A Symphony of Charge Transfer Theory, Conductive DNA Junction Modeling
and Chemical Library Design

by

Yuqi Zhang

Department of Chemistry
Duke University

Date:__________________________
Approved:

___________________________
David Beratan, Supervisor

___________________________
Katherine Franz

___________________________
Jie Liu

___________________________
Weitao Yang

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2016
ABSTRACT

A Symphony of Charge Transfer Theory, Conductive DNA Junction Modeling and Chemical Library Design

by

Yuqi Zhang

Department of Chemistry
Duke University

Date: _________________________

Approved:

___________________________
David Beratan, Supervisor

___________________________
Katherine Franz

___________________________
Jie Liu

___________________________
Weitao Yang

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the Graduate School of Duke University

2016
Abstract

Biological electron transfer (ET) reactions are typically described in the framework of coherent two-state electron tunneling or multi-step hopping. Yet, these ET reactions may involve multiple redox cofactors in van der Waals contact with each other and with vibronic broadenings on the same scale as the energy gaps among the species. In this regime, fluctuations of the molecule and its medium can produce transient energy level matching among multiple electronic states. This transient degeneracy, or flickering electronic resonance among states, is found to support coherent (ballistic) charge transfer. Importantly, ET rates arising from a flickering resonance (FR) mechanism will decay exponentially with distance because the probability of energy matching multiple states is multiplicative. The distance dependence of FR transport thus mimics the exponential decay that is usually associated with electron tunneling, although FR transport involves real carrier population on the bridge and is not a tunneling phenomenon. Likely candidates for FR transport are macromolecules with ET groups in van der Waals contact: DNA, bacterial nanowires, multi-heme proteins, strongly coupled porphyrin arrays, and proteins with closely packed redox-active residues. The theory developed here is used to analyze DNA charge-transfer kinetics, and we find that charge transfer distances up to 3-4 bases may be accounted for with this mechanism. Thus, the observed rapid (exponential) distance
dependence of DNA ET rates over distances of $\lesssim 10$ Å does not necessarily prove a tunneling mechanism.

Molecular structures that direct charge transport in two or three dimensions could help to enable the development of molecule-based electrical switches and gates. As a step toward this goal, we use theory, modeling and simulation to explore DNA three-way junctions (TWJs). Molecular dynamics (MD) simulations and quantum calculations, indicate that DNA TWJs undergo dynamic interconversion among “well stacked” conformations on the time scale of nanoseconds, a feature that makes the junctions very different from linear DNA duplexes. The studies further indicate that this conformational gating would control charge flow through these TWJs, distinguishing them from conventional (larger size scale) gated devices. Simulations also find that structures with polyethylene glycol (PEG) linking groups (“extenders”) lock conformations that favor CT for 25 ns or more. The simulations explain the kinetics observed experimentally in TWJs and rationalize their transport properties compared to double-stranded DNA. Furthermore, we redesigned DNA TWJs that have equally coupled output pathways for charge. The TWJ was also designed to switch between the two conductive states in responsive to an applied electric field.

Computationally aided drug discovery confronts the problem to balance performance and computation cost. Earlier study reveals that a less-expensive docking approach is not reliable to estimate the protein-ligand affinity in exploring drug
candidates targeting CARM1 (coactivator-associated arginine methyltransferase 1).

However, more accurate binding free energy calculation based on molecular dynamics sampling is not affordable in a high throughput screening. A truncated MD method was developed and can be used to estimate the binding free energy with similar accuracy with the full-system MD methods, while reducing the computation cost ten fold. Thus, this truncated MD method is feasible in a high throughput screening towards drug discovery.
Contents

Abstract ............................................................................................................................................. iv

List of Tables .................................................................................................................................... x

List of Figures .................................................................................................................................... xi

Acknowledgements ......................................................................................................................... xvi

1. Introduction .................................................................................................................................... 1

  1.1 Charge Transfer in DNA and Flickering Resonance ............................................................. 1

  1.2 DNA Nanostructure and DNA Electronics ........................................................................... 3

  1.3 Diversity Oriented Chemical Library Exploration for Drug Candidates ....................... 6

  1.4 List of Other Projects ............................................................................................................... 7

2. Biological Charge Transfer via Flickering Resonance ............................................................. 8

  2.1 Introduction .............................................................................................................................. 8

  2.2 Electron-Transfer Kinetics ...................................................................................................... 9

    2.2.1 2-state Electron Transfer ............................................................................................... 10

    2.2.2 A Generalized ET Rate for Multistate Resonance ..................................................... 13

  2.3 The Multisite Matching Probability ...................................................................................... 17

  2.4 Correlated vs Uncorrelated Site Energy Fluctuations ...................................................... 20

  2.5 The FR Model for DNA Charge Transport Kinetics ......................................................... 21

    2.5.1 Limitations of Current Models ....................................................................................... 21

    2.5.2 Flickering Resonance Descriptions of Charge Shift and Charge Separation in DNA ................................................................................................................................. 22

    2.5.3 Charge Shift in the FR framework .................................................................................. 25
4.4 DNA TWJ as a Charge Switch, a Field-Effect-Transistor.................................65
4.5 Conclusions and Prospectives ........................................................................71
5. Diversity Oriented Chemical Space Exploration towards Drug Discovery ..........73
  5.1 Diversity Oriented Chemical Space Search Algorithm ....................................73
  5.2 Drug Discovery Targeting CARM1 .................................................................74
  5.2.1 CARM1 plays important roles in gene regulation .......................................74
  5.2.2 Existing Libraries Targeting CARM1 .........................................................75
  5.2.3 Benchmark Test with Full-system MD Sampling .........................................75
  5.2.4 Protocols to Construct a Truncated System ...............................................79
  5.2.5 Binding Free Energies from Truncated MD is Similar to Full-system MD ......83
  5.2.6 Apply Truncated-MD to Other Protein Systems ..........................................84
  5.2.7 Truncated-MD is Easier to Converge .........................................................84
  5.2.8 Comparison of Computation Cost of Full-system MD and Truncated MD .....86
  5.2.9 Variations and Limitations of Truncated MD ...............................................87
  5.2.10 Conclusions ...............................................................................................88
References .............................................................................................................89
Biography ............................................................................................................104
List of Tables

Table 1: Computed $P_{\text{match}}$, experimental yield ratios, and fitted $\tau k_r$ values (with $k_r$ the water-trapping rate) for data of Giese et al.\textsuperscript{21} (see Figure 9 for kinetic scheme)..........................29

Table 2: Computed $P_{\text{match}}$, experimental hole arrival rates of Lewis et al.\textsuperscript{20}, and fitted $\tau$ values................................................................................................................................................32

Table 3: Quantum yields and CT rates measured experimentally for structures 1a, 1b, 1c and 1d. Data collected by Northwestern University. ........................................................................................................................................53
List of Figures

Figure 1: An overview of the flickering resonance channel and the rate expression............1

Figure 2: Design of DNA switch.................................................................4

Figure 3: The formation of a flickering resonance channel.................................9

Figure 4: The multi-state matching probability of a four-site system with nearest neighbor couplings fixed at $|V| = 10^{-3}$, $10^{-2}$, and $10^{-1}$ eV at room temperature, $\epsilon = 0.16$ eV, and equal mean energies for each site. In this analysis, the nearest-neighbor couplings are set to be constant, i.e., $PV(V)$ in Eq. 6 is a delta function peaked at $10^{-3}$, $10^{-2}$, and $10^{-1}$ eV, respectively. Note that the matching probability grows as $|V|$ grows.........................14

Figure 5: Bridge site matching probabilities (eq. (7)) decay exponentially for both the homogeneous chains (solid line) and the inhomogeneous chains (dotted line). The exponential decay constants for the matching probabilities are $0.6 – 0.7$ Å$^{-1}$ with $\sigma_E = 0.6$ eV and $V_{rms} = 0.1$ eV. Inhomogeneous chains have a smaller matching probability than homogeneous chains because $\Delta E_B = 1$ eV. Zero to three intervening bases cover a DA distance range of $3.4 - 13.6$ Å.........................................................18

Figure 6: The distribution of nearest-neighbor couplings between G and A is shown. 4,000 snapshots were sampled from a 4ns MD trajectory of B-DNA (5'-GATC-3':3'-CTAG-5'). Couplings were computed using a block-diagonalization method. The four base-pair sequence geometries were taken from the central segment of DNA from an MD simulation of a 12 base pair sequence.23 ..................................................................................................24

Figure 7: The design of Giese’s experiment. A radical cation is established on the sugar group and then charge is injected (step 2-3) onto G. Charge transfer (steps 3-4) leads to GGG oxidation. [Reproduced with permission from ref. 21 (Copyright 2001, Nature Publishing Group).] .................................................................................................26

Figure 8: Energetics of the G(A)$_n$GGG sequence. The offsets of the Gaussian peaks for a charge shift reaction are calculated using the free energy difference between G and A of 0.35 eV, $\Delta G = -0.7$ eV, and 1.2 eV (4 base pairs between D and A) $\geq \lambda \geq 0.4$ eV (nearest neighbor D and A). ..................................................................................................................27

Figure 9: Kinetic model for Giese’s experiment.........................................................28
Figure 10: Structures of (a) stilbene donor and acceptor and (b) the donor-acceptor capped DNA hairpins.\textsuperscript{20} Figure adapted from ref. \textsuperscript{20} .......................................................... 30

Figure 11: Energetics of the stilbene-capped DNA hairpins (shown schematically). The Coulomb ramp arises from (screened) electron and hole attraction. The mean values for the spectral function at each site are adopted from the analysis of Grozema \textit{et al.}\textsuperscript{22,30} and the standard deviations of these distribution functions are set to 0.6eV. The nearest-neighbor site distance was set to 3.4 Å. The increasing energy mismatch among bases for longer chains disfavors the FR mechanism at longer distances................................................................................. 30

Figure 12: Structures of DNA TWJs. (1a,1b) DNA TWJs with one extended guanine pathway and no polyethylene glycol (PEG) linking group. (1c,1d) DNA TWJs with one extended guanine pathway with one PEG linking group. (1e) A proposed DNA TWJ with guanine pathways on all three arms that may serve as a charge splitter/combiner. Structures 1a, 1b, 1c, and 1d adapted from \textsuperscript{83} were studied experimentally.......................... 40

Figure 13: Illustration of conformational switching between “T” and “Y” forms. The dotted lines represent PEG linkers.................................................................................................................. 41

Figure 14: Ratio of Sd\textsuperscript{•+} and Sa\textsuperscript{•−} absorptions at 532 nm and 572 nm for 3WJs 1a, 1b, 1c and 1d, and hairpins have the same purine pathway with TWJ 1c and 1d respectively. An exponential fit is shown with solid lines. Data from the first 50 ps are suppressed for clarity. See reference \textsuperscript{88} for additional experimental details. Data collected by Northwestern University.................................................................................................................. 45

Figure 15: (a) Structure of a DNA TWJ (1a) (adapted from\textsuperscript{83}). (b) Typical snapshot of a T-shape conformation. Backbones, linking group, and hydrogens are removed for clarity. (c) Upper: Coupling between Gs at the junctions (shown in blue oval in (a)). Once the π-stacked geometries form, the cross-junction coupling is enhanced. Bottom: Time evolution of the angles between the three DNA branches is shown with colors as in (a). .......................................................................................................................... 47

Figure 16: (Top) The unstable G-G base pair creates a new CT pathway. A detailed analysis of the new CT pathway in structure (1b). (Middle) Time evolution of the “cross junction” couplings among the guanines at the junction. The scale for the weak couplings (green and dark blue) is on the right-side vertical axis. (Bottom) Typical snapshots in different phases of the simulation. During the first 2 ns of the trajectory, all of the bases remain paired. The nearest-neighbor base coupling between G1 and G2 (yellow) in the bottom branch is the strongest. In phase II (~2 ns to ~9.7 ns), the mismatched G2-G3 base pairing breaks, forming a strong interacting π-stack with
another guanine G1 from the lower layer. Then, in phase III, fluctuations bring the guanine from the bottom branch and the right branch into close proximity, enhancing the coupling. During this time period, all of the couplings involved in the CT pathway, namely couplings between G1 and G3 (purple), G2 and G3 (light blue), and G2 and G4 (green), are considerable. Thus, a new CT pathway from G1 → G3 → G2 → G4 is activated to cross the junction. Finally, in phase IV, the cross-junction interaction is weak, and the G2-G3 base pairing reappears.

Figure 17: Time evolution of the “cross-junction” coupling of 1c. The RMS “cross junction” coupling between the two guanines at the junction (shown in the blue shadow oval) is $5.4 \times 10^{-3}$ eV during the first 0.6 ns, and increases to 0.090 eV.

Figure 18: (a) Structure of a DNA TWJ (1d) with a PEG linker (X) (adapted from\textsuperscript{83}). (b) Typical snapshot of a T-shape conformation. Backbones, linking group, and hydrogens are removed for clarity. (c) Upper: Coupling between Gs at the junctions (labeling as in (a)). Once the π-stacked geometries form, the cross-junction coupling is strongly enhanced. Bottom: Time evolution of the angles between the three DNA branches is shown with colors as in (a). This structure is locked in the T-shaped geometry after ~2 ns.

Figure 19: Typical MD snapshots representing two different “T” shaped conformations. These two snapshots are taken from different trajectories. (left) “T” shaped conformation facilitates transport along the upper-left pathway. (right) “T” shape conformation facilitates transport along the upper-right pathway.

Figure 20: Time evolution of the couplings at the junctions that are involved in the CT pathways for structure 1e. After 2 ns, the well stacked “T” shaped structure forms, and the “cross-junction” coupling is enhanced by more than 2 orders of magnitude.

Figure 21: Previously studied DNA TWJ designs of Lewis and co-workers. TWJ 1a establishes a purine pathway from the lower branch to the upper left branch, and TWJ 1b has a purine pathway that consists of a lower branch and an upper-right branch. TWJ 1c is a charge splitter design. Structures are adapted from Ref.\textsuperscript{83, 117}.

Figure 22: (a) Structure of the designed DNA TWJ charge splitter. (b) A snapshot from an MD trajectory that enables a CT pathway between A and B. (c) A snapshot from another MD trajectory with a structure that facilitates CT between A and C. After base-stacking is established, cross-junction coupling strength increase to match the nearest-neighbor couplings in B-DNA in the enabled pathways. Note that snapshots (b) and (c) come from two different MD trajectories.
Figure 23: Design of a DNA TWJ FET. The “Y” shaped structure (top) folds into two stacked “T” shaped conformations depending on the electric-field orientation.

Figure 24: Design of DNA TWJ FET with 3 PEG linkers.

