Sequence-Dependence of DX DNA Electronic Properties and Thermal Fluctuations

William B. Zhang
Biophysics Undergraduate Thesis
Department of Physics, Duke University
Durham, North Carolina 27708 USA

April 29, 2013
Acknowledgements

First and foremost, I would like to thank my advisor, David Beratan. Before I started this project, I did not know what a wavefunction was. Thank you for welcoming me into your group, for always being available to talk, for pushing me to set goals and to remain productive even when overwhelmed with other commitments, and for encouraging me to always see the broader picture and to have a plan of action.

I would also like to thank Alex Balaeff, a former postdoctoral fellow in the Beratan lab and now a professor at the University of Central Florida. Thank you for working closely with me, taking time out of your hectic schedule. You were always willing to help me with problems I encountered and for patiently explaining concepts (some very basic!) to me when I did not understand — you never considered an issue too small or trivial. Thank you. This project would not have been possible without you.

Further, I would like to thank Henry Greenside for serving on my thesis committee and for constantly encouraging me to think more carefully and to become a better scientist, Gleb Finkelstein for serving on my thesis committee on very short notice and for his helpful and insightful comments, Ravi Venkatramani for his help with coherent conductance calculations, and Chaoren Liu, Gary Zhang, BJ Williams, and Agostino Migliore for helpful discussions.

Last but not least, I would like to thank the entire physics and biophysics community at Duke University for a wonderful four years!
Abstract

The Watson-Crick base-pairing of DNA has been exploited through sticky-end cohesion and branched junctions to create complex self-assembling nanostructures. The double-crossover (DX) junction is a common motif in these structures. Interest in nanoelectronics has led to previous experimental studies of the DX structure as a nanoscale current splitter. Here, we build atomic-level models of both the original sequence and redesigned improved sequences. We produce 10 ns of molecular dynamics simulation snapshots for each sequence, which indicate a universally stable central core and fluctuating forks. We then use CNDO, a semi-empirical quantum mechanics method assuming zero differential overlap, to compute electronic structures for various segments of each system. Using the basic equation of Marcus theory, we find that our redesigned “Duke” sequence achieves a maximum cross-helical hopping rate fifty times greater than the original sequence. Our results form a foundation for atomic-level models of larger DNA nanostructures, and indicate that a careful consideration of three-dimensional geometry is crucial to sequence design in DNA nanotechnology.
## Contents

1 Introduction .................................................. 3
   1.1 Motivation for Structural Interest in DNA ...................... 3
   1.2 DNA Nanotechnology ........................................ 4
   1.3 Electronic Properties of DNA ................................. 4
      1.3.1 Effects of Base Stacking ................................ 4
      1.3.2 Effects of DNA Sequence ................................ 5
      1.3.3 Barton’s 2000 DX Experiment ............................. 6
      1.3.4 Rationale for this Project ............................... 6

2 Methods .......................................................... 8
   2.1 Structural Manipulation in VMD ............................... 8
      2.1.1 Building First DX Structure ............................. 8
      2.1.2 Subsequent DX Structures ............................... 9
      2.1.3 Solvation with Solvate Plugin ........................... 9
      2.1.4 Ionization with Meadionize Plugin ..................... 10
      2.1.5 Graphics and Analysis ................................... 11
   2.2 NAMD Molecular Dynamics .................................... 11
      2.2.1 Initial Minimization ..................................... 11
      2.2.2 Langevin Coupling ...................................... 12
      2.2.3 Water Equilibration ..................................... 12
      2.2.4 Gradual Release of Harmonic Constraints ............... 12
      2.2.5 Full Equilibration ...................................... 13
      2.2.6 Production ............................................... 13
   2.3 CNDO Quantum Calculations .................................. 14
      2.3.1 Schrödinger’s Equation .................................. 14
      2.3.2 Linear Combinations of Atomic Orbitals .................. 15
      2.3.3 Determining the Linear Coefficients .................... 15
      2.3.4 Slater Determinants .................................... 17
   2.4 Coherent Transport .......................................... 18
      2.4.1 Landauer’s Expression ................................... 18
   2.5 RMSD Trajectories ............................................ 19
      2.5.1 Aligning Molecule ....................................... 19
      2.5.2 RMSD Calculation ....................................... 19
   2.6 Molecular Orbital Couplings ................................ 19
      2.6.1 Hole Hopping in DNA .................................... 19
2.6.2 Coupling Calculation ........................................ 20
2.7 Graphics and Analysis Tools ..................................... 21
  2.7.1 Matlab .................................................... 21
  2.7.2 Microsoft Excel ........................................... 21
2.8 Other Scripting Tools and Languages .......................... 21

3 Results  .............................................................. 22
  3.1 Preliminary Distance Analysis ................................... 22
  3.2 Altered Sequence Designs ....................................... 23
  3.3 Frozen Structure Ballistic Conductance ........................ 24
    3.3.1 Electrode Contact Sites .................................. 25
    3.3.2 Conductance Measurements ................................ 25
    3.3.3 Crossover Ratio .......................................... 26
  3.4 DX Structural Conformations ................................... 27
    3.4.1 Regions of DX Structure .................................. 27
    3.4.2 Barton Sequence Structural Conformations ............... 28
    3.4.3 Duke Sequence Structural Conformations .................. 30
    3.4.4 Drastic Sequence Structural Conformations ............... 33
  3.5 HOMO Energies .................................................. 36
  3.6 Base-Base Electronic Couplings ................................ 36
    3.6.1 General Subpopulations of Couplings .................... 37
    3.6.2 Base-Dependence of Couplings ............................ 39
    3.6.3 Sequence-Dependence of Couplings ....................... 40
    3.6.4 Strongest Interhelical Couplings ......................... 40

4 Discussion ............................................................ 42
  4.1 Barton’s Sub-Optimal Guanine Placements ..................... 42
  4.2 Coherent Transport Calculations ............................... 42
  4.3 Thermal Conformations of DX Structure ....................... 43
    4.3.1 Stability of Central Core ................................ 43
    4.3.2 Behavior of Even Fork ................................... 43
    4.3.3 Behavior of Uneven Fork ................................ 43
  4.4 Crossover Couplings and Energies ............................. 44
    4.4.1 Energies ................................................ 44
    4.4.2 Couplings ................................................ 44
    4.4.3 Marcus Hopping Rates .................................... 44
  4.5 Conclusions .................................................... 45
Chapter 1

Introduction

1.1 Motivation for Structural Interest in DNA

As the primary vehicle by which organisms pass biological traits to offspring, Deoxyribonucleic acid (DNA) is fundamental to our current understanding of heritability, genetics, and evolution. It consists of two intertwining antiparallel strands in the form of a double helix. Each strand has a phosphate backbone and inward-facing nitrogenous bases which pair with bases on the opposing strand with high specificity [35]. Although there are 24 possible base pairing schemes\(^1\), the Watson-Crick scheme — adenine (A) with thymine (T) and guanine (G) with cytosine (C) — is energetically favored [47].

The fidelity of the Watson-Crick scheme is key to many of the biological functions of DNA. For example, DNA duplication occurs through an unzipping of the helix and separation of the two antiparallel strands. Each of these single strands goes on to form a new double helix, using the Watson-Crick base pairing scheme to build a new complementary strand identical to the one from which it was separated [37]. Further, the duplicate information available in complementary strands of DNA allows for excision and replacement of damaged bases [61] and correction of mismatch errors from replication [69], and many other mechanisms of repair [67].

As a result of this base-pairing specificity, the existence of an industry cheaply synthesizing oligonucleotides of arbitrary sequence, and the nanoscale size of DNA molecules, DNA is a promising material for nanotechnology. Complementary DNA sequences have the useful property of being able to self-assemble in solution through hybridization of complementary sequences [73]. This offers a convenient alternative to directly manipulating matter at the nanoscale through methods such as photolithography [51] or atomic force microscopy [50].

\(\text{If we assume that we only have the four most common bases (adenine, thymine, guanine, cytosine) and that each base can only pair with one kind of base in each scheme. Then, to find the number of possible base pairing schemes, suppose that we have the four different bases in some arbitrary fixed order on one strand. Then there is a one-to-one correspondence between permutations of the four different bases on the opposite strand and possible pairing schemes satisfying our criteria, and we know that there are } P(4, 4) = \frac{4!}{(4-4)!} = 4! = 24 \text{ possible schemes (where } P(n, k) \text{ means } n \text{ permute } k).
1.2 DNA Nanotechnology

Over the past 30 years, the field of structural DNA nanotechnology has undergone rapid development. This is in large part due to the improved experimental methods for the construction of synthetic stable branched DNA [49], which are necessary for the design and manufacture of complex structures\(^2\), and new techniques involving sticky-end cohesion [48], which allows the joining of separately assembled pieces.

Much of the motivation for this work is the possibility of constructing novel DNA-based materials by combining these features in self-assembling molecules. Polyhedral structures naturally follow from the combination of simple branched molecules, and “DNA origami” molecules of high complexity and novel shape have been synthesized [72] — Figure 1.1 is an amusing example of what has been achieved. In addition, it is possible to design sequence-dependent simple mechanical devices, notably a sequence-dependent pair of molecular tweezers constructed by Bernard Yurke and colleagues [63].

![Figure 1.1: Adapted from Figure 2 in [72]. An atomic force microscopy image of a smiley face self-assembled from a single strand of DNA.](image)

Within this field, one exciting idea is the prospect of building self-assembling nanomachines with DNA circuitry. Such structures would allow us to scale down electronics further, and could also use simpler manufacturing techniques than current electronics. The first step toward realizing this opportunity is understanding the behavior of electrons in DNA. The electronic behavior of a molecule can vary tremendously based on intrinsic and extrinsic properties such as thermal fluctuations, solvent environment, counterion placement, and monomer sequence. Therefore, in order to fully understand charge transport in DNA, it is important to understand the effects of these parameters.

