. Hole mobility in DNA A tracts. *Angew. Chem., Int. Ed.* **2006**, *45*, 7982–7985.
- (78) Ru, X.; Zhang, P.; Beratan, D. N. Assessing Possible Mechanisms of Micrometer-Scale Electron Transfer in Heme-Free *Geobacter sulfurreducens* Pili. *J. Phys. Chem. B* **2019**, *123*, 5035–5047.