Figure 25: (a) Structure of the FET TWJ design. (b) Typical snapshots of an unfolded conformation (left, snapshot taken at time 0) and two folded conformations (center and right). (c) Time evolution of the two cross-junction couplings between G1-G3 and G1-G2. The color bar at the top indicates the direction of the applied electric field. Red represents the time window with an electric field pushing the TWJ toward the bottom-right corner, blue represents the time window with an electric field pushing the TWJ toward the bottom-left corner, and grey indicates the time window with no electric field present. (d) Time evolution of the angles between the branches of the TWJ. The colors correspond to the labels in panel a.

Figure 26: Experimental design of a TWJ-FET. (1) A monolayer of single-stranded DNA (ssDNA) is attached to a gold surface. The designed TWJ has two ssDNA sticky ends that have complementary sequences with the ssDNA monolayer on the gold surface. Thus, the DNA TWJ can “stand up” on the gold surface. (2) A break junction tip is used to pull up the third branch of the DNA TWJ. (3) Due to the pulling force from the tip, the DNA TWJ folds into a “T” shape and establishes a high-conductance. (4) After applying the electric-field gate from another source (or simply changing the voltage across the molecule), the DNA TWJ folds into a different “T” shape that produces a low conductance.

Figure 27: Bind free energies for (Purple square) Sinefungin (SFG); (Green circles) Experimentally active binders (BI); (Yellow circles) Experimentally inactive binders (BI); (Red triangles) Docking predicted ligands.

Figure 28: Comparison of binding free energies from AMBER (Blue) and CHARMM (Red) force field, and binding affinity (with secondary axis) from Docking (Green).

Figure 29: Demonstration of a two-layer model for truncated MD. The docked ligand is shown in a vDW representation. The interacting shell is shown in a cpk representation and colored in yellow. The constraining layer is shown in bond representation and colored in blue.

Figure 30: Time evolutions for the RMSD of the selected residues. (Blue) full system MD; (Other colors) truncated MD with different constraint strength (yellow is the weakest...
and red is the strongest). In this case, the black line matches the full-system fluctuation.

Figure 31: Comparison of binding free energies with full-system MD (Blue), truncated MD (Red), and a linear scaling of the Truncated MD (yellow).

Figure 32: Comparison of binding free energy from full-system MD and truncated MD for AchE, 1HPV and TryP.

Figure 33: Comparison of time evolution of binding free energy with Truncated MD (blue) and full-system MD (red). Results obtained from simulations on AchE.

Figure 34: Comparison of computational cost for CARM1, TryP, AchE and 1HPV.
Acknowledgements

I’d like to thank my advisor Professor David Beratan for his encouragement, patience and generous help along this unusual journey. I’d also like to thank my committee members, Prof. Weitao Yang, Prof. Jie Liu, and Prof. Katherine Franz for many constructive discussions.

I am honored to work with a lot of brilliant scientists at Duke and also from other institutes around the world including, Prof. Alexander Balaeff, Lingyan Du, Brittany Morgan, Dr. Chetan Rupakheti, Prof. Spiros Skourtis, Prof. Qiu Wang, Dr. Xing Yin and many others. I want to give my special gratitude to Chaoren Liu and Prof. Peng Zhang, who help me in every aspect of life.

I am lucky to meet and grow with many talented and strong-minded friends at Duke. Yes, I am talking about you. Thank you for sharing your life with me.

Nevertheless, my parents Zhiyong Zhang and Liling Zhang are always supportive no matter what I am doing and what I am going to do. Finally, I want to blame my girlfriend, Mengen Huang, who should have shown up earlier in my life. My life is full of happiness and joy with you.
1. Introduction

During my graduate study, I have worked on a wide range of research topics, including charge transfer theory, DNA nanostructure designs, cheminformatics and structure-based ligand design.

The aim of this chapter is to provide (1) overviews of major projects and (2) summary of other exciting projects that are not explicitly described in this dissertation.

1.1 Charge Transfer in DNA and Flickering Resonance

Charge transport through DNA plays a central role in nucleic acid damage and repair, and it is usually modeled with a bridge-mediated tunneling mechanism (at short distances) and a multi-step hopping mechanism (at longer distances). We found that fluctuations into transient geometries that bring multiple bases into electronic degeneracy may support band-like transport, evoking a new flickering resonance (FR) mechanism (Figure 1).

$$k = \frac{1}{t_A} \times P_{match}$$

Figure 1: An overview of the flickering resonance channel and the rate expression.
In bridge mediated charge transport systems, the site energies of donor (D), bridge (B) and acceptor (A) fluctuate, and we found that these fluctuations are Gaussian distributed. We formulated the charge transfer rate as a product of the inverse of the transit time ($t_A$) through the degenerate channel and the probability ($P_{match}$) to bring the fluctuating site energies into transient degeneracy.

The matching probability ($P_{match}$) is computed by modeling the site energy distribution functions. For specific systems, we also sample the matching probability from the site energy ensemble generated using molecular dynamics and quantum methods. The transit time ($t_A$) describes ballistic-like transport through the flickering resonance channel. The rates from the FR mechanism match with experimental data for DNA charge transport.

We have found that the FR mechanism potentially plays an important role in condensed phase charge transfer systems including DNA, multi-heme proteins, bacterial nanowires and synthetic nanowires. In these systems, the energy differences, the thermal fluctuation and the electronic couplings are on the same order of magnitude, making the matching probabilities large.

Using the FR analysis, we designed and engineered a unique DNA sequence with alternating guanine-blocks that can enhance and preserve the coherent transport channel to nanometer length scales. This design was demonstrated experimentally in the group of Nongjian Tao at Arizona State University.
Publications:


1.2 DNA Nanostructure and DNA Electronics

DNA is a strong candidate for designing conductive bio-materials. Past research in conductive DNA has focused mainly on moving charge in one dimension. We have taken on the challenge of designing multi-dimensional functional DNA junctions towards establishing charge splitters and switches at the single-molecule level.

DNA three-way junctions (TWJ) are motifs that might be used to split current flowing through nucleobase stacks. Our experimental collaborators found that charge transfer in DNA TWJs is controlled by a stochastic process. With molecular dynamics (MD) simulations, we discovered that the charge transfer across the DNA three-way junctions is gated by stochastic large-scale conformational change, giving rise to the stochastic kinetic found experimentally.

Unlike charge transport in well-stacked canonical DNA duplexes, the hopping step required to cross the junction is often the conductivity bottleneck. We used semi-empirical quantum mechanical methods to calculate the electronic couplings for this
“cross-junction” step using the conformational ensemble generated from MD simulations. We found that charge can migrate across the junction only when the nucleobases near the junction fold into a canonical B-DNA like stack, suggesting a conformationally gated charge transport mechanism. To overcome this conformational-gating bottleneck, we designed a structural “extender” using a polyethylene glycol group (dashed line in the figure below) in the junction stabilizing the junction. This stabilized stacking at the junction enhanced the charge transfer efficiency, and thus we demonstrated that DNA TWJs can be used as charge splitters.

![Figure 2: Design of DNA switch.](image)

We further designed a DNA TWJ based charge switch that is responsive to applied electric fields. With a negatively charged backbone, DNA conformations can be manipulated by applied electric fields. In our design, we anchored the three termini of
the DNA TWJ to avoid the translational motion. We tuned the orientation and strength of the electric field to realize a field-induced two-state switch between the two “folded” conformations, thus enabling/disabling the AB or AC conductive pathways (Figure 2).

Publications:


Zhang, Y.; Liu, C.; Beratan, D. N. A field-effect transistor based on DNA three way junctions. Submitted to *J. Chem. Phys.* special edition on "Molecular Electronics".


1.3 Diversity Oriented Chemical Library Exploration for Drug Candidates

Many strategies to develop improved drugs and functional materials focus on a narrow “chemical space” that are closely related to known targets, despite the astronomical number of possibilities (over $10^{60}$ compounds). Based on a diversity-oriented evolutionary chemical library search algorithm (ACSESS) developed at Duke, we expanded the framework to search for Ir-containing OLED and drug candidates that target the coactivator-associated arginine methyltransferase 1 (CARM1).

CARM1 is a drug target that plays an important role in gene transcription. Typical docking methods are not sufficiently reliable to evaluate the binding affinity of candidate structures and conventional MD sampling methods are not affordable to screen a large number of ligands generated from the ACSESS algorithm. Thus, we developed a MD strategy to estimate the binding free energy with improved accuracy while reducing the computation cost ten fold.

In most protein-ligand binding scenarios, only a small fraction of protein residues (~30 in CARM1) interacts with the ligand, forming the core of the protein-ligand interaction zone. The other parts of the protein often act largely as a structural template that affects the fluctuation of the interacting core. Following this idea, we developed a strategy to create a well-crafted constraining layer of residues surrounding the interaction zone. Residues in the constraining layer are described in a manner that
aims to mimic the structural fluctuation of the full system. The simulation of this truncated system generates a conformational ensemble that can be used to calculate protein-ligand binding free energies. With this truncated simulation system, we can accelerate the binding free energy calculations more than 10 fold while still arriving at realistic binding free energies.

Manuscripts in preparation:

Lower cost sampling method for binding free energy calculation.

Computationally-assisted development of a diverse molecular library of blue emitting Ir-containing OLEDs.

### 1.4 List of Other Projects

(1) Exploration of Ir-containing OLED material with ACSESS, in collaboration with Prof. **Weitao Yang** (Duke); (2) Investigating charge transport in novel peptide nucleic acid structures, in collaboration with Prof. **David Waldeck** (University of Pittsburg) and Prof. **Catalina Achim** (Carnegie Mellon University); (3) Identification of key chemical and spatial properties for selective RNA-binding small molecules, in collaboration with Prof. **Amanda Hargrove** (Duke); and (4) Analyzing bivalent binding systems probed with AFM, in collaboration with Prof. **Eric Toone** (Duke).
2. Biological Charge Transfer via Flickering Resonance

Expanded from:


Y.Z., C.L., A.B., S.S.S., D.N.B. designed research, performed research, analyzed data and wrote the text.

2.1 Introduction

Chemical structure and, importantly, structural fluctuations determine the mechanism and kinetics of charge transfer. Redox energy fluctuations are of particular significance when transport barrier heights and the fluctuations are of similar magnitude. Indeed, the sensitivity of biological ET rates to conformational fluctuations and consequent (transient) delocalization is the topic of intense interest. Resonant enhancement of biological ET rates is consistent with a growing body of physical and structural data found in: DNA electron transfer through stacked nucleobases, extended delocalized structures of bacterial photosynthesis (including the special pair, bridging chlorophyll and pheophytin), the polaronic states of oxidized porphyrin arrays up to 7 porphyrin diameters in spatial extent, micrometer scale bacterial nanowires, multi-heme oxidoreductases, amino acid side chains in ribonucleotide reductase,
engineered protein-based hopping-chains, and centimeter scale charge-transport chains in filamentous bacteria. Here, we describe a transient or flickering-resonance (FR) mechanism for ET. The FR mechanism arises when thermal fluctuations produce geometries that enable charge delocalization across the entire structure by bringing the donor (D), bridge (B), and acceptor (A) levels into energetic degeneracy as illustrated in Figure 3. An electron or hole that is spatially localized at a starting D may move ballistically (i.e., with near constant velocity) through these fleeting structures to A. Since the probability of distorting multiple sites into energetic degeneracy is multiplicative and decays approximately exponentially with the number of sites (i.e., with distance), this transport mechanism could be mistaken for electron tunneling because of its distance dependence.

![Figure 3: The formation of a flickering resonance channel.](image)

### 2.2 Electron-Transfer Kinetics

The analysis in this chapter concerns multi-site charge transfer. Conventional biological electron-transfer theory is based on two-state dynamics, with D and A
brought into electronic resonance by medium reorganization. At resonance, the electron propagates from D to A either by tunneling (through space or via a bridge, B) i.e., non-adiabatic ET, or by strong electronic mixing among D, B, and A (adiabatic ET). Here, we explore a limit where the intrinsic energy fluctuations and/or medium reorganization create multi-state electronic degeneracy and we demonstrate that charge transport in the degenerate state results in exponential distance dependence of the ET rate in both adiabatic and non-adiabatic regimes.

2.2.1 2-state Electron Transfer

To set the stage for the generalized flickering resonance (FR) mechanism, we explain this framework in the context of two-state D-to-A electron transfer where the D and A energies are modulated by independent harmonic fluctuations. In this case, the FR picture is the standard picture of ET, where transfer takes place in thermally-populated D-A resonances, within a redox energy matching window equal to the D-to-A coupling. To frame this discussion, we adopt the simplest kinetic model for two-state ET:

\[
\begin{align*}
D^- A & \xrightleftharpoons[k_{off}]{k_{on}} R_2^+ (D^- A) \xrightarrow{1/\tau_{trans}} R_2^{+\dagger} (DA^-) \xrightarrow{1/\tau_{trap}} DA^- \\
&
\end{align*}
\]  

(1)

The overall D-to-A electron or hole transfer rate is denoted \( k_{ET} \). \( D^- \) is the initially prepared state with an electron or hole on D in a state that is not resonant (R) with A. \( R_2^+ (D^- A) \) and \( R_2^{+\dagger} (DA^-) \) denote the sub-ensemble of structures with D and A in
two-state resonance and the electron on D or A, respectively. \( \text{DA}^- \) denotes the ET product with the electron or hole on the relaxed acceptor. \( k_{on} \) is the rate to reach two-state resonance \( R_2^\pm (\text{D}^- \text{A}) \), \( 1/\tau_{\text{trans}} \) is the inverse time for transfer at resonance, and \( 1/\tau_{\text{trap}} \) is the trapping rate (considered irreversible in the context of transition state theory). The overall ET rate in this kinetic scheme is:

\[
k_{\text{ET}} = (1/\tau_*) \exp(-\Delta G_{\text{act}}/K_B T)
\]

(2)

For nonadiabatic ET, \( \Delta G_{\text{act}} = \Delta G_{\text{act}}^{\text{NA}} = (\Delta G_{f,l}^{(0)} + \lambda)^2/4\lambda \) and \( 1/\tau_* = 2\pi V^2 / \hbar \sqrt{4\pi \lambda K_B T} \). For adiabatic ET, \( \Delta G_{\text{act}} = \Delta G_{\text{act}}^{\text{AD}} \) and, if ET is limited by \( k_{on} \), then \( 1/\tau_* = \omega \) is the reaction coordinate motion attempt frequency. If ET is limited by solvent relaxation and/or by trapping, \( 1/\tau_* = 1/\tau_{\text{trap}} \) (e.g., \( \tau_{\text{trap}} \approx \tau_L \) is the longitudinal Debye relaxation time). Approximate expressions can be derived for \( \Delta G_{\text{act}}^{\text{AD}} \). For example, when \( \Delta G_{f,l}^{(0)} = 0 \),

\[
\Delta G_{\text{act}}^{\text{AD}} = (\lambda/4) - |V|^{17}.
\]