1.3 Electronic Properties of DNA

1.3.1 Effects of Base Stacking

While there has been intense debate over the mechanisms involved, there is a consensus that DNA has the ability to mediate and participate in long-range charge transport, and this has prompted interest in its possible application as a one-dimensional nanowire [60,62]. However, the conductivity of DNA has been shown to be extremely sensitive to the stacking

\(^2\)Otherwise, we could have at most very long helices with no additional connectivity.
of base-pairs [57, 58] and to conformational fluctuations [75]. Since these parameters will vary with thermal motion, it is important for any theoretical study of charge transport in DNA to include dynamics and solvent. Further, since different DNA structures will have altered base-stacking, charge transport behavior can vary dramatically between different DNA structures.

One particular DNA structure of interest is the Holliday junction [38], a biological intermediate which occurs during genetic recombination. It consists of four strands of DNA with partially complementary sequences, forming four helices which can stack in multiple conformations (see Figure 1.2). Just as nanostructures require stable branching to introduce complexity, circuits also need branching to produce complex behavior. As a ubiquitous junction whose behavior has major biological implications, it is natural for the Holliday junction to be considered as a candidate for a nanoscale current splitter.

![Holliday Junction Diagram](image)

Figure 1.2: Two schematic representations of a Holliday junction in different configurations. We see that in the isomer on the left, the red and blue branch stacks with the green and yellow branch. But in the structure on the right, the red and blue helix stacks with the red and green branch instead.

When research examined the conduction properties of a Holliday junction experimentally, however, they found that due to the rapid interconversion between different conformers, charge transport in a Holliday junction is difficult to control, and conduction along one branch will leak to all three other branches [65]. There is, however, a way to modify a four-way junction (such as a Holliday junction) which reduces these conformational fluctuations. By joining two junctions together in close proximity, we can form a double crossover (DX) structure which has more rigid base-stacking [52].

### 1.3.2 Effects of DNA Sequence

Due to delocalization of its highest occupied molecular orbital (HOMO), guanine has been theoretically predicted to have a significantly lower oxidative potential than other DNA bases [76]. This has also been experimentally confirmed in double-stranded DNA. Researchers found that when a hole-bearing guanine radical is separated by too many A-T base pairs from a guanine triplet, charge transport in a short sequence virtually disappears. On the other hand, when a longer sequence of mixed C-G and A-T base pairs is built, a significant rate of long-range transport can be observed, suggesting that guanines mediate charge transport through a hopping mechanism [59]. Moreover, this effect is magnified in guanine doublets.
and triplets, in which multiple guanine bases are stacked together, increasing the stabilization due to delocalization and further favoring the presence of a hole [54].

1.3.3 Barton’s 2000 DX Experiment

In 2000, Barton’s group examined the charge transport properties of one particular DX structure experimentally [64]. For their study, Barton chose a structural arrangement with antiparallel helices which completed an even number of half-turns between crossovers, otherwise known as the DAE motif [52]. They then selected a nucleotide sequence with strategically placed guanine doublets and triplets in an attempt to induce electrons to cross over from one helix to the other.

It has been previously established that hole migration through DNA can cause long-range oxidative damage [53]. In this study, Barton exploited this fact, using the level of oxidative damage at a base to evaluate the amount of time holes spent on it. Her group covalently attached a photo-oxidant to the 5′ end of one strand. When irradiated at 365 nm, the photo-oxidant will accept an electron from a guanine doublet, leaving a guanine radical cation. This cation can either accept an electron from another part of the molecule, or it can undergo deprotonation and become irreversibly altered. The altered version has been shown to undergo strand cleavage when treated with hot piperidine [53].

Barton’s group then performed this irradiation procedure, selectively damaging nucleobase sites where holes resided for significant periods of time. The DX structures were then treated with piperidine, causing strand cleavage at those damaged sites. The component strands of the DX assemblies were then chemically separated and loaded onto polyacrylimide gels. Examination of the proportions of each strand length then revealed the locations of the damaged sites and their relative levels of damage.

1.3.4 Rationale for this Project

Barton’s 2000 DX experiment demonstrated robust charge transport in DNA double crossover assemblies, but only along a single stacked strand [64]. Inter-helical charge transfer was not observed in Barton’s experiments. As a result, it was suggested that the two helices of a DAE DX junction are actually well-insulated from one another, with little possibility of charge crossing over.

Upon a closer examination of the sequence in Barton’s study, however, it seems that we may be able to improve upon her sequence, more fully optimizing conditions for crossover conductance. Barton’s scheme is built around the idea of strategically placed guanine doublets and triplets. From a preliminary look at her guanine placements, it is unclear whether she has fully considered the three-dimensional geometry of the twisting DNA helices. As a result, she may not have achieved the closest possible interhelical contacts between her guanine multiplets.

Therefore, our first aim is to undertake a basic distance analysis between the guanine doublets and triplets on the two helices. Once we confirm that there is room to improve on Barton’s placements, we design, build, and analyze alternative sequences which may have higher rates of charge crossover. We then capture a sample of snapshots from molecular dynamics simulations for each sequence. This will allow us to incorporate the important
effects of solvent, thermal fluctuations, and structural stabilization by counterions on the electronic properties of DX DNA. Finally, we calculate interhelical electron transfer rates using electronic couplings computed from our snapshots, and compare them for the sequences studied.
Chapter 2

Methods

2.1 Structural Manipulation in VMD

VMD\textsuperscript{1} (Visual Molecular Dynamics) \cite{55} is a program for geometric manipulation and viewing of molecules. During this project, I have used versions 1.9.0 and 1.9.1 on both Windows and on Linux.

2.1.1 Building First DX Structure

Since the geometry of the DX structure is somewhat of an intermediate between an ordinary Holliday junction and straight B-DNA, we decided to build it by combining elements of these two known structures. Specifically, our first DX structure was built by combining two straight B-DNA molecules from the model.it server\textsuperscript{2} with the sequence from \cite{64} with a Holliday junction structure Protein Data Bank file with identifier 1ZEZ. The Holliday junction structure was derived from X-ray crystallography by Franklin A. Hays and colleagues \cite{70}.

![Figure 2.1: A map of the names of atoms in a DNA nucleoside, which is simple a nucleotide without the phosphates. This figure was adapted from dA.gif on the webpage www.blc.arizona.edu/molecular_graphics/dna_structure/dna_tutorial.html.](image)

\textsuperscript{1}VMD can be downloaded at http://www.ks.uiuc.edu/Research/vmd/

\textsuperscript{2}When I used it, it was found at the webpage “hydra.icgeb.trieste.it/~kristian/dna/model_it.html”, but it is no longer accessible there. A google search suggests that it is now hosted at the address “hydra.icgeb.trieste.it/dna/model_it.html”, but I have not been able to access that page either.
First, we aligned the backbones of our B-DNA molecule with the backbone of the Holliday junction file for the nucleotides at the junction. We then noticed that the C1′ atoms (see Figure 2.1) were located at a point where the structures started to diverge dramatically. When the backbones aligned, the corresponding nitrogenous bases of the two structures were in totally different orientations. However, this did not represent a problem for us, since we wanted the bases to be stacked in the DX structure as they are in ordinary B-DNA. So, to build the junctions, we simply aligned the appropriate nucleotides at the C1′ atoms, then replaced the B-DNA ribose rings and backbone with the corresponding atoms from the Holliday junction.

![Figure 2.2: Our completed DX structure. Chain A is shown in red, chain B in yellow, chain C in green, chain D in blue, and chain E in purple.](image)

**2.1.2 Subsequent DX Structures**

Due to minor inconsistencies between different file formats and changing from two two-chain ordinary B-DNA molecules to a five-chain DAE DX molecule, building our first DX structure involved a lot of manual tweaking of structure files and renaming of atoms. Further, the initial steps to solvate the system and allow the water to equilibrate with the DNA in preparation for molecular dynamics sampling are computationally intensive, and not sequence dependent. To avoid unnecessarily repeating these time-consuming tasks and to save computational time, we built our alternative sequences on top of our original sequences, using VMD and Python to simply align and swap out individual bases after the completion of water equilibration.

**2.1.3 Solvation with Solvate Plugin**

To solve our first DX structure, we used the Solvate VMD plugin\(^3\), and immersed our DX structure in a 100Å × 100Å × 170Å box of water, yielding a system with approximately 160,000 total atoms. We chose the water box size by trying to minimize the size of the system while maintaining sufficient room for the structure to change conformations without encountering the boundaries of our system.

\(^3\)This plugin can be downloaded at http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/.
2.1.4 Ionization with Meadionize Plugin

Previous studies have found that both sodium [66] and magnesium [56] ions are important in stabilizing the structure of DNA, particularly the negatively charged phosphate groups on the backbone. While sodium ions are generally loosely associated with DNA, wandering relatively large distances, magnesium ions, with their +2 atomic charge (in elementary charge units), clamp tightly onto a particular spot on DNA and stabilize it.

While simpler simulations of ordinary B-DNA can safely ignore the effects of magnesium and simply add more sodium, the structure of the DX molecule suggests a particularly important role for magnesium. One potential source of instability in our unionized structure is the fact that we have two helices in close proximity, with negatively charged phosphates on their backbones also in close proximity. In order to stabilize this repulsive interaction, we can place magnesium ions in the interhelical space, neutralizing the excessive negative charge.

To probe appropriate locations for our magnesium ions (and later to actually place our sodium ions), we used Meadionize\(^4\), an improved version of the Autoionize plugin\(^5\) for VMD. Basically, Meadionize numerically solves the Poisson-Boltzmann equation,

\[
\nabla \cdot \left[ \epsilon (\vec{r}) \nabla \Psi (\vec{r}) \right] = -\rho^f (\vec{r}) - \sum_i c_i^\infty z_i q \lambda (\vec{r}) \exp \left[-\frac{z_i q \Psi (\vec{r})}{k_B T} \right],
\]

where \(\nabla \cdot\) is the divergence operator, \(\epsilon (\vec{r})\) is the position-dependent dielectric, \(\nabla \Psi (\vec{r})\) is the gradient of the electrostatic potential at \(\vec{r}\), \(\rho^f (\vec{r})\) is the charge density of the solute at \(\vec{r}\), \(c_i^\infty\) is the concentration of the ion \(i\) at a point infinitely far from the solute, \(z_i\) is the charge of an ion \(i\), \(q\) is the charge of a proton, \(k_B\) is the Boltzmann constant, \(T\) is the temperature, and \(\lambda (\vec{r})\) is a factor accounting for the accessibility of the position \(\vec{r}\) to the ions in solution, using the package MEAD\(^6\) for a set of points in the system in a three-dimension lattice. Then, it replaces the water molecules closest to the electrostatic potential minima with sodium ions. Meadionize also ensures that no positive ions are placed too close to each other or too close to DNA, avoiding unfavorable interactions.

We first used Meadionize to find points of electrostatic potential minima. In our test runs, we placed 148 sodium ions, effectively neutralizing the negative charge from the 148 phosphate groups. We then manually examined the results for candidate sites for magnesium placements. After experimenting with different parameters for how far the placed ions had to be from each other and from DNA, we found a good set of parameters. We allowed the sodium ions to be arbitrarily close to the DNA and required them to be at least 7.5 Å apart from each other. With this set of parameters, we were able to find many electrostatic potential minima in the interhelical space. We selected nine promising magnesium locations and manually added in the ions. We then re-solvated the magnesium-containing structure and added the remaining 130 sodium ions using more realistic Meadionize parameters (requiring them to be at least 7.5 Å from DNA and 12.5 Å from each other).

\(^4\)This plugin can be downloaded at http://www.chem.duke.edu/~ilya/software/Meadionize/docs/meadionize.html.

\(^5\)This plugin can be downloaded at http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/.

\(^6\)MEAD can be downloaded at http://stjuderesearch.org/site/lab/bashford/. It stands for Macroscopic Electrostatics with Atomic Detail.
Figure 2.3: Our fully solvated and ionized DX structure. Here, the small blue dots are water molecules, the large green spheres magnesium ions, the small yellow sphere sodium ions, and the white bonds the DNA itself.

2.1.5 Graphics and Analysis

The graphics of VMD were heavily used for geometric manipulation and viewing of every stage of our structures and simulations, as well as for producing figures for presentation. The RMSD (root mean square deviation) Trajectory tool was also heavily used in analyzing simulation results and producing plots for viewing and presentation.

2.2 NAMD Molecular Dynamics

NAMD\(^7\) (Not just Another Molecular Dynamics program) [71] is a program for running molecular dynamics simulations, especially jobs making heavy use of parallel processing. For this project, we used primarily version 2.8b1 on Linux operating systems.

2.2.1 Initial Minimization

As a result of the artificial way\(^8\) in which the DX structure was created, many energetically unfavorable interactions exist. In particular, the junctions themselves contain many strained bonds, steric clashes, and electrostatic repulsions. If we were to simply allow this system to evolve according to Newton’s laws and Maxwell’s equations, it would likely explode due to the enormous repulsive forces generated by these issues.

Therefore, to resolve these unfavorable interactions, we ran initial minimization steps on the junctions (which were artificially built and not directly based on any experimental data) and the magnesium ions (which were manually placed and allowed to be arbitrary

\(^7\)NAMD can be downloaded at http://www.ks.uiuc.edu/Research/namd/.

\(^8\)No physics were involved, only geometry.
close to the DNA). During these steps, NAMD uses a sophisticated search algorithm to explore the local parameter space\(^9\) and find an energy minimum. Once NAMD reaches an energy minimum, further directional searches should all converge back to the starting point, confirming convergence and returning a stabilized structure.\(^{10}\)

### 2.2.2 Langevin Coupling

In order to heat up our system from its frozen state, we used Langevin Coupling, which basically couples each non-hydrogen atom to a virtual thermal bath. This thermal bath contributes a random force term to the total forces acting on the atom. The magnitude of this random force term depends on the temperature specified. As a result, if the speed of the atom is less than the Boltzmann speed for the given temperature, the force will, on average, accelerate the atom. On the other hand, if the speed is above the Boltzmann speed, then the random force will, on average, slow the atom down.\(^{11}\)

### 2.2.3 Water Equilibration

Since water molecules (and dissolved ions) have fewer thermal degrees of freedom than the DX structure, they can reach equilibrium with new simulation conditions faster. Since the water equilibration requires a shorter total amount of time and greatly increases the stability of our system,\(^{12}\) it is computationally efficient to separate the water equilibration and the equilibration of the DNA structure into two runs.

To run allow the water to equilibriate, we constrained the atoms of the DNA using a harmonic restoring force.\(^{13}\) Then, we simply run molecular dynamics for 500 picoseconds, allowing the water to equilibrate.

### 2.2.4 Gradual Release of Harmonic Constraints

There are several sources of inherent instability in our DNA, such as negative charges on phosphate backbones in close contact between the two DNA helices and unstable bond angles resulting from the chimeric way in which we built our DX structure. Therefore, we must release our DNA from its harmonic constraints gradually, or the system may explode.

To do this, we first compute the expected deviation of each atom from its starting position based on the temperature. The average kinetic energy of a molecule is given by

\[^{9}\text{Parameters include bond lengths, dihedral angles, etc. — basically the geometry of the DNA structure.}\]

\[^{10}\text{It is interesting to note that these minimization steps are not actually molecular dynamics — atoms are not given velocities; the laws of physics are not applied. They are more like directed searches of a local region of parameter space.}\]

\[^{11}\text{Note that this does result in a gradual heating of an initially-minimized system, since the energy pumped into the kinetic degrees of freedom of the system by the Langevin bath will initially flow into the overcooled potential degrees of freedom. In this way, multiple steps of this random force term gradually cause the speed of the system’s average atom to reach the Boltzmann speed for the desired temperature.}\]

\[^{12}\text{For example, forming a shell around any hydrophobic portions of DNA to reduce the unfavorable interactions between the non-polar parts of DNA and the polar water.}\]

\[^{13}\text{This acts as if there were a spring attached to the atom and to the atom’s original location, so that the restoring force increases linearly with the atom’s distance from its original position.}\]
\[ K = \frac{3}{2} k_B T \]  
\[ (2.2.1) \]

where \( K \) is the average kinetic energy, \( k_B \) is Boltzmann’s constant, and \( T \) is the temperature in Kelvin. Supposing that the potential energy that is stored in our restoring force is proportional to this kinetic energy, we get that

\[ U = \frac{1}{2} k_S x^2 \propto \frac{3}{2} k_B T \]  
\[ (2.2.2) \]

where \( U \) is the potential energy stored in the harmonic restoring force, \( k_S \) is the spring constant, and \( x \) is the distance between the atom’s current position and its original position (also the amount that the “spring” has been stretched). It follows from equation 2.2.2 that

\[ x \propto \sqrt{\frac{3k_B}{4k_S} T} \propto \sqrt{\frac{1}{k_S}}; \]  
\[ (2.2.3) \]

\[ k_S \propto \frac{1}{x^2}. \]  
\[ (2.2.4) \]

Now, in our gradual release of constraints, we want the average distance of an atom from its original position to be allowed to increase linearly with time. Therefore, we see from equation 2.2.3 that for \( x \) to grow linearly, we see that we need to reduce our spring constant, \( k_S \) according to equation 2.2.4 for linearly increasing values of \( x \).

In this simulation run, we do just that. After each step of 1 nanosecond, we weaken the harmonic constraints on DNA so that the distance that the atoms are free to move increases linearly. Then, we check that the average deviations of the individual atoms have stopped increasing. This indicates that our (now very weak) harmonic restoring forces are not the limiting factor constraining the atoms — since even as we reduce the restoring force, the average distance no longer increases.

### 2.2.5 Full Equilibration

After we have released the DNA from its harmonic constraints, we are almost ready to produce snapshots for the overall goal of this project. However, there are some instabilities in our DNA that may not have been sufficiently resolved by the previous equilibration steps. For example, we expect the ends of the helices to move apart from each other as a result of the electrostatic repulsion between the phosphates. This type of large-scale motion is not captured in our earlier equilibration steps, since they are run on fairly small timescales. In order for our snapshots to be valid representations of the corresponding real system, we need the structure to be stabilized, and for its behavior to be ergodic. Therefore, we performed several “full equilibration” runs in which there were no extra constraints put on the system, and it was simply allowed to run according to Newton’s laws for a period of time.

### 2.2.6 Production

Finally, we reach the snapshot-producing stage of our molecular dynamics simulation. In this stage, NAMD is simply solving Newton’s laws for small timesteps and progressing the system.
This is different from the full equilibration stage in that in the production stage, we want the simulation to be more accurate. We are no longer simply resolving instabilities. The precise locations of atoms during this stage is now important, since they can dramatically impact our quantum calculations. Therefore, to minimize the effects of using discrete time, we significantly reduce our timestep and just allow Newton’s laws to determine the progression of the system.

### 2.3 CNDO Quantum Calculations

CNDO (Complete Neglect of Differential Overlap) is a method for determining the molecular orbitals of a system as functions of known atomic basis orbitals [39–43, 46]. Briefly, CNDO first computes the single-electron Hamiltonian for the system and constructs the Fock matrix of energies and couplings of atomic orbitals by solving the time-independent Schrödinger’s equation. Next, it solves the Roothaan-Hall equations to obtain the eigenvectors and eigenvalues of the Fock matrix — the molecular orbitals and their energies. Finally, it applies the Slater determinant to satisfy fundamental fermionic properties of electrons.