In our FR reformulation of two-state (D-to-A) ET,

\[
k_{\text{ET}} = (1/\tau) P_{\text{match}}(2)
\]

(3)

where \( P_{\text{match}}(2) \) is the probability that the D and A energy levels differ by less than the D-A electronic coupling V. If we assume independent fluctuations energy levels for the D and A sites, as well as independent fluctuations of the energies and the coupling V, then

\[
P_{\text{match}}(2) = \int_{-\infty}^{+\infty} dV \rho_1(V) \int_{-\infty}^{+\infty} dE_1 \rho_1(E_1) \int_{E_1 - |V|}^{E_1 + |V|} dE_2 \rho_2(E_2),
\]

where \( \rho_1(E_1) \) and \( \rho_2(E_2) \) are the probability distribution functions for the electronic redox energies of the electron transfer active states of sites 1 and 2. This formulation extends Hopfield’s description of ET rates in terms of Gaussian broadened electron removal and
insertion “spectral” functions. In eq. (3), $1/\tau$ is generally different from $1/\tau_*$ because $P_{\text{match}}(2) \neq \exp[-\Delta G_{\text{act}}/K_B T]$. For example (see ref. 18 SI Sect. S4), in the approximation of constant coupling ($P_Y(V) = \delta(V - |V_0|)$), the exact two-state matching probability is:

$$P_{\text{match}}(2) = \left(\frac{\theta}{\sqrt{2\pi}}\right)\exp(-\Delta G_{f,i}^{(0)} + \lambda - V_*)^2/4\lambda K_B T)$$

(4)

where $= 2|V_0|/\sqrt{2\lambda K_B T}$, $\lambda = \lambda_D + \lambda_A$ is the total reorganization energy, and $V_* \in (\pm |V_0|)$. Eq. (4) is valid in both the adiabatic ($\theta \geq 1$) and nonadiabatic ($\theta \ll 1$) regimes. From eq. (4) we see that $\tau$ in eq. (3) is related to $\tau_*$ in eq. (2) by the exact relation $1/\tau_* = (\theta/\sqrt{2\pi}) \times 1/\tau$ in any regime. In the nonadiabatic regime we have $\theta \ll 1$ and $\tau \ll \tau_*$. Further, in the non-adiabatic regime, $P_{\text{match}}(2) \simeq P_{\text{match}}^{\text{NA}}(2)$, where

$$P_{\text{match}}^{\text{NA}}(2) = \left(\frac{\theta}{\sqrt{2\pi}}\right)\exp(-\Delta G_{\text{act}}^{\text{NA}}/K_B T)$$

and $\Delta G_{\text{act}}^{\text{NA}} = (\Delta G_{f,i}^{(0)} + \lambda)^2/4\lambda$ (see ref. 18 SI section S4). Thus, we can write the non-adiabatic rate either as $k_{ET}^{\text{NA}} = (1/\tau_*)\exp(-\Delta G_{\text{act}}^{\text{NA}}/K_B T)$ with $1/\tau_* = 2\pi V_0^2/h\sqrt{4\pi \lambda K_B T}$ or $k_{ET}^{\text{NA}} = (1/\tau)P_{\text{match}}^{\text{NA}}(2)$, where, using $1/\tau_* = (\theta/\sqrt{2\pi}) \times 1/\tau$, we find $\tau = \tau_{\text{Rabi}}/\pi^2$ ($\tau_{\text{Rabi}} \equiv h/2|V_0|$ is the Rabi time). Moreover, $P_{\text{match}}^{\text{NA}}(2) = 2|V_{DA}| \times FC$, where FC is the well-known Franck-Condon factor: $FC = \langle \delta(E_{\text{final}} - E_{\text{in}}) \rangle = (\sqrt{4\pi \lambda K_B T})^{-1}\exp(-\Delta G_{\text{act}}^{\text{NA}}/K_B T)$ (see ref. 18 SI section S4). In the adiabatic limit we have $\theta \approx 1$ and $\tau \leq \tau_*$. Importantly, when $\Delta G_{f,i}^{(0)} = 0$, $P_{\text{match}}(2) = (1/3)\exp(-\Delta G_{\text{act}}^{\text{AD}}/K_B T)$, where $\Delta G_{\text{act}}^{\text{AD}} = (\lambda/4) - |V|$ and $\tau = \tau_* /3$(see ref. 18 SI section S4). As such, we have two equivalent strategies to formulate the 2-state ET rate (eqs. (2) and (3)). An advantage of using the matching probability (eq. (3)) in formulating the ET rate is that it proves a computationally accessible means to estimate ET rates without first assuming a
limiting coupling regime (i.e. adiabatic or non-adiabatic), and it enables generalization to multi-state resonance.

### 2.2.2 A Generalized ET Rate for Multistate Resonance

For two-states, the FR rate is identical to two-state ET rates (adiabatic or nonadiabatic, depending on the magnitude of $V_{DA}$). For an N-state system, the FR mechanism defines a particular ET channel (others could well co-exist). The FR kinetic model and the rate associated with FR are formulated in analogy to the above approach with the kinetic scheme:

$$
\begin{align*}
&D^−B_{N−2}A \xrightleftharpoons[k_{on}]{k_{off}} R_N^{\dag}(D−B_{N−2}A) \xrightarrow{1/\tau_{trans}} R_N^{\dag}(DB_{N−2}A−) \xrightarrow{1/\tau_{trap}} DB_{N−2}A− \\
&\text{where } R_N^{\dag}(DB_{N−2}A) \text{ and } R_N^{\dag}(DB_{N−2}A−) \text{ denote the sub-ensemble of structures with all N units (including D and A) in degeneracy and the electron on D or A, respectively.}
\end{align*}
$$

$DB_{N−2}A−$ is the product with the trapped electron on A, and $P_{match}$ is the probability of matching all N states.

The FR channel ET rate is

$$
k_{ET}^{FR} = (1/\tau_*) \exp(-\Delta G_{act(N)}/K_B T) \quad \text{or} \quad k_{ET}^{FR} = (1/\tau)P_{match}(N)
$$

$G_{act(N)}$ is the activation free energy to reach N-state D-B-A resonance and $\tau_*$ is the rate-limiting time scale associated with the kinetics ($k_{on}^{-1}$, $\tau_{trans}$, or $\tau_{trap}$ for the N-state system). In eq. (5), $\tau_{trans}$ is interpreted as the D-to-A transport time for N-state D-
**B-A** resonance with a finite lifetime. The minimum value of $\tau_{\text{trans}}$ is obtained in the infinite FR lifetime limit, which gives $\tau_{\text{trans}}^{\text{min}} \sim \tau_{\text{trans}}^{\text{bal}} \sim 2R_{DA}|V_{nn}|/\hbar$ (using a mean carrier velocity $\langle v \rangle = (1/\hbar \left[ \partial E(k)/\partial k \right] )_{E_{\text{Fermi}}} = 2r_{nn}|V_{nn}|/\hbar$ where $r_{nn}$ is the nearest-neighbor distance). For a nearest-neighbor coupling of 0.1 eV and a 3.4 Å separation among neighbors, the carrier velocity is $\sim 1$ Å/fs which places $\tau_{\text{trans}}^{\text{min}}$ in eq. (5) in the range of tens of fsec for a few redox sites. The $\tau_{\text{trans}}$ step in eq. (5) places transient carrier population on the bridge independent of the process that ultimately limits the overall rate of the FR channel (eq (5)).

![Graph](image)

**Figure 4:** The multi-state matching probability of a four-site system with nearest neighbor couplings fixed at $|V| = 10^{-3}$, $10^{-2}$, and $10^{-1}$ eV at room temperature, $e = 0.16$ eV, and equal mean energies for each site. In this analysis, the nearest-neighbor couplings are set to be constant, i.e., $P_{V}(V)$ in Eq. 6 is a delta function peaked at $10^{-3}$, $10^{-2}$, and $10^{-1}$ eV, respectively. Note that the matching probability grows as $|V|$ grows.
As in the two-state limit, we could develop a multi-state description where

\[ k_{ET}^{FR} = (1/\tau_*) \exp(-\Delta G_{act(N)}/K_B T) \]. However, the \( P_{\text{match}}(N) \) formulation is particularly convenient in terms of computation and interpretation because of its explicit treatment of the influence of nearest-neighbor coupling on the energy matching window (see Figure 4). For an N-state system with uncorrelated redox-energy fluctuations, the probability of simultaneously matching all levels, summed over all matching energies, is:

\[
P_{\text{match}}(N) = \int_{-\infty}^{+\infty} dV_P(V) \int_{-\infty}^{+\infty} dE_1 \rho_1(E_1) \int_{E_1-|V|}^{E_1+|V|} dE_2 \rho_2(E_2) \cdots \int_{\min\{E_1, E_2, \ldots, E_{N-1}\}+|V|}^{\max\{E_1, E_2, \ldots, E_{N-1}\}-|V|} dE_N \rho_N(E_N)
\]

(7)

The upper limit of integration for the Nth site takes the minimum value of all site energies for sites 1 to N-1 plus |V|, and the lower limit of the Nth site integral takes the maximum value of all site energies for sites 1 to N-1 minus |V|. In this study, the matching probabilities are calculated using numerical integration for fewer than 6 sites and are computed using Monte Carlo methods for larger systems. We can derive exact expressions for \( P_{\text{match}}(N) \) or \( \Delta G_{act(N)} \) that are valid for arbitrary inter-site couplings and can also derive corresponding relations between \( \tau_* \) and \( \tau \) (see ref. 18 SI Sect. S4). In general, \( P_{\text{match}}(N) = f \times \exp(\Delta G_{act(N)}(\{E_i\}, \{V_{\text{rms}(i)}\}, \{\lambda_i\})/K_B T) \) where, approximately, \( \propto \theta_t^{N-1} \), with \( \theta_t = 2V_{\text{rms}(i)}/\sigma_{E(i)} \), where \( \sigma_{E(i)} = \sqrt{2\lambda_i K_B T} \) is the rms fluctuation of the i-th site redox energy, and \( V_{\text{rms}(i)} \) is the i-th nearest neighbor coupling. \( \theta_t \) is interpreted as a local (nearest-neighbor) adiabaticity parameter, \( \theta_t = \tau_{FC(i)}/\tau_{Rabi(i)} \). The Franck-Condon time \( \tau_{FC(i)} = \hbar/\sqrt{2\lambda_i K_B T} \) is the time associated with the persistence of nearest-
neighbor resonance and \( \tau_{Rabi(i)} = h/V_{\text{rms}(i)} \) is the corresponding Rabi time. Therefore, 
\( f \propto \theta_i^{N-1} \) is a generalized N-state-resonance adiabaticity parameter (see ref. 18 SI Sect S4).

Even if nearest-neighbor resonances are quasi-adiabatic with \( \theta_i \) slightly less than unity, the N state resonance becomes nonadiabatic as N grows since 
\( f \propto \theta_i^{N-1} \). Thus, the effective lifetime of the N-state resonance decreases with increasing N. In eq. (6), 
\( \tau / \tau_* \propto f \) and, in a generalized nonadiabatic limit defined by \( \ll 1 \), we have \( \tau_* \gg \tau \). If the overall rate is limited by carrier transmission through the N-state resonance, \( \tau_* \) in eq. (6) is interpreted as \( \tau_{\text{trans}} \) where \( \tau_{\text{trans}} \approx \tau_{\text{trans}}^{\text{bal}} / f \) and \( \tau_{\text{trans}}^{\text{bal}} \) is in the range of tens of fsec for a few redox sites. That is, the non-adiabatic electron transmission time through the FR bridge is the band transport time increased by a non-adiabatic factor associated with degeneracy breaking.

Below, we describe a simple analytical approximation to eq. 7 for a bridge with uncorrelated fluctuations, equal average site energies, and site-energy fluctuations larger than coupling energy fluctuations (nonadiabatic limit). Importantly: i) \( P_{\text{match}}(N) \) decays exponentially with N, ii) the distance dependence of \( P_{\text{match}}(N) \) coincides with the slope of the hole transfer rates and yields in the experiments of Lewis\textsuperscript{20} and Giese\textsuperscript{21}, for reasonable values of the hole transfer rate parameters, iii) fits to experimental ET rates and yields require \( \tau \approx 1 - 15 \text{ fs} \). In order to identify the rate-limiting step associated with the kinetic scheme of eq. (5), it is necessary to consider \( \tau_* \), which is longer than \( \tau \). Different possibilities for the rate limiting steps are discussed below.
2.3 The Multisite Matching Probability

As a starting point, we assume statistically-independent fluctuations for the multiple sites in the ET system. Standard deviations ($\sigma_E$, SD) for DNA base energies estimated from QM analysis of MD sampled structures indicate $\sigma_E \sim 0.2$ eV without solvent\textsuperscript{22-24}; inclusion of solvent interactions increases these values to $\sim 0.3$-0.5 eV\textsuperscript{25-26}.

Figure 5 shows multi-site energy-matching probabilities for a four-site linear system with Gaussian fluctuations of equal mean energies ($\sigma_E = 0.16$ eV), and with nearest-neighbor couplings fixed at $|V_{nn}| = 10^{-1}$ eV, $10^{-2}$ eV or $10^{-3}$ eV. The figure indicates that the 4-site energy matching probability drops several orders of magnitude for each order of magnitude decrease in $|V|$, because smaller values of $|V_{nn}|$ require closer site-energy matching to establish flickering resonance.

The probability of matching N fluctuating site energies will decrease in a multiplicative fashion, with one multiplier per site. The energy matching probability thus drops exponentially with distance. The data in Figure 5 shows the exponential decay of energy-level matching probabilities for a homogeneous chain (solid line) and an inhomogeneous chain (dashed line). For the homogeneous chain, all sites have the same mean energy values and $\sigma_E = 0.6$ eV; each nearest-neighbor coupling has zero mean and $V_{rms} = 0.1$ eV (the standard deviation of the coupling magnitudes). For the inhomogeneous chains, D and A have the same mean energy values, while the bridging site energies are offset by 1 eV. The distance between neighboring sites is assumed to be
3.4 Å. In the case where all of the sites have the same mean energy (solid line), the exponential decay constant for the matching probability as a function of distance is ~ 0.6 Å⁻¹. In the inhomogeneous case (dashed line), the exponential decay constant is ~ 0.7 Å⁻¹. Thus, the average decay exponent increases as the energy offsets grow among the sites. The matching probability is several-fold smaller for the inhomogeneous chains compared to the homogeneous chains, although the decay of the matching probabilities with distance is not very different (because the bridging spectral functions are energy matched to one another in both cases).