#### 2.3.1 Schrödinger’s Equation

Since the concept of matter waves was first proposed by de Broglie in his 1924 PhD thesis [2, 3], the wave-particle duality of matter has been fundamental to our understanding of the universe. The general time-dependent Schrödinger’s equation, given in (2.3.1), where \( \hbar \) is the reduced Planck constant, \( \hat{H} \) is the Hamiltonian operator, which characterizes the total energy of a given wavefunction, and \( \Psi \) is the wavefunction, which is a probability amplitude describing the state of a particle and its behavior, describes the time-evolution of the quantum state of a physical system [5, 6, 23].

\[
 i\hbar \frac{\partial}{\partial t} \Psi = \hat{H} \Psi \tag{2.3.1}
\]

In addition, the solutions \( \Psi \) to the time-independent Schrödinger’s equation\(^\text{14} \) (2.3.2) are standing waves, stable states with energy \( E \).

\[
 E \Psi = \hat{H} \Psi \tag{2.3.2}
\]

For a system consisting of one electron and one proton, the Hamiltonian is given by equation (2.3.3),

\[
 \hat{H} \left( \vec{r}, \vec{R} \right) = - \frac{\hbar^2}{2m} \nabla_\vec{r}^2 - \frac{\hbar^2}{2M} \nabla_\vec{R}^2 - \frac{e^2}{4\pi\varepsilon_0 \left| \vec{r} - \vec{R} \right|}, \tag{2.3.3}
\]

where \( \vec{r} \) is the position of the electron, \( \vec{R} \) is the position of the proton, \( m \) is the mass of an electron, \( M \) is the mass of a proton, \( \varepsilon_0 \) is the vacuum permitivity, \( \nabla_\vec{z}^2 \) is the value of the

\(^\text{14} \)Note that this equation assumes a time-independent Hamiltonian.
Laplacian operator\textsuperscript{15} at $\vec{x}$, and $e$ is the elementary charge. Solving the time-independent Schrödinger equation for this Hamiltonian, we find that the solutions correspond to the atomic orbitals of hydrogen.

From previous spectroscopic data and theoretical calculations involving modifications of the equations for hydrogen, the atomic orbitals for the various types of atoms in our system — Carbon, Nitrogen, Hydrogen, Oxygen, Phosphorus — are known. Let’s call the wavefunctions corresponding to the orbitals for each atom in our structure “atomic basis functions”, and write them as $\psi_1, \ldots, \psi_n$. These atomic basis functions form an orthonormal basis for the vector space corresponding to all possible standing waves in a given molecule; these basis functions themselves do not represent the possible states for electrons in a molecule. During the covalent bonding process, the atomic basis orbitals of constituent atoms can interact and form more energetically favorable molecular orbitals \[7, 8, 12, 17–22, 24–26, 28–34, 36\].

### 2.3.2 Linear Combinations of Atomic Orbitals

In general, molecular orbital wavefunctions can be approximated by a linear combination of atomic orbital wavefunctions from the constituent atoms of the molecule [13]. In other words, the molecular orbitals in a given system can be described by

$$
\varphi_i = \sum_{j=1}^{n} c_{ij} \psi_j = c_{i1} \psi_1 + \ldots + c_{in} \psi_n,
$$

where $\varphi_i$ is the $i$th molecular orbital, $\psi_j$ is the $j$th atomic basis function, and $c_{ij}$ is the linear coefficient corresponding to the $i$th molecular orbital and the $j$th atomic basis function.

### 2.3.3 Determining the Linear Coefficients

Since the atomic basis functions $\psi_1, \ldots, \psi_n$ are known based on the types of atoms and their locations in the molecule, we only need to determine the coefficients $c_{ij}$ to find the molecular orbitals. They can be found by first computing the Fock matrix,

$$
F = \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{bmatrix},
$$

where $f_{ij}$ is given by

$$
f_{ij} = \int_{-\infty}^{\infty} \psi_i \hat{H} \psi_j.
$$

It is now clear that each diagonal element $f_{ii}$ of the Fock matrix represents the energy of the atomic orbital $\psi_i$ and that each off-diagonal element $f_{ij}$, where $i \neq j$, represents the electronic coupling between the atomic basis orbitals $i$ and $j$.

Next, we see that an arbitrary wavefunction

\[\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\] in Cartesian coordinates.
\[ \varphi_i = \sum_{j=1}^{n} c_{ij} \psi_j \]  

(2.3.7)
can be written as an abstract vector in terms of our atomic basis functions:

\[ \varphi_i = \begin{bmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{in} \end{bmatrix}. \]  

(2.3.8)

Moreover, the Fock matrix is an approximation to the true Hamiltonian of our system, and can therefore be used to compute the energy of an arbitrary wavefunction in terms of its component atomic basis functions. To do so, we simply left-multiply the Fock matrix by the transpose of the vector in equation 2.3.8 and right-multiply by the vector itself. Then we see that

\[
E_{\varphi_i} = \varphi_i^T F \varphi_i = \begin{bmatrix} c_{i1} & \cdots & c_{in} \end{bmatrix} \begin{bmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & \ddots & \vdots \\ f_{n1} & \cdots & f_{nn} \end{bmatrix} \begin{bmatrix} c_{i1} \\ \vdots \\ c_{in} \end{bmatrix}.  
\]  

(2.3.9)

In order for a wavefunction to be a standing wave (and therefore a molecular orbital) in our molecular system, it must be an eigenvector of the Fock matrix — our energy operator. So the issue of finding the molecular orbitals for our system becomes an eigenvector problem. We simply need to diagonalize the Fock matrix, yielding

\[
S^{-1} F S = D  
\]  

(2.3.10)

where \( F \) is the Fock matrix for our system, \( S \) is the orthonormal matrix whose column vectors are the eigenvectors of \( F \), and \( D \) is the diagonal matrix whose entries are the eigenvalues of \( F \), corresponding to the energies of each molecular orbital. Equation 2.3.10 is a special case of the Roothaan-Hall equations where the basis set is orthonormal [26,27].

Thus, we see that the column vectors of \( S \),

\[
S = \begin{bmatrix} s_{11} & \cdots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nn} \end{bmatrix};  
\]  

(2.3.11)

\[ \varphi_i = \begin{bmatrix} s_{i1} \\ s_{i2} \\ \vdots \\ s_{in} \end{bmatrix}, \]  

(2.3.12)
give the molecular vector representations of the molecular orbitals of our system. To obtain them as wavefunctions, we simply rewrite them in terms of their coefficients and the atomic basis functions:
\[ \varphi_i = \sum_{j=1}^{n} s_{ij} \psi_j. \]  

(2.3.13)

### 2.3.4 Slater Determinants

The molecular orbitals we obtained in the previous section, \( \varphi_1, \ldots, \varphi_n \), correspond to those of a single-electron system. For a multi-electron system we can multiply our single-electron functions together and assign them to different electrons as an anzatz for the joint wavefunction \([9–11,15,16]\):

\[ \chi(\vec{x}_1, \ldots, \vec{x}_n) = \varphi_1(\vec{x}_1) \varphi_2(\vec{x}_2) \cdots \varphi_n(\vec{x}_n). \]  

(2.3.14)

However, this guess, known as the Hartree product, does not satisfy the Pauli exclusion principle \([1,4]\) or the antisymmetry properties of fermions.

The Pauli exclusion principle states that no two fermions can occupy the same wavefunction. Mathematically, we can realize this as the constraint that

\[ \varphi_i(\vec{x}) = \varphi_j(\vec{x}) \]  

for \( i \neq j \) implies that

\[ \chi(\vec{x}_1, \ldots, \vec{x}_n) = 0. \]  

(2.3.16)

Further, the fundamental antisymmetry property of fermions requires that when two fermions are exchanged in a multi-fermionic wavefunction, the wavefunction is negated, as shown in equation 2.3.17.

\[ \chi(\vec{x}_1, \ldots, \vec{x}_i, \ldots, \vec{x}_j, \ldots, \vec{x}_n) = -\chi(\vec{x}_1, \ldots, \vec{x}_j, \ldots, \vec{x}_i, \ldots, \vec{x}_n) \]  

(2.3.17)

Moreover, identical fermions are indistinguishable. When we exchange two fermions, the results of any measurements of the system cannot be altered. For the two-particle case, we see that taking the determinant of the matrix

\[ \frac{1}{\sqrt{2}} \begin{bmatrix} \varphi_1(\vec{x}_1) & \varphi_2(\vec{x}_1) \\ \varphi_1(\vec{x}_2) & \varphi_2(\vec{x}_2) \end{bmatrix} \]  

(2.3.18)

yields an expression with the required properties:

\[ \chi(\vec{x}_1, \vec{x}_2) = \frac{1}{\sqrt{2}} \left| \begin{bmatrix} \varphi_1(\vec{x}_1) & \varphi_2(\vec{x}_1) \\ \varphi_1(\vec{x}_2) & \varphi_2(\vec{x}_2) \end{bmatrix} \right| = \frac{1}{\sqrt{2}} (\varphi_1(\vec{x}_1) \varphi_2(\vec{x}_2) - \varphi_1(\vec{x}_2) \varphi_2(\vec{x}_1)). \]  

(2.3.19)

More generally, we can construct an appropriate fermionic wavefunction by applying the Slater determinant \([14]\):

\^16 An English version of Fock’s paper can be found in \([68]\).