Figure 5: Bridge site matching probabilities (eq. (7)) decay exponentially for both the homogeneous chains (solid line) and the inhomogeneous chains (dotted line). The exponential decay constants for the matching probabilities are 0.6 – 0.7 Å⁻¹ with $\sigma_E = 0.6$ eV and $V_{rms} = 0.1$ eV. Inhomogeneous chains have a smaller matching probability than homogeneous chains because $\Delta E_B = 1$ eV. Zero to three intervening bases cover a DA distance range of 3.4 - 13.6 Å.
The simple exponential distance dependence of the energy matching probabilities found in Figure 5 is readily understood using an approximate model. Consider a chain with site energies that are described as independent Gaussian random variables with zero means and identical nonzero standard deviations, $\sigma_E$. The joint probability density function for the N site-redox-energies is

$$\rho(E_1) \times \rho(E_2) \times \ldots \times \rho(E_N)$$

where $\rho(E) = \left( \frac{1}{\sqrt{2\pi\sigma_E^2}} \right) e^{-E^2/2\sigma_E^2}$. An upper bound to the energy-matching probability is approximately $P_{\text{match}}(N) \approx \left( \int_0^{V_{\text{rms}}} dE \rho(E) \right)^N$. Since $V_{\text{rms}}/\sigma_E \approx 1/6$ in the example above, we can expand the exponential functions in terms of $(E/\sigma_E)^2$. Retaining the zeroth-order term, we find

$$P_{\text{match}}(N) \approx \left( \sqrt{\frac{2}{\pi}} \left[ \frac{V_{\text{rms}}}{\sigma_E} \right] \right)^N$$

or

$$P_{\text{match}}(N) \approx e^{-\Phi R}, \quad \Phi = \frac{1}{\Delta R} \ln \left[ \sqrt{\frac{2}{\pi}} \left( \frac{V_{\text{rms}}}{\sigma_E} \right) \right]$$

for a distance $\Delta R$ between adjacent sites (corrections to this result are found by retaining higher-order terms in $(E/\sigma_E)^2$). For the inhomogeneous case of Figure 5, a similar argument leads to $P_{\text{match}}(N) \approx e^{-E_B^2/2\sigma_E^2} e^{-\Phi R}$ where $E_B$ is the mean energy gap between the D/A states and the N-2 B sites. Using the values of Figure 5 (typical for DNA), i.e., $V_{\text{rms}}=0.1$ eV and $\sigma_E=0.6$ eV, we find that $\Phi \approx 0.6$ Å⁻¹ for a 3.4 Å spacing between bases. This is an excellent approximation to the numerical values found in Figure 5. For the inhomogeneous system in Figure 5, the decay of the D to A rate for the superexchange mechanism would be proportional to the square of the B-mediated D-A tunneling matrix element $V_{DA}^2 \propto e^{-\beta R_{DA}}$, where $\beta=(2/3.4)$ Å⁻¹ ln[(ΔE/B/V rms)] in the
perturbative limit. Substituting $V_{\text{rms}} = 0.1 \text{ eV}$ and $\Delta E_B = 1 \text{ eV}$ gives $\beta \approx 1.4 \text{ Å}^{-1}$. As such, for the parameters used in the illustration, the FR mechanism has much softer distance dependence than the superexchange rate. This simple analytical argument is strictly valid for $V_{\text{rms}}/\sigma_E < 1$ (the generalized nonadiabatic limit described above), and it can be used to assess the relative distance dependence of superexchange compared to FR as a function of $V_{\text{rms}}$, $\sigma_E$, and $\Delta E_B$. An approximately exponential distance decay of the matching probability is also observed in molecular dynamics simulations and is discussed below. This simple analysis thus captures the source of the matching probability distance dependence. The crucial point from the above analysis is that FR transport requires all N-2 bridging level energies to match each other, and the D and A energies; this matching probability drops exponentially with distance.

### 2.4 Correlated vs Uncorrelated Site Energy Fluctuations

The illustrative FR computations above assume that site-energy fluctuations for each site are uncorrelated Gaussians with $\sigma_{E_i} \sim \sqrt{2 \lambda_i K_B T}$. Numerical simulations of Elstner demonstrate that the site-energy fluctuations of neighboring bases in DNA are strongly correlated. Mixed quantum/classical simulations indicate approximately Gaussian distributed site-energy distributions and nearest-neighbor electronic coupling distributions. However, the site-energy fluctuation correlations break the simple connections between $\sigma_{E_i}$, $\lambda$, and $K_B T$, and produce values of $\lambda \approx 1.2 \text{ eV}$. Below,
we compute the level-matching probabilities $P_{\text{match}}(N)$ using molecular dynamics simulations that include correlated energy-level fluctuations.

2.5 The FR Model for DNA Charge Transport Kinetics

2.5.1 Limitations of Current Models

A large body of experimental data indicates two regimes of DNA charge transfer: a regime with strong (exponential) dependence of the rate/yield on distances up to \( \sim 3\text{-}7 \) base pairs and a much weaker dependence for longer distances.\(^{36}\) Theoretical models have focused on describing short-distance transport with a tunneling model and longer-distance transport with a multi-step or multi-range hopping approach.\(^{37,38}\) This framework has challenges associated with understanding transient B population in an apparent tunneling regime and with describing the transition between strong and weak distance dependent regimes.\(^{4,20,22-23,31,39-42}\)

Early theoretical studies did not adequately reproduce the experimental transition between strong and weak distance dependence for DNA ET. For example, studies of Grozema \textit{et al.} produced a transition to hopping at shorter distances than was indicated in the experiments of Lewis \textit{et al.}\(^{22}\) Bixon and Jortner noted that the distance of the DNA charge-transfer yield transition (from the rapid to slow distance dependence reported by Giese) can be described by a tunneling/nearest-neighbor coupling model
without thermal fluctuations. However, their model could not account for the weak bridge-length dependence observed for the yield and for the rate ratios.\textsuperscript{40} Some of these concerns were alleviated by introducing multi-range hopping.\textsuperscript{41} As well, Ratner and coworkers recently used a mixture of superexchange and multi-step hopping (with transient charge trapping on the B \textsuperscript{31}) to describe the transition from strong to weak ET distance dependence in the experiments of Lewis, Wasielewski and co-workers. Barton and coworkers suggested that coherent transport could be operative over large distances.\textsuperscript{4} Beratan and co-workers found that thermal fluctuations of simulated DNA base energies cause transient delocalization among the bases.\textsuperscript{23} Finally, Fiebig, Lewis, and coworkers found electron population on the photoexcited hole D (Sa) prior to detecting hole population on the hole A (Sd) in the shorter DA distance systems, which is unexpected in a pure tunneling model.\textsuperscript{20,42} It is possible that this short time regime electron population on the photoexcited Sa species arises from charge transfer in the prepared photoexcited state.\textsuperscript{39}

### 2.5.2 Flickering Resonance Descriptions of Charge Shift and Charge Separation in DNA

The HOMO energies of DNA bases are often used to estimate site energies in simple transport models.\textsuperscript{43-44} The site energies of purines are higher than those of pyrimidines, so hole transfer is dominated by the purines. It is found that thermal fluctuations of the DNA bases produce distributions of base (gas phase) HOMO energies.
that are found to be approximately Gaussian distributed. QM/MM simulations that include solvent interactions further broaden the site-energy distributions, producing increased probabilities of level matching even for sites with different mean energies (i.e., with different redox potentials). DNA sampling and energy computation protocols based on a model AT sequence, including correlated site-energy fluctuations, are described below (and refs. 22-23, 26, 29-30, 45-46), and these data are required for the FR kinetic analysis discussed below.

The coupling among nearest-neighbor sites in DNA bases is very well studied. In the two DNA experimental systems considered below, (Figure 7 and Figure 10), prior studies found that the neighboring (A) site energies are strongly correlated, but are weakly correlated with the neighboring guanines and stilbenes. To include the effects of bridge (A) site-energy correlations in our analysis of the FR mechanism we performed MD simulations on a model bridge system, double stranded poly-adenine (5’-(A)8-3’5’-(T)8-3’). (We performed 10ns MD simulations and saved 10,000 structural snapshots for further analysis). CNDO calculations were performed on the snapshots to obtain adenine site energies for each snapshot. The MD/CNDO energies used for the central four bridging adenines thus capture the site-energy correlations. To model the uncorrelated D and A energy fluctuations, we assumed site energy fluctuations with \( \sigma_E = 0.6 \text{ eV} \). We then shifted the sampled D, B, and A energies for the two systems as indicated in Figure 8 and Figure 11 to account for the distance dependent reorganization.
energy and the electron-hole attraction energies. Here, we calculated the nearest-neighbor coupling interaction between intra-strand G and A using a block diagonalization method based on structural snapshots computed earlier; the result is shown in Figure 6. The coupling is Gaussian distributed with a standard deviation of 0.095 eV and an average coupling magnitude of ~ 0.075 eV is consistent with a previously reported value of 0.069 eV.

Figure 6: The distribution of nearest-neighbor couplings between G and A is shown. 4,000 snapshots were sampled from a 4ns MD trajectory of B-DNA (5'-GATC-3':3'CTAG-5'). Couplings were computed using a block-diagonalization method. The four base-pair sequence geometries were taken from the central segment of DNA from an MD simulation of a 12 base pair sequence.

To search for energy matching geometries, the nearest-neighbor electronic couplings (V) were assumed to be identical and were sampled from a normal distribution with zero mean and standard deviation of 0.1 eV for each MD snapshot. This nearest-neighbor coupling distribution is typical of base-base (see Figure 6 and ref.
and base-stilbene interactions.\textsuperscript{22} If the offset between the maximum and the minimum values of all sampled site energies for one geometry was found to be within a window of $\pm V$ (i.e., the site energies all fall within the energy matching window), we considered that particular snapshot to be energy matched and to support FR transport. We then sampled structures until the multi-site matching probability $P_{\text{match}}(N)$, computed as just described, converged.

### 2.5.3 Charge Shift in the FR framework

We used the MD derived parameters in the FR model to describe hole transfer and trapping in DNA systems studied by Giese.\textsuperscript{21} As indicated in Figure 7, a hole is injected to a G site via photoexcitation of a 4'-acylated nucleotide. From there, the hole can either be lost to solvent in a chemical reaction that converts the G to an 8-oxoG, or it can propagate across an AT bridge to a lower-energy triple-G state and exit to solvent from that unit (Figure 7). The location of the 8-oxoG is then revealed through strand cleavage with piperidine, and the chemical-yield ratio $P_{\text{GGG}}/P_G$ for the cleavage products at the injection-site G or the destination site GGG is measured using gel electrophoresis.\textsuperscript{21} Thus, the $P_{\text{GGG}}/P_G$ ratio provides a measure of the G$\rightarrow$GGG charge transfer rate compared to the G$\rightarrow$8-oxoG oxidation rate. The distance dependence of the $P_{\text{GGG}}/P_G$ ratio is studied by varying the number of (AT) base pairs from 1 to 16 between the G and GGG traps.
Figure 7: The design of Giese’s experiment. A radical cation is established on the sugar group and then charge is injected (step 2-3) onto G. Charge transfer (steps 3-4) leads to GGG oxidation. [Reproduced with permission from ref. 21 (Copyright 2001, Nature Publishing Group).]

Three spectral functions were introduced for the donor G, the bridge adenines (A), and the acceptor GGG (Figure 8). The energy gap between the peaks of the spectral functions was set to ΔG+\lambda where ΔG is the free energy of charge shift between the sites and \lambda is the reorganization energy. ΔG is the charge shift reaction free energy for donor G to As or GGG. The free energy change for G\textsuperscript{+}\textsuperscript{*}A to GA\textsuperscript{+}\textsuperscript{*} ET is +0.35 eV\textsuperscript{20} and from G\textsuperscript{+}\textsuperscript{*}(GGG) to (GGG)\textsuperscript{+}\textsuperscript{*} is -0.7 eV.\textsuperscript{50} Typical values for DNA ET range from 0.4 eV (nearest neighbor) to 1.2 eV (4 B units between D and A)\textsuperscript{46,51}. Smaller free energy differences between G and GGG states were reported as well (ca. 0.08 eV).\textsuperscript{52} Decreased trapping free energies can also be accommodated in the FR model.
Figure 8: Energetics of the G(A)₃GG sequence. The offsets of the Gaussian peaks for a charge shift reaction are calculated using the free energy difference between G and A of 0.35 eV, $\Delta G = -0.7$ eV, and 1.2 eV (4 base pairs between D and A) $\geq \lambda \geq 0.4$ eV (nearest neighbor D and A).

The FR probabilities for Giese’s structures are calculated using sampling from both MD and distribution functions using the energetics shown in Figure 8 and a normally distributed energy matching window randomly selected from a Gaussian distribution with zero mean and standard deviation of 0.1 eV, as described above with G-A couplings of 0.043 eV and A-A couplings of 0.08 eV. The FR charge transfer rate is modeled with eq. (6), where $\tau$ is an adjustable parameter used to fit the absolute values of the yields.

To compare our simulations with the experimental chemical-yield ratios ($P_{GGG}/P_G$), we use the kinetic model suggested by Jortner et al. In this model, the water trapping rate for the holes on GGG is assumed to be much larger than the back CT rate, i.e., all of the holes that arrive at GGG are trapped. Therefore, the yield ratio is determined by the ratio of the charge transport rate and the trapping rate of the single G,
\[ \frac{P_{GGG}}{P_G} = \frac{k_{CT}}{k_r}. \]  We set the trapping rate \( k_r \) to be \( \sim 2.0 \times 10^9 \text{ s}^{-1} \), consistent with the earlier estimate of Jortner et al.\textsuperscript{53}  Giese et al. estimated the water trapping rate to be as low as \( \sim 6.0 \times 10^4 \text{ s}^{-1} \)\textsuperscript{54} based on a tunneling rate through two B sites of \( 2.5 \times 10^6 \text{ s}^{-1} \)\textsuperscript{55}

![Figure 9: Kinetic model for Giese’s experiment](image)

The computed and experimental chemical-yield ratios are shown in Table 1. In the short-distance regime, the calculated chemical-yield ratios have an exponential decay parameter of \( \beta \sim 0.5 \text{ Å}^{-1} \), in good agreement with experimental data. As the chain length grows beyond 3-4 base pairs, the hole transfer presumably becomes dominated by incoherent (hopping) transport, and the FR mechanism is no longer relevant.
Table 1: Computed $P_{\text{match}}$, experimental yield ratios, and fitted $\tau k_r$ values (with $k_r$ the water-trapping rate) for data of Giese et al.\textsuperscript{21} (see Figure 9 for kinetic scheme)

<table>
<thead>
<tr>
<th>Bridge units</th>
<th>$P_{\text{match}}$</th>
<th>$P_{GAG}/P_G$</th>
<th>$\tau k_r \times 10^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$4.8 \times 10^{-3}$</td>
<td>250</td>
<td>1.9</td>
</tr>
<tr>
<td>2</td>
<td>$8.5 \times 10^{-4}$</td>
<td>30</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>$2.1 \times 10^{-4}$</td>
<td>4</td>
<td>5.2</td>
</tr>
<tr>
<td>4</td>
<td>$6.0 \times 10^{-5}$</td>
<td>3.5</td>
<td>1.7</td>
</tr>
</tbody>
</table>

2.5.4 Application of the FR Model to DNA Hairpins

We also used the coupling and site-energy distribution functions described above in the FR model to explore charge transfer in stilbene-capped DNA hairpins studied by Lewis, Wasielewski, and Ratner\textsuperscript{20, 42} (see Figure 10). In their experiments, hole transfer occurs between the photoexcited stilbene electron donor $S_d$ and electron acceptor $S_a$ separated by 1 to 6 AT bridging base pairs.\textsuperscript{19} In contrast to the Giese system, hole transfer between the stilbenes produces the charge-separated state $S_a^-/S_d^+$. The electron-hole attraction in that state creates a $\mathbf{D-A}$ charge separation energy ramp (Figure 11, cf. ref. \textsuperscript{22}), in addition to the redox potential difference between the stilbenes and adenines.
Figure 10: Structures of (a) stilbene donor and acceptor and (b) the donor-acceptor capped DNA hairpins. Figure adapted from ref. 20.

Figure 11: Energetics of the stilbene-capped DNA hairpins (shown schematically). The Coulomb ramp arises from (screened) electron and hole attraction. The mean values for the spectral function at each site are adopted from the analysis of Grozema et al. 22, 30 and the standard deviations of these distribution functions are set to 0.6eV. The nearest-neighbor site distance was set to 3.4 Å. The increasing energy
mismatch among bases for longer chains disfavors the FR mechanism at longer distances.

The mean energies of the spectral functions at each site are computed from the analysis of Grozema et al.: $<E_i> = E_{ion}(A) - E_{el.aff}(Sa) - E_{exc}(Sa) - E_{solv.aff}(A^+ + Sa^-) - E_{elst}(A \cdot Sa + Sa \cdot A^-) + \lambda_{inner}$. $E_{ion}(A)$ is the “gas-phase” ionization potential of adenine in the DNA stack, $E_{el.aff}(Sa)$ is the vacuum electron affinity of Sa, $E_{exc}(Sa)$ is the optical excitation energy, and $E_{elst}(A^+ + Sa^-)$ is the D/A Coulomb interaction in the charge-separated state. The standard deviations of these energy distribution functions are set to 0.6 eV, similar to the values from Elstner’s QM/MM analysis of DNA. If the excited state has exciplex character with partial delocalization onto the DNA, the site energies should be decreased.

Using the protocol described to treat site-energy correlations, we calculated the matching probability for DNA hairpins (with 1 to 3 intervening ATs). The energy ramp produces large energy gaps between sites for systems with more than three AT bridging units, making the probability negligible to form a degenerate geometry at these larger distances. We therefore focus our study of the FR mechanism on Sa(A)$_n$Sd ($n = 1, 2, 3$) DNA hairpins. In the energy matching probability analysis for the Sa(A)$_n$Sd DNA, the nearest neighbor coupling is taken to be normally distributed around zero with standard deviation of 0.1 eV. The FR charge transfer rate is modeled with eq. (6), where $\tau$ is an
adjustable parameter used to fit the absolute values of the rates. The computed $\beta$-value is consistent with the measured kinetic data (Table 2).

Table 2: Computed $P_{\text{match}}$, experimental hole arrival rates of Lewis et al.\textsuperscript{20}, and fitted $\tau$ values

<table>
<thead>
<tr>
<th>Bridge units</th>
<th>$P_{\text{match}}$</th>
<th>$k_{\text{arrival}}^{\text{Expr}}$ ns$^{-1}$</th>
<th>$\tau$, fs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$7.6 \times 10^{-3}$</td>
<td>580</td>
<td>13.1</td>
</tr>
<tr>
<td>2</td>
<td>$2.1 \times 10^{-4}$</td>
<td>27</td>
<td>7.8</td>
</tr>
<tr>
<td>3</td>
<td>$5.3 \times 10^{-5}$</td>
<td>3.4</td>
<td>15.6</td>
</tr>
<tr>
<td>4</td>
<td>$1.3 \times 10^{-5}$</td>
<td>1</td>
<td>13.0</td>
</tr>
</tbody>
</table>

2.5.5 Rate Limiting Time Scales

The distance dependence of the FR matching probability $P_{\text{match}}(N)$ is consistent with the measured distance dependence of the charge shift yields (Table 1) and charge separation rates (Table 2) at short distances. We have not assumed a particular ET regime (adiabatic or nonadiabatic), and it is remarkable that the slope of the distance dependence of DNA rates and yields is predicted successfully by the MD derived $P_{\text{match}}(N)$. Our only fitting parameter is $\tau$ in eq. (6). Recall that the time scale of the rate-limiting step ($\tau_*$ in eq (6)) in the generalized kinetic scheme of eq. (5), is related to $\tau$ by

$$(\tau_*/\tau) \propto \prod_{i=1}^{N-1} \theta_i^{-1}$$

where $\theta_i \approx V_{\text{rms}}/\sigma_E$, regardless of the reaction adiabaticity.

Assuming $\theta_i \approx 1/3 - 1/6$ for the DNA systems studied here, $\tau_*$ is in the tens of femtoseconds to picoseconds range for the photoinduced charge-separation experiments,
and $\tau_*$ is in the hundred of nanoseconds to tens of microseconds for the charge shift (yield) experiments. Which step(s) in eq. (5) determine $\tau_*$? We do not expect $\tau_{\text{trap}}$ to depend strongly on distance, so it is not likely to be equal to $\tau_*$. $\tau_{\text{trans}}$ could be the origin of $\tau_*$ in the charge separation experiments because the fitted $\tau$ (Table 2) is as fast as tens of femtosecond. However, we also expect $k_{\text{on}}$ to be distance dependent (eq. (5)), because $\Delta G_{\text{act}(N)}$ is distance dependent and so is the attempt frequency to form the N-state flickering resonance. Therefore, $k_{\text{on}}^{-1}$ could determine $\tau_*$ in both experiments, and it could be the rate-limiting step for FR (we expect that $k_{\text{on}}$ to differ for charge-shift and charge-separation reactions). For ultra-fast ET, it is possible that the time scale determining the ET rate prefactor will differ for different observation time windows, as ET-active species can be depleted and replenished on multiple time scales.

In summary, the FR mechanism is compatible with the slopes of the distance dependences of the rates and yields in the charge-separation and charge shift experiments at shorter-distances. However, our analysis cannot resolve whether the FR channel is rate limited by the time scale of D-B-A transport in the N-state resonance conformation or by the activation rate to the N-state resonance conformation. Regardless of the rate-limiting kinetic step, the signatures of the FR mechanism are the exponential rate decay with distance and a transient carrier population on the bridge.
2.6 Relationship of FR to Other Transport Mechanisms

2.6.1 Resonant Tunneling

Resonant tunneling across single-molecule junctions\textsuperscript{56} arises when energy eigenvalues of the molecular states fall in the voltage gap between the two electrodes, and the electrode-bridge couplings are sufficiently strong that the lifetime of the electron on the bridge is much shorter than the charge trapping time on the bridge. The approach to and persistence of the resonant configuration distinguishes FR from resonant tunneling. In resonant tunneling between electrodes, the resonant configuration may exist for the lifetime of the poised voltage.\textsuperscript{57} In the FR case, thermal inner and outer sphere fluctuations create and destroy the electronic resonant configurations that live for the fs timescale, a time scale that is several orders of magnitude shorter than the lifetime of the electrode voltage-induced resonance. In the context of solution phase ET, resonant tunneling is usually invoked when the intermediate bridge state is resonant with D and A states for essentially all ensemble members, and the D-to-B and B-to-A couplings are sufficiently strong that the lifetime of the electron in the bridge is much shorter than the bridge trapping time.\textsuperscript{3}
2.6.2 Multirange Hopping, Mixed Tunneling/Hopping, and FR

The FR mechanism produces distance dependences consistent with those observed in the experiments of Giese, Lewis, Wasielewski, and co-workers. However, FR transport may co-exist with other transport mechanism, including pure superexchange,\textsuperscript{38} nearest-neighbor and multi-range hopping,\textsuperscript{40-41} and mixed hopping/superexchange.\textsuperscript{31, 58} As such, our analysis elucidates a viable transport mechanism, but it does not assure that this mechanism is the exclusive functional mechanism. In the FR model, we have not allowed for carrier thermalization or polaron formation on the bridge.

2.6.3 Transport with Extended Coherence

Barton and coworkers have reported a body of transport data supporting coherent (resonant) transport over 7 or 8 adenines,\textsuperscript{4} and they suggest that a coherent transport mechanism may grow in significance as the transport distance increases. Indeed, it was found that CT rates are not determined predominately by the length of the DNA, but rather by its flexibility.\textsuperscript{4, 59} In the language of the FR model, this finding could arise from correlations among the site-energy fluctuations that would enhance the energy matching probabilities. As such, the FR model creates a conceptual framework in which to pose and test hypotheses regarding the role of extended-range coherences in DNA charge transport.
2.6.4 Coexistence of Transport Mechanisms

FR may be plausibly competitive with superexchange\textsuperscript{24}, because it is expected to have a decay exponent less than or equal to that found for superexchange ET. However, as the D-A distance grows, thermally activated hopping takes over,\textsuperscript{60} and all rate processes that drop exponentially with distance will make an insignificant contribution to the charge transport. Our analysis focuses on the key multi-site energy matching probabilities (based on analytical analysis and on atomistic DNA simulations) that determine the “prefactor” for the FR charge-transfer rate.

2.7 Conclusions

The FR probability drops approximately exponentially as a function of distance, so FR transport has a signature that can be mistaken for tunneling. This theoretical framework provides a scheme to quantify ideas of how stacking dynamics may influence the coherence of charge transfer. The FR mechanism is consistent with observed ET rates in hole-transfer experiments of Giese and of Lewis. FR explains exponential distance decay of rates in the presence of transient D and A carrier populations that do not sum to unity as they must in the superexchange regime.

The redox energy matching probability depends critically on nearest neighbor couplings and energy gaps, as well as on their fluctuations. Level matching can be substantial, even when the D, B, and A sites are non-resonant on average. Fluctuations
are expected to depend upon temperature, structure, flexibility, and coupling pathway interferences. As such, one should be able to manipulate the energy-matching propensity. In particular, the FR model predicts approximately exponential distance dependence for the ET rate with decay exponent \( \Phi \sim \ln(\sigma E/V_{rms}) \sim \ln(\sqrt{2\lambda kT}/V_{rms}) \). This characteristic behavior provides a specific testable signature for the FR mechanism.

Disrupting energy-level matching during electron transmission, combined with relaxation, can produce multi-range hopping or mixed hopping and superexchange.\(^{31, 41}\) Future studies need to track the carrier dynamics that follows the disruption of resonance, in order to set more precise bounds on the admixture of coherent, incoherent, and mixed transport. The lessons of the present study are that FR: (1) obviates the need to invoke tunneling in short distance DNA ET, (2) rationalizes the observation of electron arrival at the A without the concomitant reduction of the D population in a DNA hairpin that also displays steep distance dependent ET rates, (3) suggests new approaches to controlling ET by varying site energy and coupling distributions. Building further upon the isomorphism between energy and electron transfer kinetics,\(^{16}\) it seems likely that the framework described here may be of use for analyzing coherences among excitonic states of artificial and natural light-harvesting complexes.\(^{48}\).

Based on:


Y. Z. did the simulation.

A. K. T and A. P. N. S. made the molecules.

R. Y. measured the spectrum.

Y. Z. and D. N. B. wrote the text.

All the authors involved in editing and discussion.

3.1 Conformationally Gated Charge Transfer in DNA Three-Way Junctions

3.1.1 Introduction

DNA is a target structure for examining charge transport in self-assembled organic structures.⁴,⁶⁸ To explore DNA constructs for the kinds of functionality present in conventional electronics,⁶⁹-⁷¹ the examination of junction structures is essential. DNA
three-way junctions (TWJs),\textsuperscript{21, 70, 72-79} G4 junctions,\textsuperscript{69, 80-82} and double-crossover assemblies,\textsuperscript{79, 82} are thus promising candidates. Long-distance radical cation migration\textsuperscript{74} and Förster resonance energy transfer\textsuperscript{73} in DNA TWJs have been examined previously. However, little is known regarding the nature of charge transport in structures with the kinds of large-amplitude structural fluctuations present in these TWJs.

Lewis, Wasielewski and co-workers recently found that charge migrates through the DNA TWJs shown in Figure 12.\textsuperscript{83} For distances of \~10 bases, hopping among purines is believed to dominate the charge transfer (CT).\textsuperscript{21, 36, 41-42, 60, 77, 84-85} Thus, purine networks are appealing frameworks for splitter/combiner designs. Charge separation dynamics and quantum yields were measured using transient absorption spectroscopy for structures 1a, 1b, 1c and 1d shown in Figure 12.\textsuperscript{83} The quantum yields for these DNA TWJs are relatively low compared to the yields in double-stranded DNA with similar sequences and donor-acceptor distances, while the charge transfer times are of the same order of magnitude for both systems. Importantly, CT is not complete in the 7 ns measurement window for TWJ 1c and 1d, and the hole trapping kinetics are well-fit using a zero-order kinetic scheme, suggesting that CT may be gated by conformational changes.\textsuperscript{83}
Figure 12: Structures of DNA TWJs. (1a,1b) DNA TWJs with one extended guanine pathway and no polyethylene glycol (PEG) linking group. (1c,1d) DNA TWJs with one extended guanine pathway with one PEG linking group. (1e) A proposed DNA TWJ with guanine pathways on all three arms that may serve as a charge splitter/combiner. Structures 1a, 1b, 1c, and 1d adapted from 83 were studied experimentally.