\^17 An English translation of Pauli’s fundamental paper can be found at \([45]\) as part of the collection “The Old Quantum Theory” \([44]\).
\[
\chi(\vec{x}_1, \ldots, \vec{x}_n) = \frac{1}{\sqrt{n!}} \left| \begin{array}{cccc}
\varphi_1(\vec{x}_1) & \cdots & \varphi_n(\vec{x}_1) \\
\vdots & \ddots & \vdots \\
\varphi_1(\vec{x}_n) & \cdots & \varphi_n(\vec{x}_n)
\end{array} \right|.
\]  
(2.3.20)

Here, we see that our fermionic indistinguishability, antisymmetry, and Pauli exclusion properties are guaranteed by the properties of matrices and determinants. For example, switching two rows of the matrix corresponds to exchanging two electrons and changes the sign of the determinant, but does not alter the result of measurements of the system. Further, if any two of the columns are identical, corresponding to two electrons occupying the same orbital, the matrix is nonsingular and the determinant is zero, as required. Thus, we have obtained a joint distribution of electrons with required properties.

\section*{2.4 Coherent Transport}

To calculate charge transport rates in the coherent transport regime, we used CNDO to output the Fock matrix for a segment of the DX molecule, then used an NEGF method to estimate conductance in the case of electrodes in direct contact with our molecule.

\subsection*{2.4.1 Landauer’s Expression}

Coherent conductance means that particles traveling between two sites will remain in phase — in this situation, it means that the charge transport occurs in a single superexchange step, tunneling under an energy barrier. This contrasts with hopping, in which the charge moves between sites through multiple tunneling events, stopping in a series of local energy minima.

For this calculation, we specifically assume ballistic conductance, and can therefore use the Landauer expression. This basically means that electrons are passing through a medium with negligible resistivity due to scattering. In this case, the mean free path of an electron, the average distance it will travel before being scattered, is significantly larger than the distance it must tunnel (in one step) to travel between two sites.

To calculate the conductance between different bases in a frozen DX structure, we used a Non-Equilibrium Green’s Function (NEGF) method to calculate the conductance of a molecule according to the Landauer equation:

\[
I(V) = \frac{q}{h} \int T(E) \left[ f_L(E) - f_R(E) \right] dE,
\]

where \( q \) is the charge of an electron, \( h \) is Planck’s constant, \( T(E) \) is the transmission probability as a function of the energy of the electron, and \( f_L(E) \) and \( f_R(E) \) are the Fermi functions for our two simulated electrodes.
2.5 RMSD Trajectories

To generate RMSD (root mean square distance) trajectories for our molecular dynamics snapshots, we used the built-in RMSD Trajectory Tool in VMD.

2.5.1 Aligning Molecule

Since we want to examine the conformational changes in our molecule, we want to neutralize the effects of translational motion. To do this, we simply compare each frame of our simulation to the first frame, and calculate the RMSD for our domain of interest. Then, the average vector from an atom in our domain in the first frame to the corresponding atom in our current frame tells us the average translational motion of our domain since the first frame. We simply subtract this vector from every atom in our current frame, effectively negating this movement. This aligns every subsequence frame of our simulation to the first frame.

2.5.2 RMSD Calculation

Now, to actually compute the RMSD from the first frame for a selection, we simply compute the distance between each atom in our selection in our current frame, post-alignment, and the corresponding atom in the first frame. We then average the values of all of the atoms in our selection, and are done. To plot an RMSD trajectory, this is repeated for all ten thousand snapshots.

2.6 Molecular Orbital Couplings

2.6.1 Hole Hopping in DNA

Since it has been established that hole-hopping is a major mechanism for long-range charge transport in DNA, we believe that the HOMO-HOMO couplings between two bases, which are proportional to current in the Fermi Golden Rule,

\[ k_{i\rightarrow j} = \frac{2\pi}{\hbar} |V_{if}|^2 \delta(E_i - E_f), \]

where \( k_{i\rightarrow j} \) is the rate of hopping from site \( i \) to site \( j \), \( \hbar \) is the reduced Planck’s constant, \( k_B \) is Boltzmann’s constant, \( E_i \) and \( E_f \) are the energies of the system in the initial and final states, and \( V_{if} \) is the electronic coupling between the initial and final molecular orbitals. Since it is known that long-range charge transport through DNA occurs through a hole hopping mechanism, this is a more physically meaningful way of computing conductance through the DX molecule. Further, in the case of electron transport, we can use the basic equation of Marcus theory,

\[ k_{ET} = \frac{2\pi}{\hbar} |V_{if}|^2 \frac{1}{\sqrt{4\pi \lambda k_B T}} \exp\left(-\frac{(\Delta G^0 + \lambda)^2}{4\pi \lambda k_B T}\right), \]

where \( \Delta G^0 \) is the reaction free energy.
to calculate the actual Golden Rule hopping rate. In equation 2.6.2, \( k_{ET} \) is the electron transfer rate, \( \hbar \) is the reduced Planck’s constant, \( k_B \) is Boltzmann’s constant, \( \lambda \) is the reorganization energy, \( \Delta G^\circ \) is the change in free energy between the initial and final states, and \( V_{if} \) is the electronic coupling between the initial and final molecular orbitals.

### 2.6.2 Coupling Calculation

In order to compute the HOMO-HOMO coupling between a given pair of nucleobases, A and B, we first isolate each single base and use CNDO to compute and diagonalize the Fock matrix for a single base. Using the coefficients from the \( S \) matrix in equation 2.3.10, we can construct the highest occupied molecular orbital for each base as a linear combination of known atomic basis functions. We know that the HOMOs of A and B can be written as vectors in the abstract space of our basis functions:

\[
M_A = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}; \quad (2.6.3)
\]

\[
M_B = \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix}. \quad (2.6.4)
\]

Next, we use CNDO to compute the electronic structure for the two-base system. This time, we only care about the Fock matrix and the couplings between the atomic orbitals of the two different nucleobases,

\[
F_{A+B} = \begin{bmatrix} f_{1,1} & \cdots & f_{1,m} & \cdots & f_{1,m+n} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
f_{m,1} & \cdots & f_{m,m} & \cdots & f_{m,m+n} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
f_{m+n,1} & \cdots & f_{m+n,m} & \cdots & f_{m+n,m+n} \end{bmatrix}, \quad (2.6.5)
\]

where \( m \) is the number of atomic basis functions for nucleobase A, \( n \) is the number of atomic basis functions for nucleobase B, and \( f_{i,j} \) is the coupling between atomic basis functions \( i \) and \( j \) in the combined basis. Since the HOMOs are constructed as linear combinations of these atomic orbitals, to find the HOMO-HOMO couplings, we simply write

\[
r = M_A^T F_{A\rightarrow B} M_B, \quad (2.6.6)
\]

where \( r \) is our desired HOMO-HOMO coupling and \( F_{A\rightarrow B} \) is the \( m \times n \) truncated Fock matrix with only couplings between between atomic basis functions of different nucleobases:

\[
F_{A\rightarrow B} = \begin{bmatrix} f_{1,m+1} & \cdots & f_{1,m+n} \\
\vdots & \ddots & \vdots \\
f_{m,m+1} & \cdots & f_{m,m+n} \end{bmatrix}. \quad (2.6.7)
\]
2.7 Graphics and Analysis Tools

2.7.1 Matlab
Matlab version R2010b was used to construct Figure 3.1 and to produce the conductance estimates in section 3.3.

2.7.2 Microsoft Excel
Microsoft Excel 2010 was used to produce several graphs and charts.

2.8 Other Scripting Tools and Languages
Python and Tool Command Language were used to manipulate file formats and structures.
Chapter 3

Results

3.1 Preliminary Distance Analysis

In order to assess the quality of guanine placements in Barton’s sequence, we did a comprehensive analysis of the distances between every pair of interhelical residues. Our results are shown in Figure 3.1.

Figure 3.1: The results of an exhaustive distance analysis between each residue on both helices. We see that Barton’s guanine doublet and triplet placements effectively miss the dark blue regions of closest interhelical contact. Each axis represents an ordering of all the nucleobase sites on one helix, so each point in the plane represents a distance between two interhelical sites.
3.2 Altered Sequence Designs

In response to finding Barton’s sequence sub-optimal, we proposed alternate sequences which may be more conducive to charge transport. Figures 3.3 and 3.4 show the two sequences we designed, a “Duke” sequence which rearranges the bases in Barton’s sequence to optimize guanine placements while retaining the same number of G-C and A-T base pairs, and a “Drastic” sequence which consists of alternating A-T base pairs on one helix and entirely G-C pairs on the other. The “Duke” sequence attempts to isolate the effects of guanine placement, while the “Drastic” sequence, which energetically favors holes in the G-C strand, is an attempt to push electrons to cross over from the A-T strand to the G-C strand by any means available. Figure 3.2 also shows the original Barton sequence structure.

Figure 3.2: Two representations of the DX structure with Barton’s sequence. Here, the red residues are guanines, the light blue ones are cytosines, the light green adenines, and the yellow ones thymines.
3.3 Frozen Structure Ballistic Conductance

As a first-order approximation to the charge-transport properties of our DX structure, we computed the ballistic conductance of our frozen structure. Since the actual conductance rates of a DX structure would be significantly affected by solvent dynamics, thermal fluctuations in structure, and counterions in solution, it is important to note that these results are only a rough approximation to the DX structure’s actual behavior.
3.3.1 Electrode Contact Sites

For our ballistic conductance calculation, we essentially calculated the current for the situation in which two electrodes are attached to parts of our DNA, a small bias is applied, and charge is allowed to flow from one electrode, through a segment of our molecule, to the other electrode. To do this, we needed to choose electrode contact sites.

Since we want to explore the possibility of charge crossing over from one helix to the other, we will place our electrodes at the point of closest interhelical contact — at one of the junctions. Figure 3.5 shows the bases to which electrodes were coupled for this calculation.¹

![Barton Sequence](image1)

![Duke Sequence](image2)

Figure 3.5: The two DX structures for which we calculated ballistic transport rates. The green site indicates where one electrode was connected. Then, the other electrode was connected to each of the base pairs in the red region and conduction from the green site was calculated. The electrodes were coupled to the non-hydrogen atoms of the nitrogenous bases only, with a coupling constant of 1.

3.3.2 Conductance Measurements

Figures 3.6 and 3.7 show in tabular form the conductance between the green site² in figure 3.5 and each base pair in the red regions.

We see that for both structures, conductance from our starting site to its immediately adjacent neighbors (either the next base on the same helix or directly across on the other helix) is on the order of $10^{-9}$ siemens, but that this drops off dramatically with increasing distance.

¹We also only computed the electronic structure for the sites at which we attached the electrodes and two flanking basepairs, since CNDO has an unfortunate limit on the size of the system for which it can compute electronic structures.

²This site also corresponds to the blank position in figures 3.6 and 3.7).
<table>
<thead>
<tr>
<th>Site</th>
<th>Conductance (S)</th>
<th>Site</th>
<th>Conductance (S)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.66E-09</td>
<td>2</td>
<td>9.20E-09</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.86E-14</td>
<td>5</td>
<td>2.65E-15</td>
<td>6.31E-06</td>
</tr>
<tr>
<td>4</td>
<td>4.20E-10</td>
<td>6</td>
<td>2.80E-10</td>
<td>1.48E-05</td>
</tr>
<tr>
<td>7</td>
<td>4.16E-15</td>
<td>8</td>
<td>3.24E-11</td>
<td>1.29E-04</td>
</tr>
<tr>
<td>9</td>
<td>4.18E-15</td>
<td>10</td>
<td>3.42E-10</td>
<td>6.28E-06</td>
</tr>
</tbody>
</table>

Figure 3.6: Conductance calculations for ballistic transport in Barton’s frozen structure. Red cells denote guanine-cytosine base pairs.

<table>
<thead>
<tr>
<th>Site</th>
<th>Conductance (S)</th>
<th>Site</th>
<th>Conductance (S)</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.47E-09</td>
<td>2</td>
<td>5.38E-09</td>
<td>2.15E-05</td>
</tr>
<tr>
<td>3</td>
<td>1.15E-13</td>
<td>5</td>
<td>2.76E-14</td>
<td>7.41E-04</td>
</tr>
<tr>
<td>4</td>
<td>3.72E-11</td>
<td>7</td>
<td>6.46E-15</td>
<td>9.92E-05</td>
</tr>
<tr>
<td>6</td>
<td>6.52E-11</td>
<td>8</td>
<td>3.52E-11</td>
<td>5.90E-05</td>
</tr>
<tr>
<td>10</td>
<td>1.55E-10</td>
<td>11</td>
<td>2.67E-15</td>
<td>1.72E-05</td>
</tr>
</tbody>
</table>

Figure 3.7: Conductance calculations for ballistic transport in the Duke frozen structure. Red cells denote guanine-cytosine base pairs.

### 3.3.3 Crossover Ratio

In order to directly compare the magnitude of charge crossing over to the other helix to the amount of charge continuing in the same helix, we simply take the ratio of conductances for two interhelical base pairs directly adjacent from each other: for example, sites 2 and 3, or 4 and 5.
Figure 3.8: A plot of the crossover conductance divided by the straight stack conductance for a number of sites.

When we plot these ratios and attempt to find a linear fit in figure 3.8, we do not see any clear trend. But we do confirm that our Duke structure has a moderately higher crossover ratio, and that in general, the crossover conductances are roughly four orders of magnitude smaller than straight stack conductances.

### 3.4 DX Structural Conformations

After properly equilibrating and stabilizing our structure as described in the results, we ran several consecutive production steps for each structure. For the Barton sequence, a run of 9.2 nanoseconds and 10400 snapshots was completed, with the first 2400 snapshots taken every 0.5 picoseconds, and the last 8000 being taken every 1 picosecond. For the Duke and Drastic sequences, a run of 10.0 nanoseconds and 10000 snapshots was completed, with snapshots being taken every 1 picosecond for the entire run.

#### 3.4.1 Regions of DX Structure

Due to its large size and complexity, the DX structure has several domains, which can fluctuate and change conformations more or less independently. It is therefore easier to analyze the DX structure’s thermal conformations piecemeal. In Figure 3.9, we divide the structure into three domains, the “uneven fork”, the “central core”, and the “even fork”, and highlight them in yellow, red, and blue, respectively.
3.4.2 Barton Sequence Structural Conformations

For Barton’s sequence, we see in figure 3.10 that the central core is generally extremely stable. All snapshots are within, roughly, 2 Å of the initial structure, and in general remain between 1 Å and 2 Å from it, with no large scale changes. On the other hand, the uneven fork and even fork regions vary greatly, fluctuating between 3 Å and 7 Å RMSD from the initial structure after the 5000th snapshot. This is primarily due to the fact that the ends of the forks are free to move, while the central core is constrained on both ends.
Figure 3.11: Selected conformations of the “even fork” region of Barton’s structure.

Snapshot 480 in figure 3.11 shows the structure of the even fork prior to any significant large-scale motion. Snapshot 4941 shows the even fork opening up during the small plateau between snapshots 4800 and 5800, while snapshot 8420 shows the fork close to its peak separation. Snapshot 9597 shows the even fork returning to a more closed conformation.

Figure 3.12: Selected conformations of the “central core” region of Barton’s structure.

As we can see from all the snapshots in figure 3.12, as well as the RMSD plot in figure 3.10 the central core region is extremely stable. While it undergoes minor twists and helical strain (likely caused by the movement of the forks), its fundamental structure remains unchanged throughout our simulation.
Figure 3.13: Selected conformations of the “uneven fork” region of Barton’s structure.

Snapshot 530 in figure 3.13 shows the structure of the uneven fork prior to any significant large-scale motion. Snapshot 6226 shows a stable closed structure after the fork has opened once. Note the position of the second magnesium from the left between the phosphate backbones of the two helices. Snapshot 6646 lies in a cluster of low-separation structures, and is even more closed than snapshot 6226. Note that in snapshot 6646, the second magnesium from the left is now attached to the backbone of the shorter helix and sitting in a major groove of the longer helix. Finally, snapshot 7886 shows the peak opening of uneven fork.

3.4.3 Duke Sequence Structural Conformations

Figure 3.14: A plot of the RMSD at each snapshot for each region of the Duke structure.
As in Barton’s sequence, we see in figure 3.14 that the central core of the Duke sequence is also similarly stable. An interesting difference, on the other hand, is the behavior of the forks. In the Barton sequence, the forks fluctuated between 3 Å and 7 Å RMSD from the initial structure after the 5000th snapshot. Here, however, the even fork of the Duke sequence only moves between 2 Å and 6 Å distance from the original structure, while the uneven fork displays a particularly large scissor-like motion, varying between 2 Å and 10 Å distance.

Figure 3.15: Selected conformations of the “even fork” region of the Duke structure.

As usual, the first snapshot, snapshot 100, shows the structure of the even fork before any large-scale motion. Snapshots 4986 and 7184 show the even fork in a slightly open conformation. Interestingly, they also reveal that the rightmost magnesium ion in snapshot 100 has escaped the structure entirely. Snapshot 7642 shows the maximum opening of the even fork of the Duke structure.

\[3\]

While it may appear to be close, or attached, to the “bottom” of the structure, a three-dimensional view reveals that it is not in the plane of the DX structure, and is quite far away.
The stability of the Duke central core is similar to that of the Barton central core, and appears to be largely unaffected by the additional magnesium ion floating above it\textsuperscript{4}.

---

\textsuperscript{4}Again, note that this magnesium ion is not in the plane of the DX structure.
snapshot 7479 shows a different one. In snapshot 5442, the two helices separate in the plane of the DX structure, more or less. But in snapshot 7479, the top helix bends out of the plane of the structure, while the bottom helix bends “down” in the plane of the structure. Snapshot 9624 shows the uneven fork returned to a closed conformation.

3.4.4 Drastic Sequence Structural Conformations

![Drastic Sequence RMSD vs. Time](image)

Figure 3.18: A plot of the RMSD at each snapshot for each region of the Drastic structure.

In figure 3.18, we see that the Drastic sequence has indeed done something drastic. While it displays the same stability in its central core as the other two sequences, we also see this level of stability in the Drastic even fork! In contrast, the uneven fork of the Drastic sequence reaches nearly 12 Å RMSD from the uneven fork of the initial structure.
In figure 3.19, snapshot 428 and 4128 show conformations very similar to the initial structure of the even fork. Snapshot 5763 is also similar, except that the right-most magnesium ion is located exclusively on the longer helix, rather than in the interhelical space. In snapshot 9999 (the final one of our simulation), the magnesium ion has just escaped, but the even fork remains closed.

In figure 3.20, we see that the central core of the Drastic sequence remains quite stable throughout the entire simulation.
Figure 3.21: Selected conformations of the “uneven fork” region of the Drastic structure.

Snapshot 750 of figure 3.21 shows the Drastic uneven fork just before it starts to open up around snapshots 800-1100. Snapshot 1295 shows an interesting conformation in the semi-open state, where the ends are being held together by the left-most magnesium, but there is a gap in the middle. It seems that the second magnesium ion from the left is pulling the major groove of the shorter helix together, without doing much to attract the backbone of the longer helix. Snapshot 8162 shows the peak RMSD for this simulation, and we see that the uneven fork is open wider than we’ve seen in any other sequence. Finally, snapshot 8598 shows a more closed conformation at a local minimum in RMSD, found between two iterations of structures like snapshot 8162.
3.5 HOMO Energies

After our molecular dynamics runs were complete, CNDO was used to compute the electronic structures of every purine (guanine or adenine) nucleobase alone. By diagonalizing the Fock matrix obtained, we can find the energies of the molecular orbitals of each nucleobase. By counting the number of electrons in each type of purine, we can fill up the molecular orbitals from lowest energy, and find the HOMO\(^5\). The distributions of energies are down in figure 3.22. We see that the energies of each type of nucleobase are approximately symmetric and normally distributed, with the energies of guanines entered at roughly \(-7.5\) eV and the energies of adenines at roughly \(-7.8\) eV.

---

\[\text{Figure 3.22: This figure shows the distribution of energies of the highest occupied molecular orbitals of the isolated purines of all three sequences. On the left, we see a histogram of all of the energies. In the center, we see only the energies of guanines. And on the right, we see the energies of the adenines alone.}\]

3.6 Base-Base Electronic Couplings

In order to find the electronic couplings between pairs of purines, we also computed the electronic structure of pairs of purines together. The Fock matrix can then tell us the electronic couplings between the atomic basis orbitals of the two nucleobases. Since the coefficients for the linear combination of atomic basis orbitals constituting the molecular orbitals are known from the single base CNDO calculations, we can use them to find the couplings between molecular orbitals.

To understand the joint distribution of couplings and interbase distance\(^6\), we generated three-dimensional histograms. To view all the couplings together on a reasonable scale, we took the natural logarithm of their magnitudes.

---

\(^5\)Or alternatively, we can simply search for the sizeable HOMO-LUMO gap.

\(^6\)In this study, interbase distance is measured by the distance between the centers of mass of the two nucleobases.
3.6.1 General Subpopulations of Couplings

Figure 3.23: A histogram of the interbase distances and nonzero electronic couplings for all purine-purine pairs.

Figure 3.23 shows the joint histogram of the natural logarithm of the magnitude of the HOMO-HOMO electronic coupling and the interbase distance. We see that there are three major peaks, and we will attempt to identify the subpopulations contributing to each of these peaks.
Figure 3.24: A histogram of the interbase distances and nonzero electronic couplings various subpopulations of purine-purine pairs. In the top left corner, we have the distribution for adjacent bases in the same DNA strand. In the top right, we have the distribution for non-adjacent bases in the same DNA strand. In the bottom left, we have the distribution for bases in the same helix, but not the same strand. And in the bottom right, we have the distribution for bases in different helices.

In figure 3.24, we see that the adjacent base couplings account for the peak closest to the top in figure 3.23, while non-adjacent bases in the same stack account contribute to the lowest peak. We also see that cross-stack couplings contribute to both the lowest peak and the second peak. To view cross-helical couplings, we actually need to expand our axes, since the vast majority of them are very small and are between two bases which are quite far apart. The cross-helical peak was shown in the figure 3.23 because non-zero cross-helical couplings are also extremely uncommon compared to the other types. As a result, the group of cross-helical couplings could not be seen on a histogram containing all the other types of
couplings.

### 3.6.2 Base-Dependence of Couplings

![Figure 3.25](image)

Figure 3.25: This figure shows the distribution of couplings between different types of bases. On the left, we see the distribution of all the guanine-guanine couplings. In the center, we have the distribution of all the adenine-adenine couplings. And on the right, we have the distribution of all the guanine-adenine couplings.

We see from figure 3.25 that the distribution of couplings is fairly similar for G-G pairs and G-A pairs. However, it appears that A-A pairs have fewer couplings in the leftmost peak. However, this is probably an artifact of our sequence design. Since we were trying to get charge to cross over from one helix to another, we made use of guanine triplets and doublets, and sometimes placed guanines next to adenines. But we almost never had two adjacent adenines in the same strand. Therefore, few of these very favorable interactions are found in the A-A couplings, and the A-A histogram seems to lack the peak closest to the top in the overall histogram (figure 3.23).
3.6.3 Sequence-Dependence of Couplings

Figure 3.26: This figure shows the distribution of coupling in our each of our three sequences. On the left, we have the couplings from Barton’s sequence. In the center, we have the Duke sequence. And on the right, we have results from the Drastic sequence.

Here, we see that the overall population of couplings is very similar for our three sequences, since they each contain adjacent couplings on the same DNA strand, non-adjacent couplings, and cross-stack couplings, which combine to form the three main peaks.

3.6.4 Strongest Interhelical Couplings

Figure 3.27: This figure shows the distribution of nonzero interhelical couplings for each sequence. From left to right, we have Barton’s sequence, the Duke sequence, and the Drastic sequence.

Above, in figure 3.27, we have distributions of the interhelical couplings for each sequence. From the plots, we cannot see much detail — only that the Duke and Drastic sequences have a somewhat broader cloud of interhelical couplings than Barton’s sequence.

To get a more quantitative perspective, we created figure 3.28. This figure collects every interhelical pair of bases with a nonzero coupling in at least one snapshot, and lists, from left to right, the two bases involved, the average magnitude of the couplings (including the ones
<table>
<thead>
<tr>
<th>Bases</th>
<th>Average</th>
<th>V</th>
<th>σ of V (eV)</th>
<th>Snapshot Share</th>
<th>Max Coupling (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barton</td>
<td>A15_G_E21_A</td>
<td>6.67E-11</td>
<td>1.52E-09</td>
<td>0.0055</td>
<td>7.94E-08</td>
</tr>
<tr>
<td>Barton</td>
<td>A16_G_E19_A</td>
<td>1.93E-09</td>
<td>1.04E-08</td>
<td>0.0854</td>
<td>2.44E-07</td>
</tr>
<tr>
<td>Barton</td>
<td>A17_G_E19_A</td>
<td>2.82E-10</td>
<td>4.36E-09</td>
<td>0.0124</td>
<td>2.19E-07</td>
</tr>
<tr>
<td>Barton</td>
<td>A26_G_E8_G</td>
<td>4.21E-12</td>
<td>4.29E-10</td>
<td>0.0001</td>
<td>4.38E-08</td>
</tr>
<tr>
<td>Barton</td>
<td>A15_G_E21_A</td>
<td>6.67E-11</td>
<td>1.52E-09</td>
<td>0.0055</td>
<td>7.94E-08</td>
</tr>
<tr>
<td>Barton</td>
<td>A16_G_E19_A</td>
<td>1.93E-09</td>
<td>1.04E-08</td>
<td>0.0854</td>
<td>2.44E-07</td>
</tr>
<tr>
<td>Barton</td>
<td>A17_G_E19_A</td>
<td>2.82E-10</td>
<td>4.36E-09</td>
<td>0.0124</td>
<td>2.19E-07</td>
</tr>
<tr>
<td>Barton</td>
<td>A26_G_E8_G</td>
<td>4.21E-12</td>
<td>4.29E-10</td>
<td>0.0001</td>
<td>4.38E-08</td>
</tr>
<tr>
<td>Barton</td>
<td>Sums</td>
<td>2.29E-09</td>
<td>1.67E-08</td>
<td>0.1034</td>
<td>5.86E-07</td>
</tr>
<tr>
<td>Duke</td>
<td>D19_G_D18_G</td>
<td>8.88E-10</td>
<td>4.60E-09</td>
<td>0.1088</td>
<td>1.51E-07</td>
</tr>
<tr>
<td>Duke</td>
<td>C10_G_D18_G</td>
<td>5.27E-10</td>
<td>3.28E-09</td>
<td>0.1128</td>
<td>1.72E-07</td>
</tr>
<tr>
<td>Duke</td>
<td>C10_G_C11_G</td>
<td>7.26E-10</td>
<td>3.86E-09</td>
<td>0.0858</td>
<td>7.75E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>C10_G_C12_A</td>
<td>7.32E-12</td>
<td>4.64E-10</td>
<td>0.0003</td>
<td>3.97E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>C9_G_C11_G</td>
<td>5.17E-09</td>
<td>4.38E-08</td>
<td>0.0998</td>
<td>1.75E-06</td>
</tr>
<tr>
<td>Duke</td>
<td>C9_G_C12_A</td>
<td>5.90E-10</td>
<td>6.00E-09</td>
<td>0.0284</td>
<td>2.10E-07</td>
</tr>
<tr>
<td>Duke</td>
<td>A14_G_C12_A</td>
<td>8.15E-11</td>
<td>1.32E-09</td>
<td>0.0092</td>
<td>4.80E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>A14_G_C13_A</td>
<td>4.69E-12</td>
<td>2.43E-10</td>
<td>0.0005</td>
<td>1.64E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>C1_G_C20_G</td>
<td>4.19E-10</td>
<td>3.02E-09</td>
<td>0.0456</td>
<td>1.04E-07</td>
</tr>
<tr>
<td>Duke</td>
<td>C1_G_B13_G</td>
<td>1.14E-12</td>
<td>1.10E-10</td>
<td>0.0002</td>
<td>1.10E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>B12_G_C20_G</td>
<td>9.48E-12</td>
<td>2.15E-10</td>
<td>0.0039</td>
<td>1.12E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>B12_G_B13_G</td>
<td>7.80E-10</td>
<td>7.20E-09</td>
<td>0.0621</td>
<td>3.90E-07</td>
</tr>
<tr>
<td>Duke</td>
<td>B11_G_B13_G</td>
<td>1.11E-11</td>
<td>5.96E-10</td>
<td>0.0008</td>
<td>4.55E-08</td>
</tr>
<tr>
<td>Duke</td>
<td>Sums</td>
<td>9.22E-09</td>
<td>7.47E-08</td>
<td>0.5582</td>
<td>3.03E-06</td>
</tr>
<tr>
<td>Drastic</td>
<td>D19_G_C11_A</td>
<td>3.53E-13</td>
<td>2.95E-11</td>
<td>0.0002</td>
<td>2.87E-09</td>
</tr>
<tr>
<td>Drastic</td>
<td>C10_G_C11_A</td>
<td>1.80E-09</td>
<td>8.85E-09</td>
<td>0.1692</td>
<td>3.79E-07</td>
</tr>
<tr>
<td>Drastic</td>
<td>C9_G_C11_A</td>
<td>7.79E-09</td>
<td>2.94E-08</td>
<td>0.2730</td>
<td>6.22E-07</td>
</tr>
<tr>
<td>Drastic</td>
<td>B12_G_B13_A</td>
<td>2.40E-10</td>
<td>2.28E-09</td>
<td>0.0320</td>
<td>8.90E-08</td>
</tr>
<tr>
<td>Drastic</td>
<td>B11_G_B13_A</td>
<td>1.24E-11</td>
<td>9.25E-10</td>
<td>0.0006</td>
<td>8.91E-08</td>
</tr>
<tr>
<td>Drastic</td>
<td>Sums</td>
<td>9.85E-09</td>
<td>4.15E-08</td>
<td>0.4750</td>
<td>1.18E-06</td>
</tr>
</tbody>
</table>

Figure 3.28: A table of several statistics for our nonzero interhelical couplings.

that are zero), the standard deviation of that mean, the proportion of snapshots in which that coupling was nonzero, and the highest coupling observed between those two bases.
Chapter 4

Discussion

4.1 Barton’s Sub-Optimal Guanine Placements

We see in figure 3.2 that several of her guanine doublets and triplets (shown in red) are located on the ends of the helices, away from the junctions. Since these ends often move apart as a result of electrostatic repulsion, we do not expect much charge to crossover at those locations. In fact, only one of her three interhelical guanine doublet-triplet pairs is located in the stable central core, and that the doublet of that pair is facing away from the triplet! In addition to identifying problems with Barton’s sequence through visual inspection, we have also confirmed quantitatively in figure 3.1 that her placements are suboptimal.

In light of her sub-optimal design, we set out to produce improved sequences, shown in figures 3.3 and 3.4. With our Duke sequence, we placed guanines in optimal positions, while in our Drastic sequence, we tried to focus on energetics, creating one helix entirely out of AT repeats and the other entirely from stacked G-C pairs. Since the intrahelical A-A hopping competes with the cross-helix A-G hopping, we disrupted the purine-purine stacking by making the sequence from AT repeats. Perhaps by making intrahelical hopping less favorable, we can improve the relative probability of interhelical hopping.

4.2 Coherent Transport Calculations

The results from our coherent transport calculations suggest that we have made some improvement over Barton’s structure at the crossovers themselves. We see in figures 3.6 and 3.7 that the Duke sequence has a marginally higher conductance to sites 1 and 5, and a significantly higher one for site 3. The ratio column also shows that for the sites 3 and 5 capture a slightly larger proportion of the current in the Duke structure than their intrahelical competitors.

However, our coherent transport calculations show only marginal gains at the crossovers themselves. As we move further away from those sites, the Duke sequence no longer offers any clear advantage over the Barton one. Moreover, the ratios show that even though we have made marginal improvements, the crossover conductance is still 4-6 orders of magnitude lower than the straight stack conductance.

Since this calculation was made in a frozen structure without any dynamics, solvent, or
counterions, it is only a rough estimate. In this model, current crossing from one helix to the other is at least four orders of magnitude less likely than current continuing in the same helix. But ballistic conductance is not the best model for long-range charge transport in DNA. It was been shown that hole hopping is a more realistic mechanism [59], so we analyzed our dynamic snapshots under that paradigm.

4.3 Thermal Conformations of DX Structure

Interestingly, our three sequences behaved very differently in our molecular dynamics simulations. While all three had extremely stable central cores, the behaviors of the even and uneven forks varied.

4.3.1 Stability of Central Core

While all three structures showed similar dynamics in the central core — essentially small fluctuations and no large-scale conformational changes — this behavior is critical to the behavior of the overall molecule. The stability of the central core is critical to our analysis in that it insulates the fluctuating forks from each other. Without it, we could not break our structure into three pieces for analysis, and would have a much more complicated situation.

4.3.2 Behavior of Even Fork

In Barton’s sequence, the two forks seemed to cover the same RMSD range. This suggests that in some sense, the even fork is bending more than the uneven fork. Since the uneven fork is longer, if both forks bend open in a scissor-like motion at the same angle, the RMSD for the uneven fork would be greater. Since their RMSDs are more or less equal, the even fork must have a slightly greater maximum angle of bending.

The reason for this behavior in the Barton structure is unclear, but may be linked to the stability of the rightmost magnesium ion in the snapshots in figure 3.11. Unlike the Duke and Drastic sequences (see figures 3.15 and 3.19), Barton’s sequence keeps that magnesium in the even fork through the entire simulation. Perhaps an extra magnesium there can stabilize the additional bending and helical twisting.

On the other hand, in both the Duke and Drastic sequences, the even fork maintains a lower RMSD compared to the uneven fork (see figure 3.14). In the Duke sequence, we see that the fork still opens and closes as in the Barton structure, but that it does so to a lesser degree. In the Drastic sequence, however, the fork remains as shut and stable as the central core for the entire simulation! It is not clear if this divergent behavior is due to random thermal fluctuations or if it is a direct consequence of our sequence designs.

4.3.3 Behavior of Uneven Fork

In all three structures, the uneven fork opens and closes, driven by the opposing forces of electrostatic repulsion between phosphates on the backbone and the attraction between the magnesium ions and the phosphates. As shown in figures 3.13, 3.17, and 3.21, the Barton
uneven fork opens the least, the Drastic uneven fork opens the widest, and the Duke uneven fork is somewhere in between. The reasons for the discrepancies between the sequences are unclear.

From what we do know, however, it is clear that the magnesiums play a crucial role in the scissor-like motion experienced by the forks. It would be very interesting to repeat our simulations with only sodium in solution and observe the dynamics (if those structures are stable and do not explode).

4.4 Crossover Couplings and Energies

4.4.1 Energies

In figure 3.22, we confirmed the well-documented fact (cite) that guanines are the most favorable sites for holes to reside on. We see that the HOMO energies of guanines and adenines were both symmetrically distributed and centered at -7.5 eV and -7.8 eV, respectively. We then note that this energy is the energy of an electron occupying the HOMO. If a hole is on the base, 7.5 eV or 7.8 eV of energy is required to ionize that electron. Therefore, having the higher HOMO electron energy gives the lower ionization potential. So we see that hole hopping on guanines is indeed energetically favored, and that the energy gap between guanine and adenine is approximately 0.3 eV.

4.4.2 Couplings

From our coupling calculation results in figures 3.23 and 3.24, we see that there exist four major subpopulations of couplings: adjacent base pairs with staked purines in the same helix, adjacent base pairs with unstacked purines in the same helix, non-adjacent bases in the same helix, and bases in different helices. While the interhelical couplings are generally several orders of magnitude lower than the others, certain conformations have revealed couplings on the order of $10^{-6}$ eV, which is comparable to some adjacent base couplings in the same stack, and many cross-stack couplings in the same helix!

Therefore, we see that although the two helices of the DX structure are generally insulated from each other, through careful sequence design we can improve the crossover probabilities. In addition, since we have not yet taken into account energetics, it may be possible to further improve the likelihood of charge crossing over.

4.4.3 Marcus Hopping Rates

We can pick out the electronic couplings with the largest magnitude for the Barton and Duke sequences from figure 3.28, we see that the Duke maximum coupling is approximately one order of magnitude larger.

$$|V|_{\text{max-Barton}} = 2.44 * 10^{-7} \text{ (eV)}$$
$$|V|_{\text{max-Duke}} = 1.75 * 10^{-6} \text{ (eV)}$$

Next, we can use a literature value for the reorganization energy and use our computed values for the change in free energy from figure 3.22.
Substituting these values into the Marcus rate expression (see equation 2.6.2, we compute the maximum interhelical hopping rates for each sequence (for the best choice of two bases, in the best possible conformation),

\[ k_{\text{max-Barton}} = \frac{2\pi}{\hbar} |V|_{\text{max-Barton}} \frac{1}{\sqrt{4\pi \lambda k_B T}} \exp \left( -\frac{(\Delta G_{G-G}^\circ + \lambda)^2}{4\pi \lambda k_B T} \right) = 1117 \text{ (s}^{-1}) ; \]

\[ k_{\text{max-Duke}} = \frac{2\pi}{\hbar} |V|_{\text{max-Duke}} \frac{1}{\sqrt{4\pi \lambda k_B T}} \exp \left( -\frac{(\Delta G_{G-G}^\circ + \lambda)^2}{4\pi \lambda k_B T} \right) = 57512 \text{ (s}^{-1}) , \]

and find that our Duke sequence yields a 50-fold improvement over the original Barton sequence.

4.5 Conclusions

In this project, we have given a description of the thermal conformations and the base-base electronic couplings for three different sequences of DX DNA. Since many DNA nanostructures — such as Paul Rothemund’s DNA origami [72], Thom LaBean’s grids [74], and Chengde Mao’s polyhedra [77] — contain the DX junction as a basic element, we believe that our results can serve as a foundation for modeling larger structures and understanding their behavior.

In addition, we have shown that Barton’s original sequence missed the optimal guanine placements for inducing charge crossover, and that a redesigned sequence can significantly improve the chances of charge crossover. In doing so, we have illustrated the danger in designing helical DNA sequences based on linear diagrams, and have demonstrated a more accurate method of understanding the three-dimensional geometry of DNA structures.
Bibliography


[77] Zhang, Chuan and Ko, Seung Hyeon and Su, Min and Leng, Yujun and Ribbe, Alexander E and Jiang, Wen and Mao, Chengde, Symmetry controls the face geometry of DNA polyhedra, Journal of the American Chemical Society 131 (2009), no. 4, 1413–1415.