Assuming a hopping mechanism, 41, 60, 84-85 the CT rate is proportional to the square of the electronic coupling between hopping sites. In the DNA TWJs, the nearest-neighbor electronic couplings within the same branch are found in simulations to be as strong as ~0.08eV, similar to values in B-DNA.23, 45 However, these couplings between bases in different branches (“cross-junction” couplings) depend on the structure of the junction. Thus, the bottleneck for CT in DNA TWJs may involve charge hopping through the junction gated by conformational changes. The aim of our study is to explore this hypothesis.
Figure 13: Illustration of conformational switching between “T” and “Y” forms. The dotted lines represent PEG linkers.

We used molecular dynamics (MD) simulations to model the structural fluctuations of DNA TWJs. The junctions are found in simulations to undergo large-scale conformational changes on the nanosecond time scale. The simulations indicate that two of the three DNA strands can form a B-DNA like stack, while the third branch can adopt a position nearly perpendicular to the stack, thus forming a “T” shaped structure, see Figure 13. Conformational switching among “T” structures may thus gate charge flow through the junctions. Structures with polyethylene glycol (PEG) linkers are found to remain in the “T” shape geometries as long as 25 ns. We examined the electronic properties of the different conformations and found that B-DNA like stacks have inter-branch couplings that are enhanced by several orders of magnitude compared to the disrupted stacks. These findings indicate that charge flow through the DNA TWJ will likely be gated by conformational changes.
3.1.2 Methods

The DNA 3-way junction structures (Figure 12) were constructed by combining 3 B-DNA segments generated from x3DNA, and the PEG linker was added to structures 1c, 1d and 1e. The conformational ensembles were generated with extended MD simulations (up to 25ns) using the NAMD program with the CHARMM force field. Electronic structure calculations were performed on the MD snapshots using a semiempirical INDO/s method. The electronic couplings between sites were computed with a block-diagonal The DNA TWJ structures (Figure 12) were constructed as described below. Initial conformations of each structure were assembled by combining three individual canonical double stranded B-DNA segments that were generated using the x3DNA web-application. These B-DNA segments were placed in a Y-like geometry in a plane and then rotated to find the positions that bring the 3' O and 5' P atoms from different segments within covalent bond distances. The backbones from different B-DNA segments were linked into a single covalent chain. A PEG linking group was introduced in structures 1c, 1d and 1e (Figure 12). A range of different initial conformations did not substantially change the results. MD simulations were performed using the NAMD program with the CHARMM force field, an NPT ensemble, PME, and periodic boundary conditions. The CHARMM force field tends to under-estimate the \( \pi \)-interaction between the bases with a rather small error. The DNA TWJs were solvated in a box of TIP3 water molecules with 15 Å from the TWJ to the edge of the box.
Cl$^-$ and Na$^+$ ions were added to maintain the 100 mM ionic strength used in experiments, and extra one Na$^+$ per base were added to neutralize the negative charges on the DNA backbones. We also increased the NaCl concentration to 1 M, and the results did not change qualitatively.

For each structure, three MD simulations (8 ns each) were run at three temperatures (278 K, 298 K and 318 K). An extended 25 ns simulation was also run at 298 K. 20 separate trajectories (3 ns each) were computed to explore the conformational ensemble. As many as 25,000 coordinate frames (1 frame/ps) were saved from each trajectory. The sampled snapshots are uncorrelated in their electronic structure as the 1 ps window significantly exceeds the electronic coupling fluctuation time scale for DNA, which is estimated to be 10-100 fs.$^{26,29}$

For each snapshot, the atomic coordinates of the purine bases in the junction were extracted, and dangling bonds were capped with hydrogen atoms. The electronic structure analysis was performed on the base stack, with backbone atoms and waters described using point charges. The electronic properties were analyzed using semiempirical quantum methods (INDO/s in the CNDO program$^{71}$). The couplings between the nucleobases were calculated with a block-diagonalization method.$^{41,75}$
3.1.3 Results and Discussion

The experiments of Lewis, Wasielewski’s and co-workers find that the CT kinetics across the TWJ is very different for junctions with and without PEG linkers. Without PEG, the CT experiment is challenging to reproduce. With PEG linkers present, the CT kinetics is very similar to that found in duplex DNA with similar sequences (shown in Figure 14). Our simulations indicate that the difference in the CT behavior in the two classes of structures arises from the distinctive conformational fluctuations displayed in TWJs with and without the PEG linkers. Moreover, TWJs with PEG are able to stay in a CT-active “T” form for tens of ns, making the CT in these stacked structures during the lifetimes of the stacked geometry very similar to CT in duplex DNA.
Figure 14: Ratio of Sd\textsuperscript{\textasciitilde} and Sa\textsuperscript{\textasciitilde} absorptions at 532 nm and 572 nm for 3WJs 1a, 1b, 1c and 1d, and hairpins have the same purine pathway with TWJ 1c and 1d respectively. An exponential fit is shown with solid lines. Data from the first 50 ps
are suppressed for clarity. See reference 88 for additional experimental details. Data collected by Northwestern University.

3.1.3.1 DNA TWJ without a Linking Group: Rapid Switching between Conformations

In structure 1a (no PEG), the purine pathway was designed to direct charge into the upper-left branch. In contrast, structure 1b uses a G-G mismatch at the junction that was intended to direct charge onto the upper-right branch. In the MD simulations of 1a and 1b, significant conformational changes were observed on the ns time scale as indicated in Figure 15. In some trajectories, three arms of the TWJ fluctuate out of the plane and later return to the plane74.
Figure 15: (a) Structure of a DNA TWJ (1a) (adapted from\textsuperscript{83}). (b) Typical snapshot of a T-shape conformation. Backbones, linking group, and hydrogens are removed for clarity. (c) Upper: Coupling between Gs at the junctions (shown in blue oval in (a)). Once the $\pi$-stacked geometries form, the cross-junction coupling is enhanced. Bottom: Time evolution of the angles between the three DNA branches is shown with colors as in (a).
In structure 1a, the computed coupling between the two guanines at the junction is enhanced by one order of magnitude when a T conformation forms. The steric repulsion at the junction avoids the angles between arms to reach 180°C. However, in the T conformation, the “cross-junction” coupling is 3 orders of magnitude weaker than nearest-neighbor couplings in well-stacked B-DNA.

In structure 1b, we found conformational changes around T-like structures, similar to those found in structure 1a. The G-G mismatch base pair is found to disconnect and reform on the ns timescale in some of the simulations. When the guanines are not hydrogen bonded, they sometimes form a zipper-like 36 G-G stacking pattern that enhances interactions between the two branches to create a CT pathway (Figure 16).
Figure 16: (Top) The unstable G-G base pair creates a new CT pathway. A detailed analysis of the new CT pathway in structure (1b). (Middle) Time evolution of the “cross junction” couplings among the guanines at the junction. The scale for the weak couplings (green and dark blue) is on the right-side vertical axis. (Bottom) Typical snapshots in different phases of the simulation. During the first 2 ns of the trajectory, all of the bases remain paired. The nearest-neighbor base coupling between G1 and G2 (yellow) in the bottom branch is the strongest. In phase II (~2 ns to ~9.7 ns), the mismatched G2-G3 base pairing breaks, forming a strong interacting π-stack with another guanine G1 from the lower layer. Then, in phase III, fluctuations bring the guanine from the bottom branch and the right branch into close proximity, enhancing the coupling. During this time period, all of the couplings involved in the CT pathway, namely couplings between G1 and G3 (purple), G2 and G3 (light blue), and G2 and G4 (green), are considerable. Thus, a new CT pathway from G1→G3→G2→G4 is activated to cross the junction. Finally, in phase IV, the cross-junction interaction is weak, and the G2-G3 base pairing reappears.
The DNA TWJs simulated here reveal the dynamical nature of their base stacking. Indeed, the experimental kinetic measurements find that CT rates in 1a and 1b do not follow simple kinetics (in contrast to B-DNA linked species) and are difficult to reproduce (see Figure 14). To stabilize base stacking and to increase electronic coupling, a PEG linker was embedded in the second generation of TWJ structures. The kinetics of charge transfer in these PEG-linked structures is reported in ref. 83, and the simulations of these structures are reported below.

3.1.3.2 DNA TWJ with a Linker: Stabilizing “T” Conformations

Simulations of DNA TWJ structures with a PEG linking group (Figure 12 1c and 1d) indicate that both dynamics and CT characteristics of these strain-relieved systems are quite different from those revealed for systems without PEG. MD simulations of structures 1c and 1d show that distances between terminal oxygen atoms of the PEG groups vary from ~3 Å up to 17 Å. The PEG linkers accommodate strain induced at the DNA TWJ, thus slowing conformational interconversions between “T” structures and making π-stacked geometries more stable.

In one MD trajectory of TWJ 1c, the structure stacks in a “T” shape conformation 0.6 ns after starting in a nearly “Y” shaped geometry. The RMS “cross junction” coupling between the two guanines at the junction is $5.4 \times 10^{-3}$ eV during the first 0.6 ns, and then
increases to $9.0 \times 10^{-2}$ eV (Figure 17). Once the $\pi$-stack forms, CT is expected to proceed with almost the same rate as for transport in canonical B-DNA with the same sequence.

![Figure 17: Time evolution of the “cross-junction” coupling of 1c. The RMS “cross junction” coupling between the two guanines at the junction (shown in the blue shadow oval) is $5.4 \times 10^{-3}$ eV during the first 0.6 ns, and increases to 0.090 eV.](image)

In the first ~1.7 ns of MD simulations for structure 1d, the RMS averaged “cross-junction” G1 - G2 coupling is $5.4 \times 10^{-3}$ eV, similar to that in structure 1c. After ~1.7 ns, the structure is locked into a “T” conformation (Figure 18, panel b), and the G1 - G2 coupling increases to an average of ~0.12 eV, close to the coupling between nearest bases in B-DNA$^{23,45}$. Thus, when the stable “T” conformation forms, the electronic coupling
across the TWJ becomes very similar to that in B-DNA. Once π-stacked geometries are formed, the bottleneck in the CT pathway is likely to be the hopping step between the two mismatched Gs, characterized by RMS coupling of $\sim 2.6 \times 10^{-3}$ eV.

Figure 18: (a) Structure of a DNA TWJ (1d) with a PEG linker (X) (adapted from$^{83}$). (b) Typical snapshot of a T-shape conformation. Backbones, linking group, and hydrogens are removed for clarity. (c) Upper: Coupling between Gs at the
junctions (labeling as in (a)). Once the \( \pi \)-stacked geometries form, the cross-junction coupling is strongly enhanced. Bottom: Time evolution of the angles between the three DNA branches is shown with colors as in (a). This structure is locked in the T-shaped geometry after \( \sim 2 \) ns.

We extended the MD simulations of 1c and 1d to times as long as 25 ns, and the “T” conformation persisted. In some MD trajectories, formation of a stable “T” configuration was not observed.

Our simulations replicate the key experimental findings. That is, the PEG-modified structures transfer charge with kinetics similar to that found in duplex DNA with similar sequences (see Table 3). The quantum yields for CT in the PEG-modified structures are found experimentally to be lower than is found in duplex DNA (Table 3). This observation is probably explained by the simulation result that indicates only a portion of TWJs remain in CT-active well-stacked conformations during the measurement time window of the experiment. The stochastic, conformationally gated CT mechanism is expected to produce zero-order CT kinetics, as is found experimentally in these systems.

<table>
<thead>
<tr>
<th>Structures</th>
<th>Quantum yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c</td>
<td>6.3 ± 1.7%</td>
</tr>
</tbody>
</table>
3.1.3.3 A Proposed DNA TWJ Charge Splitter/Combiner

In the TWJ structures considered above, charge is directed toward one of two branches because of the asymmetric nature of the purine pathways. A next step in the design of TWJs is the splitter/combiner motifs with equivalent branches (Figure 12 1e).

The DNA TWJ has a G-G base pair on the lower branch near the junction, providing hopping routes into both upper branches. Both upper branches are designed with energy-favored purine hole pathways. In the MD simulations, a stable “T” shaped conformation appeared after several nanoseconds (similar to structures 1c and 1d). These conformations were found to have two orientations (Figure 19), which allow charge to penetrate into both upper arms. In our simulations, the structure is locked in the “T” geometry for 25 ns.
Figure 19: Typical MD snapshots representing two different “T” shaped conformations. These two snapshots are taken from different trajectories. (left) “T” shaped conformation facilitates transport along the upper-left pathway. (right) “T” shape conformation facilitates transport along the upper-right pathway.

In one trajectory leading to a T-like conformation, as shown on the left side of Figure 19, the lower and the upper-left branches form a B-DNA-like stacked geometry after ~0.3 ns. Once formed, the cross-junction coupling between G2 and G4 is 0.097 eV and the G1-G2 coupling is 0.12 eV, similar to the nearest-neighbor couplings in B-DNA. Consequently one expects that charge will flow through the junction to the acceptor on the upper-left branch with a rate similar to that in well-stacked B-DNA. An alternative MD trajectory was also obtained that was locked in a “T” conformation (right side of Figure 19) that aligns two other strands (Figure 19).
Figure 20: Time evolution of the couplings at the junctions that are involved in the CT pathways for structure 1e. After 2 ns, the well stacked “T” shaped structure forms, and the “cross-junction” coupling is enhanced by more than 2 orders of magnitude.

These simulations illustrate the key dynamical aspects of charge splitter/combiners at the single-molecule scale. Note that we were not able to observe switching between the two kinds of stable “T” conformations in the designed splitter, probably because the switching time scale is beyond the reach of the MD simulations implemented here.

3.1.4 Conclusions

Simulated DNA TWJs are found to fluctuate among multiple “T” structures. The coupling across the junctions is enhanced when a “T” structure establishes B-DNA like π-stacking. This is found in the simulations only when PEG linkers are embedded in
DNA. Indeed, the more poorly stacked structures have nearest-neighbor couplings that are several orders of magnitude weaker than in B-DNA. These simulations are consistent with measured CT rate and yield data. Structures 1a and 1b are found to undergo rapid switching among different “T” conformations. Even while in “T” conformations, the “cross-junction” couplings are several orders of magnitude smaller than the nearest-neighbor couplings in B-DNA. Structures with PEG linkers (1c and 1d) are locked into well-stacked “T” conformations, where the “cross-junction” couplings are similar to the nearest-neighbor couplings in B-DNA. As a result, the PEG linked structures should have CT kinetics similar to that found in B-DNA with a corresponding purine pathway. We also explored the CT pathways in a possible DNA TWJ charge splitter/combiner, 1e. The simulations indicate that charge can indeed proceed to both of the upper arms of the TWJ, thus suggesting that DNA-based splitter/combiners should be within reach.
4 A DNA field-effect transistor

Based on:


Y. Z. did the simulation.

Y. Z. and D. N. B. wrote the text.

All the authors involved in editing and discussion.

4.1 Introduction

Non-silicon based electronics concepts\textsuperscript{89-92} have attracted great attention, and nucleic acids provide popular nanoscale building blocks. DNA is not only crucial for life as the information carrier, but also can be engineered as a charge conducting material.\textsuperscript{4, 24, 93-94} Indeed, nucleic acids are effective hole conductors.\textsuperscript{4, 20-21, 43, 53, 95-99} The ability of DNA to self-assemble into extended nanostructures by design makes DNA a particularly exciting candidate for nanoscale electronics.\textsuperscript{100-105}

Several mechanisms have been proposed to describe charge transport in DNA.\textsuperscript{18, 37, 41, 106-110} In the short distance regime and high-barrier regime, superexchange is expected to dominate the charge transport.\textsuperscript{110-111} As the distance becomes very large, the multi-step hopping dominates the transport.\textsuperscript{41, 108, 112} For lower barrier transport, the
mechanisms at short distance can involve a combination of coherent and incoherent processes. Indeed, charge delocalization across 6 AT base pairs has been reported in simulations. Moving the energy of a hole donor close to the bridge states favors charge delocalization. Thermal fluctuations can activate transient resonance among donor/acceptor and bridge states, producing ET rates that decay exponentially as a function of distance, imitating one of the familiar hallmarks of tunneling. Importantly, signatures of coherent charge transfer across more than 10 stacked guanine-rich sequences were found recently.

Several strategies were developed recently to move beyond 1D constructs for charge transport in DNA. These include DNA double-crossover junctions, three-way junctions (TWJ) and G-quadruplex junctions. The dynamics of charge motion in these branched structures can be monitored with time resolved spectroscopy or single-molecule methods. In most designs, transport through the branching region of the molecule is believed to limit the transport kinetics. Both experimental and theoretical results suggest that efficient charge transport across the junctions requires pi-stacking through the junction region. While charge transport through these novel assemblies has been demonstrated, effective strategies to gate the flow of charge through these novel junctions (i.e., to direct the charge down one pathway vs an alternative pathway) are limited.
Figure 21: Previously studied DNA TWJ designs of Lewis and co-workers. TWJ 1a establishes a purine pathway from the lower branch to the upper left branch, and TWJ 1b has a purine pathway that consists of a lower branch and an upper-right branch. TWJ 1c is a charge splitter design. Structures are adapted from Ref. 83, 117.

Recent studies indicate that CT can proceed across the DNA TWJs shown in Figure 21: Previously studied DNA TWJ designs of Lewis and co-workers. TWJ 1a establishes a purine pathway from the lower branch to the upper left branch, and TWJ 1b has a purine pathway that consists of a lower branch and an upper-right branch. TWJ 1c is a charge splitter design. Structures are adapted from Ref. 83, 117, establishing a prototype charge splitter design.\textsuperscript{83} The polyethylene glycol (PEG) linkers indicated as X/Y increase the mobility of the junction and are found to be crucial in charge transport, presumably because they enable assembly of an extended pi-stack. In TWJ 1a, the
stilbene Sa species on the lower branch is photoexcited, and injects a hole into the DNA. Following the purine pathway, this migrating hole hops upward across the junction to oxidize the stilbene Sd on the upper-left branch. Similarly, oxidation is observed on the upper-right branch in TWJ 1b via the purine pathway, including a cross-strand hopping step between a G-G mismatched base pair. TWJ 1c is a prototype design for a charge splitter. Moreover, experimental and theoretical results have found that the CT kinetics in these TWJs is controlled by a conformationally-gated stochastic mechanism indicated in Figure 13. ¹¹⁷

In the previous charge-splitter design, charge transport can lead to oxidation on both branches. ¹¹⁷ However, the nature of the transport is not equivalent because charge flowing into the upper right arm of Figure 21 is disrupted by the non-Watson-Crick G-G base pair mismatch. Here, we propose a modified DNA three-way junction design, where charge can flow into the two pathways with equal coupling. Moreover, we are able to control the conformation of the DNA TWJ, by applying an electric field. The field switches the pi-stacking to favor one or the other of the two charge-transfer pathways. This field-modulated pathway structure enables facile switching of charge flow through a DNA molecular junction.
4.2 Methods

Starting conformations for each structure were built using the x3DNA web-application\textsuperscript{85} as described in reference.\textsuperscript{117} Beginning the MD simulations with different initial conformations did not change the results substantially. MD simulations were performed using the NAMD program\textsuperscript{86} with the CHARMM force field.\textsuperscript{78, 87} The DNA TWJs were solvated in a water box and neutralized with Na\textsuperscript{+}, and NaCl was maintain as the values indicated below (ranging from 0.0 to 0.4 M, as described below). The electronic properties were analyzed using semiempirical quantum methods (INDO/s with the CNDO program\textsuperscript{125}). The couplings between the nucleobases were calculated with a block-diagonalization method.\textsuperscript{41, 75} (See reference\textsuperscript{117} for computational details).

For TWJ 2, three MD simulations (8 ns each) were run at three temperatures (278 K, 298 K and 318 K). Eight separate trajectories (8 ns each at 298K) were computed to explore the conformational ensemble. As many as 8,000 structures (1 structure/ps) were saved from each trajectory.\textsuperscript{29, 33} For TWJ 3, 10 MD simulations (56 ns each) were run at 298K with an electric field (using the eField option in NAMD) in two different orientations. The fields were applied with the time sequence indicated in Figure 5. The 56 ns simulation includes the following steps: (1) starting from a “Y” shaped structure, the electric field was turned on in one orientation (L) for 4 ns, followed by 4 ns of relaxation (electric field off); (2) the electric field in another orientation (R) was turned on for 12 ns, followed by 4 ns of relaxation; (3) the electric field in the L orientation was
turned on for 12 ns followed by a 4 ns of relaxation; (4) the electric field was turned on in
the R orientation for 12 ns followed by 4 ns of relaxation. The L and R field orientations
are defined using the center of the three anchoring positions and the upper-left (L) or
upper-right (R) branches.

4.3 DNA TWJ as a Charge Splitter

Earlier DNA TWJ designs (1a, 1b and 1c in Figure 21) do not have equivalent CT
kinetics for the two output pathways. That is, transport from the lower strand to the
upper-left or to upper-right strand of Figure 21 are not identical. CT into the upper right
branch is disrupted by the weak electronic coupling through the G-G mismatch, which
is computed to be 2-3 orders of magnitude smaller than the nearest-neighbor base-to-
base pi-pi coupling in stacked B-DNA. Thus, the CT rate for the hopping step between
the two mismatched Gs is expected to be 4-6 orders of magnitude slower than the other
hopping steps in the CT pathway.

We now simulate the performance of a newly designed DNA TWJ charge splitter
(Figure 22) that removes the impediment of the G-G mismatch and also offers
equivalence among the two output transport pathways.
Figure 22: (a) Structure of the designed DNA TWJ charge splitter. (b) A snapshot from an MD trajectory that enables a CT pathway between A and B. (c) A snapshot from another MD trajectory with a structure that facilitates CT between A and C. After base-stacking is established, cross-junction coupling strength increase to match the nearest-neighbor couplings in B-DNA in the enabled pathways. Note that snapshots (b) and (c) come from two different MD trajectories.

MD simulations beginning with a “Y” shaped conformation (TWJ 2) fluctuate and form two stacked conformations as shown in Figure 22b and Figure 22c. In the snapshot shown in Figure 22b, branches A and B stack into a “T” shaped B-DNA like structure, thus enabling CT between the donor Sa on branch A and acceptor Sd on branch B. Similarly, in another snapshot from a different MD trajectory (Figure 22c), branch A stacks with branch C, and a strong CT pathway between Sa on branch A and Sd on branch C is established. These “T” conformations have a lifetime of more than 50 ns in our simulations,\textsuperscript{117} longer than the CT time which is believed to be around 20 ns.\textsuperscript{83}
We further investigated the “cross-junction” couplings for G₁-G₂ and G₁-G₃ as shown in Figure 22. For a trajectory shown in Figure 22b, branches A and B stack after ~3.5ns. The cross-junction coupling between G₁ and G₂ is strongly enhanced, with an averaged coupling absolute value of 0.07 eV, similar to the nearest-neighbor coupling value in B-DNA.⁴⁵ In another trajectory (Figure 22c), stacking along the alternative pathway appears after ~6.0ns, and the cross-junction coupling between G₁ and G₃ increases to an averaged absolute value of 0.065 eV.

Our findings indicate that TWJ 2 can access two CT pathways that enable charge migration from the lower branch to the upper-left and upper-right branches. Importantly, both pathways are expected to transport charge as efficiently in these stacked structures as is found in canonical B-DNA. However, no more than one CT pathway is found to be enabled at any time, and the switching between the two pathways is governed by stochastic fluctuations of the system. In the following section, we explore strategies to control the output pathway for the charge by applying an electric field to make DNA TWJs candidates for transistor-like function.

### 4.4 DNA TWJ as a Charge Switch, a Field-Effect-Transistor

The negatively charged DNA backbone makes it possible to manipulate the conformation of DNA TWJ by applying electric fields.¹²⁶-¹²⁹ The findings of Kawai et al. show that DNA translocation through a nanopore may be slowed by applying a
transverse electric field. Wierzbinski et al. found that the conformation of double stranded DNA changes with the applied voltage between an STIM tip and a gold substrate. Motivated by these prior findings, we designed a field-effect transistor (FET) based on the multiple states of the TWJ motif (Figure 23). This design includes one unfolded state (U) and two folded states (L and R). We demonstrated that we can accomplish the switching process between the three states as indicated in Figure 3 by applying electric fields in different orientations.

Figure 23: Design of a DNA TWJ FET. The “Y” shaped structure (top) folds into two stacked “T” shaped conformations depending on the electric-field orientation.
To realize the conformational changes between all three states (Figure 23), we introduced additional backbone flexibility near the junction by adding two more PEG groups than in prior structure, as shown in Figure 24. The C8 and C9 atoms of Sd, at the end of each branch, are constrained to prevent translational motion of the TWJ in the electric field.

![Diagram of DNA TWJ FET with 3 PEG linkers](Image)

**Figure 24:** Design of DNA TWJ FET with 3 PEG linkers.

We began the simulation with an unfolded “Y” conformation and applied a uniform electric field along the direction from the center of the TWJ towards one of the
upper arms, as indicated in step U→L of Figure 23. The electric field remained on for 4 ns and we allowed the system to relax for 4 ns. We found the structure folded into a “T” shaped conformation (L) and that the cross-junction coupling increased significantly, to 0.02 eV. We then applied an electric field at an angle of 120° with respect to the initial field in step L→R. After 12 ns of simulation with the reoriented electric field, and 4 ns of simulation with the field turned off, the TWJ folded into another T shape (R). The cross-junction coupling (G1-G2) increased to an average value of 0.03 eV. The time evolutions of the two cross junction couplings (G1-G2 and G1-G3) are shown in Figure 25. We repeated steps R→L and L→R; the TWJ responded to the electric fields and switched between the two folded conformations (see Figure 25).
Figure 25: (a) Structure of the FET TWJ design. (b) Typical snapshots of an unfolded conformation (left, snapshot taken at time 0) and two folded conformations (center and right). (c) Time evolution of the two cross-junction couplings between G1-G3 and G1-G2. The color bar at the top indicates the direction of the applied electric field. Red represents the time window with an electric field pushing the TWJ toward the bottom-right corner, blue represents the time window with an electric pushing the TWJ toward the bottom-left corner, and grey indicates the time window with that no electric field is present. (d) Time evolution of the angles between the branches of the TWJ. The colors correspond to the labels in panel a.

The electric field guided folding brings the magnitude of the cross-junction couplings in the well-stacked conformations into the range of values found typically for nearest neighbor base-to-base couplings in B-DNA (0.06 eV). With the applied electric field turned on, the conformational changes are usually completed within 5 ns, while it
takes more time to disrupt the initial stacking. Due to the time-scale limitations of our simulations, the success ratio for this folding process is around 40%. We expect higher switching probabilities in experiments with longer field turn-on times. We estimate the switch ratio for this device to be as large as \( \sim 10^7 \) Hz and the on/off ratio to be around \( 10^4 \).

In the simulations, we choose the electric field to be 0.02 kcal/(mol Å e). When we increased the electric field strength to 0.05 kcal/(mol Å e), the base pairing and the stacking were distorted. When the electric field was decreased to 0.01 kcal/(mol Å e), it was too weak to drive the conformational changes within the 16 ns time period we used in our simulation (12 ns applied electric field and 4 ns relaxation). We studied the influence of ionic strength on the simulations (varying it from 0 M to 0.4 M). The results and conclusions did not change qualitatively. With lower ionic strength, the TWJ responded to the electric field more rapidly.

We designed a FET TWJ structure that responds to applied electric fields with large-scale conformational changes that establish/break pi-coupling pathways. This coupling pathway switching enables controlled charge flow through branched nucleic acid structures. A proposed experimental design to realize these DNA-FETs is shown in Figure 26. The gating electric field could be integrated into a break-junction tip structure, and could therefore produce non-linear current-voltage dependences. Indeed, Wierzbinski et al. concluded that applied voltages change nucleic acid conformations in STM break-junction experiments.\(^{129}\)
Figure 26: Experimental design of a TWJ-FET. (1) A monolayer of single-stranded DNA (ssDNA) is attached to a gold surface. The designed TWJ has two ssDNA sticky ends that have complementary sequences with the ssDNA monolayer on the gold surface. Thus, the DNA TWJ can “stand up” on the gold surface. (2) A break junction tip is used to pull up the third branch of the DNA TWJ. (3) Due to the pulling force from the tip, the DNA TWJ folds into a “T” shape and establishes a high conductance. (4) After applying the electric-field gate from another source (or simply changing the voltage across the molecule), the DNA TWJ folds into a different “T” shape that produces a low conductance.

4.5 Conclusions and Prospectives

The prospect for electrical circuitry assembled from nucleic acids is intriguing and is motivated by the first attempt of Barton et al. to assemble DNA double-crossover junctions.82,115 Promising splitter designs include recent DNA TWJ and G-quadruplex
structures. With the development of the DNA origami techniques,\textsuperscript{105} DNA assemblies of diverse structure can be accessed and this toolbox seems to place key circuit elements into reach. Moreover, non-biological nucleic acids, such as peptide nucleic acids\textsuperscript{98, 130-132}, glycol nucleic acids,\textsuperscript{133} and locked-DNA,\textsuperscript{134-135} provide structural diversity with rich electronic and dynamical properties.

The charge splitter field-effect transistor designs described here provide a seemingly accessible target toward nanoscale, self-assembling circuit elements.
5. Diversity Oriented Chemical Space Exploration towards Drug Discovery

In collaboration with Lingyan Du from Prof. Qiu Wang’s lab and Dr. Chetan Rupakheti.

Manuscripts in preparation.

5.1 Diversity Oriented Chemical Space Search Algorithm

Most of functional materials and FDA approved drug are narrowed in a tiny fraction of the synthetically approachable small molecular universe (SMU). To find the hidden gems in the unexplored chemical space, our colleague Dr. Virshup et al. developed the Algorithm for Chemical Space Exploration with Stochastic Search (ACSESS).\textsuperscript{136} This algorithm is capable to generate a more diverse library than conventional genetic algorithms aiming at chemical library designs. Dr. Rupakheti et al. further developed the ACSESS framework to include a screening engine to optimize varieties of properties (Property-Optimizing ACSESS, PO-ACSESS),\textsuperscript{137-139} e.g., singlet excitation energy, oscillator strength, binding affinity, etc. The details of ACSESS and PO-ACSESS appear in references.\textsuperscript{136-137, 139}
5.2 Drug Discovery Targeting CARM1

To explore potential drug candidates targeting coactivator-associated arginine methyltransferase 1 (CARM1), we implemented different methods to assess the binding affinity into the ACSESS framework.

Previous study led by Dr. Rupakheti discovered a library of 12 compounds with high binding affinity estimated by the Docking algorithm. Unfortunately, after synthesis and bio-activity determination by Prof. Wang’s lab, none of the 12 compounds inhibits the protein’s activity. After a detailed study with molecular dynamics (MD) sampling, we found that Docking is not reliable to estimate the binding affinity for CARM1.\textsuperscript{139}

Meanwhile, MD methods are much more expensive (several days) than the Docking algorithm (several minutes). Thus, it is not computationally feasible to implement the MD method into the ACSESS framework to screen thousands of compounds generated. Here, we developed a truncated-MD strategy that can estimate the binding free energy accurately while reducing the computation cost tenfold.

5.2.1 CARM1 plays important roles in gene regulation

CARM1 involves in multiple biological processes by methylating various substrates\textsuperscript{140} including RNA binding proteins, RNA slicing factors,\textsuperscript{140} histone acetyltransferase CBP/p300\textsuperscript{141} and histone H3R17.\textsuperscript{142} Moreover, CARM1 dysregulation has been found in many cancer and developmental diseases,\textsuperscript{143} making it a plausible
drug target. The methylation process of CARM1 involves a coactivator SAM (turns into SAH after the methylation) and the kinetic process is not controversial. However, there is no known cellular active small molecule inhibitor for CARM1. Here, with structural-based computationally facilitated drug discovery strategy, we aim to generate a chemical library of potential binders targeting at the coactivator pocket of the CARM1 for further experimental investigation.

5.2.2 Existing Libraries Targeting CARM1

Collaborated with Prof. Wang’s lab, we have a pilot library of molecules which potentially presents inhibitory activities towards CARM1. Molecules in this library are structural similar with a reported pan-methyltransferase inhibitor. Among this library, three of the molecules possess inhibitory activity towards CARM1. With the structural scaffold based on the first library, the second library is generated from ACSESS framework with binding affinity assessed by Docking algorithm. The 12 compounds in the second library have been synthesized and tested experimentally, unfortunately none of them can inhibit CARM1 mediated methylation.

5.2.3 Benchmark Test with Full-system MD Sampling

As mentioned previously, binding affinity assessed by docking is not reliable towards screening inhibitor for CARM1. We applied a more comprehensive MD
strategy\textsuperscript{148-149} to estimate the binding free energy. With the MD strategy, we are able to generate a conformational ensemble that consists of thousands of snapshots. Based on the conformation ensemble, we calculate an ensemble averaged binding free energy with the MMGBSA (Molecular Mechanics / Generalized Born and Surface Area Continuum Solvation) methods\textsuperscript{150-152}. Detailed computation procedure is described in the following paragraph.

The protein structure is based on PDB: 5DXJ and 5DWQ.\textsuperscript{153} The coactivator pocket is defined by a $15 \, \text{Å} \times 15 \, \text{Å} \times 17.25 \, \text{Å}$ space originally occupied by the coactivator-mimic sinefugin (SFG) in the crystal structure.\textsuperscript{153} The initial protein-ligand complex conformation is obtained by docking the ligand into the predefined protein pocket with AutoDock vina\textsuperscript{154}. Five best docking conformations are selected. Amber\textsuperscript{155} is used to generate force field parameters. MD simulations are performed with NAMD package.\textsuperscript{86} For each initial conformation, a 15 ns of MD simulation is performed and snapshot were saved every 10 ps. As many as 1,500 snapshots are saved for each initial docking conformation. MMGBSA calculation was performed with MMPBSA.py\textsuperscript{152} on each snapshot after 1 ns (1,400 in total). An averaged binding free energy is obtained for each initial docking conformation. The lowest binding free energy among the 5 docking conformations represents the binding affinity of the protein-ligand pair.
Following the procedure described above, we performed the full-system MD sampling for the molecules described in the previous section and the results are shown in Figure 27.

![Graph showing binding free energies for different compounds.]

**Figure 27: Bind free energies for (Purple square) Sinefungin (SFG); (Green circles) Experimentally active binders (BI); (Yellow circles) Experimentally inactive binders (BI); (Red triangles) Docking predicted ligands.**

The Sinefungin (SFG) has a binding free energy of -92.66 kcal/mol, indicating a strong binding with the designated protein pocket. Binding free energies for the three active binders in the experimental set (BI113, BI50 and BI73) are all lower than -60 kcal/mol, suggesting a moderate binding. Three compounds from the experimentally inactive binder set were also examined. Two of them (BI26 and BI39) showed a less-favored binding. BI71 has a binding free energy of -80.42 kcal/mol. BI71 is structurally
similar with BI73 and the binding free energy based on MD simulation can not
distinguish between the two molecules.

The results suggest that: (1) All the active binders (SFG, BI113, BI50 and BI73)
have a binding free energies less than -60 kcal/mol; (2) Some false positive hits (e.g. BI71)
indicate binding affinity may be not the only factor that determines the binding.

![Graph](image)

**Figure 28:** Comparison of binding free energies from AMBER (Blue) and
CHARMM (Red) force field, and binding affinity (with secondary axis) from Docking
(Green).

We also compared the binding free energies based on MD simulation with
CHARMM force field as shown in Figure 28. CHARMM36\textsuperscript{156} force field was used for the
protein and force field for the ligand generated by online server provided by CGenFF\textsuperscript{157}.
CHARMM force field agrees well with AMBER force field with a systematic shift in
binding free energy. Both methods are very different from docking method, which is less consistent with the experimental results.

5.2.4 Protocols to Construct a Truncated System

The MMGBSA binding free energy based on full-system MD is more reliable than binding affinity predicated by docking. However, MD methods are not affordable in the high throughput screening. Here, we developed a protocol to construct a truncated MD system to reduce the computation cost.

For most of protein-ligand scenarios, only a small fraction of residues (often < 50) are interacting with the ligand. The other residues in the protein often act as a template that limits the fluctuation of the interacting zone. The aim of truncated MD is to reduce the computation cost while mimicking the fluctuation of the full-system. We constructed a two-layer model that can represent: (1) Detailed interaction between the ligand and the interacting residues in the core (interacting layer); (2) Limited fluctuation constrained by a constraining layer consisting of the outer-shell residues.
Figure 29: Demonstration of a two-layer model for truncated MD. The docked ligand is shown in a vDW representation. The interacting shell is shown in a cpk representation and colored in yellow. The constraining layer is shown in bond representation and colored in blue.

Parameters in this truncated MD model include: the number of residues in each layer and the strength of the constraints applied to the constraining layer. Both parameters are extracted from full-system MD simulations for specific proteins as described below.

A 15 ns full-system MD simulation with both protein and ligand (often the native ligand within the crystal structure) was performed. To define the residues in the interacting layer, we count the appearance of heavy atoms (i.e., non-hydrogen atoms)
that are within 5 Å of any atoms of the ligand in each snapshot. If more than 2 heavy
atoms on average appears in the interacting zone (< 5 Å) of the ligand, this residue is
included in the interacting layer. Similarly, the residues in the constraining layer are
defined as more than 2 heavy atoms are within 5 Å of any atoms in the interacting layer
on average among all the snapshots.

After the construction of the two-layer model, we constrained all the atoms in the
constraining layer and C and N atom on the backbone in the interacting layer. Four sets
of different applied constraint strength (0.0005, 0.001, 0.005, 0.01 kcal/mol * Å²) are
tested with MD simulations. The RMSD (root-mean-square deviation of atomic positions)
of the residues in the two layers are recorded (an example to construct truncated MD for
CARM1 was shown in Figure 30). The trajectory colored in blue was the RMSD from the
full-system MD. The trajectory colored in yellow is performed with low constraint
strength (0.001 kcal/mol * Å²) and the strongest constraint appeared in the trajectory
colored in red. The trajectory colored in black (0.005 kcal/mol * Å²) reproduced the full-
system fluctuation, thus 0.005 kcal/mol * Å² is the optimal constraint strength for
CARM1.
Figure 30: Time evolutions for the RMSD of the selected residues. (Blue) full system MD; (Other colors) truncated MD with different constraint strength (yellow is the weakest and red is the strongest). In this case, the black line matches the full-system fluctuation.

After obtaining the necessary parameters, we can set up a truncated MD in the following steps: (1) Dock the unknown ligand into the predefined protein pocket and save 5 initial protein-ligand conformations; (2) Truncate the unrelated part of the protein and only keep the two layers of residues; (3) Add harmonic restraints to all the atoms in the constraining layer and C and N atoms of the backbone of the residues in the interacting layer; (4) Perform MD simulation on the truncated system; (5) MMGBSA binding free energy calculation on the snapshots collected in (4) and the lowest represents the binding free energy for the ligand.
5.2.5 Binding Free Energies from Truncated MD is Similar to Full-system MD

With the protocol described in the previous section, we computed the binding free energies based on truncated MD. The binding free energies based on truncated MD and full-system MD are very similar (Figure 31).

![Figure 31: Comparison of binding free energies with full-system MD (Blue), truncated MD (Red), and a linear scaling of the Truncated MD (yellow).]

The binding free energies based on Truncated MD differ with the free energies based on full-system MD by a linear scaling factor (Figure 31). We can conclude that the binding free energy based on truncated MD can qualitatively reproduce the binding free energy based on full-system MD. Thus, truncated MD can be used to investigate protein-ligand binding with the accuracy comparable with full-system MD methods and much lower computation cost.
5.2.6 Apply Truncated-MD to Other Protein Systems

We applied the truncated MD strategy to other binding targets including Acetylcholinesterase (AchE)\textsuperscript{158}, HIV-1 protease (1HPV)\textsuperscript{159} and Trypsin (TryP)\textsuperscript{160}. The binding free energies based on truncated MD strategy are comparable to the results from the full-system MD method as shown in Figure 32.

![Figure 32: Comparison of binding free energy from full-system MD and truncated MD for AchE, 1HPV and TryP.](image)

5.2.7 Truncated MD is Easier to Converge

In the full-system MD, the unrelated part of the protein sometimes makes the simulation very hard to converge to equilibrium. We divided the 14 ns production simulation time period into seven 2 ns time segments, and we computed the binding
free energies for each time segments (Figure 33). The binding free energies based on truncated MD often reach equilibrium within 3 ns. However, some trajectories from the full-system MD cannot reach equilibrium within the 15 ns simulation time.

![Graph showing binding free energy over time for truncated and full-system MD](image)

**Figure 33**: Comparison of time evolution of binding free energy with Truncated MD (blue) and full-system MD (red). Results obtained from simulations on AchE.

Thus, within 3 ns of truncated MD simulation, we can obtain a converged binding free energy. However, more simulation time is needed to obtain a converged binding free energy in the full-system MD.
5.2.8 Comparison of Computation Cost of Full-system MD and Truncated MD

We compared the computation cost to sample the conformational ensemble with MD methods for all 4 proteins mentioned above (CARM1, AchE, HIVP and TryP) as shown in Figure 34. All the computation cost comparisons are based on a single node (Intel® Xeon® CPU E5-2630 v3 @ 2.40GHz) performance.

![Figure 34: Comparison of computational cost for CARM1, TryP, AchE and 1HPV.](image)

The cpu hours needed to complete 1 ns of full-system MD simulation varies for different protein systems. For larger proteins like CARM1 and AchE, 1 ns simulation takes ~100 cpu hours (~4 days for 15 ns MD simulations on a 16-core computing node). TryP and 1HPV are smaller in size, and ~60 cpu hours are needed for 1ns MD simulation (~2.5 days for 15 ns MD simulation on a 16-core computing node). The computation cost
for truncated MD does not change much from protein to protein. As in most cases, the two-layer truncated protein-ligand systems are similar in size. 1 ns of truncated MD simulation need 10-20 cpu hours (~3 hours for 3ns MD simulation on a 16-core computing node).

Thus, we accelerate the rate-limiting step in the binding free energy calculation from several days to several hours, giving rise to MD-based high throughput screening.

5.2.9 Variations and Limitations of Truncated MD

The essential idea of the truncated MD is to mimic the fluctuation of the full system. The parameters in the truncated MD can be tuned to fit more complicated scenarios. For a more flexible protein binding domain, more residues should be included in the interacting layer and the constraining layer by enlarging the cutoff distance (5 Å in this study). The constraint strength parameters can also be customized differently, i.e., different residues can be constrained differently.

As only a small fraction of the protein is kept in the truncated system, the truncated MD cannot reproduce the situation where binding induced conformational largely contributes to the binding free energy.
5.2.10 Conclusions

Here, we developed a Truncated-MD method that can be used to estimate the protein-ligand binding free energy. This method often results in less than 10% difference in binding free energy calculations. We reduce the computation cost to sample the conformational ensemble with MD methods from several days to several hours, making it possible to implement the Truncated MD method into the ACSESS framework. This is a pioneering attempt to explore the chemical library with MD approaches.
References


Biography

Gary was born in Sichuan, China in October, 1991. Both of his parents are high school chemistry teachers. Even though Gary attempts to decline the strong influence, chemistry is already coded in his gene. Gary obtained his Bachelor in chemistry from Nankai University in 2011 and hopefully gets a Ph.D. in chemistry before 4pm, Nov. 16th, 2016. During his graduate career, he purposed a new flickering resonance charge transfer mechanism in biological systems and pioneered the design of conductive junctions based on DNA nanostructures. He is now focusing on exploration of energy capture and conversion materials as well as possible drug leads in the vastness of chemical space. He has authored and co-authored 11 peer-reviewed journal articles and received the James F. Bonk Fellowship and the Marcus Hobbs Fellowship.

Publication list: