Modeling Electrical Transport through Nucleic Acids

Jianqing Qi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

M. P. Anantram, Chair

Lih Y. Lin

Georg Seelig

Program Authorized to Offer Degree:
Department of Electrical Engineering

University of Washington

Abstract

Modeling Electrical Transport through Nucleic Acids

Jianqing Qi

Chair of the Supervisory Committee:
Professor M. P. Anantram
Department of Electrical Engineering

Nucleic acids play a vital role in many biological systems and activities. In recent years, engineers and scientists have been interested in studying their electrical properties. The motivation for these studies stems from the following facts: (1) the bases, which form the building blocks of nucleic acids, have unique ionization potentials. Further, nucleic acids are one of the few nanomaterials that can be reproducibly manufactured with a high degree of accuracy (though admittedly their placement at desired locations remains a challenge). As a result, designed strands with specific sequences may offer unique device properties; (2) electrical methods offer potential for sequencing nucleic acids based on a single molecule; (3) electrical methods for disease detection based on the current flowing through nucleic acids are beginning to be demonstrated.

While experiments in the above mentioned areas is promising, a deeper understanding of the electrical current flow through the nucleic acids needs to be developed. The modeling of current flowing in these molecules is complex because: (1) they are based on atomic scale contacts between nucleic acids and metal, which cannot be reproducibly built; (2) the conductivity of nucleic acids is easily influenced by the environment, which is constantly changing; and (3) the nucleic acids by themselves are floppy.

This thesis focuses on the modeling of electrical transport through nucleic acids that are connected to two metal electrodes at nanoscale. We first develop a decoherent transport

model for the double-stranded helix based on the Landauer-Büttiker framework. This model is rationalized by comparison with an experiment that measured the conductance of four different DNA strands. The developed model is then used to study the: (1) potential to make barriers and wells for quantum transport using specifically engineered sequences; (2) change in the electrical properties of a specific DNA strand with and without methylation; (3) difference in electrical conduction accompanying the conformational change of DNA between A- and B- forms; (4) role of water in influencing the tight-binding Hamiltonian and orbital distribution in DNA strands; and (5) transport properties of RNA:DNA hybrids. While the results in this thesis demonstrate that experiments can be understood with the developed model, it also reveals that further development of *ab initio* methods to include the role of environment and vibrations without oversimplifying assumptions can lend further insight.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

A: adenine

C: cytosine

CC: coupled cluster

CI: configuration interaction

Cm: methylated cytosine

DFT: density functional theory

DNA: deoxyribonucleic acid

DOS: density of states

ds: double-stranded

EA: electron affinity

G: guanine

GGA: generalized gradient approximation

HF: Hartree-Fock

HOMO: highest occupied molecular orbital

IP: ionization potential

LDA: local-density approximation

LSDA: local spin-density approximation

LUMO: lowest unoccupied molecular orbital

MCBJ: mechanically controlled break junction

MD: molecular dynamics

MM: molecular mechanics

NAB: Nucleic Acid Builder

NTCDA: 1,4,5,8-naphthalene-tetracarboxylic-dianhydride

PCM: polarizable continuum model

PCR: polymerase chain reaction

PNA: peptide nucleic acid

PTCDA: 3,4,9,10-perylene-tetracarboxylic-dianydride

QM: quantum mechanics

RNA: ribonucleic acid

ss: single-stranded

STM: scanning tunneling microscope

T: thymine

TB: tight-binding

U: uracil

WBL: wide band limit

XC: exchange-correlation

ACKNOWLEDGMENTS

I would like to thank my supervisor Prof. M. P. Anantram for his encouragement and guidance during my Ph.D stage. I would also like to thank our experimental collaborators Prof. Joshua Hihath, Dr. Juan Manuel Artés and Yuanhui Li at University of California, Davis for their experimental support and intensive discussion. I also appreciate the help from our theoretical collaborator Dr. Niranjan Govind at Pacific Northwest National Laboratory. Special thanks to my committee members - Prof. David Masello from Department of Chemistry, Prof. Lih Y. Lin and Prof. Georg Seelig from Department of Electrical Engineering. My gratitude also goes to my colleagues and friends who have been helping me with advice on both work and life. Last but not least, I want to thank my family for being supportive for so many years.

DEDICATION

to my family

Chapter 1
INTRODUCTION

Nucleic acids have two broad categories, the deoxyribonucleic acids (DNA) and the ribonucleic acids (RNA), both of which are essentially important biological molecules. In this chapter, I will first describe the physical structure of DNA and RNA, and the motivation that we study the electrical transport through them. I will also give a big picture about the current research status in this field. I will cover the objectives and topics of my thesis as well.

1.1 Physical structure of nucleic acids

Both DNA and RNA consist of various numbers of nucleotides and come up with different conformations. The nucleotides are the building blocks of the nucleic acids and form the basis of the large number of sequences. Each nucleotide is made of three parts - a phosphate group, a sugar ring and a nucleobase [1]. The alternating phosphate groups and sugar rings connect together through the strand and constitute the backbones, which provide stability to the structure. In Figure 1.1, we show an example of a DNA building block.

In Figure 1.1, the phosphate group is shown as the blue block, the sugar ring in yellow and the nucleobase in pink. The phosphate-sugar backbone is common for all normal DNA nucleotides, but the nucleobases may vary from nucleotides to nucleotides. For DNA, there are two groups of nucleobases - the purines, which include adenine (A) and guanine (G), and the pyrimidines, which include thymine (T) and cytosine (C). In the example of Figure 1.1, the nucleobase we show is G. In Figure 1.2, we show the four nucleobases (A, G, T, C) terminated with hydrogen atoms. The atoms in the rings are numbered.

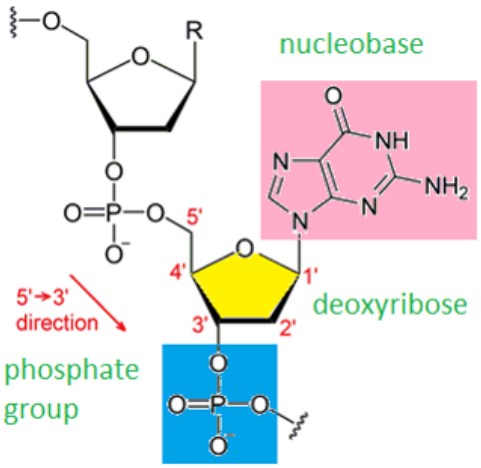
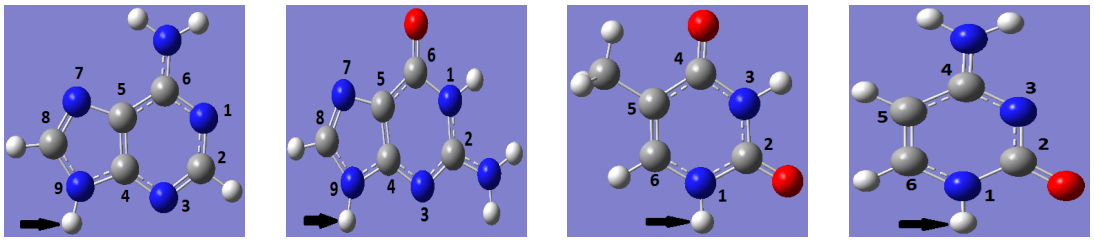


Figure 1.1: An example of a DNA building block, which is made of a phosphate (blue region), a sugar ring (yellow region), and a nucleobase (pink region) [2].

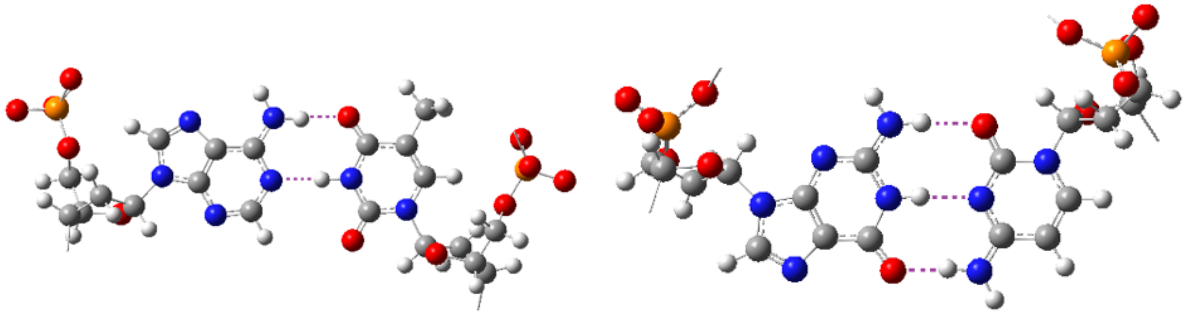


(a) (b) (c) (d)

Figure 1.2: Structures of the hydrogen atoms terminated DNA nucleobases. (a) A; (b) G; (c) T; and (d) C. Black arrows indicate the hydrogen atoms that replace the backbones.

DNA is a polymeric chain. Usually, a DNA molecule exists in nature with a double-stranded (ds) helical structure [3]. The two strands align in antiparallel direction, with the $5'$ end terminated with a phosphate group and the $3'$ end terminated with a hydroxyl group. In a normal dsDNA, A is always paired with T with two hydrogen bonds and G is always paired with C with three hydrogen bonds - this is the well known complementary base pairing, as shown in Figure 1.3. The complementary pairing renders DNA two unique and appealing properties - the molecular recognition and the self-assembly [4, 5]. Molecular recognition

refers to a molecule's ability to interact with other molecules or substrates through selective bonds, and self-assembly is a process in which a molecular component spontaneously organizes itself in supermolecular aggregates and forms an organized structure under proper situations. With these two properties, DNA molecules have a large potential to be used as active components in nanodevices.



(a) (b)

Figure 1.3: Illustration of the complementary base pairing. (a) $A : T$ complementary base pairing; (b) $G : C$ complementary base pairing.

In living cells, the DNA double helical structures mainly adopt three forms - B-form, A-form and Z-form, among which B-form is the most abundantly found one in nature. Both B-form and A-form DNA molecules are right-handed helices. In B-form DNA, the twist angle between two nearest stacking base pairs is 36°, indicating that to construct a unit cell to form a periodic structure analogous to solid state systems, ten bases are needed. The distance between two nearest neighboring stacking bases is around 3.4 $\mathring{A}$ and the diameter of the helix is about 2 nm. Different from B-form DNA, an A-form DNA has a twist angle of about 32.7°, the distance between two nearest neighboring bases on one strand is around 2.3 $\mathring{A}$, and the diameter of the helix is approximately 2.3 nm. In addition, a B-form DNA has a tilt angle between the base pairs and its axis of only $-1.2°$, while that of an A-form DNA is 19°. In Figure 1.4, the side and top views of a 13-base pair DNA in B-form and A-form are shown.

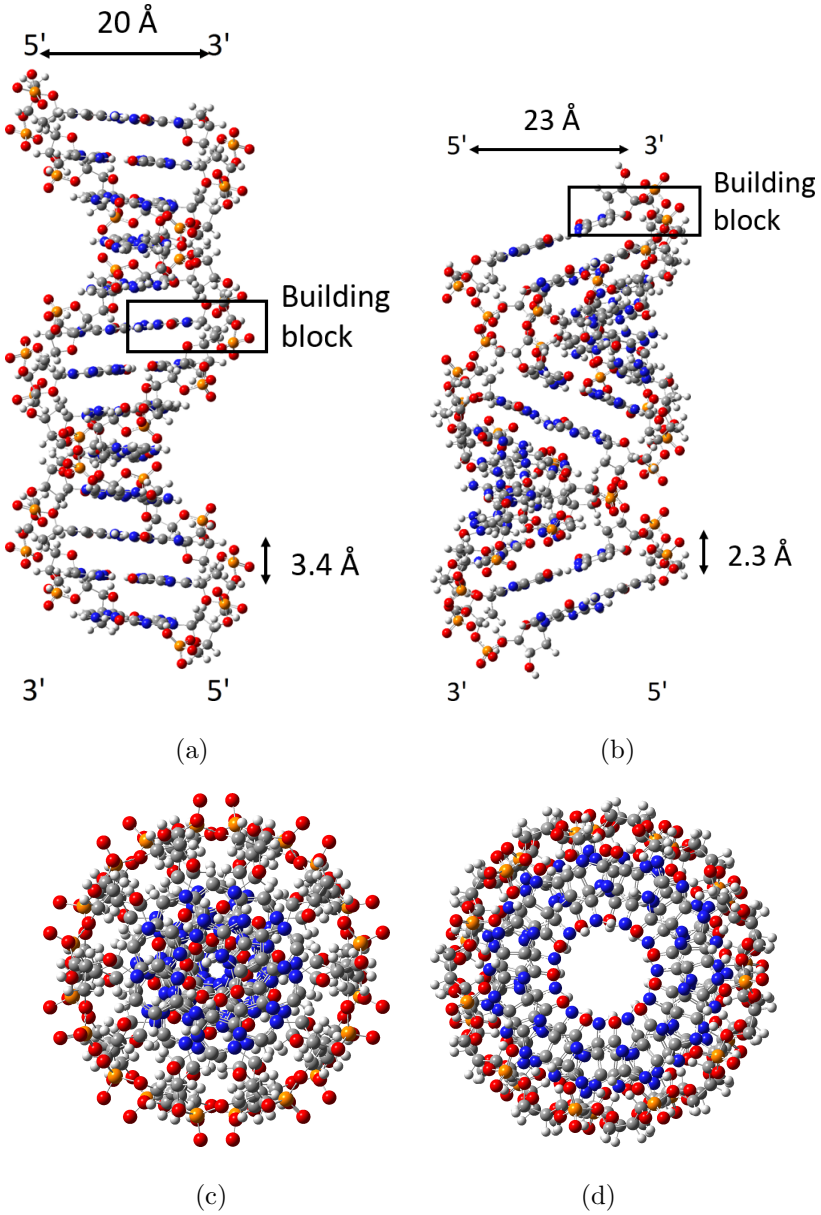


Figure 1.4: Side and top views for a 13-base pair DNA in B-form and A-form. (a) Side view in B-form; (b) side view in A-form; (c) top view in B-form; (d) top view in A-form. The building block, distance between the two nearest neighboring stacking bases, and diameter of the helix in the side views of B-form and A-form DNA are labeled.

Normally, a B-form DNA has more than 13 water molecules nearby while A-form DNA has about $5-10$ water molecules [6]. Change in humidity can cause geometry transitions between a B-form DNA and an A-form DNA. Unlike B-form and A-form DNA, Z-form DNA has a left-handed double helical structure, with a twist angle of 60° per two base pairs, a stacking distance between the two nearest neighboring bases of 3.7 $\mathring{A}$ and a helix diameter of 1.8 nm.

Similar to DNA, RNA is also composed of different numbers of nucleotides with three components - a phosphate group, a sugar ring and a nucleobase, as shown in Figure 1.5. Despite the similarity to DNA, the structure of RNA is different from DNA in a few ways: (1) RNA has a $2'-OH$ in the sugar ring, which makes it more active and less stable. (2) Instead of the T base in DNA, RNA has the nucleobase uracil (U). Together with A, G and C bases, U forms the building blocks for the RNA strand. In Figure 1.6, we show the structure of the U base. (3) Similar to DNA, RNA can also adopt the double helical structures, such as A-form and A′-form. However, the conformations of RNA molecules are much more flexible and complex than DNA and they usually come in single strands and can adopt various complicated structures.

1.2 Why is charge transport through nucleic acids important?

Nucleic acids are best known as genetic carriers in many biological systems [6]. During the last few decades, the possibility of using nucleic acids in nanotechnology has been gaining attention from chemists, physicists, as well as engineers. In recent years, the study of the electronic structures of the nucleic acids have led to applications in many areas such as molecular devices, the third generation sequencing techniques and disease detection, making the understanding of the electron motion along nucleic acids a meaningful one.

Over the last few decades, the semiconductor industry has followed a path of decreasing feature size. If this trend continues, the feature size of the electronic components will approach the atomic limit, at which point the functionalities of the device will be limited by some fundamental problems, such as charge leakage [7]. Molecular electronics, which uses

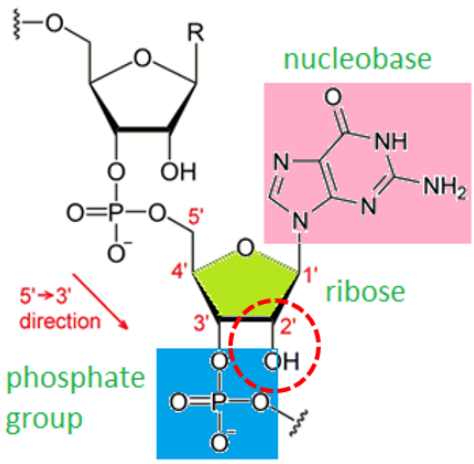


Figure 1.5: An example of a RNA building block, which is made of a phosphate group (blue region), a ribose sugar ring (yellow region), and a nucleobase (pink region) [2]. The hydroxyl group has been circled with the red dashed line to highlight the difference from the structure of a DNA building block.

a single molecule as an active component, may be considered as a candidate at this scale. The first molecular device can be traced back to 1974 when Aviram and Ratner proposed a current rectifier consisting of a donor π system and an acceptor π system [8]. Since then, molecular electronic devices have been increasingly attracting attention from researchers working on both top-down and bottom-up approaches [9, 10, 11]. Molecular electronics that have the potential to act as photodiode [12], transistors [13], rectifier [14] and switches [15] have been experimentally identified.

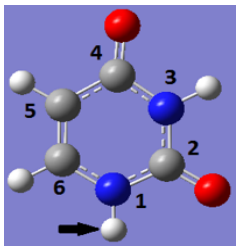


Figure 1.6: Structure of the hydrogen atom terminated nucleobase, U. Black arrow indicates the hydrogen atom which replaces the backbone.

Among various molecular candidates, DNA has the attractive features of molecular recognition and self-assembly as aforementioned. The study of the electronic properties of DNA started as early as in 1960s, when Eley and Spivey proposed that the dsDNA with paired bases arranged along the helix axis with an interplane spacing of 3.4 $\mathring{A}$ could behave like a one-dimensional aromatic crystal, and the π orbital in the stacking bases could overlap effectively [16]. Later experiments demonstrated that the double-helix could support charge migration over distances ranging from tens to hundreds of angstroms [17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. From an electrical device perspective, double barrier resonant tunneling structures [27], spin specific electron conductor [28, 29], field effect transistor [30], trinary logic [31], optomechanical molecular motor [32], negative differential resistance [33], detection of lesions by repair proteins [34] and doping [35] are in principle possible with DNA.

Along with the studies from the electronics perspective, the intensive investigation of long-distance charge migration through nucleic acids is further motivated by their various applications in other areas such as medical sciences. Recent work shows that the understanding of the electronic properties of nucleic acids can bring new techniques for the third generation sequencing technologies. Current sequencing methods are mainly based on light emission [36], which requires polymerase chain reaction (PCR) amplification. An alternative sequencing idea is to measure the transverse current through a single-stranded (ss) DNA as it translocates a nanogap or nanopore [36, 37, 38, 39, 40, 41, 42]. The distinguishable electronic structures of different nucleotides result in different electrical/ionic currents, which can be used as electronic signatures for the nucleotides and make the third-generation sequencing technique possible.

Additionally, recent experiments also reveal the possibility of detecting diseases by measuring the electrical conductance of DNA strands with mutated bases [43, 44, 45]. Many diseases are associated with the mutation in nucleotides, which causes the electronic properties to be different from the properties of a normal one. This property can be used to tell the mutated DNA from a native one. The study of this area may open new perspectives in medical application.

Further advancements in these fields require the clear understanding of the charge transport mechanism through nucleic acids. These exciting topics stimulate an intense and continuing interest in biochemical, physics and engineering communities.

1.3 Current status

Charge migration through DNA is classified into two broad categories: charge transfer and charge transport [46]. Charge transfer refers to the transfer of charge between the donor and the acceptor intercalators placed at the two ends of DNA strands, while charge transport experiments involve measuring the current-voltage characteristics of a DNA molecule placed between two metal electrodes.

Currently, the charge transfer mechanism has been studied widely by both experimentalists and theorists, and the understanding of the transfer mechanism has reached a consensus. Based on the work of Barton and coworkers [47, 48], Giese and coworkers [19, 22, 49], Jortner and coworkers [50, 51, 52], et al., the overall charge transfer for a DNA strand is a multistep hopping process between G bases, with the G bases serving as stepping stones because of its lowest ionization potential (IP) among the four bases. On the basis of this picture, two principal mechanisms regarding charge transfer through DNA have been established. The first one is referred to as the superexchange mediation, which consists of a coherent tunneling between G bases separated by three or fewer $A:T$ base pairs. This mechanism is characterized by the exponential dependence of the transfer rate on the distance between donor and acceptor. Thermally induced hopping via A bases is involved when the number of $A:T$ base pairs is more than three.

While the studies for charge transfer mechanism seems to have reached a general picture, direct measurements for the conductance of DNA molecules remains a controversial issue. The first direct electrical transport measurement was performed in 1998 [53]. Since then, great progresses have been made. Nowadays, some basics about charge transport through DNA have been understood experimentally, such as the barrier effect of the $A:T$ base pairs [54]. However, as far as modeling goes, there is little work that can match experiments across

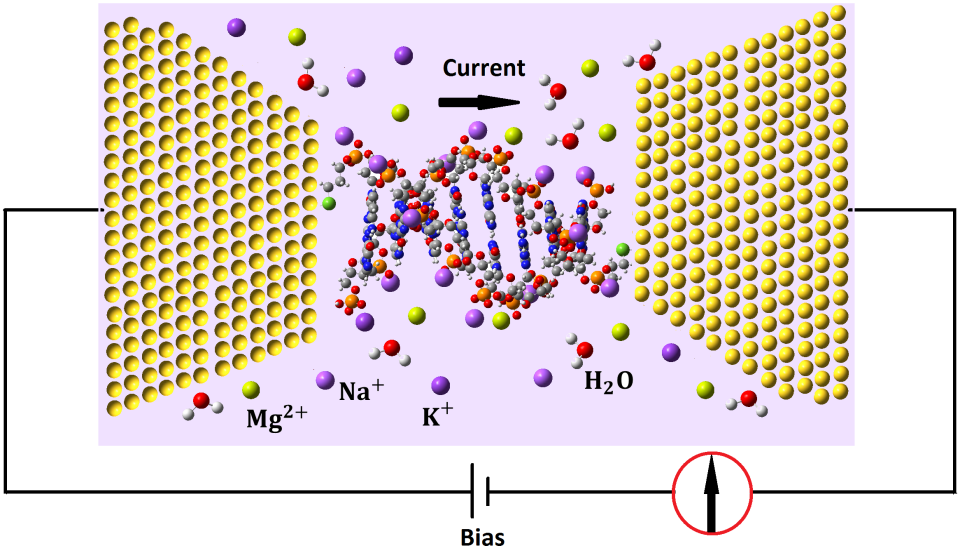


Figure 1.7: A metal-molecule-metal junction formed by a single DNA molecule bounded to two gold electrodes. Also present are water molecules, three types of counterions, Na^+, K^+ and Mg^{2+}, and the thiol groups which connect DNA to the gold surfaces.

sequences. Explaining experiments and making predictions remain a challenge. Compared with nanoengineered materials such as nanotubes and nanowires that have comparable dimensions, there are lots of confusions in this area. Most of the transistors today have a channel length of less than 35 nm (the cross section is wider), and there are many modeling tools available to model them. However, the problem of modeling DNA is much more difficult. What are the reasons?

1.4 Challenges

In Figure 1.7, we show the fundamental components of a molecular electronics made from DNA, which has two metal electrodes and one DNA molecule bridged across the electrodes with linker groups. Also present are water molecules, and three types of counterions, Na^+, K^+ and Mg^{2+}.

The modeling of the conductivity of such a DNA device is a complex problem because many factors contribute to the charge transport process. These factors include: (1) the

DNA molecule itself. The length, sequence and conformation (A-form, B-form or Z-form) of the molecule play an intrinsic role in determining the conducting properties. For the rigid transistors, the structures are much simpler. However, for DNA, the combination of the four different building blocks can give a very large number of sequences with various lengths in different conformations. This is also one of the interests in this area. The sequence is important [49, 55] because of the different IPs of the four nucleobases [56] and sequence dependent conduction [54, 57]. Conformation changes (lattice vibrations/deformation) can also modify the conductivity of DNA molecules. Theoretical calculations have confirmed that changes in the distance and the angle between consecutive bases can cause variability in conduction channels by influencing the hybridization between the π orbitals of adjacent bases [27]; (2) interaction with the electrodes, which is determined by the nature of electrical contact [58], and connection configuration between the molecule and the contacts [59, 60]. In a molecular electronics, the electrodes may be made of metals such as gold and platinum. To connect the molecule to the electrode, the molecule is usually terminated with chemical groups such as $-SH$, $-NH_2$ and $-COOH$. For a gold electrode, one of the most commonly used terminating groups is the thiol group, $-CH_2 - CH_2 - CH_2 - SH$. The materials of electrodes, terminating group, and the connection configuration between them determine the coupling strength between the molecule and the electrode, and the alignment of molecular energy levels of the molecule with respect to the Fermi levels of the electrodes [61]; (3) interaction with the surrounding environment. As a floppy molecule, the melting temperature of the DNA is only $60 - 80°$C, depending on the length and sequence. The soft-mattered nature of DNA makes its conductivity susceptible to the local environment, including the effect of impurities, electron-phonon due to the lattice vibrations, and electric field from moving water molecules and ions when the molecule is placed in solution. In wet DNA, water molecules are critical in influencing the molecular structure and hence the electronic properties. For example, an A-from DNA has less water molecules around than a B-form DNA [6]. The change in the hydration level may result in the structural transition between an A-form and a B-form DNA and hence different conductances. At different hydration levels,

different local density of states (DOS) are obtained [62]. In addition, because the backbone of a DNA molecule is negatively charged in its phosphate groups, it is believed that the molecule is surrounded by cations such as Na^+, K^+, Mg^{2+} and H_3O^+. Previous studies have shown that variations in cation position [62] and type [63] can modulate the HOMO and LUMO levels. Experiments in Reference [64] have demonstrated that the conductance of DNA molecules strongly depends on the conformation and local environment.

Even if one only focuses on modeling transport inside DNA (such as Figures 1.4(a) and 1.4(b)), it is not an easy problem. There are two length scales when we model the conductance inside DNA. The first one is the strong bonding between neighboring atoms, as in silicon or carbon nanotube, where the bonding distances are usually $1 - 2$ $\mathring{A}$. Even if these are the only bonds that one has to model, the problem is more difficult than that in silicon/carbon nanotubes, because usually a DNA molecule is not periodical. Besides, there is a second distance one should take into account - the distance between the two stacking bases, which is about 3.4 $\mathring{A}$ for B-form DNA and 2.3 $\mathring{A}$ for A-form DNA. Since most of the atomic wavefunctions typically decays quite a bit at these scales, this distance introduces a weak yet important electronic coupling into the system. With these two distances, we have a strong regime and a weak regime, making the modeling of transport inside a DNA challenging and the match of the experiments not straightforward.

As a result of these factors, there are large uncertainties in the experimental results. Consequently, the conductivity of DNA has been a controversial problem for years and the reported conducting properties vary from experiments to experiments. Researchers have found that DNA can either be an insulator [53, 65], a semiconductor [66, 67], an ohmic conductor [68, 69], or a superconductor [70, 71], depending on a variety of conditions. Furthermore, these factors make it much more challenging to model DNA than modeling transistors. Our level of understanding of this area is significantly less mature compared to other more rigid molecular structures.

1.5 Objectives of this thesis

This thesis is concerned with the modeling of electrical transport through nucleic acids including both DNA and RNA, which are bridged in between two metal electrodes. The objectives are two-folded. Firstly, a model and computer codes based on the Green's function have been developed to calculate the transmission and conductance through the two-probe system. This model has been demonstrated to be able to explain the observations in experiments. Secondly, using this model, various electronic phenomena through nucleic acids have been characterized, which is useful for understanding, controlling and engineering the conductance with desired requirements. The contribution from a few factors to the conductance of nucleic acids have been studied, including sequence (specially, the barrier role of T base and $A:T$ base pair in G-rich strands), decoherence, methylation of the cytosine base, conformation (B-form and A-form), water molecules and RNA:DNA hybrids. We use relatively large computational models that are based on the combination of *ab initio* calculations and Landauer-Büttiker formalism.

The content of this thesis is based on the following papers:

1. Jianqing Qi, M. P. Anantram, The influence of water on charge transport through DNA, in preparation.

2. Jianqing Qi, Niranjan Govind, M. P. Anantram, The role of cytosine methylation on charge transport through a DNA strand, accepted.

3. Juan Manuel Artés, Yuanhui Li, Jianqing Qi, M. P. Anantram and Joshua Hihath, Conformational gating of DNA conductance, submitted.

4. Jianqing Qi, M. P. Anantram, (Invited) Electron flow in GC8 DNA strand, ECS Transactions, vol. 64, no. 12, pp. 1-9, 2014.

5. Jianqing Qi, M. P. Anantram, Charge transport through methylated DNA strand, in MRS Proceedings, vol. 1689, 2014.

6. M. P. Anantram and Jianqing Qi, Modeling of electron transport in biomolecules: Application to DNA, International Electron Device Meeting (IEDM), 2013 IEEE International,

pp. 32.3.1 - 32.3.4.

7. Jianqing Qi, Neranjan Edirisinghe, M. Golam Rabbani, and M. P. Anantram, Unified model for conductance through DNA with the Landauer-Buttiker formalism, Physical Review B, vol. 87, no. 8, pp. 085404, 2013.

8. Jianqing Qi, Md. Golam Rabbani, Suranga Edirisinghe, and M. P. Anantram, Transport of charge in DNA heterostructures, Nanotechnology (IEEE-NANO), 11th IEEE Conference on, pp. 487-491, 2011.

Chapter 2

METHODS AND MODELS

In order to investigate the electronic properties of devices at nanoscale, several theoretical tools have been developed during the last few decades. The modeling of electrical transport in DNA molecules connected to contacts is in early stages, compared to solid-state nano-transistors. The objective of this chapter is to review and present the approaches for transport problems in a molecule bridged across two metal contacts.

2.1 Methods overview

Transport problems concern with the flow of electrons in response to the applied biases. A few methods are available to model the charge transport through a general system, including the Boltzmann transport theory, the drift-diffusion model, the rate equation and the quantum transport approach. The Boltzmann transport theory has a long history, which was first proposed in 1870s for a dilute gas with non-interacting particles [72]. Later on, it was extended to simulate the electron/phonon transport in solids [73, 74]. The Boltzmann transport theory is described by the following equation,

$$\frac{\partial f}{\partial t} + \frac{\vec{p}}{m}\cdot\bigtriangledown f + \vec{F}\cdot\frac{\partial f}{\vec{p}} = (\frac{\partial f}{\partial t})_{coll.} \tag{2.1}$$

where f is the probability density function, t is the time, $\vec{p}$ is the momentum, m is the mass of the particle and $\vec{F}$ is the external force on the particle. "coll." on the right-hand side stands for "collision", describing the scattering events.

Another method is the drift-diffusion model, which can be derived from the Boltzmann equation [75]. For an electron, the current density given by the drift-diffusion model is,

$$J_n = qn\mu_n\varepsilon + qD_n\frac{dn}{dx} \tag{2.2}$$

where q is the charge for an electron, n is the electron density, μ_n is the mobility of the electron, ϵ is the electric field, D_n is the diffusivity of the electron, and x is the distance of the electron movement. Similarly, the current density for a hole is,

$$J_p = qp\mu_p\varepsilon - qD_p\frac{dp}{dx} \tag{2.3}$$

where p is the hole density, μ_p is the mobility of the hole, and D_p is the diffusivity of the hole. In Equations. 2.2 and 2.3, the first term is the drift component driven by the electric field, and the second term is the diffusion component due to the gradient of the particle.

The drift-diffusion model shown in Equations. 2.2 and 2.3 are probably not a suitable method for the nucleic acids problem, because the nucleic acids have various sequences and it is difficult to come up with one drift-diffusion equation that could handle all the cases. Also, there might be tunneling involved in the electron flow through the nucleic acids, making the application of both the drift-diffusion model and the Boltzmann transport theory (2.1) in modeling transport through nucleic acids infeasible.

The third method is the rate equation for hopping transport, which has the form [76],

$$\frac{\partial f_i}{\partial t} = \sum_i [f_j(1-f_i)W_{ji} - f_i(1-f_j)W_{ij}] \tag{2.4}$$

where f_i is the electron occupation number at site i, or more generally, energy level i. W_{ij} is the transition probability. This method is valid only when the energy levels of the system are very sharp. There has been work that models DNA using a rate equation involving hopping and tunneling [50, 51, 52, 77, 78], depending on the sequence involved.

In most cases, when modeling the electrical transport through nucleic acids, we use quantum transport approach, which in principle can handle all the aforementioned cases. A detailed description of the quantum transport equation based on the Landauer-Büttiker formalism will be presented later in Section 2.3. In this section, we highlight the essentials of the method used in this thesis.

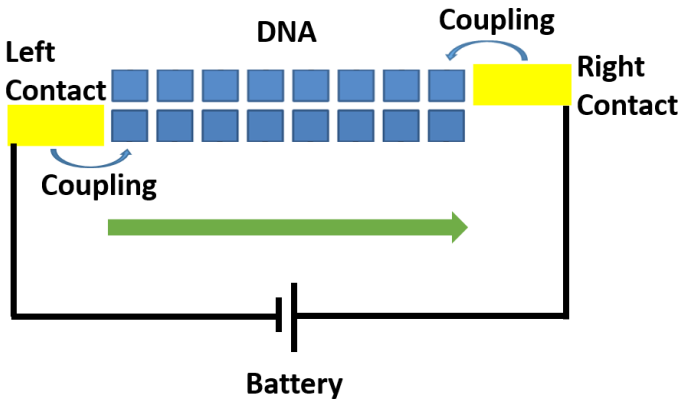


Figure 2.1: Sketch of a typical DNA device, which involves a DNA molecule at nanoscale and two macroscopic electrodes. The blue blocks represent DNA bases in the double-helical structure, and yellow blocks represent the metal electrodes.

In Figure 2.1, we show the sketch of a typical DNA device, which involves a DNA molecule at nanoscale and two macroscopic electrodes. The blue blocks represent DNA bases in the double-helical structure, and yellow blocks represent the metal electrodes.

In order to get the information of electron flow through such as a molecular device, we need to know the electronic properties of DNA, the effects due to the macroscopic contacts, and the coupling between the molecule and contacts. A few groups [27, 46, 58, 63, 79, 80, 81, 82, 83, 84, 85, 86] have examined the conductance of DNA molecules lying between two metal contacts with the coordinates frozen based on the Landauer-Büttiker formalism.

The modeling of conductance in wet DNA lying between metal contacts is a more difficult problem because the surrounding environment (water molecules, ions and conformation) fluctuates with time. In this case, recent work by References [87, 88, 89, 90] uses a combination of quantum mechanics (QM)/molecular mechanics (MM) to evaluate the role of fluctuating environment on DNA transport. These calculations freeze the location of the atoms at each sampling point in the molecular dynamics (MD) simulation, and calculate the phase-coherent conductance using the Landauer-Büttiker approach.

In review of the previous work regarding the charge transport problem in nucleic acids, it is found that most of the works are based on the simplified systems, where an ideal

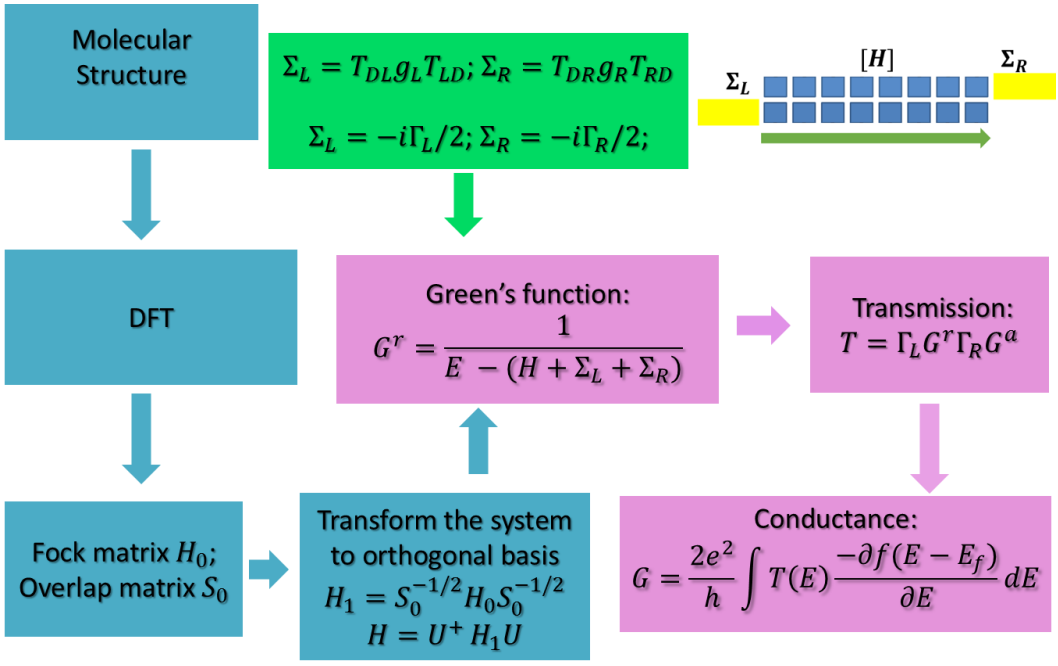


Figure 2.2: Flow chart of the simulation method for coherent transport, which is based on DFT, the Landauer-Büttiker picture and the Green's function approach.

$poly\{G:C\}/poly\{A:T\}$ system is often considered, or only a single or very few number (typically $3-4$) base/base pair are included. Experimentally, it is very difficult to measure the conductance through short strands at such length scales.

My work is inspired by device modeling approaches over the last twenty years to understand transport in nanostructures. I try to reproduce and explain the experimental trends by using quantum transport with decoherence. To begin with, I first show the flow chart of the simulation method for coherent transport in nucleic acids in Figure 2.2. This method is based on the density functional theory (DFT), the Landauer-Büttiker picture and the Green's function approach.

The methods shown in Figure 2.2 involves two major steps: (1) obtaining the Hamiltonian which describes the electronic structure of the molecule. To get a comprehensive understanding of the molecule, an *ab initio* calculation is often involved. In my model, this steps involves the generation of the molecular structure, implementing *ab initio* calcu-

lations, and transforming the system into an orthogonal basis. If one is only interested in a few energy levels such as the highest occupied molecular orbital (HOMO) or the lowest unoccupied molecular orbital (LUMO), a model Hamiltonian is preferred; (2) carrying out the transport calculation to obtain the transmission and conductance, which includes the modeling of the effect from the electrodes, the interaction between the molecule and the electrodes, and the computation of the transmission/conductance using Green's function and the Landauer-Büttiker formalism. In the remainder of this chapter, I will discuss these procedures in detail.

2.2 Hamiltonian of the nucleic acids

The evaluation of the Hamiltonian for the molecule (also referred to as 'device molecule') lying between the metal contacts is a central problem in molecular devices. Currently, two broad classes of theoretical approaches are used to establish the Hamiltonian of a device molecule - the *ab initio* calculations and the model Hamiltonian methods [91]. *Ab initio* approaches [27, 63, 84, 92] are powerful tools to obtain the accurate description of the electronic structure for a molecule, and have been extensively used in a wide variety of quantum chemistry/physics problems. The drawback of the *ab initio* approach is the expensive computational cost, the high requirements of memory and disk space. Also, the *ab initio* approaches do not predict the correct HOMO-LUMO gaps in most cases. So, they must be used with care in studying transport properties. As an alternative, the model Hamiltonians that account for only one or two energy levels on each base and their interactions with energy levels in neighboring bases [58, 80, 81, 82, 83, 84, 85, 86] have found wide spread use. In the model Hamiltonian approach, the HOMO-LUMO gaps and Hamiltonian parameters are required as an input. In my thesis, both the *ab initio* calculations and the model Hamiltonian are considered.

2.2.1 Full Hamiltonian from ab initio calculations

The term '*ab initio*' means 'from the beginning' in Latin. In quantum chemistry and quantum physics, it means obtaining the information about the electronic structure of an atom, molecule or material by solving the Schrödinger's equation within a set of approximations without fitting the parameters to experimental data. The detailed derivation and development of *ab initio* algorithms can be found in many quantum chemistry/physics books and papers [93, 94], and are beyond the scope of my thesis. Here I briefly review a few key concepts of the *ab initio* method based on Reference [93].

The time-independent Schrödinger's equation is,

$$\mathbf{H}\Psi(\mathbf{R},\mathbf{r}) = E\Psi(\mathbf{R},\mathbf{r}) \tag{2.5}$$

where $\mathbf{H}$ is the Hamiltonian of the many-body system, Ψ is the many-body wavefunction which is a function of the positions of the nuclei $\mathbf{R}$ and electrons $\mathbf{r}$, and E is the total energy of the system. To solve Equation 2.5 exactly is a daunting task. A good approximation is the Born-Oppenheimer approximation, which is based on the fact that the nucleic are much heavier than electrons (the mass of a proton is ~ 1836 times larger than that of an electron). With a much larger mass, the nuclei travel much more slowly than the electrons and can be viewed as stationary with the electrons moving around. With this assumption, the Born-Oppenheimer approximation neglects the coupling between the nuclear and the electronic motions, which allows the wavefunction to be written as the product of the nuclear and electronic functions,

$$\mathbf{H}\Psi(\mathbf{R},\mathbf{r}) = \Psi_n(\mathbf{R})\psi(\mathbf{R},\mathbf{r}) \tag{2.6}$$

where $\Psi_n(\mathbf{R})$ denotes the M - nuclear wavefuction and $\psi(\mathbf{R},\mathbf{r})$ is the wavefunction for the N electrons. Within the Born-Oppenheimer approximation, the Hamiltonian of the N electrons is,

$$H_e = -\sum_{i=1}^{N} \frac{\hbar^2}{2m} \nabla_i^2 + \frac{1}{2} \sum_{i<j}^{N} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{N} V_{ext}(\mathbf{r}_i) \tag{2.7}$$

where the first term is the kinetic energy of the electrons, the second is the electron-electron interaction energy, and the third is the interaction due to the external potential, which is usually the summation of contribution from the M ions, $V_{ext}(\mathbf{r}) = \sum_{A=1}^{M} \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|}$.

The electron-electron interaction term in Equation 2.7 introduces complexity to obtain the exact solution of the equation, by making the many-body problem not separable into single-particle equations. Only very few extremely simple systems, such as the hydrogen atom, can be solved exactly with Equation 2.7. To tackle the many-body problem, many methods have been proposed. The simplest one is the Hartree-Fock (HF) method, which is based on the expansion of the wavefunction in Slater determinants and neglects the correlation between electrons. To account for the electron correction, several post-HF methods such as the configuration interaction (CI), coupled cluster (CC), Møller-Plesset perturbation theory (MP2, MP3, MP4, MP5, MP6) have been developed.

DFT is a different method from the HF theory, in which the electron correlation effect is included with the electron density. The establishment of DFT is based on the first and the second Hohenberg-Kohn theorems, which state that: (1) the electron density $\rho(\mathbf{r})$ determines the ground state properties of the system; and (2) the electron density that minimizes the energy is the density for the ground state. The second theorem is a variational principle.

With these theorems, we have the Kohn-Sham equation,

$$[-\frac{\hbar^2}{2m} \nabla^2 + V_{eff}(\mathbf{r})]\psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \tag{2.8}$$

This is a single-particle equation, which reduces the complexity of the many-body problem in Equation 2.7 greatly. $V_{eff}(\mathbf{r})$ is the effective single-particle potential with three components,

$$V_{eff}(\mathbf{r}) = V_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r\prime})}{|\mathbf{r} - \mathbf{r\prime}|} d\mathbf{r\prime} + V_{XC}[\rho(\mathbf{r})] \tag{2.9}$$

with

$$\rho(\mathbf{r}) = \sum\nolimits_{i=1}^{N} |\psi_i(\mathbf{r})|^2 \tag{2.10}$$

The second term in Equation 2.9 represents the electron-electron Coulomb repulsion, which refers to the Hartree term. The third term $V_{XC}[\rho(\mathbf{r})]$ is the exchange-correlation (XC) potential, which includes the effect of electron correlation, $V_{XC}[\rho(\mathbf{r})] = \frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})}$. If the XC potential is given, the Kohn-Sham equation can be solved in a self-consistent manner with an initial guess of the electron density $\rho(\mathbf{r})$.

A central problem with DFT is the evaluation of the XC functional. In most cases, this term is unknown and has to be derived with a set of approximations. A few widely used approximations are the local-density approximation (LDA), the local spin-density approximation (LSDA), and the generalized gradient approximation (GGA). Modern DFT also concerns with the hybrid functionals, among which B3LYP is the most popular one.

Nowadays, quite a few quantum chemistry and solid-state physics computer programs are available to carry out *ab initio* calculations, such as GAMESS [95], VASP [96], Q-Chem [97], ADF [98], SIESTA [99], Gaussian [100] and NWChem [101]. In my thesis, I mainly use the Gaussian 09 software package [100] for the DFT calculations.

As shown in the flow chart 2.2, the first step of the computation for DFT calculations is to generate the molecular structure. For nucleic acids involved in single-molecule conductance experiments, the molecules usually consists of hundreds of atoms, sometimes even more than one thousand atoms. In my work, the atomic coordinates of most of the molecular structures are generated with the Nucleic Acid Builder (NAB) software package [102].

The next step is to use quantum chemistry to calculate the Hamiltonian and overlap matrices. In this step, the choice of XC functionals and basis functions are critical as the results from DFT calculations usually depend on them. After the Hamiltonian and overlap matrices H_0 and S_0 are obtained, we could further transform them into an orthogonal basis using Löwdin transformation [46, 103],

$$H_1 = S_0^{-\frac{1}{2}} H_0 S_0^{-\frac{1}{2}} \tag{2.11}$$

Each diagonal sub-block of H_1 corresponds to the representation of a single nucleobase. A unitary transformation is further performed,

$$H = U^{\dagger} H_1 U \tag{2.12}$$

where U is a block diagonal matrix. To obtain U, we first diagonalize every diagonal sub-block of H_1 and then arrange the eigenvectors in the order of DNA bases. While the diagonal blocks of H_0 are full matrices, the diagonal blocks of H are diagonal matrices and the dimension of the diagonal block matrices is equal to the number of orbitals used to represent that base. Physically, the diagonal blocks of H correspond to the localized energy levels of each DNA base and the off-diagonal blocks correspond to interactions between different bases. If we look at the values of the off-diagonal elements, we find that the interaction corresponding to energy levels at the two nearby bases is large while that corresponding to two far away bases is relatively small. From our calculation, the hopping integral for two Gs in the first nearest neighboring positions in a B-form DNA is as large as 100 meV and that for two Gs in the second nearest neighboring positions is only a few milli electron volts.

Because DFT calculation for long DNA strands is very time-consuming, researchers alternatively first carry DFT calculations on short strands (typically two to five base pairs), and extract the Hamiltonian and overlap matrices for the bases and their interactions with neighbors. One can then construct a larger matrix corresponding to the Hamiltonian for a longer strand from the sub-Hamiltonians obtained from the aforementioned calculations on short strand. When constructing the matrices, one considers only the interaction between the nearest neighboring bases, and a block tridiagonal matrix representation of the Hamiltonian/overlap matrices is obtained [104, 105]

The representation of H in Equation 2.12 can be truncated to obtain the parameters, such as the onsite energies of a single DNA base and the hopping integrals between two

neighboring DNA bases for the interested energy levels, to build the model Hamiltonian.

2.2.2 Model Hamiltonian

Despite the great success of DFT in both quantum chemistry and quantum physics, they have downside - computational costs and bandgap values are prime examples. Also, the large-scale Hamiltonian matrix from the DFT calculations consequently results in an increased amount of computational time in the Green's function calculation stage. Thus, the application of DFT is usually limited to small systems. Although with the rapid development of the computer technology, researchers are able to obtain the electronic properties of a relatively large system, DFT calculations for systems containing thousands of atoms are still a daunting problem.

An alternative approach to the establishment of the Hamiltonian is to construct a matrix with a few dominant energy levels (usually around the HOMO or LUMO levels) of the system. This is the model Hamiltonian approach, which is also referred to as the tight-binding (TB) model. In a TB model, only a few charge transport parameters, such as the onsite energies and hopping amplitudes, need to be considered [58, 79, 80, 81, 82, 83, 85, 86, 106]. The advantage of this method is that it may be able to capture the essential physics of a system with a much lower computational cost compared with the *ab initio* approaches, and opens up the possibility of studying large systems with a huge number of atoms. However, this model depends sensitively on the choice of TB parameters which are strongly influenced by the environment. A popular approach is to combine *ab initio* approaches with model Hamiltonian calculations and tweak parameters to understand experiments, *i.e.* one can first extract the charge transport parameters from *ab initio* calculations and then build a TB model.

The establishment of the TB model for DNA has gone through several stages [46, 107, 108]. The simplest TB model for DNA starts with the one-dimensional wire, which considers a base pair as an individual site and describes the coupling between two nearest neighboring sites using a hopping amplitude. A two-channel model was then proposed to account for the

distinction between two bases in a base pair and the orientation effect for hopping integrals. The backbone effect was also included with the fishbone model, which is based on the wire model but uses additional sites and hopping amplitudes to represent the backbone and its interconnections with the bases. Combining the strengths of the two-channel model and the fish model, a ladder model was built to account for the distinction between two bases within a base pair and the effect from the backbone. Sketches of the wire model, two-channel model, fishbone model and ladder model are shown in Figures 2.3(a), 2.3(b), 2.3(c) and 2.3(d), respectively.

In our work, we analyze the TB parameters to deepen the understanding of the charge transport in nucleic acids. The TB model we adopt is from Reference [46] and is similar to the two-channel model discussed above. However, unlike the classic two-channel model, we also include the coupling between two bases in the second neighboring positions from the two complementary strands in this model. A sketch of the this model is illustrated in Figure 2.3(e). A detailed description of Figure 2.3(e) will be provided after the mathematical expression of the model is developed.

The mathematical representation of the TB model in Figure 2.3(e) can be obtained by truncating the unitary Hamiltonian H as shown in Equation 2.12,

$$H_{TB} = \sum_i \epsilon_i c_i^\dagger c_i + \sum_{\langle i,j \rangle} t_{i,j}^{intra} (c_i^\dagger c_j + H.c.) + \sum_{\langle i,j \rangle} t_{1i,j}^{inter} (c_i^\dagger c_j + H.c.) + \sum_{\langle\langle i,j \rangle\rangle} t_{2i,j}^{inter} (c_i^\dagger c_j + H.c.) \quad (2.13)$$

In the above equation, $c_i^\dagger$ (c_i) is the creation (annihilation) operator at the ith base, $H.c.$ is the Hermitian conjugate, ϵ_i is the onsite energy of the ith base, $t_{i,j}^{intra}$ is the intra-strand coupling between two nearest neighboring bases within one strand, $t_{1i,j}^{inter}$ is the inter-strand coupling between two nearest neighboring bases from the two complementary strands (i. e., the coupling for two bases within one base pair), and $t_{2i,j}^{inter}$ is the inter-strand coupling between two second nearest neighboring bases from the two strands. The notation $\langle i,j \rangle$ and $\langle\langle i,j \rangle\rangle$ in the summation depict the two bases in the first and second nearest position, respectively [46].

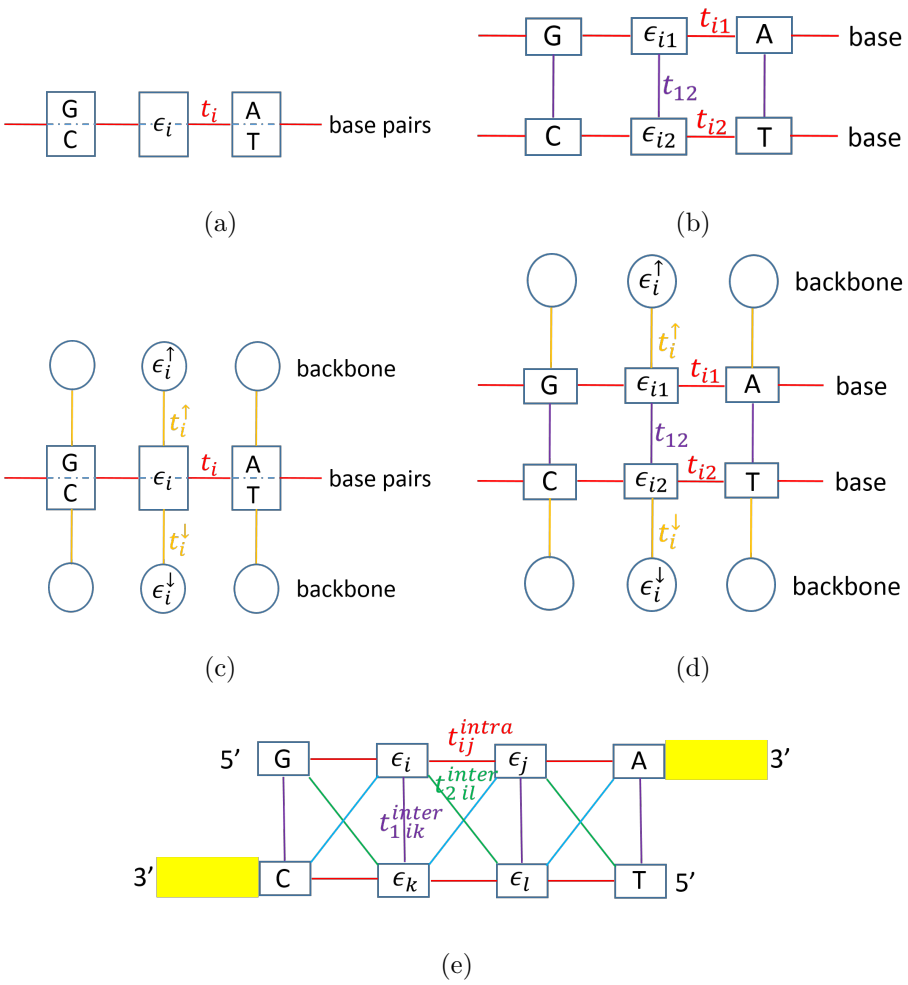


Figure 2.3: Five TB Hamiltonian models for a dsDNA. (a) 1-D wire model; (b) two-channel model; (c) fishbone model; (d) ladder model. These four models are from Reference [107]. Onsite energies and hopping integrals are marked here. (e) TB model for a dsDNA from Reference [46] used in our work. Different colors correspond to the onsite energies and hopping integrals between the two neighboring bases both within and between strands, with red color for the nearest intra-strand hopping, purple for the nearest inter-strand hopping, green for the second nearest inter-strand hopping from 5′ end to 5′ end, and blue for the second nearest inter-strand hopping from 3′ end to 3′ end. Yellow blocks represent the gold electrodes.

In Figure 2.3(e), we illustrate the interactions in the TB model with a four-base pair DNA strand. The solid lines represent the hopping between two neighboring bases both within and between the strands, with red for the nearest intra-strand hopping, purple for the nearest inter-strand hopping, green for the second nearest inter-strand hopping from $5'$ to $5'$ ends and blue for the second nearest inter-strand hopping from $3'$ to $3'$ ends. The TB parameters ϵ_i, $t_{i,j}^{intra}$, $t_{1i,j}^{inter}$ and $t_{2i,j}^{inter}$ have also been labeled in the corresponding position. As we will see in the later discussion, the inter-strand couplings for two second nearest neighboring bases, $t_{2i,j}^{inter}$, depends on the orientation of the strand. To mimic the direct conductance measurement experiment, we include two gold electrodes in the setup, as shown by yellow blocks at the $3'$ ends in Figure 2.3(e).

2.3 Transport calculation

As the dimension of an electronic device shrinks to the nanometer scale, the mean free path of the electron is comparable to or even much larger than the size of the device, leading to the occurrence of quantum mechanical effects. For example, the coupling between the device and the electrode, which is negligible in a macroscopic system, becomes important in determining the conductance of the device. At such a small scale, the classical Ohm's law breaks down and new theories for modeling the conductance of the devices, and explaining the observations from experiments are developed. A successful approach in this area is the Landauer-Büttiker theory, which has proved to be a computationally efficient and useful tool for charge transport in nanodevices. Within the framework of the Landauer-Büttiker theory, the conductance of the device is related to the transmission probability of the electron traveling through the system [109, 110]. In this section, I will present the approach for transport calculation in nucleic acids based on the Landauer-Büttiker formalism.

2.3.1 Landauer-Büttiker formalism

The energy level diagram of a typical molecular device formed by a metal-molecule-metal junction is depicted in Figure 2.4. Here, HOMO and LUMO are analogous to the valence

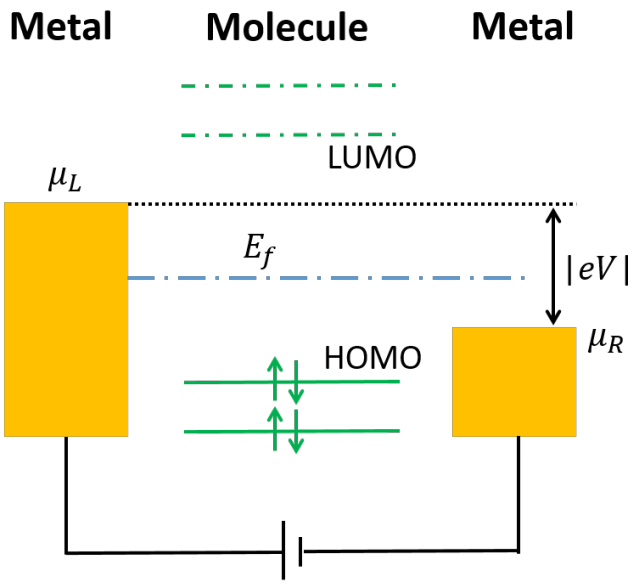


Figure 2.4: Energy level diagram of a molecular device formed by a metal-molecule-metal junction. Energy levels are marked with lines.

and conductance bands in a semiconductor. μ_L and μ_R are the chemical potentials of the left and right electrodes. The alignment of the chemical potentials of the two electrodes can be adjusted by the external battery, $\mu_L - \mu_R = qV$. E_f is the Fermi level, which usually lies in the HOMO-LUMO gap and depends sensitively on the surface condition and work function of the electrodes [61].

A general expression for the electric current through the molecular device shown in Figure 2.4 is given by the following equation [111],

$$I = \frac{2q}{h} \int_{-\infty}^{+\infty} T(E) \left[f(E - \mu_L) - f(E - \mu_R) \right] dE \tag{2.14}$$

where q is the charge of an electron, h is the plank constant, $T(E)$ is the transmission probability for an electron with energy E, and $f(E - \mu_{L/R})$ is the Fermi function describing the distribution of an electron with energy E in the left (right) electrode, which is given by,

$$f(E-\mu_{L/R}) = \frac{1}{\exp\left(\frac{E-\mu_{L/R}}{k_B T}\right)+1} \tag{2.15}$$

where k_B is Boltzmann's constant and T is the temperature.

For the low bias transport or the linear response, the difference between the chemical potentials of the two electrodes is rather small. The linear response conductance is,

$$G = G_0 \int_{-\infty}^{+\infty} T(E) \frac{-\partial f(E-E_f)}{\partial E} dE \tag{2.16}$$

where $G_0 = \frac{2q^2}{h} \simeq 7.75 \times 10^{-5} \Omega^{-1}$ is the quantum conductance.

At low temperature and when it is a good approximation to assume that the transmission does not change significantly at energies where $\frac{df}{dE}$ is appreciable, the linear response conductance can be further approximated as,

$$G \simeq \frac{2q^2}{h} T(E_f). \tag{2.17}$$

For a coherent process, where the electron transmits from the left electrode to the right electrode without any incoherent scattering, $T(E_f)$ is the total transmission for the electron to go from the left electrode to the right electrode. In decoherent transport, the electron may experience inelastic scattering due to electron-phonon interactions and the electrical field from the ambient environment, which is common for devices at room temperature. In this case, the electron can lose phase information to the environment and as a result, Equations 2.16 and 2.17 no longer hold. In Section 2.3.3, I will discuss a phenomenological approach - the Büttiker probes, to include the decoherence effect.

2.3.2 Green's function approach

The transmission coefficient in Equation (2.14) is usually found with the Green's function approach. In this section, I will briefly review the Green's function method with a two-probe system, and then discuss several approaches that evaluate the self-energy matrices to include the contact effect.

2.3.2.1 Green's function for a two-probe system

The retarded Green's function G^r of a molecular device is defined as,

$$(EI - H)G^r = I \tag{2.18}$$

Here, H is the Hamiltonian matrix, I is an identity matrix of the same size of H. With a non-orthogonal basis, we also have to include the overlap matrix.

In a two-probe system illustrated in Figure 2.4, the entire system can be partitioned into three parts - the left electrode, the device molecule in between the two electrodes, and the right electrode. Accordingly, the Hamiltonian matrix H, and the retarded Green's function G^r can also be partitioned into three parts in the same manner. Adopting the TB model, which only considers the interaction between the two nearest neighboring blocks and neglects the interaction between the left electrode and right electrode, we have the following equations [112],

$$H = \begin{bmatrix} H_{LL} & H_{LM} & 0 \\ H_{ML} & H_{MM} & H_{MR} \\ 0 & H_{RM} & H_{RR} \end{bmatrix} \tag{2.19}$$

and

$$G^r = \begin{bmatrix} G^r_{LL} & G^r_{LM} & G^r_{LR} \\ G^r_{ML} & G^r_{MM} & G^r_{MR} \\ G^r_{RL} & G^r_{RM} & G^r_{RR} \end{bmatrix} \tag{2.20}$$

In the above equations, the matrices H_{LL}, H_{MM}, H_{RR}, H_{LM}, H_{ML}, H_{RM} and H_{MR} represent the sub-Hamiltonian for the left electrode, the device molecule, the right electrode, and the coupling between the molecule and the electrodes. Similarly, the sub-blocks, G^r_{LL}, G^r_{MM}, G^r_{RR}, G^r_{LM}, G^r_{ML}, G^r_{RM}, G^r_{MR}, describe the Green's function for the left electrode, the device molecule, the right electrode, and the interaction between the molecule and the electrodes. Since the Hamiltonian of the whole system is Hermitian, we have $H_{LM} = H^\dagger_{ML}$

and $H_{RM} = H_{MR}^{\dagger}$. The left (right) electrode is usually modeled as a semi-infinite ballistic lead. Thus, the sizes of the sub-matrices H_{LL}, H_{RR}, H_{ML}, H_{MR} and their corresponding Green's function sub-matrices are infinite.

Equation (2.18) is transformed into,

$$\begin{bmatrix} EI_{LL} - H_{LL} & -H_{LM} & 0 \\ -H_{ML} & EI_{MM} - H_{MM} & -H_{MR} \\ 0 & -H_{RM} & EI_{RM} - H_{RR} \end{bmatrix} \begin{bmatrix} G^r_{LL} & G^r_{LM} & G^r_{LM} \\ G^r_{ML} & G^r_{MM} & G^r_{MR} \\ G^r_{RL} & G^r_{RM} & G^r_{RR} \end{bmatrix} = \begin{bmatrix} I_{LL} & 0 & 0 \\ 0 & I_{MM} & 0 \\ 0 & 0 & I_{RR} \end{bmatrix} \tag{2.21}$$

Solving Equation 2.21, we have the following equation for the retarded Green's function of the molecule,

$$G^r_M(E) = \frac{1}{EI_{MM} - H_{MM} - \Sigma^r_L - \Sigma^r_R} \tag{2.22}$$

where Σ^r_L and Σ^r_R are the retarded self-energy matrices for the left and right electrodes, respectively. For the left electrode,

$$\Sigma^r_L(E) = H_{ML} g_L(E) H_{LM} \tag{2.23}$$

where $g_L(E)$ is the retarded surface Green's function for the left semi-infinite electrode.

$$g_L(E) = \frac{1}{EI_{LL} - H_{LL}} \tag{2.24}$$

The self-energy matrix and the surface Green's function for the right electrode can be found similarly,

$$\Sigma^r_R(E) = H_{MR} g_R(E) H_{RM} \tag{2.25}$$

$$g_R(E) = \frac{1}{EI_{RR} - H_{RR}} \tag{2.26}$$

Two more quantities, Γ_L and Γ_R, which represent the broadening of the molecular energy levels when the molecule is placed near the electrodes, can be calculated by,

$$\Gamma_{L(R)} = i(\Sigma^r_{L(R)} - \Sigma^a_{L(R)}) \tag{2.27}$$

Here, $\Sigma^a_{L(R)}$ is the advanced self-energy of the left (right) electrode, $\Sigma^a_{L(R)} = \left(\Sigma^r_{L(R)}\right)^\dagger$.

Once the self-energies Σ^r_L and Σ^r_R are found, the coherent transmission can be computed with

$$T(E) = Trace\left(\Gamma_L G^r_M \Gamma_R G^a_M\right) \tag{2.28}$$

where $G^a_M = (G^r_M)^\dagger$, is the advanced Green's function. The electric current through the device can be obtained consequently using Equation (2.14).

The DOS is calculated from [113],

$$D(E) = -\frac{1}{\pi}\sum_\alpha Im\left(G_{MM}(E)\right) \tag{2.29}$$

where the summation is performed over all the relevant atomic orbitals α.

The time-consuming parts in the computation are: (1) calculation of the relevant elements of the Green's function in Equation 2.22 (note that we do not need the whole inverse); and (2) the accurate evaluation of self-energies Σ_L and Σ_R using Equations 2.23 and 2.25. Computing the self-energies for realistic electrodes is challenging. To obtain an accurate result, the inclusion of electrode surface atoms with optimized structure and their influence on the device molecule are required. I will now present two methods for modeling the self energies due to the electrodes with a set of approximations.

2.3.2.2 Recursive algorithm

The first method applies the recursive technique by making use of the semi-infinite properties of the macroscopic electrodes [113, 114]. Take the left electrode for example. The g_L in Equation 2.23 is an infinite dimensional matrix representing the isolated left contact. However, in the calculation of Σ_L^r in Equation 2.23, we only require those elements of g_L connecting the molecule to the left contact, which are given by H_{ML}. The only non-zero elements of H_{ML} are those that connect the molecule to the left contact. Ideally, we can further assume that only atoms from the nearest neighboring layers of the electrode and the device molecule interact with each other. As a result, we only need to know the edge element of the surface Green's function g_L.

For simplicity, we define $A = EI - H$ [113]. Similar to the Hamiltonian H, A can be partitioned into the three components that represent the left electrode, the device molecule, and the right electrode,

$$A = \begin{bmatrix} A_{LL} & A_{LM} & 0 \\ A_{ML} & A_{MM} & A_{MR} \\ 0 & A_{RM} & A_{RR} \end{bmatrix} \tag{2.30}$$

The contribution from the left electrode is,

$$A_{LL} = EI_{LL} - H_{LL} \tag{2.31}$$

We have,

$$A_{LL} g_L = I_L \tag{2.32}$$

The semi-infinite left lead in a layered structure together with the device molecule is shown in Figure 2.5, with the label starting from the layer closest to the device molecule.

A detailed derivation of the surface Green's function for the contact shown in Figure 2.5 can be found in Reference [113]. Here, I only give the final result,

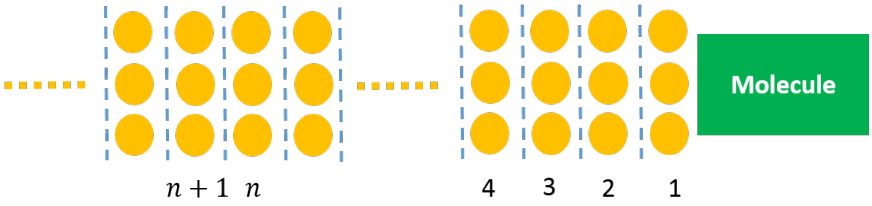


Figure 2.5: The semi-infinite left lead in a layered structure together with the device molecule.

$$[A_L^{11} - A_L^{12} g_L^{11} A_L^{21}] g_L^{11} = I_{LL}. \tag{2.33}$$

The superscripts 1 and 2 correspond to layers 1 and 2 in Figure 2.5. g_L^{11} is a square matrix representing only the surface Green's function of the left contact at layer 1. A_L^{12} is a matrix connecting layers 1 and 2.

A similar approach applies to the derivation of the surface Green's function for the right electrode.

2.3.2.3 Wide band limit

The second method we adopt is the wide band limit (WBL) in which the energy-dependent part of $\Gamma_{L/R}$ is neglected [115, 116, 117, 118, 119]. This approximation is valid when the DOS of the metal electrode is dominated by a wide band of states and is nearly a constant. For gold electrodes, the DOS is dominated by the s band [120]. Within this assumption, a very simple self-energy representation of the contacts can be obtained. The real part of the self-energy $\Sigma_{L/R}^r$ are disregarded and the imaginary part of the self-energy is controlled by the broadening matrix $\Gamma_{L/R}$ through the following equation,

$$\Sigma_{L/R}^r = -\frac{i\Gamma_{L/R}}{2} \tag{2.34}$$

where $\Gamma_{L(R)}$ is treated as an energy-independent parameter. Mathematically, the coupling matrix is diagonal with the non-zero elements representing the coupling strength. This

approximation has been adopted by other researchers in charge transport through DNA [82, 121, 122], especially at small biases as long as the value of $\Gamma_{L(R)}$ is tuned to match experiments.

2.3.3 Büttiker probes

For short organic molecules which are connected to two metal electrodes, charge transport can be viewed as a coherent process where the electrons tunnel through the molecules before getting scattered by the phase-breaking events. However, for long molecules, decoherence is expected to become significant due to the interaction between electrons and the lattice vibrations of the molecule as well as ambient electromagnetic fields. Decoherence in dry DNA molecule mainly arises from time-dependent fluctuations where the electrons will lose phase information to the environment.

In this work, we model the decoherence using the phenomenological Büttiker probes [123, 124]. We note that the method of including decoherence is not unique. Five recent papers include the effect of decoherence in wet DNA using completely different approaches [80, 81, 82, 121, 125]. Reference [125] represents the influence of atomic charges from the DNA backbone, water molecules and counterions with an empirical force field using a hybrid QM/MM framework. Reference [121] uses a new statistical decoherence model developed by their group. References [80] and [81] use the harmonic phonon bath to describe the dissipative environment. Reference [82] includes the effect of energetic vibronic coupling via a full-fledged nonequilibrium Green's function approach. In our approach the DNA strands are fixed and charge transport in DNA is a decoherent process. A self-energy is used to represent the decoherence. A similar approach has been adopted by Reference [126] with a simplified model Hamiltonian.

Büttiker probes are fictitious probes that extract electrons from the device and re-inject them after going through the phase breaking events, as illustrated in Figure 2.6. In contrast to the coherent process, which involves a simple two-probe system, charge transport in this case can be viewed as the averaged result of the scattering events between multiple probes

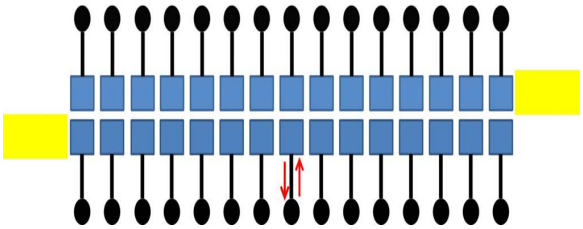


Figure 2.6: Sketch of a device with Büttiker probes. Yellow regions indicate the ends connecting to the electrodes. Blue regions indicate the molecular components (such as DNA bases), and black dots are Büttiker probes. Red arrows describe the behavior of electron at Büttiker probes: first extracted from the device and then re-injected back.

of an electron and the current is calculated with the effective transmission which takes the scattering events among all the probes into account. An algorithm for including the Büttiker probes is given by the D'Amato-Pastawski model [127]. The main effect of including Büttiker probes is to broaden the energy levels.

To include Büttiker probes, one can start with modifying the Green's function for a two-probe system, Equation (2.22). Equation (2.22) is only valid in the coherent case. For decoherent transport, an additional term, Σ_B, which describes the effect of Büttiker probes, should be added. In an orthogonal basis, Equation (2.22) can be modified to be,

$$G^r = \frac{1}{E - H_M - \Sigma_L^r - \Sigma_R^r - \Sigma_B} \tag{2.35}$$

As before, $\Sigma_{L(R)}$ represents the coupling of the DNA to left (right) electrode through which charge enters (leaves) the DNA. Different from the coherent transport, now we have an additional term, Σ_B, which is the self-energy due to the phase-breaking Büttiker probes. Specifically, we use Σ_i to represent the decoherence at the ith probe, $\Sigma_B = \sum_i \Sigma_i$. In our model, Σ_i is controlled by an energy-independent coupling strength between the probe and the coherent system Γ_i, $\Sigma_i = -i\Gamma_i/2$. In the calculations, we attach the Büttiker probes to the energy levels of the bases that correspond to the diagonal elements of H.

The next step is to use the Green's function approach to calculate the conductance for various coupling strengths of the Büttiker probes. In the remainder of this sub-section, we

summarize the steps of obtaining the effective transmission and conductance after including the effect of Büttiker probes using the D'Amato-Pastawski model [127].

Assuming that the total number of sites (for example, DNA bases) is N (say, $N = 30$), the number of Büttiker probes is $N_b = N - 2$. At the low bias region, the current at the i th probe is,

$$I_i = \frac{2q}{h} \sum_{j=1}^{N} T_{ij}[\mu_i - \mu_j], i = 1, 2, \cdots, N \tag{2.36}$$

The factor 2 is used to account for the spin degeneracy. In the above equation, T_{ij} is the transmission between the ith and jth probes calculated by $T_{ij} = \Gamma_i G^r \Gamma_j G^a$, where G^a is the advanced Green's function, $G^a = (G^r)^\dagger$. Because the transmission coefficients are reciprocal, we have $T_{ij} = T_{ji}$ [73].

Using the condition that the net current is zero at each Büttiker probe, Equation (2.36) gives us N_b independent formulas, from which we can express the chemical potential of the ith Büttiker probe, μ_i,

$$\mu_i - \mu_L = [\sum_{j=1}^{Nb} W_{ij}^{-1} T_{jR}](\mu_R - \mu_L), i = 1, 2, \cdots, N_b \tag{2.37}$$

In the above equation, W^{-1} is the inverse matrix of W, whose elements are given by [128], $W_{ij} = [(1 - R_{ii})\delta_{ij} - T_{ij}(1 - \delta_{ij})]$, where R_{ii} is the reflection probability at probe i, which is given by $R_{ii} = 1 - \sum_{j \neq i}^{N} T_{ij}$.

The currents at the left electrode I_L and right electrode I_R are not zero. Because of the conservation of the electron number, $I_L + I_R = 0$. Expressing the current at the left electrode in terms of the difference between the chemical potentials at the two electrodes μ_L and μ_R,

$$I_L = \frac{2q}{h} T_{eff}(\mu_L - \mu_R). \tag{2.38}$$

Comparing Equation (2.38) with the expression for the current at the left electrode given by Equation (2.36) and making use of Equation 2.37, we can write the effective transmission between left electrode and right electrode as,

$$T_{eff} = T_{LR} + \sum_{i,j=1}^{N_b} T_{L,i} W_{ij}^{-1} T_{j,R} \tag{2.39}$$

Here, the first term T_{LR} is the coherent transmission from left electrode to right electrode while the second term describes the contribution from the decoherence.

To summarize, we show the flow chart of implementing Büttiker probes approach based on the D'Amato-Pastawski model in Figure 2.7.

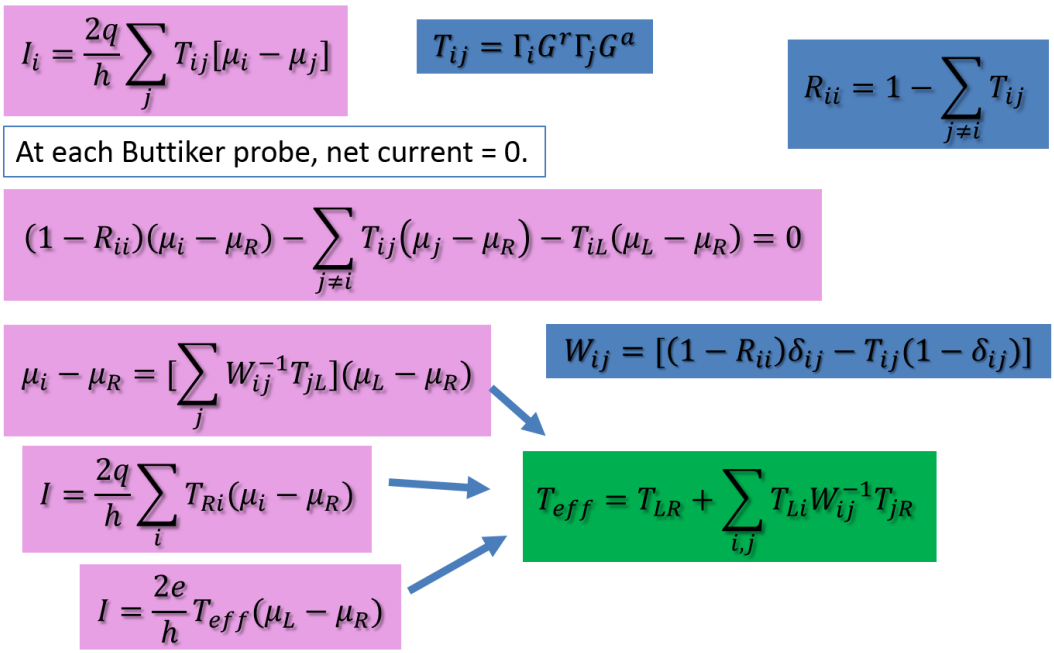


Figure 2.7: Flow chart of the implementation of Büttiker probes approach based on the D'Amato-Pastawski model.

In the flow chart of Figure 2.2 which shows the steps of calculating the transmission and conductance of a molecular device, two modifications are made. One is the self-energies due to Büttiker probes in the retarded Green's function and the other is the modification in the transmission equation due to the decoherent contribution. The flow chart of simulating the decoherent transport is shown in Figure 2.8 with the two new terms marked with dashed circles.

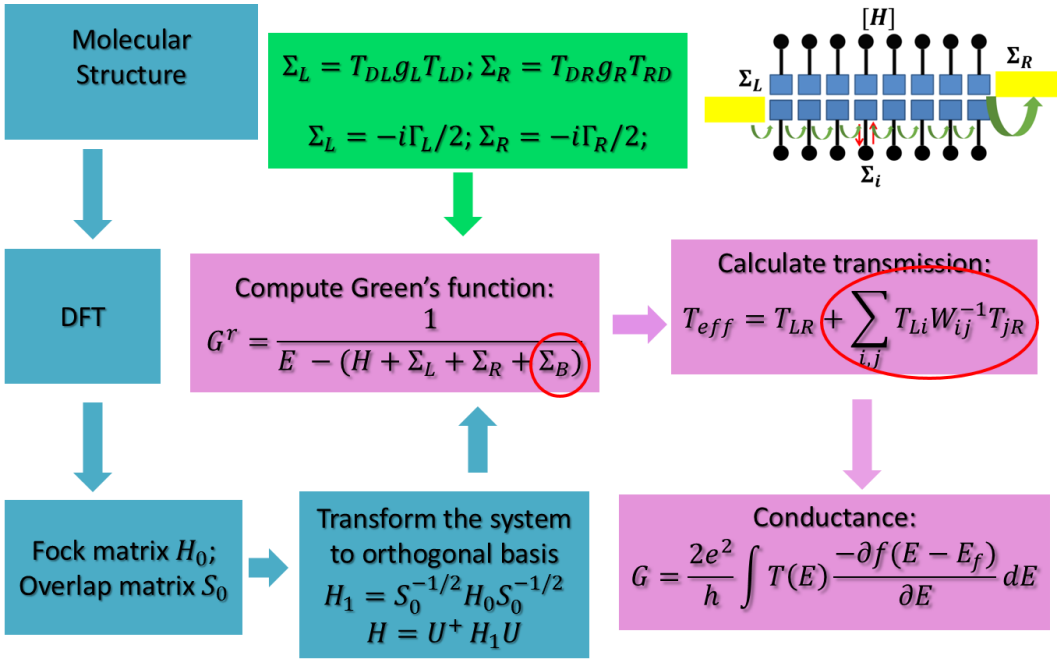


Figure 2.8: Flow chart of the simulation method for decoherent transport, which is based on DFT, the Landauer-Büttiker picture and the Green's function approach. Compared with the coherent transport, two modifications are made - (1) the self-energies due to Büttiker probes in the retarded Green's function; (2) the modification in the transmission equation due to the decoherent contribution. Both are marked with the dashed circles.

The Matlab code implementations of the algorithms shown in Figures 2.2 and 2.8 are provided in Appendix A.

Chapter 3

ELECTRICAL TRANSPORT IN DNA HETEROSTRUCTURES

One of the motivations for the study of the electrical transport through nucleic acids stems from the fact they offer a unique framework for nanoscale electronic devices and have potential to be used as active components in electrical circuits. In this chapter, I will present the transport results based on a series of DNA heterostructures with the B-form configuration in dry state to illustrate the idea of using nucleic acids as devices. We will start with the calculation of the IP values of the single bases and base pairs in B-form DNA. Then we study the possibility of building quantum barriers/wells and superlattices. We first consider transport in a single-stranded $poly\{G\}(T)_N poly\{G\}$ sequence. We show that T plays a role as a barrier and the transmission decreases exponentially with increasing number of Ts. We also investigate the transmission and conductance through superlattices, $poly\{GT\}$ and $poly\{GGTT\}$. Minibands are observed in these structures. We then consider transport through the double-stranded $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$ and $poly\{(G:C)(A:T)\}$ to see how the single-stranded results change. The effect of coupling between DNA molecule and metal contacts is also explored. In this chapter, we present examples for calculations without the backbone for B-form DNA.

3.1 Ionization potentials of single B-form DNA bases and base pairs in gas phase

In nature, DNA is often present in the aqueous phase. Most conductance measurement experiments for DNA are also performed in solution, with water molecules and counterions. Due to its floppy nature, the conformation and conductivity of DNA can be easily affected by the surrounding environments. Nevertheless, to reveal the intrinsic property of DNA

conductance, it is useful to study the electronic structure of DNA in dry state, excluding the effect from the external environment. In experiment, a dry DNA strand can be obtained by making use of the nitrogen gas [54]. The strand is then further stabilized by polycation spermidine, which is expected to neutralize the negative charges in the phosphate groups without disturbing the base pairing and base stacking of B-form DNA [129, 130]. In this manner, the perturbation from motions of counterions and water molecules can be reduced to the minimum.

When a molecule is placed near a metal electrode, charge will flow between the molecule and the electrode to keep the system in equilibrium. Therefore, it is meaningful to look at the HOMO and LUMO levels of the molecule. In Table 3.1, we show the HOMO, LUMO and IP values for the four isolated DNA bases (without backbone and terminated with a hydrogen atom) calculated with Gaussian 09, using B3LYP XC functional and 6-31G basis set [131, 132]. We also show the HOMO, LUMO and IP of the single $G:C$ and $A:T$ base pairs in B-form configuration. The IP values from our calculations are consistent with that in Reference [56].

Table 3.1: HOMO, LUMO levels and IP values for the four single DNA bases and two single base pairs in B-form configuration calculated with B3LYP/6-31G. Unit, eV.

Base	A	T	G	C	$A:T$	$C:C$
HOMO	-6.14	-6.63	-5.69	-6.05	-5.95	-5.21
LUMO	-0.51	-1.22	-0.13	-0.91	-1.17	-1.29
IP	7.98	8.62	7.52	8.26	7.68	7.11

Table 3.1 shows that the HOMOs of the four single DNA bases follow the order $T < A < C < G$, and the HOMOs of the single base pairs follow the order $(A:T) < (G:C)$. The order of the IP values for the four bases is $G < A < C < T$, and for the base pairs it is $(A:T) < (G:C)$. Specifically, the IP of G is around 1 eV lower than that of T. Considering the definition that IP refers to the energy required to remove an electron from a molecule, we

can draw the conclusion from Table 3.1 that G bases are preferred by hole transport, while T bases act as barriers in hole transport because they have the highest IP values. Similarly, $G:C$ base pairs are preferred in hole transport over the $A:T$ base pairs because of their lower IP.

The distinct energy levels and IP of the building blocks provides a framework to construct DNA sequences with interesting device properties. One can construct quantum wells and barriers from DNA that have resonances akin to double barrier resonant tunneling diodes and superlattices [27, 133]. Also, because these building blocks can be reproducibly engineered in arbitrary sequences, one can build devices with sequence-dependence, analogous to the for long studied sequence dependent charge transfer in DNA in chemistry [49].

In the following sections of this chapter, we consider dry DNA heterostructures with B-form configuration in both single- and double-stranded forms. The purpose is to investigate the possibility of using DNA as active components in electrical circuits by modeling the conduction in barriers and superlattices, and studying the effect of the DNA molecule-metal contact coupling.

3.2 ***Heterostructures*** $poly\{G\}T_Npoly\{G\}$

We first model the transmission through a barrier for holes constructed from the sequence $poly\{G\}T_Npoly\{G\}$, where N is the number of Ts, varying from 0 to 8. We treat the 10 bases in the middle as the device part and construct the two semi-infinite contacts by repeating the 10 bases in a unit cell of G. Then we calculate the surface Green's function as discussed in Chapter 2. A similar method is applied to the single-stranded structures $poly\{GT\}$, $poly\{GGTT\}$ and double-stranded structures $poly\{(G:C)\}(A:T)_Nploy\{(G:C)\}$, $poly\{(G:C)(A:T)\}$. We construct the contact leads by repeating the 10 bases in a unit cell of $poly\{GT\}$, and the 20 bases in a unit cell of $poly\{GGTT\}$ (10 G bases and 10 T bases) respectively. The same approach applies to the double-stranded helices $poly\{(G:C)\}(A:T)_Nploy\{(G:C)\}$ and $poly\{(G:C)(A:T)\}$.

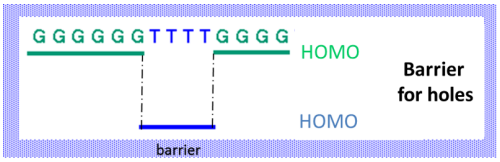
We show the transmission through $poly\{G\}T_Npoly\{G\}$ at the energy points around the

HOMO levels as a function of the different number of Ts in Figure 3.1. Also presented is the energy digram for $poly\{G\}T_Npoly\{G\}$ where T_N forms a barrier for hole transport.

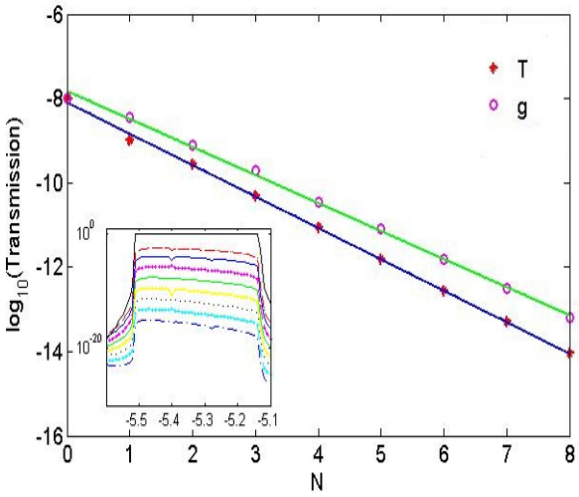
We find the width of the conduction channel around HOMO to be about 370 meV (inset plot of Figure 3.1(b)), which agrees with the earlier work [27]. We plot the transmission versus N at energy $E = -5.44\ eV$ (red stars and blue curve). The values of transmission decrease exponentially with the increasing number of Ts. Thus we conclude that T is a large barrier. From the fitting equation at the energy point $E = -5.44\,eV$, $log_{10}(Transmission) = -1.86N - 0.24$, we find that the addition of each T can cause a reduction in transmission by approximately 72 times. This can be understood by considering the difference in HOMO levels of G and T shown in Table 3.1. In our calculation, the HOMO level of an isolated G base is about $-5.69\ eV$, while the HOMO level of an isolated T base is about $-6.63\ eV$. This indicates that for hole transport T represents a barrier. As the valence band offset between G and T is approximately 1 eV (comparison of IPs leads to a similar conclusion), we test the possibility of modeling the transmission through the $poly\{G\}T_Npoly\{G\}$ by a simple square potential barrier, $poly\{G\}g_Npoly\{G\}$, where g refers to the modified G whose onsite energy has been decreased by 1 eV. The resulting transmission is significantly higher and the agreement with the square potential barrier is poor (magenta circles and green curve in Figure 3.1).

We also model the conductance through these sequences. We focus on the low bias limit and evaluate the linear response conductance at 300 K. The Fermi level E_f in the electrodes and the molecular energy levels are critical in understanding the flow of charge through the molecule. Experimentally, it is difficult to determine the position of E_f because of the complicated details of the molecule-contact coupling. Here, we explore the conductance when the Fermi level is in the HOMO vicinity. The conductance through $poly\{G\}(T)_Npoly\{G\}$ at energy points around the HOMO level in terms of the number of Ts is shown in Figure 3.2 (inset).

Again, we see that with the increasing number of Ts, the conductance decreases exponentially. Similar to the case for transmission in $poly\{G\}(T)_Npoly\{G\}$, we also choose an



(a)



(b)

Figure 3.1: (a) Energy digram for $poly\{G\}T_N poly\{G\}$, where T_N forms a barrier for hole transport; (b) common logarithm of transmission v.s. N for $poly\{G\}T_N poly\{G\}$ (red stars) and $poly\{G\}g_N poly\{G\}$ (magenta circles) when $E = -5.44\ eV$. The fitting relation for Ts is $log_{10}(Transmission) = -1.86N - 0.24$ (blue line), indicating that with the addition of each T, the transmission decreases about 72 times. The fitting relation for the modified Gs is $log10(Transmission) = -1.69N + 0.44$ (green line). The poor agreement between the blue line and green line is because of the weak coupling between G and T. Inset, transmission v.s. energy (in unit of eV) for $poly\{G\}T_N poly\{G\}$. The width of the conduction channel is about 370 meV. With the increasing number of Ts, the magnitude of transmission decreases. From top to bottom, $N = 0$ (black solid), 1 (red points), 2 (blue solid), 3 (magenta stars), 4 (green x), 5 (yellow circles), 6 (black dots), 7 (cyan plus), 8 (blue dashdot).

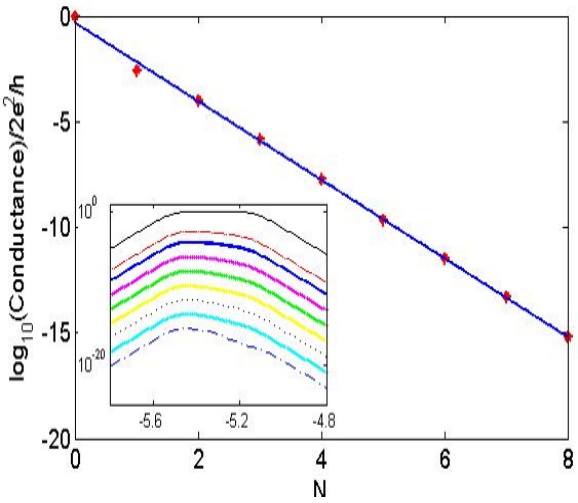
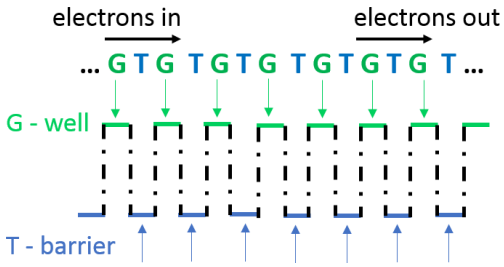


Figure 3.2: Common logarithm of conductance v.s. N through $poly\{G\}(T)_N poly\{G\}$ when $E = -5.44\ eV$ (red stars). The fitting relations are $log_{10}(Conductance) = -1.86N - 0.32$ (blue line). Inset, conductance v.s. Fermi energy (in unit of eV) for $poly\{G\}(T)_N poly\{G\}$ at low bias limit. With the increasing number of Ts, the conductance decreases. From top to bottom, $N = 0$ (black solid), 1 (red points), 2 (blue solid), 3 (magenta stars), 4 (green x), 5 (yellow circles), 6 (black dots), 7 (cyan plus), 8 (blue dashdot).

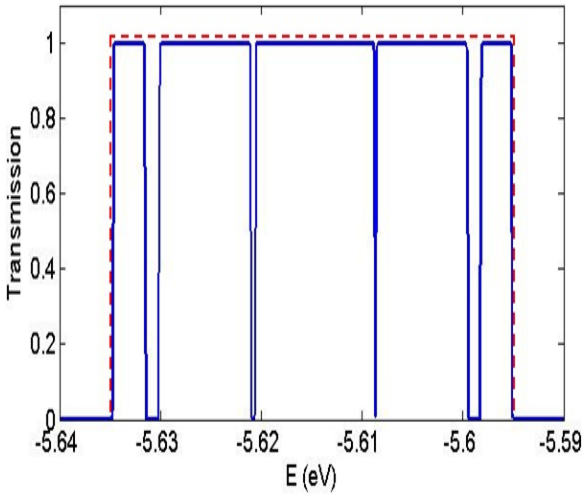
energy point at $E_f = -5.44\ eV$ and plot the conductance versus the number of Ts. From the fitting equation, we again see the "barrier effect" of T - with the addition of one T, the conductance decreases approximately by 72 times. The experimental results show an exponentially decreasing relation between the conductance through a $G:C$ rich DNA molecule as the number of $A:T$ base pairs increases [54].

3.3 Superlattices $poly\{GT\}$ *and* $poly\{GGTT\}$

Once the magnitude of T as a barrier was ascertained, we used this property to construct the following superlattices: (1) $poly\{GT\}$; and (2) $poly\{GGTT\}$, with the aim of observing minibands in these structures. Minibands are expected because G serves as a quantum dot for holes while T is a barrier connecting the quantum dots, akin to that seen in semiconductor superlattices [134]. In Figure 3.3, we present the energy digram and transmission of $poly\{GT\}$ at the energy close to the HOMO level.



(a)



(b)

Figure 3.3: (a) Qualitative figure of the energy diagram of $poly\{GT\}$, where G bases act as wells and T bases act as barriers; (b) transmission v.s. energy for $poly\{GT\}$. The width of the conduction channel is about 40 meV. The blue solid curve represents the calculated transmission result, while the red dashed curve shows the ideal conduction channel. The four dips in the blue curve are due to the variations in the eigenvalues of the Hamiltonian for the bases. The dips will disappear if we construct the structure with Hamiltonians of identical eigenvalues.

A miniband with a width of approximately 40 meV is predicted in the superlattices $poly\{GT\}$. Compared to the width of 370 meV in $poly\{G\}$, the role of T as a barrier for hole transport is clear. It is worth noting that the four dips in the blue solid curve are not real gaps. We have confirmed that the occurrence of these dips is due to the variations in the eigenvalues of the Hamiltonian for the bases. In the ideal situation, the eigenvalues for all G bases and all T bases should be identical. However, when we construct the DNA structures, small numerical errors may exist in parameters such as the distance and the angle between adjacent bases, which may cause variations in the eigenvalues of bases.

We also show the DOS for $poly\{GT\}$ in Figure 3.4. For clarity, we only show the DOS at the first four bases in the molecular device. DOS at other bases provides similar information.

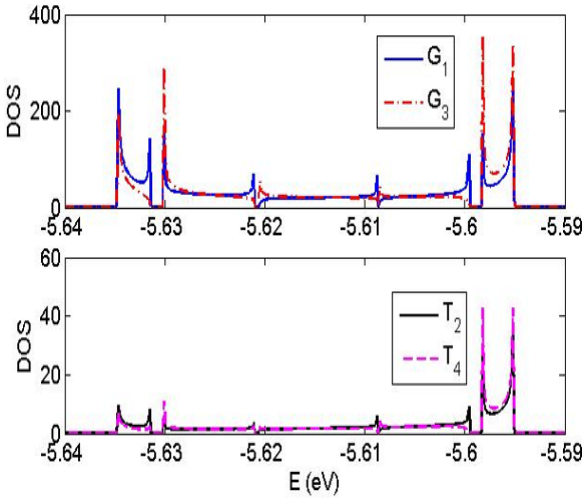
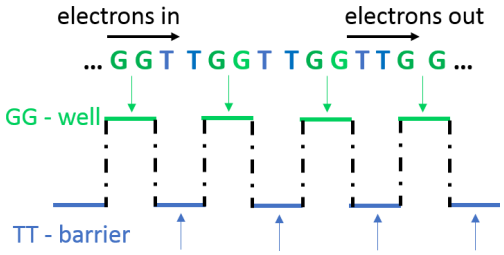


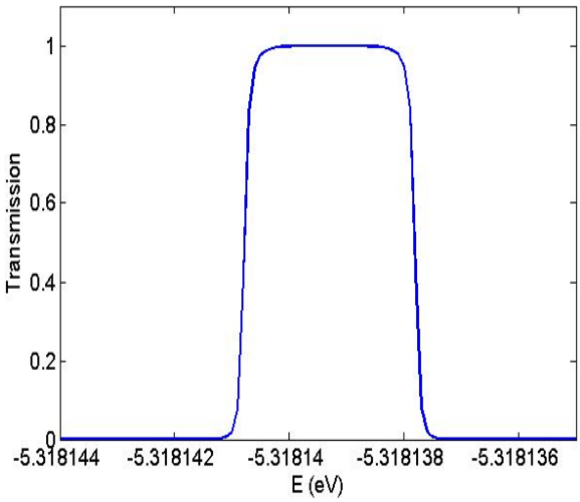
Figure 3.4: DOS for $poly\{GT\}$ at the first four bases in the molecular device. Upper, DOS for the first two G bases in the molecular device; lower, DOS for the first two T bases in the molecular device. The subscripts of G and T reference the base number in Figure 3.3(a), counting from left to right.

DOS indicates the number of states per energy interval at each energy level that can be occupied in various spatial regions. The transmission at a given energy depends on the availability of delocalized states at that energy. The transmission in Figure 3.3(b) corresponds to the DOS in Figure 3.4.

In Figure 3.5, we show the transmission through $poly\{GGTT\}$ at energy points that are close to the HOMO level. In the case of $poly\{GGTT\}$, the width of the miniband is only 3 μeV, which is about thirteen thousand times smaller than the bandwidth of $poly\{GT\}$. This narrow width is rationalized by noting that the transmission through $poly\{G\}TTpoly\{G\}$ is 72 times weaker than the transmission through $poly\{G\}Tpoly\{G\}$ (Figure 3.1(b)).



(a)



(b)

Figure 3.5: (a) Qualitative figure of the energy diagram of $poly\{GGTT\}$, where GG bases act as wells and TT bases act as barriers; (b) transmission v.s. energy for $poly\{GGTT\}$. The width of the miniband is only 3 μeV because of the strong barrier effect of TT.

The DOS for $poly\{GGTT\}$ is also calculated, as shown in Figure 3.6. For clarity, we only show the DOS for the first eight bases (the first four G bases with the first four T bases) in

the molecular device. Similar to the observation in the case of $poly\{GT\}$, we also find that the transmission in Figure 3.5(b) follows the trend of the DOS in Figure 3.6.

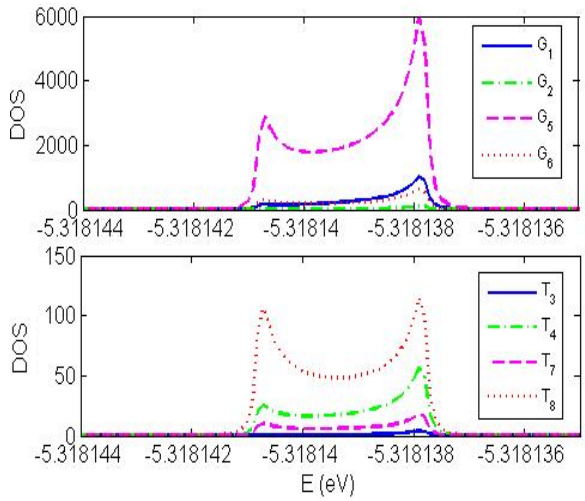


Figure 3.6: DOS for $poly\{GGTT\}$ at the first eight bases in the molecular device. Upper, DOS for the first four G bases in the molecular device; lower, DOS for the first four T bases in the molecular device. The subscripts of G and T reference the base number in Figure 3.5(a), counting from left to right.

In Figure 3.7, we present the conductance versus Fermi energy in the HOMO vicinity through $poly\{GT\}$ and $poly\{GGTT\}$. In our calculation, HOMO value for $poly\{GT\}$ and $poly\{GGTT\}$ is $-5.60\ eV$ and $-5.32\ eV$, respectively. Compared to the conduction through $poly\{G\}$ the maximum value for $poly\{GT\}$ decreases about 3 times, while for $poly\{GGTT\}$ the conductance we obtain is on the order of $10^{-5}\ G_0$, which is reasonable because of the strong barrier effect of T.

3.4 Heterostructures $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$

In Sections 3.2 and 3.3, we study the transport properties of the DNA heterostructures and superlattices built from single strands. We now study the construction of tunable barriers and superlattices using engineered DNA sequences with the dsDNA.

The IP value of $G:C$ is 0.57 eV smaller than that of $A:T$ (Table 3.1). As a result,

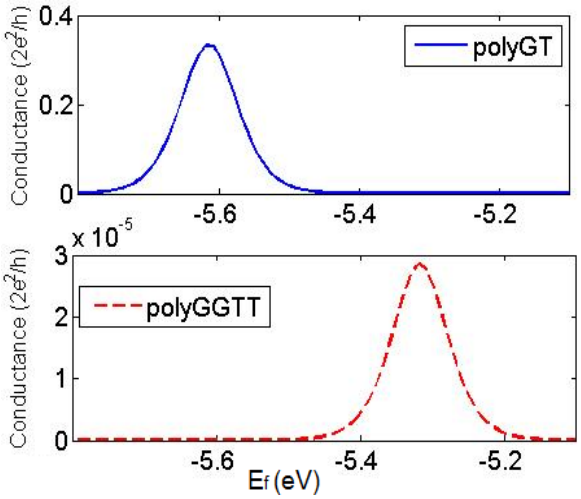


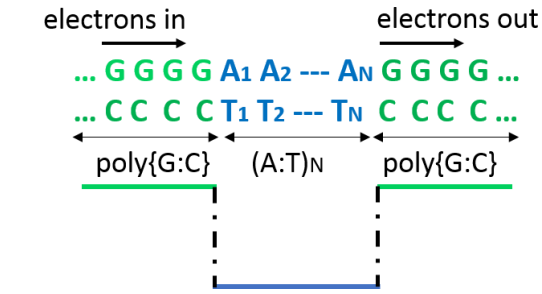
Figure 3.7: Conductance v.s. Fermi energy through $poly\{GT\}$ (upper) and $poly\{GGTT\}$ (lower). Compared to $poly\{G\}$, the maximum value of conductance for $poly\{GT\}$ decreases about 3 times, and the magnitude of conductance for $poly\{GGTT\}$ is only on the order of $10^{-5}G_0$. The reason is the large barrier effect of T.

a strand that consists of $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$ will behave as a barrier for holes. Our calculated transmission for such a strand as a function of N, the number of $A:T$ forming the barrier is shown in Figure 3.8. Also shown is the energy diagram of $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$, where the $A:T$ base pairs form a barrier for holes to flow between the two poly$\{(G:C)\}$ regions. The calculations in Figure 3.8(a) use the TB model (all the other figures in this chapter use the full Hamiltonian).

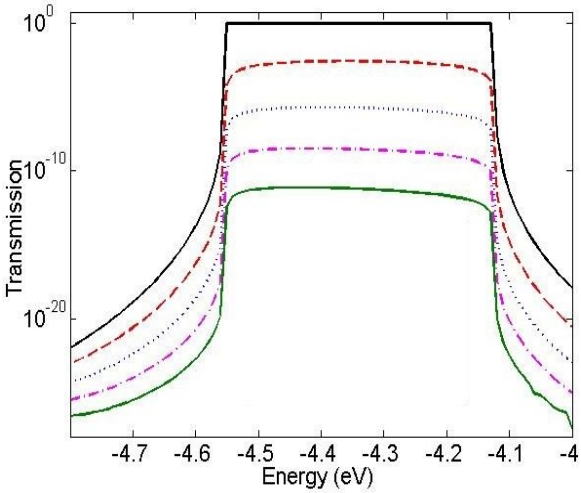
The transmission decreases exponentially with an increase in the length of the barrier N. When there are no $A:T$s, the transmission is unity, as expected, within the valence band, which has a bandwidth of only 420 meV. As the number of $A:T$s increases from 1 to 4, the transmission decreases exponentially.

3.5 Superlattice $poly\{(G:C)(A:T)\}$

Similar to the single G and T, we can build superlattice with the periodic repetition of $(G:C)(A:T)$ with the fact that the $A:T$ base pair acts as a barrier for hole transport.



(a)



(b)

Figure 3.8: (a) A suitably designed double strand is a barrier for hole transport. The $A:T$ base pairs form a barrier for hole to flow between the two poly$\{G:C\}$ regions; (b) transmission for $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$. From top to bottom, $N = 0,\ 1,\ 2,\ 3,\ 4$. The black curve is the transmission for the poly$\{(G:C)\}$ strand without $A:T$, which is unity in the valence band. As the number of $A:T$ (N) increases, the transmission in the entire valence band exponentially decreases.

The energy diagram and transmission of $poly\{(G:C)(A:T)\}$ structure consisting of barriers $(A:T)$ separated by wells $(G:C)$ are shown in Figure 3.9(a).

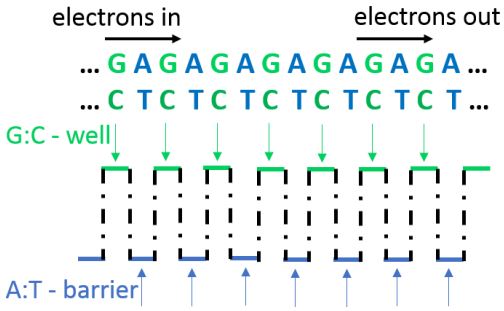
Based on the knowledge of superlattices in semiconductor heterostructures, one would expect a miniband whose width depends on the strength of the barrier $(A:T)$ and energy level of the wells $(G:C)$. Figure 3.9(b) shows the highest lying valence miniband. The width of the miniband is a little larger than 10 meV, which is indicative of the strength of strong barriers between consecutive Gs.

3.6 Metal-DNA-metal interface

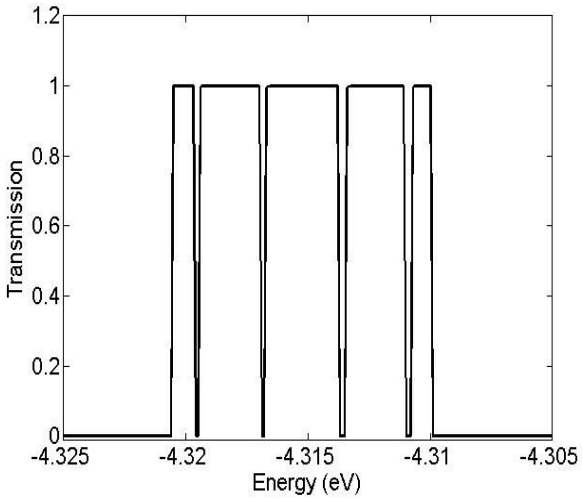
Finally, we explore the role of the coupling between the DNA device and the metal contacts in influencing the transmission. The strength of the coupling between the molecular device and the contacts is important in determining the flow of the current. In the previous part, we construct the two contacts with repetitions of the unit cell of the DNA molecule, either with the unit cell being the same as that of the device such as in $poly\{G\}$, $poly\{GT\}$, $poly\{GGTT\}$, $poly\{(G:C)\}$ and $poly\{(G:C)(A:T)\}$, or with the unit cell different from the device unit such as in $poly\{G\}(T)_N poly\{G\}$ and $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$ when $N \neq 0$. In both cases, the couplings between the device and the contacts are chosen to be perfect, producing a transmission with the maximum value almost equal to 1. However, in an experiment, usually a molecular device is connected to two metal contacts. Experimentally, it is difficult to control the coupling between the molecule and the metal contacts.

In order to study how the metal-DNA-metal contact coupling influences the transport, we consider a unit cell of $poly\{GT\}$ connected to two metal contacts and study the transmission at energy around HOMO level by treating the coupling with various strengths, ranging from the weak-coupling limit to the strong-coupling limit. The transmission results at $\Gamma_L = \Gamma_R =$ 10 meV, 50 meV, 100 meV, 300 meV and 500 meV are presented in Figure 3.10.

We see that when the coupling increases from 10 meV to 50 meV and then to 100 meV, the magnitude of the transmission keeps increasing. When the coupling is increased further,



(a)



(b)

Figure 3.9: (a) Qualitative figure of the energy diagram of $poly\{(G:C)(A:T)\}$ strand, which behaves as a quantum well superlattice. Effective potential energy digram shows $A:T$ forms a barrier for holes and $G:C$ forms the wells in the superlattice; (b) transmission v.s. energy. The width of the miniband is approximately 10 meV.

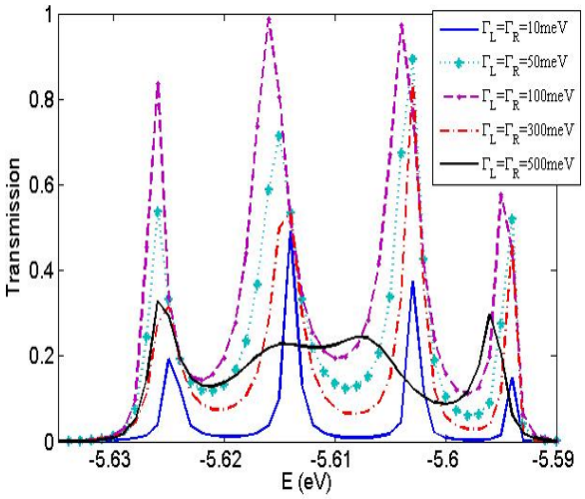


Figure 3.10: Transmission v.s. Fermi energy through $poly\{GT\}$ with the strength of coupling between DNA molecule and the contact varying from 10 meV, 50 meV, 100 meV, 300 meV and 500 meV. The magnitude of transmission first increases and then decreases when the strength of the coupling increases from the weak limit to the strong limit.

to 300 meV and 500 meV, the magnitude of the transmission decreases. This is expected because the transmission is determined by the product of the DOS and the coupling. As the coupling increases, DOS in the molecule broadens. When the coupling is not too strong, the broadening is small; hence the DOS does not change much. As the coupling increases from the weak limit, the transmission increases. However, when the coupling is strong enough, the DOS will broaden significantly and its value is very small, causing a reduction in the transmission. The transmission results under different coupling strengths can be used to compare with the experimental results.

3.7 Summary

In this chapter, we show examples where engineered DNA heterostructures behave as barriers and superlattices. We study the electrical transport in dry B-form single DNA strands, $poly\{G\}(T)_N poly\{G\}$, $poly\{GT\}$ and $poly\{GGTT\}$, and double strands $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$ and $poly\{(G:C)(A:T)\}$. By studying the transmission and conductance

through $poly\{G\}(T)_N poly\{G\}$, we confirmed that T is a barrier for hole transport, which has been observed experimentally. The transmission and conductance decrease exponentially when the number of Ts is increasing. By calculating the fitting equation at energy point when $E = -5.44\ eV$, we find that one more T can cause 72 times reduction in transmission and conductance. For superlattices, $poly\{GT\}$ and $poly\{GGTT\}$, we find minibands with a width of about 40 meV and 3 μeV respectively. Compared to the bandwidth in $poly\{G\}$, which is around 370 meV, the narrow bandwidths in $poly\{GT\}$ and $poly\{GGTT\}$ are attributed to the large barrier effect of T. For the double-stranded structures, minibands with widths of 420 meV and 10 meV are found in $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$ and $poly\{(G:C)(A:T)\}$, respectively.

We also investigate the influence of coupling between DNA molecule and the metal contacts. By modeling the transmission through a unit cell of $poly\{GT\}$ connected to metals contacts at different coupling strengths, we show that when the coupling increases from the weak limit, the magnitude of transmission also increases. When the coupling increases further into the strong limit, the transmission begins to decrease. This is because the transmission is determined by the product of the DOS and the coupling. At small coupling the DOS does not broaden a lot and the coupling dominates, while at strong coupling the value of DOS is very small because of the large broadening and therefore the transmission decreases.

While promising, the experimental realization of device concepts discussed above needs careful attention because weak electronic coupling between bases makes the conductance susceptible to defects and variations in local environment.

Chapter 4

UNIFIED MODEL FOR CONDUCTANCE THROUGH DNA WITH THE LANDAUER-BÜTTIKER FORMALISM

4.1 Introduction

In this chapter, we begin to look at the electrical transport through strands beyond the model structures discussed in Chapter 3. We use the model developed in Chapter 2 to understand the charge transport mechanism and explain the measurement data in experiments. Previous work has qualitatively described transport in DNA, but explaining the experiments of conductance is still a challenge. Here we model the conductance of four different dry DNA strands considered in the experiments of Reference [54], and provide an explanation for the low bias conductance. We show that a DFT-derived Hamiltonian when coupled with a phenomenological model is capable of providing an explanation of the experiments on the electrical conductance of DNA. We also point out the importance of decoherence in $A:T$ base pairs, and predict the difference between the decoherence rates in $A:T$ and $G:C$ base pairs for dry DNA that are necessary to explain the experiments involving the four different strands. The rates for decoherence are not clear from previous work, and we hope that our work will provide some guidelines for future comparison.

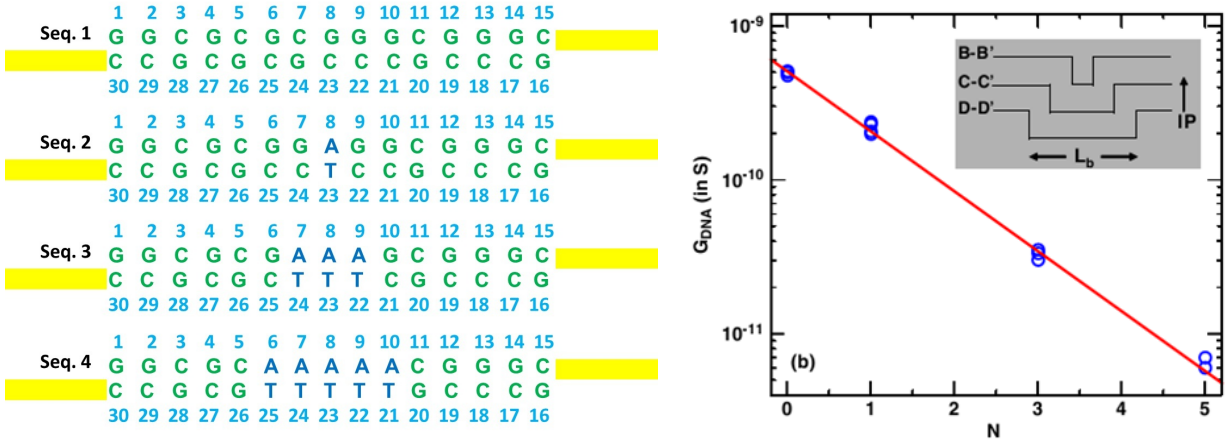
Specifically, in this chapter we model the zero-bias conductance for four different DNA strands within Landauer-Büttiker framework. Our approach consists of three elements: (1) *ab initio* calculations of DNA; (2) Green's function approach for transport calculations; and (3) the use of two parameters to determine the decoherence rates. We first study the role of the backbone. We find that the backbone can alter the coherent transmission significantly at some energy points by interacting with the bases, though the overall shape of the transmission stays similar for the two cases. More importantly, we find that the

coherent electrical conductance is significantly smaller than what the experiments measured. We consider DNA strands under a variety of different experimental conditions and show that even in the most ideal cases, the calculated coherent conductance is much smaller than the experimental conductance. To understand the reasons for this, we carefully look at the effect of decoherence. By including decoherence, we show that our model can rationalize the measured conductance of the four strands, both qualitatively and quantitatively. We find that the effect of decoherence on $G:C$ base pairs is crucial in getting agreement with the experiments. However, the decoherence on $G:C$ base pairs alone does not explain the experimental conductance in strands containing a number of $A:T$ base pairs. Including decoherence on $A:T$ base pairs is also essential. By fitting the experimental trends and magnitudes in the conductance of the four different DNA molecules, we estimate that the deocherence rate is 6 meV for $G:C$ and 1.5 meV for $A:T$ base pairs.

4.2 Experimental strands considered

The four strands we study consist of the following 15-base pair dsDNA molecules in B-form configuration [54]: (a) $5'-GGCGCGCGGGCGGGC-3'$; (b) $5'-GGCGCGGAGG$ $CGGGC-3'$; (c) $5'-GGCGCGAAAGCGGGC-3'$; (d) $5'-GGCGCAAAAACGGGC-3'$, as shown in Figure 4.1(a). Conductances of these strands were measured under dry condition at room temperature. Contact coupling was formed between the DNA and the metal electodes through thiol groups at the $3'$ ends, because thiol groups are expected to provide a strong coupling between the DNA molecule and the metal contact. The polycation spermidine was used to stabilize the negatively charged dsDNA. In these systems, both the base pairing and base stacking of the B-form DNA are expected to remain unaffected [129, 130]. Nitrogen was further used to dry the DNA molecules, which allowed the repeatable conductance measurements without degradation. In this experiment, a clear exponential dependence on numbers of $A:T$ base pairs was observed, as shown in Figure 4.1(b), consistent with the general DNA conduction picture that the G base has the lowest IP value.

We choose these four strands for the following reasons: (1) the 15-base pair DNA strand



(a) (b)

Figure 4.1: (a) Structures of the four dithiol-derivatized dsDNA molecules used in Reference [54]; (b) measured conductance (open circles) of the four strands v.s. the number of the $A:T$ base pairs (N). An exponential dependence of conductance on N has been fitted (solid line), indicating that the $A:T$ provides a barrier [54].

is within the computational limit of DFT; (2) the middle part of strands changes from $G:C$ in $A:T$ sequentially, allowing the role of $G:C$ and $A:T$ to be studied more obviously; (3) the conductance was measured in dry state, where the influence from the motion of water molecules and counterions was eliminated. In this manner, the perturbation from the surrounding environment is reduced.

The method for modeling transport through DNA lying between two metal contacts has been discussed in detail in Chapter 2. As discussed in Chapter 2, we use two broadening parameters, Γ_L and Γ_R, to represent the coupling strengths for the left and right contacts. In this work, we vary Γ_L and Γ_R between 50 and 100 meV, and find that there is no significant change in our results. For the results presented here, we set $\Gamma_L = \Gamma_R = 100\ meV$.

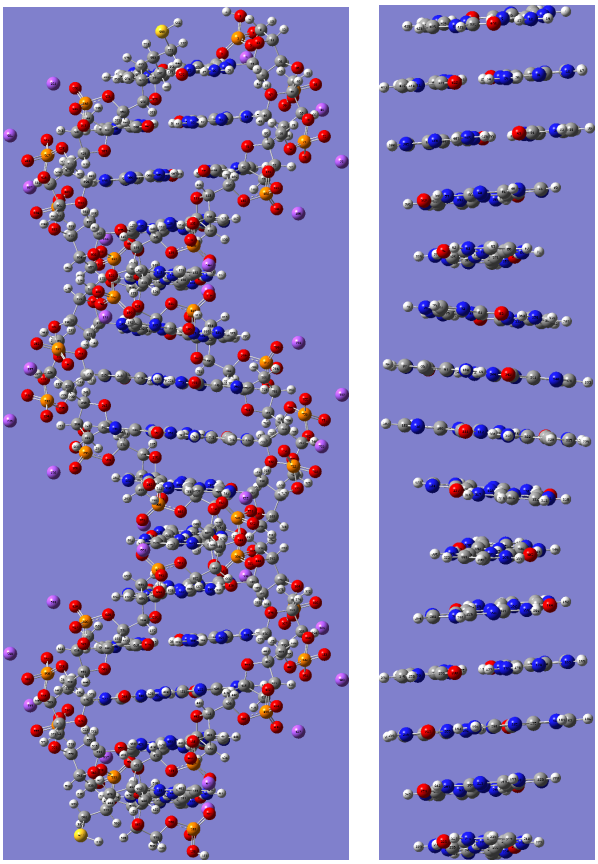
In the following sections, we will analyze the effect of two factors in DNA charge transport using the Landauer-Büttiker formalism, namely, the backbone and the decoherence. In Section 4.3, we compare the transmission for the four strands with and without the backbones in the phase-coherent limit. We find that the backbone does not change the transmission

qualitatively. However, it can affect the transport significantly at some energy points by interacting with the atoms in bases. In Section 4.4, we discuss the role of decoherence. We observe that the coherent transmission is too small to explain the experiment. We infer that the low value of conductance can be due to a number of reasons: poor coupling to contacts, position of Fermi energy or incorrect bandgap. We show below that even for the best possible values of these parameters, the phase-coherent conductance is smaller than the experimental conductance. We then show that by including appropriately chosen decoherence rates the calculated conductance values are comparable to the experiment. The strength of the decoherence rate is difficult to compute *ab initio*. Here we have considered a variety of values to give some insight into the ones which may be experimentally feasible. In the process of modeling decoherence, we have also found that it is possible to explain the experiments only if we assume that there is also decoherence in the $A:T$ barriers. That is, the $A:T$ barriers cannot be assumed to be static barriers.

4.3 Effect of backbone

In this section, we study the role of DNA backbone on charge transport using the strands discussed in Section 4.2. To study the effect of the backbone, we have calculated transmission for two cases - DNA strands with and without backbones in the phase-coherent limit. For the strands with backbone, we use the positively charged sodium ions Na^+ to neutralize the backbone. To determine the geometry of the DNA strand after being neutralized, we first carry DFT calculation on a short strand containing five G stacking bases with sodium ions placed nearby the phosphate group by setting the initial distances based on Reference [92]. The atom coordinates for DNA molecule are fixed, while the positions of Na^+ ions are relaxed. We approximate the optimized position of the sodium ion around the backbone of the middle G in this short strand to be the positions of the thirty sodium ions in the long 15-base pair strand. We then carry out DFT calculation on the 15-base pair strand to obtain the self-consistent Hamiltonian and overlap matrices. For the strands without backbone, we simply delete the backbone and then terminate the base with hydrogen atoms

using GaussView 5 [135]. An example of the 15-base pair Seq. 1 with and without backbone is shown in Figure 4.2.



(a) (b)

Figure 4.2: Structures for Seq. 1 with sequence $5' - GGCGCGCGGGCGGGC - 3'$. (a) Seq. 1 with backbone terminated with the thiol groups at the $3'$ ends and neutralized by the Na^+ counterions around the phosphate groups in the backbone; (b) Seq. 1 without backbone with the bases terminated with hydrogen atoms.

The transmission of the four DNA strands is calculated in the phase-coherent limit as shown in Figure 4.3. For the strands with backbone, we model the contacts via two approaches - charge is injected and extracted at the: (1) base and backbone (blue solid) and (2) base only (magenta dash). Transmission for the strands without the backbone is shown by the red dash-dot curve of Figure 4.3. We note that the overall shapes of transmission with

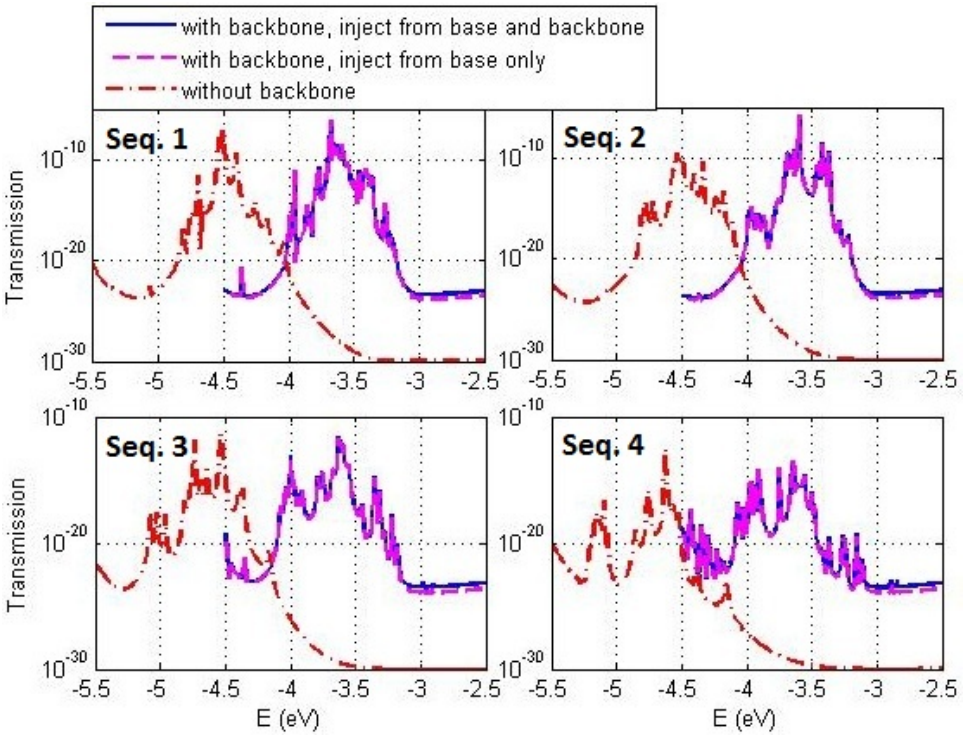


Figure 4.3: Transmission v.s. energy for the four DNA strands with and without the backbones in the phase-coherent limit. The transmission for the strands with the backbones is calculated by two methods - charge is injected and extracted at the: (1) base and backbone (blue solid); and (2) base only (magenta dash). Transmission for the strands without the backbone is shown by the red dash-dot curve. The transmission through the base only is slightly smaller than that through the base and backbone, and the overall shapes of transmission with and without the backbone qualitatively seem to be similar but the two curves are shifted in the energy axis.

and without the backbone qualitatively seem to be similar but the two curves are shifted in the energy axis.

To obtain a quantitative understanding, we shift the energy axis of the transmission for the strand without the backbone to the right-hand side by 0.85 eV for Seq. 1 and 2, and 1 eV for Seq. 3 and 4, as shown in Figure 4.4. A close inspection of Figure 4.4 reveals that the difference in magnitude of transmission for the strands with and without the backbones can be up to 5 orders for Seq. 1, 6 orders for Seq. 2, 7 orders for Seq. 3 and 8 orders for

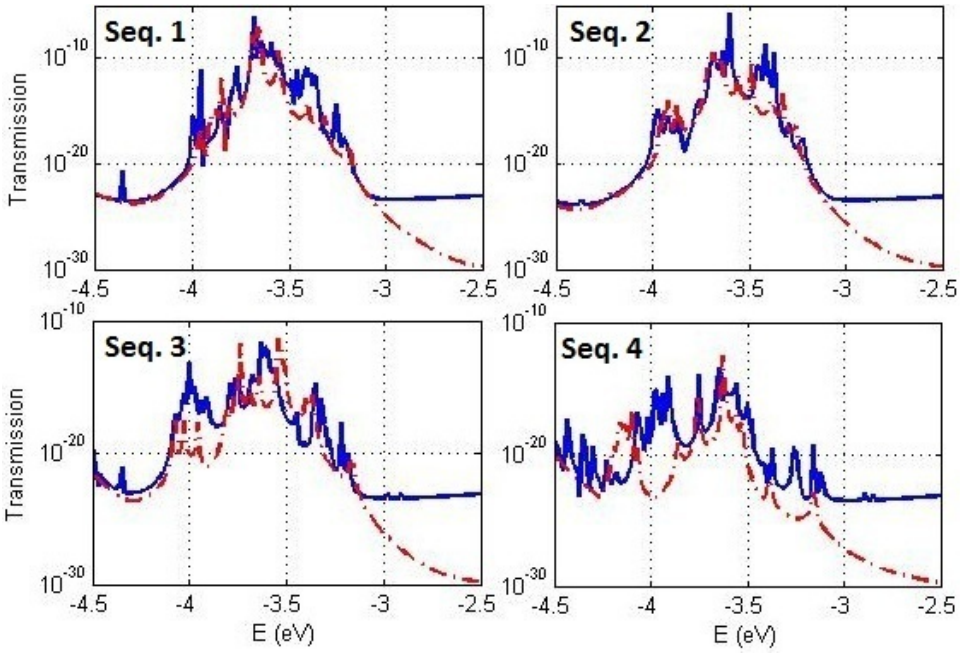
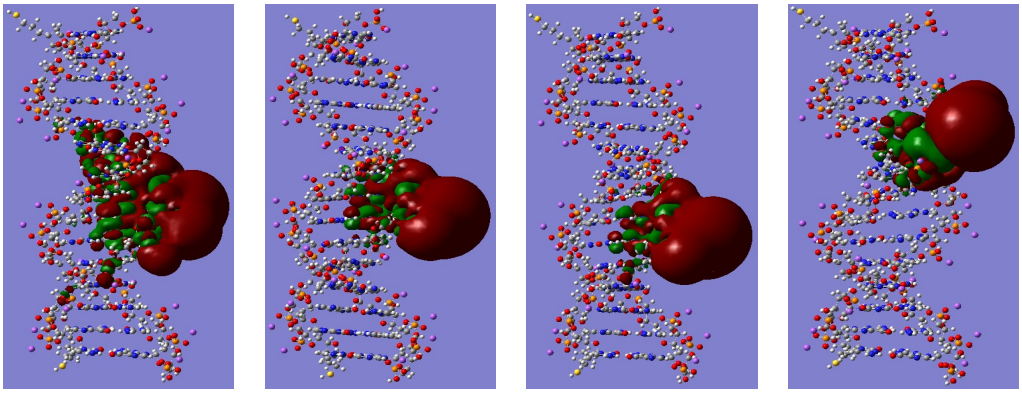


Figure 4.4: Transmission v.s. energy for the four DNA strands with and without backbone in the phase-coherent limit. The transmission for the strands with backbone is shown in blue curves, and that for strands without backbone is shown in red with the energy axis shifted 0.85 eV for Seq. 1 and Seq. 2, and 1 eV for Seq. 3 and Seq. 4. The difference in magnitude of transmission for the strands with and without the backbones can be significant at some energy points.

Seq. 4 at some energy points.

We further study the effect of backbone by plotting the HOMO orbitals for the four strands. HOMO orbitals for DNA strands with the backbones have been studied by a few groups [62, 122, 136]. It has been shown that the spatial distribution of HOMO orbitals depends on a variety of conditions, such as the position of counterions relative to the phosphate groups [62] and the hydration levels [136]. In addition, Reference [122] states that the HOMO orbitals can temporarily have a large weight on the backbones as a function of time. In our calculation, we find that the HOMO orbital is distributed over both the bases and the backbone, as shown in Figure 4.5.

We conclude from the above observations related to transmission (Figure 4.4) and wave



(a) (b) (c) (d)

Figure 4.5: Iso-surfaces of the HOMO orbitals for (a) Seq. 1; (b) Seq. 2; (c) Seq. 3; and (d) Seq. 4. Red color represents the positive part of the wave function and green represents the negative part of the wave function. The HOMO orbital is distributed over both the bases and the backbone.

function (Figure 4.5) that while most of the contribution to charge transport comes from the bases, the phosphate groups, sugar rings and sodium ions on the backbone alter the electronic structure sensitively by interacting with the bases. This leads to mismatches in the coherent transmission for strands with and without the backbone at some energy points while keeping the overall shape similar. We also note that the coherent transmission through the HOMO-LUMO gap also depends on the inclusion of the backbone. This finding indicates that when modeling the coherent charge transport in DNA molecules, one should take into account the effect of backbone. In the remainder of this chapter, we will focus on the strands with the backbones, injecting and extracting from and into the strands via both backbone and base.

4.4 Effect of decoherence

In this section, we examine the effect of decoherence on DNA conductance using the same strands as in the previous section. Besides the interaction between the molecule and the metal contacts, the position of the Fermi level E_f relative to the molecular energy levels is

also important in understanding the flow of charge. E_f depends sensitively on the surface condition and work function difference between the two materials. Experimentally, it is very difficult to determine the position of E_f. Here, we explore the conductance when the Fermi level is in the HOMO vicinity. At the low bias region, only electrons with energies nearby the Fermi level contribute to the transport. Thus, the transmission $T(E) \cong T(E_f)$. From DFT calculations, the HOMO levels of the 4 strands have been determined to be $-3.18\ eV$, $-2.84\ eV$, $-2.91\ eV$, $-2.85\ eV$, respectively.

The zero-bias conductance versus Fermi energy is shown in Figure 4.6. Comparing the conductances of Seq. 2, Seq. 3 and Seq. 4, we find that the $A:T$ base pair plays the role of a barrier. This is consistent with the fact that in our calculation the IP of an isolated $G:C$ base pair is $7.12\ eV$ while that of the $A:T$ base pair is $7.68\ eV$ The similarity of the conductance for Seq. 1 and Seq. 2 can be attributed to the fact that Seq. 1 has a $G:C$ base pair in the sixth position and a $C:G$ base pair in the seventh position which can cause interstrand hopping, and result in lower transmission while in the Seq. 2 both the sixth and seventh positions are placed with $G:C$ base pairs. However, the presence of $A:T$ base pair in the eighth position cancels out this advantage. More importantly, in this coherent model, we find that the conductivity is much smaller than the experimental results. At the low bias region, the experimental value [54] of the conductance for Seq. 1, Seq. 2, Seq. 3 and Seq. 4 is around 5×10^{-10} S, 2×10^{-10} S, 3.5×10^{-11} S and 6×10^{-12} S, respectively. However, in the coherent case of our calculation, the conductance is only about $10^{-24}S$ for Seq. 1 and 10^{-28} S for Seq. 2, Seq. 3 and Seq. 4 when the Fermi level is nearby the HOMO. For Seq. 1, the coherent result is about 10^{14} times smaller than the experiment; for Seq. 2, 10^{18} times smaller; for Seq. 3, 10^{17} times smaller and for Seq. 4, 10^{16} times smaller.

The huge difference between the coherent conductance and experimental values indicates that there are other mechanisms that play a role in determining the conductance. We rule out the location of Fermi energy, because irrespective of where the Fermi energy is in Figure 4.6 (including well within the HOMO band), the conductance is significantly smaller than in the experiments. We also rule out the nature of the coupling to contacts because we have

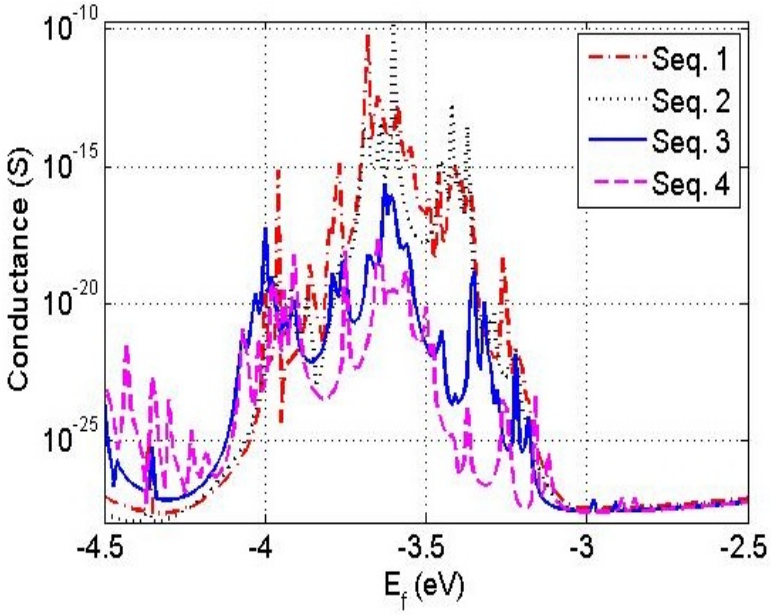


Figure 4.6: Conductance v.s. Fermi energy for the four DNA strands without any decoherence. The coherent conductance is orders of magnitude smaller than experiment irrespective of the locations of Fermi levels for all the four strands.

verified that our results change by small amounts when the coupling to the contacts Γ_L and Γ_R is varied from 50 to 100 meV. Another reason for the larger experimental conductance could be the possibility of having a large number of strands, which is extremely unlikely in the break junction geometry of Reference [54].

The mismatch in energy level between neighboring bases combined with the small interbase coupling leads to the smaller than experimental conductances in the phase coherent calculations. Decoherence, due to interaction with the environment, broadens these energy levels and can lead to a larger conductance. To account for decoherence, we use Büttiker probes as presented previously.

As shown in Figure 4.6, for hole transport, $G:C$ base pair is preferred while $A:T$ base pair is considered as a barrier because the transmission decreases with increasing number of $A:T$ base pairs. As the transit time through a rigid barrier $A:T$ is small, we first include decoherence only on the $G:C$ base pairs with the coupling strength $\Gamma_i = 5\ meV$, 6 meV

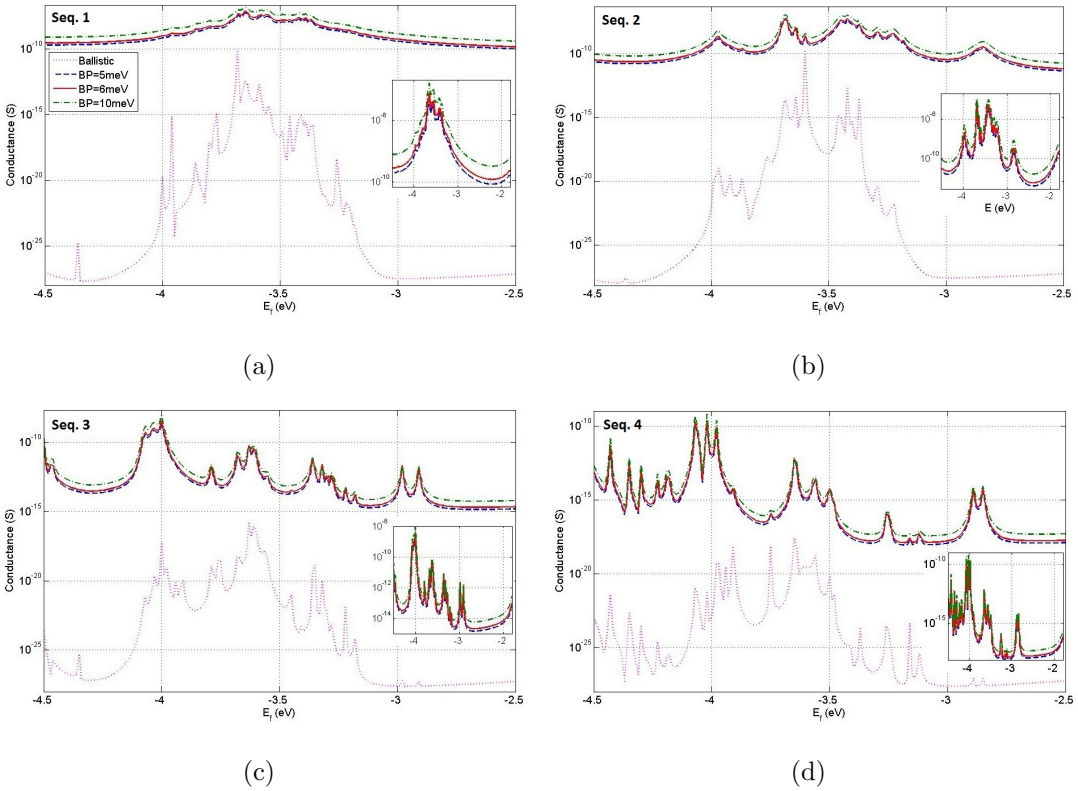


Figure 4.7: Conductance v.s. Fermi energy with 5 meV, 6 meV and 10 meV decoherence only on $G:C$ base pairs for (a) Seq. 1; (b) Seq. 2; (c) Seq. 3; and (d) Seq. 4. The conductance values for Seq. 3 and Seq. 4 around their HOMO levels are still too small to explain the experiment, suggesting that the decoherence on $A:T$ barrier is also important.

and 10 meV, as shown in Figure 4.7.

Compared with the ballistic transport, one can find that after including Büttiker probes the conductance is smoother for the four strands. The small peaks in the ballistic conductance are smeared out due to the broadening of energy levels. For Seq. 1, we find that when we add 5 meV decoherence only on the $G:C$ base pairs, the conductance increases by around 10^{15} times around their HOMO levels, while for Seq.2, Seq. 3 and Seq. 4, the increase is 10^{18}, 10^{16} and 10^{13} respectively. The tremendous enhancement in the conductance after including Büttiker probes suggests the crucial role that decoherence has in charge transport through DNA molecules.

In addition, we find that for strands containing $A:T$ base pairs, the same values of decoherence rates can enhance the conductance more effectively for Seq. 2 than Seq. 3 and Seq. 4. We observe that when the decoherence on $G:C$ base pairs is 5 meV, as the number of $A:T$ base pairs increases from 1 to 5, the increase in conductance decreases from around 10^{18} to 10^{13}. This is because in Seq. 3 and Seq. 4, where the numbers of $A:T$ are 3 and 5, respectively, the wider barriers begins to play a larger role than in Seq. 2. We also find that when the decoherence on $G:C$ base pairs changes from 5 meV to 10 meV, the conductance increases around 4 times for Seq. 1 and Seq. 2 around the HOMO levels. However, for Seq. 3 and Seq. 4, when the decoherence value on $G:C$ varies from 5 meV to 10 meV, the conductance only increases 1.5 times. More importantly, for Seq. 1 and Seq. 2, as the Büttiker probe coupling strength changes, we can fit the experimental conductance by setting the Fermi energy to be a particular value. For instance, when the decoherence on $G:C$ is 5 meV if we set the Fermi level to be -3.06 eV for Seq. 1 and -2.82 eV for Seq. 2, the computed conductance is comparable to that of the experiment. However, for Seq. 3 and Seq. 4, where the wide $A:T$ barriers are present, irrespective of the strength of decoherence chosen (varying from 0 to 10 meV), the conductances are still too small when compared with the experiments. For example, for Seq. 3 the best conductance value is 10 times smaller than the experiment and for Seq. 4, the best value is 100 times smaller. This indicates that the decoherence on $A:T$ barriers is also important, that is, the $A:T$ barrier is not a static barrier for hole transport. For this reason, we study the effect of decoherence on $A:T$ base pairs to get better agreement with the experiments.

Figure 4.8 shows the conductance versus Fermi energy with different decoherence values on both $G:C$ and $A:T$ base pairs.

It is found that for Seq. 2, the decoherence on $G:C$ base pairs is more important than that on $A:T$ base pairs. For example, when the decoherence on $A:T$ base pair is fixed at 2 meV, the conductance increases 1.5 times with the decoherence on $G:C$ base pairs increasing from 5 meV to 6 meV. However, if the decoherence on $G:C$ base pairs is fixed at 6 meV, the conductance stays unchanged at the HOMO level when the decoherence on

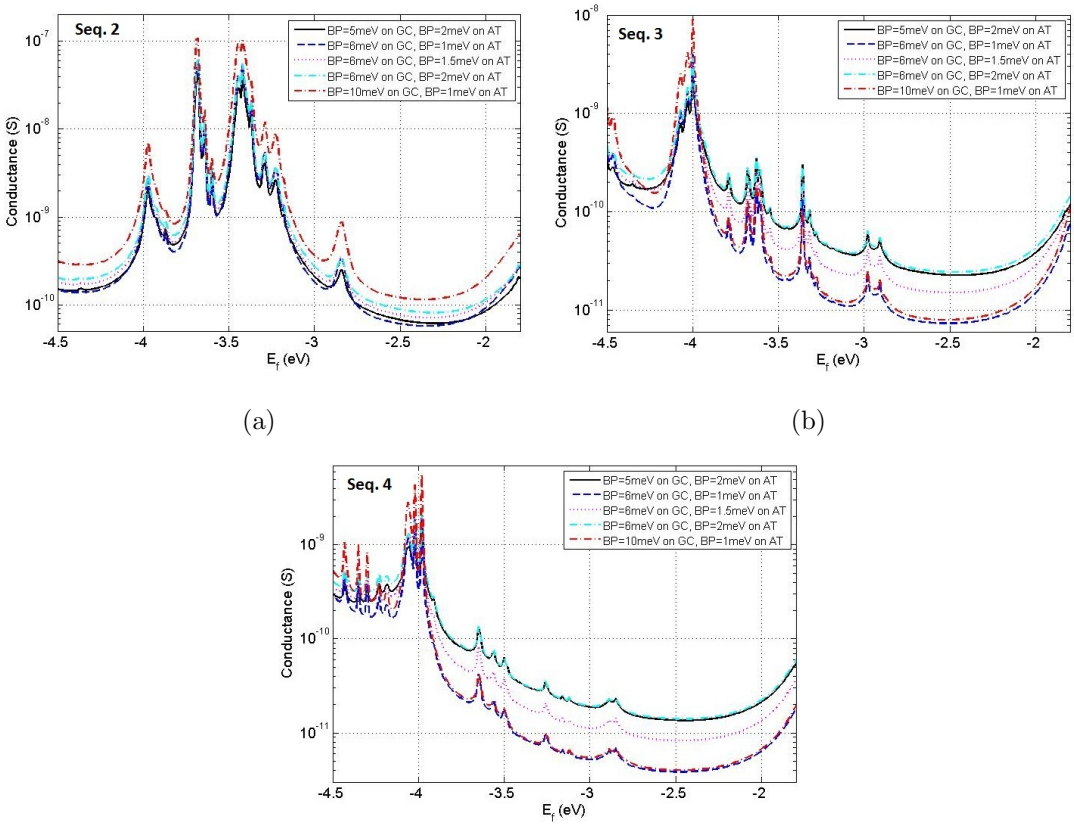


Figure 4.8: Conductance v.s. Fermi energy with different decoherence values on both $G:C$ and $A:T$ base pairs for (a) Seq. 2; (b) Seq. 3; and (c) Seq. 4. Decoherence on $A:T$ base pairs can increase the conductance effectively, especially for Seq. 3 and Seq. 4.

$A:T$ base pair is varied from 1 meV to 2 meV. In contrast, for Seq. 3 and Seq. 4, the decoherence on $A:T$ base pairs are much more important. Take Seq. 3 for example. If the decoherence on $A:T$ base pairs is fixed at 2 meV and the decoherence on $G:C$ is varied from 5 meV to 6 meV, the conductance only changes around 1.1 times. Similarly, if the decoherence on $A:T$ is fixed at 1 meV and the decoherence on $G:C$ is changed from 6 meV to 10 meV, the conductance also only changes around 1.1 times. However, if the decoherence on $G:C$ is fixed at 6 meV and the decoherence on $A:T$ is changed from 1 meV to 2 meV, the conductance increases around 2.8 times.

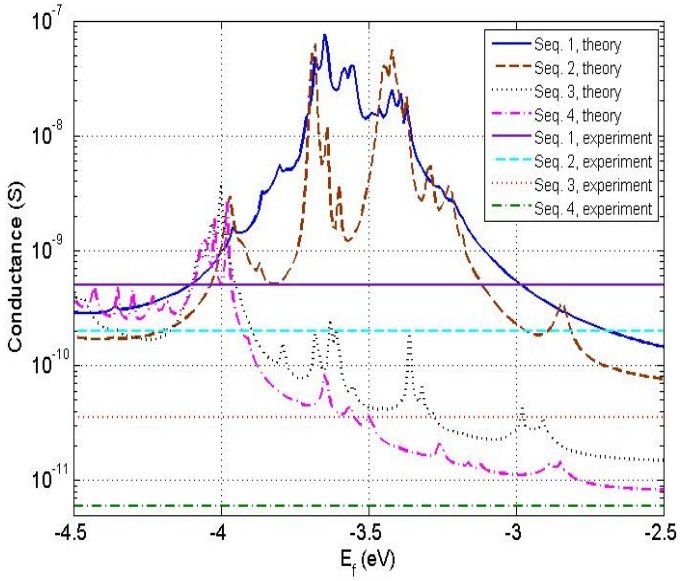


Figure 4.9: Conductance v.s. Fermi energy with 6 meV decoherence on $G:C$ and 1.5 meV decoherence on $A:T$. The conductance has been increased effectively. The conductance values for the 4 strands around their HOMO levels are quantatively comparable to the experimental results, demonstrating the importance of adding decoherence on both $G:C$ base pairs and $A:T$ barrier.

We find that for the four strands a decoherence strength of 6 meV on $G:C$ base pairs and 1.5 meV on $A:T$ base pairs give a relatively good agreement with experiments. The results are summarized in Figure 4.9, where the experimental values of conductance from Reference [54] are also shown, with horizontal lines.

We find that for the four strands, the computed conductance lies in the ball park of the experimental values. That is, the qualitative trend of the conductance variation for the four sequences matches well with the experiments (Conductance decreases from Seq. 1 to Seq. 4). More importantly, in contrast to the phase coherent results in Figure 4.6, where the conductance values are orders of magnitude smaller than the experiments, the conductance values in Figure 4.9 are now comparable to the experiments. This qualitative match holds well irrespective of the Fermi energy as can be seen in Figure 4.9. Although for Seq. 4, the computed conductance is about 2 times larger than the experimental result around the

HOMO level, we still think that the fitting is good because of the uncertain factors (such as Fermi energy location) in the experiments.

These results indicate that the decoherence is important in experiments involving dry DNA, and this might be one of the reasons why the theory has difficulty in explaining experiments. Recently, there have been proposals to calculate the conductance of DNA by calculating the phase coherent conductance for many points (each of which provides a static set of distinct coordinates) along the trajectory of and MD simulation and averaging the coherently calculated conductance [137]. These coordinates include the effect of time varying ions and water molecules, apart from thermally induced lattice vibrations. It would be interesting to compare results from the two methods for the case of dry DNA to see if one can match experiments on the linear response conductance and current-voltage characteristics.

4.5 Summary

In this chapter, we have modeled the zero-bias conductance of dry DNA using a combination of DFT and the phenomenological Büttiker probes to account for decoherence. We first explore the effect of the backbone in charge transport by comparing the coherent transmission for strands with backbones and strands whose backbones have been deleted. We find that the DNA backbone can affect the coherent transmission significantly at some energy points. However, the overall shapes of the transmission are similar for the strands with backbones and strands without backbones. More interestingly, we find that the calculated conductance using phase coherent transport is orders of magnitude smaller than experiments, and that including the effect of decoherence on $G:C$ is crucial because it broadens the energy levels of DNA and significantly enhances the conductance. By comparing with experiments, we find that quite a large decoherence is required even on the $A:T$ base pairs, even though they behave as barriers in the energy ranges of importance. It is also worth noting that for the $G:C$ rich strands (strands which contain one $A:T$ base pair), the decoherence on $G:C$ base pairs is more important than that on $A:T$ base pairs, while for the strands containing three or five $A:T$ base pairs, the decoherence on $A:T$ plays a more significant role.

By analyzing the four different DNA strands, it has been determined that the decoherence strength due to Büttiker probes is approximately 6 meV for $G:C$ base pairs and 1.5 meV for $A:T$ base pairs. While the phenomenological Büttiker probes is able to explain the experiments qualitatively at low biases, a more accurate description of vibronic coupling would be necessary to explain experiments at large biases. This is a significantly more difficult problem requiring knowledge of both the vibronic modes and the coupling strength in DNA, which is well beyond the extent of this work.

This chapter is based on a paper by Jianqing Qi, et al., published in *Physical Review B* [138].

Chapter 5

ROLE OF CYTOSINE METHYLATION ON CHARGE TRANSPORT THROUGH A DNA STRAND

Cytosine methylation has been found to play a crucial role in various biological processes, including a number of human diseases. The detection of this small modification remains challenging. In this chapter, we computationally explore the possibility of detecting methylated DNA strands through direct electrical conductance measurements. Using DFT and the Landauer-Büttiker method, we study the electronic properties and charge transport through an eight-base pair methylated DNA strand and its native counterpart. We first analyze the effect of cytosine methylation on the TB parameters of two DNA strands and then model the transmission of the electrons and conductance through the strands both with and without decoherence. We find that the main difference of the TB parameters between the native DNA and the methylated DNA lies in the onsite energies of (methylated) cytosine bases. The intra- and inter- strand hopping integrals between two nearest neighboring guanine base and (methylated) cytosine base also change with the addition of the methyl groups. Our calculations show that in the phase-coherent limit, the transmission of the methylated strand is close to the native strand when the energy is near HOMO level and larger than the native strand by 5 times in the bandgap. The trend in transmission of coherent transport also holds in the presence of the decoherence. The lower conductance for the methylated strand in the experiment is suggested to be caused by the more stable structure due to the introduction of the methyl groups. We also study the role of the XC functional and the effect of contact coupling by choosing coupling strengths ranging from weak to strong coupling limit.

5.1 Introduction

Cytosine methylation refers to the replacement of the hydrogen atom by a methyl group at the 5th position of the pyrimidine ring [139], as shown in Figure 5.1. This small modification has been found to be important in changing physical properties of DNA. For example, it can increase the melting temperature of the duplex [140], the molecular polarizability of the pyrimidine [141] and enhance the stability of the double strands. Recent study has also revealed that the methylation of cytosine bases has an essential effect on the separation of DNA strand [142]. More importantly, it plays a crucial role in many biological and cellular processes such as gene expression regulation [143, 144], cellular differentiation [145], X-chromosome inactivation [146], embryonic development [147, 148, 149, 150, 151] and genomic imprinting [152]. As a consequence, an increasing number of human diseases [153], including cancer [154, 155], have been found to be associated with the abnormal activities of DNA methylation.

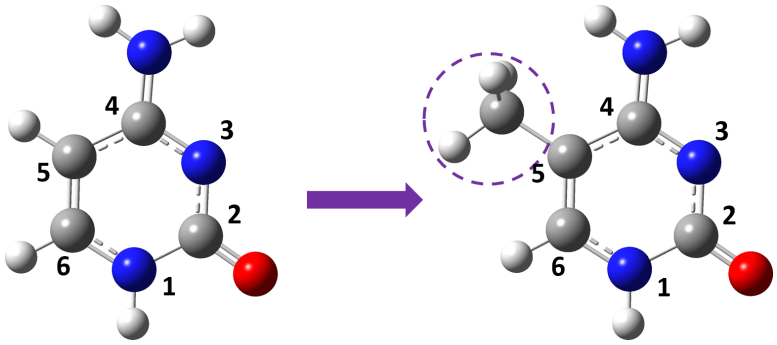


Figure 5.1: Structure of a native cytosine base (left) and a 5′-methylcytosine base (right). On the right side, the hydrogen atom at the 5′ position is replaced by a methyl group.

Targeting and identifying the methylated DNA is important for disease diagnosis and drug design. The main detection techniques used today involve PCR-based methods [156], which usually require chemical reactions [157, 158] to maintain the methylation feature. With the rapid development and progress in single-molecule electrical approaches [10, 54, 57, 159], new techniques utilizing the electronic properties of DNA bases have also been used

to experimentally detect the methylated strands [39, 41, 42, 43, 44, 45, 160, 161]. These pioneering initial works show that it may be possible to detect methylated sequence/base by direct conductance measurement.

Despite the recent progress in experiment, understanding charge transport through a methylated DNA sequence/base from a theoretical perspective is lacking. In this chapter, we study charge transport through an eight-base pair methylated DNA strand and its native counterpart used in Reference [44]. We aim to provide an understanding of the effect of cytosine methylation on charge transport through DNA strand. To this end, we use the combination of DFT and the Landauer-Büttiker approach.

DFT is a powerful tool to obtain the accurate description of the electronic structure of molecules and materials. However, it is well known that in DFT calculations the results may depend on the choice of XC functionals and basis sets. In addition, the lack of the derivative discontinuity and self-interaction errors in standard functionals like GGAs and hybrid-GGAs can lead to erroneous single-molecule transport predictions [162]. In this chapter, we also study the effects of methylation on the eight-base pair DNA strand with the B3LYP and CAM-B3LYP functionals.

The remainder of this chapter is organized as follows: we first present DNA strands used in this work - an eight-base pair $G:C$-rich B-form strand and its methylated counterpart. Next, we show the results from our calculations. We first examine the electronic properties of a single methylated base and base pair by comparing their IP, EA (electron affinity), hydrogen bonding and the HOMO orbital distribution with the native counterparts. We then study the structure of the eight-base pair strands. We first look at the effect of methylation on the TB parameters, including the onsite energies of the guanine and cytosine bases and hopping integrals between the first and second nearest neighboring bases. We further investigate the influence of cytosine methylation on the electronic properties of DNA by studying the transmission and conductance in both phase-coherent and decoherent cases. We also study the role of XC functionals on DNA charge transport. Then, we investigate the effect of contact coupling to understand the effects of the delocalization and self-interaction errors

from XC functionals on the transmission. Finally, we summarize our findings in the last section.

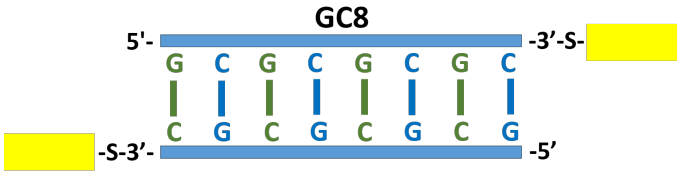
5.2 Experimental strands considered

In this chapter, we mainly focus on the following two DNA sequences, each of which has eight base pairs: (1) the native strand, $5'-GCGCGCGC-3'$ ($GC8$), where G refers to the guanine base and C refers to the native cytosine base as defined in Chapter 1; (2) the methylated counterpart, $5'-GCmGCmGCmGCm-3'$ ($GCm8$), where Cm refers to the methylated cytosine base. In the experiment, thiol groups are used to provide interactions between the molecule and the gold electrodes at the $3'$ ends. The sketches of these two strands with thiol groups and gold electrodes are shown in Figure 5.2. This choice of sequence is motivated by the experiment in Tao's group [44], where the scanning tunneling microscope (STM) break junction technique is used and conductance distribution histograms are obtained. The purpose of the experiment is to detect disease by comparing the conductance of strands containing mutated cytosine with the ones containing native cytosines. Different from the experiment in Reference [54], the experiment here is done in solution, where numerous water molecules and counterions are expected to surround the DNA molecules. However, in order to only focus on the intrinsic properties of the methylated DNA, we exclude the effect of water molecules and counterions.

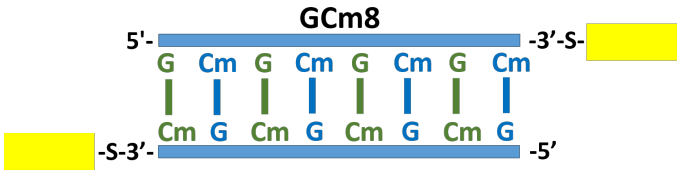
The measured conductance results from the experiments in Reference [44] are shown in Table 5.1, where six sets of experiments are involved, in each of which the conductance value of $GCm8$ is lower than that of native $GC8$. One can also see that the measured conductance value for the same molecule varies from experiment to experiment. This fluctuation indicates that the conductivity of DNA suffers from the variation in the local environment.

5.3 Method

The approach we use in this chapter has been discussed in detail in Chapter 2, following the steps in flow charts shown in Figures 2.2 and 2.8. Here we describe the structures of the



(a)



(b)

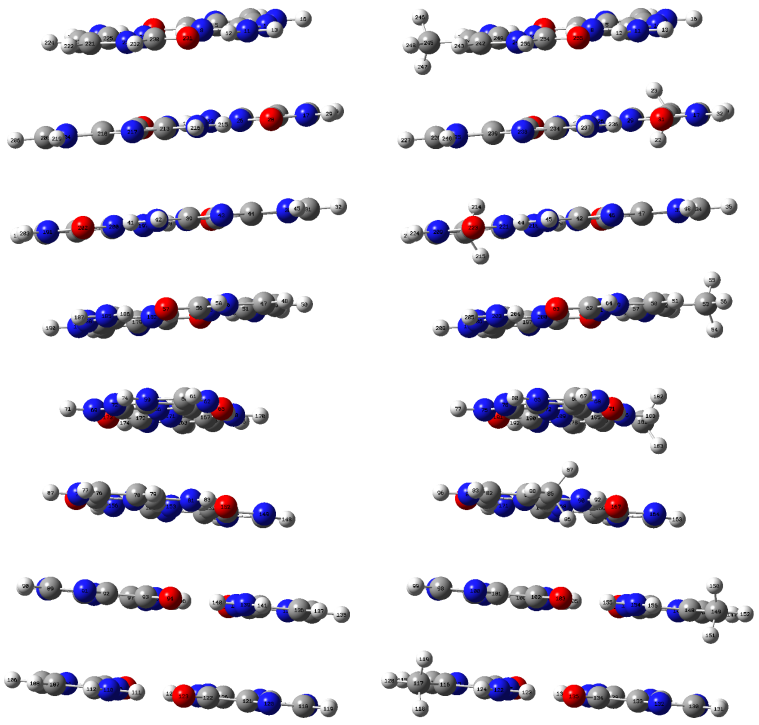
Figure 5.2: DNA strands used in Reference [44]. (a) $GC8$; (b) $GCm8$. Also presented are the thiol groups and the gold electrodes.

Table 5.1: Conductance values for six experiments on each molecule. μ is the mean for all experiments and σ is the standard deviation. This data is from Reference [44].

	1	2	3	4	5	6	μ	σ
$GC8$ $(\times 10^{-4} G_0)$	2.62	2.41	2.27	2.26	2.22	2.18	2.33	0.163
$GC8m$ $(\times 10^{-4} G_0)$	2.23	2.20	2.09	2.00	1.95	1.41	1.98	0.299

strands used in current calculation. The atomic coordinates of the B-form DNA strands are generated with NAB, as discussed in Chapter 2. Because the conduction through a B-form DNA strands is mainly caused by the overlap of π orbitals in the nucleobases [16] and our prior work shows that in the vicinity of the HOMO level the transport is mainly through the bases [138], we delete the backbone to reduce the expensive computational cost and terminate the nucleobases with hydrogen atoms. For the methylated strands, MD simulations show that the methyl groups do not have large influence on the Watson-Crick base pairing for B-form DNA [163]. Thus, we assume that the methylated strand maintains the same B-form as

the native one. We first replace the hydrogen atom in the 5th position with a $-CH_3$ methyl group, then relax the methyl group at the B3LYP/6-31G(d) [131, 132, 164, 165] level to get the optimized structure of the methylated cytosine. The structures used in our calculation in this chapter are shown in Figure 5.3.



(a) (b)

Figure 5.3: Simplified DNA strands used in our calculations in this chapter, where the backbones have been deleted and the bases are terminated with hydrogen atoms. (a) $GC8$; (b) $GCm8$.

Next, we carry out the DFT calculations using the B3LYP XC functional and 6-31G basis set. After the full Hamiltonian with an orthogonal basis H is obtained, we extract the TB parameters at the HOMO level to build a model Hamiltonian, following Equation 2.13. Specifically, we extract the onsite energies of a single DNA base and the hopping integrals between two neighboring DNA bases, as shown in Figure 2.3(e). We emphasize that we use the full Hamiltonian to compute the transport properties of the DNA strands, while the TB

Hamiltonian is used to further our understanding of the charge transport mechanism. We finally calculate the transmission and conductance of the molecules using Equations 2.28, 2.39 and 2.16. To account for decoherence, we use the Büttiker probes, as discussed in Chapter 2.

5.4 Results and discussion

5.4.1 Electronic properties of a single methylated base and base pair

To study the role of the methyl groups in DNA, we first compare the electronic properties of a single Cm base and a single $(G:Cm)$ base pair with their native counterparts (C and $G:C$). The atomic coordinates of these structures are shown in Appendix B. The IP and EA values for C, Cm, $G:C$ and $G:Cm$ calculated with B3LYP/6-31G in gas phase are shown in Table 5.2. Hydrogen bondings between the base pairs $G:C$ and $G:Cm$ are also presented.

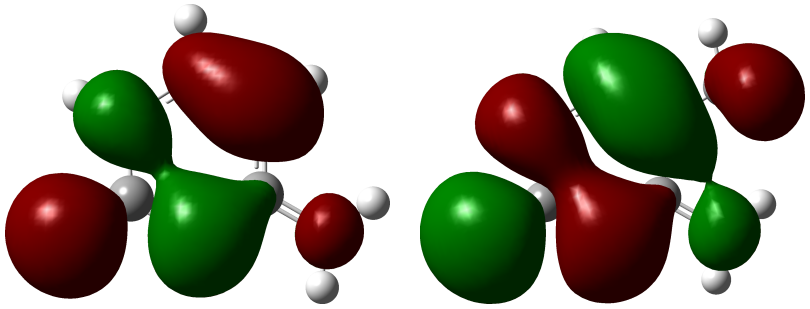
Table 5.2: IP and EA for a single C base, a single Cm base, a single $G:C$ base pair and a single $G:Cm$ base pair. The calculation is based on B3LYP/6-31G. Hydrogen bondings between the base pairs $G:C$ and $G:Cm$ are also presented.

Quantities (eV)	IP	EA	hydrogen bonding
C	8.277	1.291	-
Cm	7.997	1.300	-
$G:C$	7.115	0.707	1.524
$G:Cm$	7.064	0.736	1.547

One can see that the addition of the methyl group decreases the IP by 280 meV for the single C base and 51 meV for the single $G:C$ base pair. However, the presence of the methyl group has only a small influence on the EA. EA increases by 9 meV for a single base, and 29 meV for a single base pair. Therefore, we expect the methyl groups plays a more important role in the hole transport. We also report the strength of the hydrogen bonding

between the two complementary bases within a $G : C$ and $G : Cm$ base pair, as shown in Table 5.2. The hydrogen bonding for the methylated base pair is 23 meV stronger than that of the native base pair. This suggests that the methylated base pair has a more rigid structure and is more stable when it interacts with the water molecules and counterions. Thus, it suffers from less fluctuations when placed in solution.

We also investigate the HOMO orbital distribution of the single C base and the single Cm base, as shown in Figure 5.4. We find that the presence of the methyl group can induce a charge re-distribution on the base. The HOMO resides partially on the $-CH_3$ group.

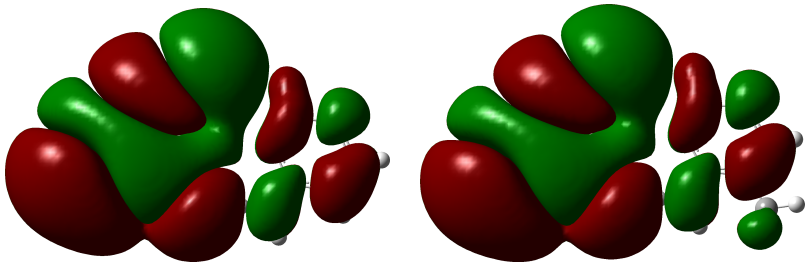


(a) (b)

Figure 5.4: HOMO orbital distribution for (a) a single C base; and (b) a single Cm base calculated by B3LYP/6-31G in gas phase with iso-value = 0.02. The presence of the methyl group can induce a charge re-distribution on the base. The HOMO resides partially on the $-CH_3$ group.

We also study the HOMO orbital distribution of the methylated base pair and compare it with the native one, as shown in Figure 5.5. It is well accepted that the G base has a lower IP and it is easier for the hole to reside on it. Thus, in both Figures 5.5(a) and 5.5(b), a large portion of the HOMO orbital locates on the G base. We also observe that the methyl group induces a change in the charge re-distribution of the C base.

The comparisons of the IP, EA, hydrogen bonding and HOMO orbital distribution between the single methylated cytosine base and base pair with their native counterparts suggest that the introduction of the methyl group can change the electronic properties of a



(a) (b)

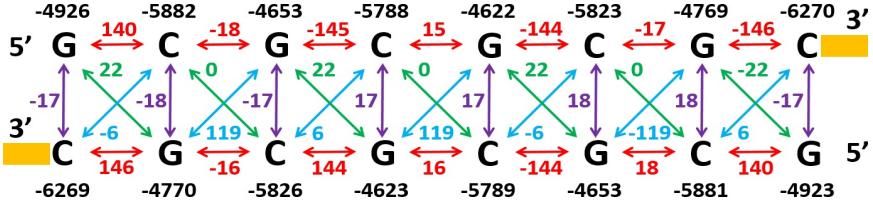
Figure 5.5: HOMO orbital distribution for (a) a single $G : C$ base pair; and (b) a single $G : Cm$ base pair calculated by B3LYP/6-31G in gas phase with iso-value = 0.0005. A large portion of the HOMO orbital is located on the G base in both base pairs. The methyl group induces a change in the charge distribution of the C base.

DNA. Therefore, we expect that this small modification also influences the charge transport properties in DNA. In the following sections, we focus on the two eight-base pair strand, $GC8$ and $GCm8$. We first compare the TB parameters of the HOMO level between $GC8$ and $GCm8$, and then compare the transmission and conductance of $GC8$ and $GCm8$ for both phase-coherent and decoherent cases. Two different XC functionals, the B3LYP and the CAM-B3LYP based range-separated functionals, are considered.

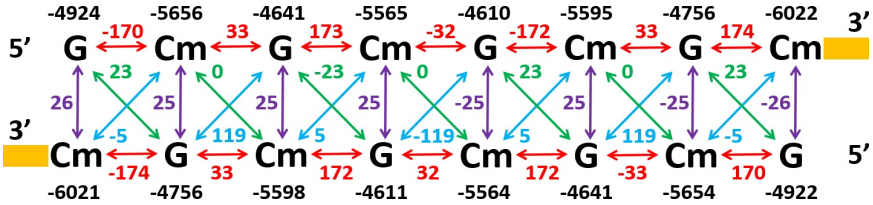
5.4.2 Tight-binding parameters of the HOMO orbital

In this part, we analyze the TB parameters of the HOMO level for $GC8$ and $GCm8$, including the onsite energy of each base and hopping integrals between two neighboring bases. Onsite energies and hopping integrals are two important factors in determining the charge transport properties of a DNA strand. These parameters can be extracted from the Hamiltonian H in Equation (2.12) using the TB representation in Equation (2.13). The result with B3LYP is shown in Figure 5.6, where the numbers shown in black are the onsite energies, the arrows and numbers shown by other different colors correspond to the hopping integrals between two neighboring bases both within and between the complementary strands.

Our values of the TB parameters are comparable to those in previous work [51, 77, 166,



(a)



(b)

Figure 5.6: TB parameters of the HOMO level obtained from DFT calculation with B3LYP/6-31G for (a) $GC8$; and (b) $GCm8$ (Unit, meV). Arrows and numbers shown in various colors follow the representations in Figure 2.3(e).

167, 168]. Each base on a single strand is separated from its neighbors by 3.4 $\mathring{A}$. This large inter-base distance makes the hopping integral between bases small. To appreciate this, we note from Figure 5.6 that the TB parameter to hop between consecutive bases is smaller than 200 meV. In comparison, nanotubes and nanowires have TB parameters that are almost ten times larger ($2-3\ eV$). As DNA strands are floppy and melt at low temperatures, vibrational coupling modifies electronic transport significantly as the length of DNA increases.

The significant differences between the parameters of $GC8$ and $GCm8$ at HOMO level are the onsite energies of Cs and Cms, and the hopping integrals between two nearest neighboring G and C (or Cm). For example, in Figure 5.6 we can see that the values of the onsite energies for Cs in $GC8$ are around $-5800\ meV$, while the values for Cms in $GCm8$ are around $-5600\ meV$, which are approximately 200 meV higher than the native cytosine bases. The absolute values of the intra-strand hopping integrals between the two nearest neighboring G and Cm from the $5'$ to $3'$ ends and from $3'$ to $5'$ ends increase by 30 meV and 15 meV respectively, compared to $GC8$, while the absolute values of the inter-strand

hopping integrals between the two nearest neighboring G and Cm increase by about 8 meV. It is also interesting to see that the absolute values of the intra-strand hopping between the two nearest neighboring G and C (or Cm) in the $5'$ to $3'$ and $3'$ to $5'$ directions are different by almost 9 (5) times for $GC8$ ($GCm8$), which reveals the orientation-dependent nature of electron hopping in the double helical structure.

5.4.3 Transport properties

The above discussions show that the small modification on cytosine bases can cause a change in TB parameters at the HOMO level of a DNA strand. To understand the role of methylation on transport properties, we calculate the transmission and conductance of $GC8$ and $GCm8$ in both phase-coherent and decoherent cases, using the full Hamiltonian from the DFT calculation and focus on the energy region which is in the vicinity of the HOMO level. In our calculations, the HOMO energies for $GC8$ and $GCm8$ computed with B3LYP/6-31G are -4.519 eV and -4.512 eV, as shown in Table 5.3. We set the coupling strengths to be 100 meV for $\Gamma_{L(R)}$ and 5 meV for Γ_i (see Equation 2.35 in Section 2.3.3).

5.4.3.1 Phase-coherent transport

We first consider the phase-coherent transport in $GC8$ and $GCm8$. The transmission and conductance of the two strands in the HOMO vicinity using B3LYP are shown in Figure 5.7 and its inset. We find that the transmission of $GCm8$ is very close to that of $GC8$ nearby the HOMO level, while away from the HOMO (when $E > -4$ eV), the transmission of $GCm8$ is larger than $GC8$ by about 5 times in the bandgap. As one moves into the HOMO band, the transmission of both strands are oscillatory, and the transmission values for $GCm8$ can be larger than $GC8$ in some energy windows and smaller in other windows. The ratio of the transmission between $GCm8$ and $GC8$ in the energy window from -5.5 eV to -4 eV ranges from 3.37×10^{-5} to 1.42×10^{5} times.

Here, we provide a discussion based on the picture for conduction that is used in semiconductors. Needless to say, there are assumptions in borrowing this picture for conduction in

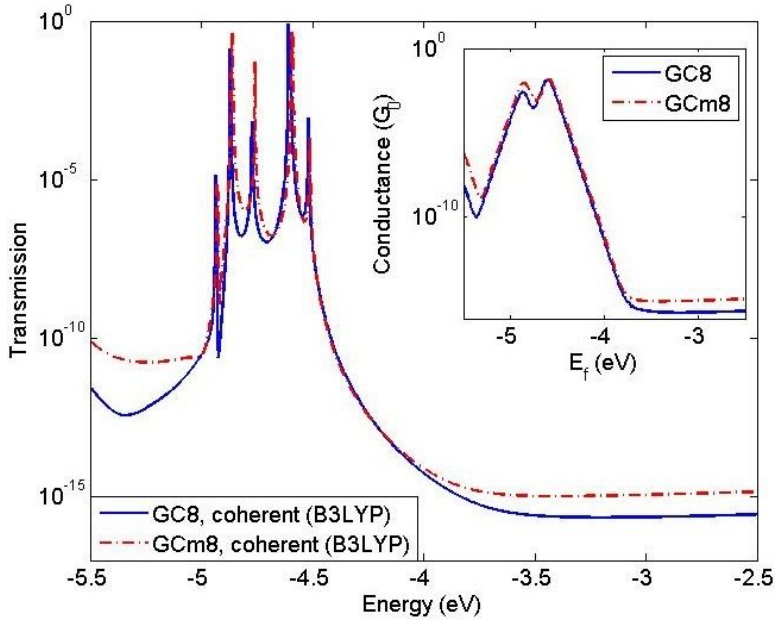


Figure 5.7: Transmission v.s. energy for $GC8$ (blue solid) and $GCm8$ (red dash) in phase-coherent transport calculated with B3LYP. Inset, conductance v.s. Fermi energy for $GC8$ (blue solid) and $GCm8$ (red dash).

DNA because dipoles on the molecule-contact interface and detailed geometry of the contact atoms are not taken into account. However, it is of some value to ferret out an answer from this analysis. To proceed, we also show the linear response conductance as a function of Fermi energy in the inset of Figure 5.7, with a shift in the energy axis by $|\mathrm{IP}_{GCm8} - \mathrm{IP}_{GC8}|$ for $GCm8$. We note that the conductance of $GCm8$ is larger than that of $GC8$ by about 2 times for Fermi energy locations nearby the HOMO level and 5 times away from the HOMO in the bandgap.

It is also interesting to note that the conductance plot in the inset of Figure 5.7 points to the fact that if the Fermi energy is shifted by as small as 50 meV (say, due to location of contact atoms), the conductance can change by as much as $5 - 10$ times. It is of relevance to note that in the experiments the conductance fluctuates between various measurements, though the mean conductance for $GC8$ and $GCm8$ are well defined. Apart from the changes in the coupling between the contact and DNA, Figure 5.7 also shows that variations in the

location of the Fermi level by as small as 50 meV may be responsible for the fluctuations in conductance between different measurements. A calculation that is able to account for the environmental variations which shift the Fermi energy will be useful to yield further insight.

5.4.3.2 Phase-decoherent transport

In Chapter 4, we have suggested that decoherence due to an electron's interaction with the surrounding environment plays a large part in determining DNA conductance. Now we study the influence of decoherence on charge transport through $GC8$ and $GCm8$ to see if our conclusions in the above section change. We first set the decoherence rate to be 5 meV, a typical value in carbon nanotubes at room temperature [169, 170, 171] for both strands. The transmission and conductance for $GC8$ and $GCm8$ calculated with B3LYP after including decoherence with a rate of 5 meV is shown using blue solid and red dashed curves in Figure 5.8.

We find that for both $GC8$ and $GCm8$, after including decoherence, the peaks of the transmission become wider and the transmission values at the peaks are smaller compared to that in phase-coherent transport, due to the broadening of the localized energy levels, while away from the peaks the transmission values increase. We also note that decoherence can decrease the difference between $GC8$ and $GCm8$ for both transmission and conductance. This effect is more obvious in the HOMO-LUMO gap. For example, the transmission of $GCm8$ with 5 meV decoherence is only 1.2 times larger in the HOMO-LUMO gap, compared to 5 times in coherent transport. Similarly, the conductance of $GCm8$ with the inclusion of 5 meV decoherence is 1.5 times larger than $GC8$ near the HOMO.

As has been pointed out by several groups, the addition of the methyl groups in the cytosine bases can increase the molecular polarizability and help stabilize the DNA structure [141, 172, 173, 174, 175, 176, 177, 178, 179, 180]. Using DFT calculation with B3LYP/6-31G, we find that the hydrogen bonding for a single methylated $G:C$ base pair is 23 meV stronger than the unmethylated one, as shown in Table 5.2. This suggests that the methylated DNA has a more rigid structure and undergoes less fluctuation in the presence of water molecules,

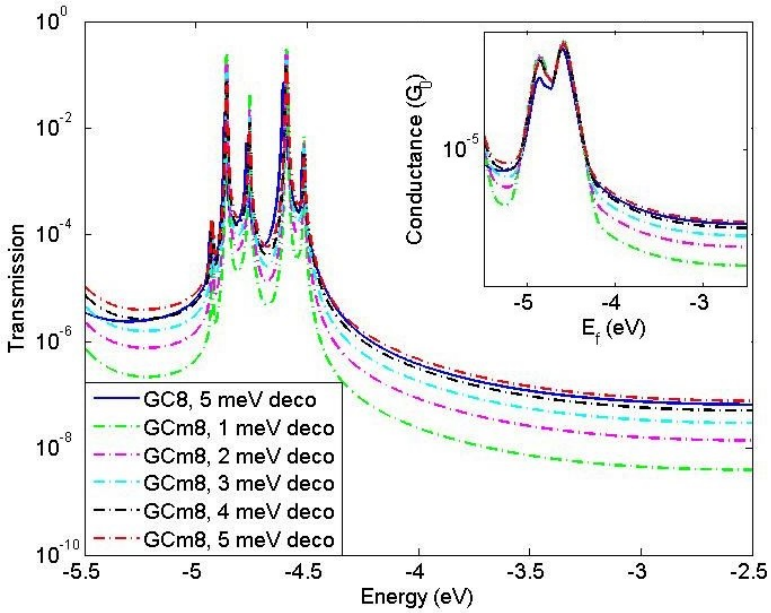


Figure 5.8: Transmission v.s. energy for $GC8$ and $GCm8$ calculated with B3LYP after including decoherence with a rate of 5 meV for $GC8$ and $1-5$ meV for $GCm8$. Inset, conductance v.s. Fermi energy for $GC8$ and $GCm8$. Blue solid, $GC8$ with a decoherence rate of 5 meV; green dashed, magenta dashed, cyan dashed, black dashed and red dashed are for $GCm8$ with a decoherence rate of 1 meV, 2 meV, 3 meV, 4 meV and 5 meV, respectively.

counter ions and other impurities, which may result in a smaller decoherence rate than the native DNA strand. To investigate the effect of the change in molecular stability, we study the transport as a function of the decoherence rate for $GCm8$. Since the precise value of the decoherence rate due to the structural alteration is hard to determine, we treat it as a variable. The transmission and conductance for $GCm8$ with decoherence rate ranging from 1 meV to 4 meV are shown in a series of dashed curves in Figure 5.8.

Interestingly, as the decoherence rate of $GCm8$ drops below 5 meV, the transmission of $GC8$ is larger than $GCm8$ by a few times, which is in qualitative agreement with the experimental observation that the conductance of $GC8$ is 1.2 times larger than $GCm8$ on average. Take the energy point at -4 eV for example. As the decoherence rate of $GCm8$ decreases from 5 meV to 4 meV, the transmission of $GC8$ is now 1.2 times larger than

$GCm8$. As the rate further decreases to 3 meV, 2 meV and 1 meV, the transmission is 2.1, 4.4, 15.7 times larger, respectively. For the conductance plots in the inset of Figure 5.8, we find a similar trend. With this analysis, we suggest that the enhanced stability of the molecular structure induced by the methyl groups may contribute to the lower conductance for $GCm8$ in the experimental measurement.

5.4.4 Transport study with CAM-B3LYP based range-separated functional

As DFT calculations are sensitive to the choice of the XC functionals, it is crucial to investigate this effect on transport. In addition to B3LYP, we consider the CAM-B3LYP [181] based range-separated functional. The choice of the latter is motivated by recent advances in range-separated DFT [182, 183, 184, 185, 186, 187], where the range-separation parameters are tuned to obey Koopmans' theorem, i.e. making the HOMO as close as possible to the IP of the neutral molecule.

Rather than using the default parameters for CAM-B3LYP in Gaussian 09, we tune the relevant parameters [187, 188] to satisfy Koopman's theorem, that is, to make the HOMO and IP of a neutral molecule to satisfy |HOMO + IP| = 0.

In a range-separated functional, the exchange term is described with two parts [187],

$$\frac{1}{r_{12}} = \frac{\alpha + \beta erf(\mu r_{12})}{r_{12}} + \frac{1 - [\alpha + \beta erf(\mu r_{12})]}{r_{12}} \tag{5.1}$$

The first term is the short-range interaction, treated with DFT exchange and the second term is the long-range interaction, treated with HF exchange. α and β are parameters which control the component of DFT and HF exchange, with $\alpha + \beta = 1$, $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ [189]. We can find the optimal values of α, β and μ by minimizing the following function for |HOMO + IP|,

$$\begin{aligned} J(\alpha, \mu) &= |IP_{SCF}(\alpha, \mu) - IP_{Koopmans}(\alpha, \mu)| \\ &= |E_{SCF}^{cation}(\alpha, \mu) - E_{SCF}^{neutral}(\alpha, \mu) + \epsilon_{HOMO}^{neutral}(\alpha, \mu)| \end{aligned} \tag{5.2}$$

By varying the values of α, β and μ, we find that when $\alpha = 0.2$, $\beta = 0.8$, $\mu = 0.15$, the |HOMO + IP| function in Equation (5.2) for $GC8$ can be minimized. This set of parameters is consistent with the choice of PBE-based range-separated hybrid density functional for 3,4,9,10-perylene-tetracarboxylic-dianydride (PTCDA) and 1,4,5,8-naphthalene-tetracarboxylic-dianhydride (NTCDA) molecules [186]. Since the structure of $GCm8$ differ from $GC8$ only by the methyl groups, we assume this set of parameters also works for $GCm8$. A comparison of the physical quantities, HOMO, IP and |IP + HOMO| for $GC8$ and $GCm8$ calculated with functionals B3LYP and the tuned CAM-B3LYP are shown in Table 5.3 and Table 5.4, respectively.

Table 5.3: Physical quantities computed with B3LYP/6-31G (Unit, eV).

Molecule	HOMO	IP	|IP + HOMO|
$GC8$	-4.519	5.366	0.847
$GCm8$	-4.512	5.364	0.852

Table 5.4: Physical quantities computed with the tuned CAM-B3LYP/6-31G (Unit, eV).

Molecule	HOMO	IP	|IP + HOMO|
$GC8$	-6.012	6.014	0.002
$GCm8$	-6.005	5.990	0.015

We can see that with both functionals, the HOMOs of $GC8$ and $GCm8$ are very close (only differ by 7 meV) due to the structural difference of the two strands. Compared to B3LYP, the effect of the tuned CAM-B3LYP is to shift the HOMO for both strands by around 1.5 eV. IP of $GC8$ is also similar to that of $GCm8$ with the two different functionals. As expected, |IP + HOMO| for $GC8$ improve significantly with the tuned CAM-B3LYP functional, which indicates the efficacy of the parameters, α, β and μ. It is also worth noting

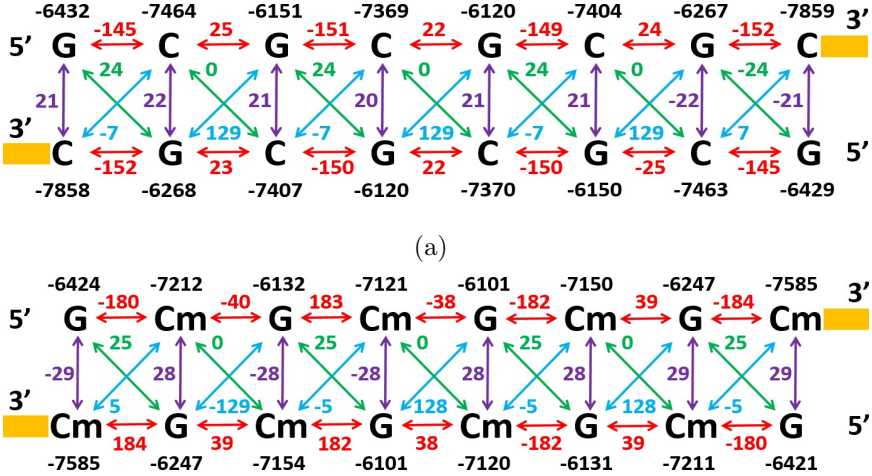


Figure 5.9: TB parameters of the HOMO level obtained from DFT calculation with the tuned CAM-B3LYP/6-31G for (a) $GC8$; and (b) $GCm8$ (Unit, meV). Arrows and numbers shown in various colors follow the representations in Figure 2.3(e).

that the |IP + HOMO| for $GCm8$ with the tuned CAM-B3LYP is only 15 meV, which is slightly smaller than the one calculated with B3LYP, suggesting that the range-separation parameters determined for $GC8$ also hold for a different strand with a similar structure ($GCm8$).

As a comparison to the results computed with B3LYP, we also show the TB parameters with the tuned CAM-B3LYP in Figure 5.9. Similar to the findings with B3LYP, one can see that with the tuned CAM-B3LYP, the main influence of cytosine methylation is to change the onsite energies of cytosine bases by about 200 meV and to modify the hopping integrals between the two nearest neighboring G and C (Cm) by approximately the same amount as that in B3LYP. Comparing the quantities computed by B3LYP with the corresponding ones computed by the tuned CAM-B3LYP, we find that the main difference between the two sets of the parameters lie in the onsite energies of the G and C (Cm) bases, shifting by around 1500 meV, while the hopping integrals are of the same order of magnitude, only deviating by a few milli-electron volts.

We also show the transmission and conductance for $GC8$ and $GCm8$ in coherent transport computed with the tuned CAM-B3LYP in Figure 5.10. The main observation in the results with B3LYP also holds here with CAM-B3LYP - the transmission of $GCm8$ is close to $GC8$ near the HOMO level and larger than $GC8$ away from the HOMO. Different from B3LYP, the conductance values of $GCm8$ is now 8 times larger than $GC8$ near the HOMO, compared to 2 times in the above discussion for B3LYP. We attribute this finding to the larger difference in IP between $GC8$ and $GCm8$ with the tuned CAM-B3LYP, as shown in Table 5.4.

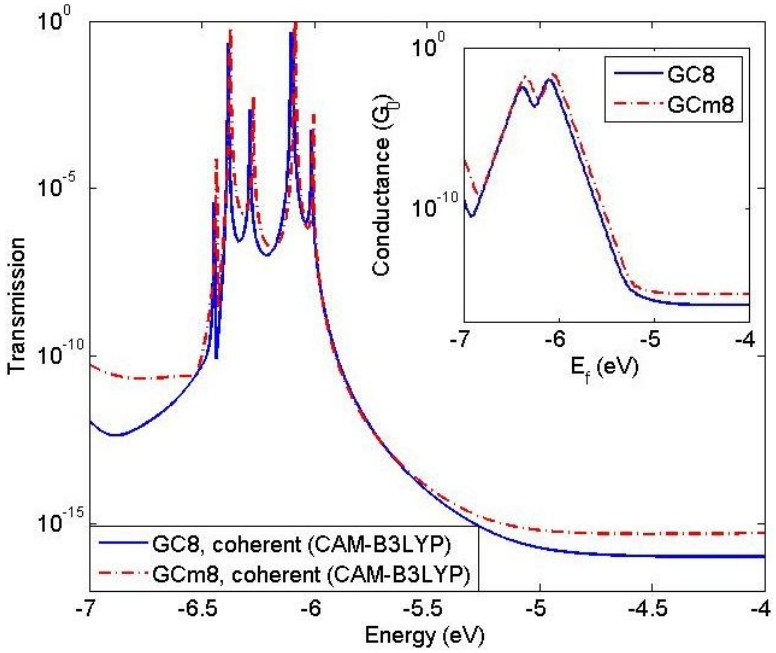


Figure 5.10: Transmission v.s. energy for $GC8$ (blue solid) and $GCm8$ (red dash) in phase-coherent transport calculated with the tuned CAM-B3LYP. Inset, conductance v.s. Fermi energy for $GC8$ (blue solid) and $GCm8$ (red dash).

Comparing the transmission and conductance in Figures 5.7 and 5.10, we conclude that the transport properties in the coherent case calculated by B3LYP and the tuned CAM-B3LYP mainly differ in the energy scale. This is in agreement with the observations in the TB parameters at the HOMO level shown in Figures 5.6 and 5.9 that the onsite energies of DNA bases with the two different functionals differ by as much as 1500 meV, while the

intra- and inter-strand hopping integrals are quite similar.

In Figure 5.11, we show the transmission and conductance for $GC8$ and $GCm8$ calculated by the tuned CAM-B3LYP with the inclusion of 5 meV decoherence for $GC8$ and $1-5$ meV for $GCm8$. We find that the transmission and conductance through $GC8$ and $GCm8$ with the tuned CAM-B3LYP is qualitatively similar to that with B3LYP, except for the shift in the energy scale. The observation of the lower conductance for $GCm8$ in the experiments may be due to the more stable molecular structure of $GCm8$ induced by the methyl groups also hold here. We point out that in this case the conductance of $GCm8$ with the decoherence rate of 5 meV is about 4 times larger than $GC8$ at the same energy point near the HOMO level, due to the fact that the IP of $GCm8$ is 24 meV smaller than $GC8$ with the tuned CAM-B3LYP.

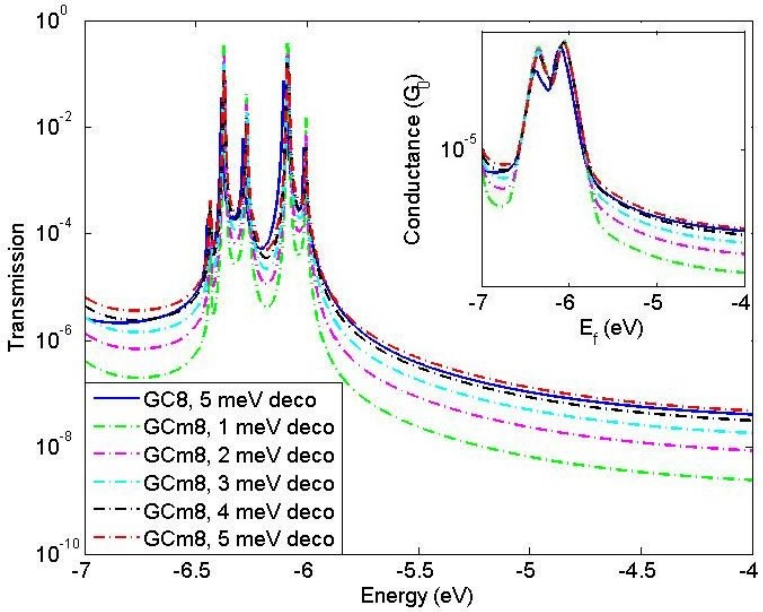


Figure 5.11: Transmission v.s. energy for $GC8$ and $GCm8$ calculated with the tuned CAM-B3LYP after including decoherence with a rate of 5 meV for $GC8$ and $1-5$ meV for $GCm8$. Inset, conductance v.s. Fermi energy for $GC8$ and $GCm8$. Blue solid, $GC8$ with a decoherence rate of 5 meV; green dashed, magenta dashed, cyan dashed, black dashed and red dashed are for $GCm8$ with a decoherence rate of 1 meV, 2 meV, 3 meV, 4 meV and 5 meV, respectively.

5.4.5 Effect of contact coupling

Based on the observations in Sections 5.4.3 and 5.4.4, we find similar transport results with B3LYP and the tuned CAM-B3LYP XC functionals in both phase-coherent and decoherent transport. In this section, we further study the effect of functionals on DNA charge transport with various coupling strengths to the contacts. The broadening of the molecular levels due to the coupling to contacts plays an important role in determining the electron flow through a molecule [61]. According to the TB parameters in Sections 5.4.2 and 5.4.4, the hopping integral value in a DNA molecule ranges from a few milli-electron volts to hundreds of milli-electron volts. A normal coupling strength due to the electrical contacts should be comparable to the hopping integral. Thus, we choose three values of coupling strengths, $\Gamma = 1\ meV$ (weak coupling limit), 100 meV (normal coupling) and 10 eV (strong coupling limit) and assume $\Gamma_L = \Gamma_R = \Gamma$ to study the transmission of $GC8$ and $GCm8$ with a decoherence rate of 5 meV, as shown in Figure 5.12.

To quantify our understanding, the energy scale for transmission computed with the tuned CAM-B3LYP has been shifted by $|\mathrm{HOMO_{CAM\text{-}B3LYP}} - \mathrm{HOMO_{B3LYP}}|$ in Figure 5.12. For $GC8$, as the coupling strength increases, the transmission increases drastically with both functionals. This effect is more evident when the coupling strengths are in weak regime. For example, as the coupling strengths increases from 1 meV to 100 meV, the transmission of $GC8$ is enhanced by as much as 25 times.

It is reasonable to assume that, in the weak molecule-electrode coupling limit, the misaligned molecular levels (for example, from a B3LYP calculation) with respect to the electrodes would result in an incorrect molecular response as charge is transferred from the electrodes into the molecule which should be observable. However, our transmission results with both B3LYP and the tuned CAM-B3LYP yield almost identical results in the vicinity of the HOMO irrespective of the strength of the contact coupling even though the delocalization and self-interaction errors are minimized in the tuned range-separated XC forms [183, 184, 185, 186, 187]. We find similar results for $GCm8$. The natural follow-on question

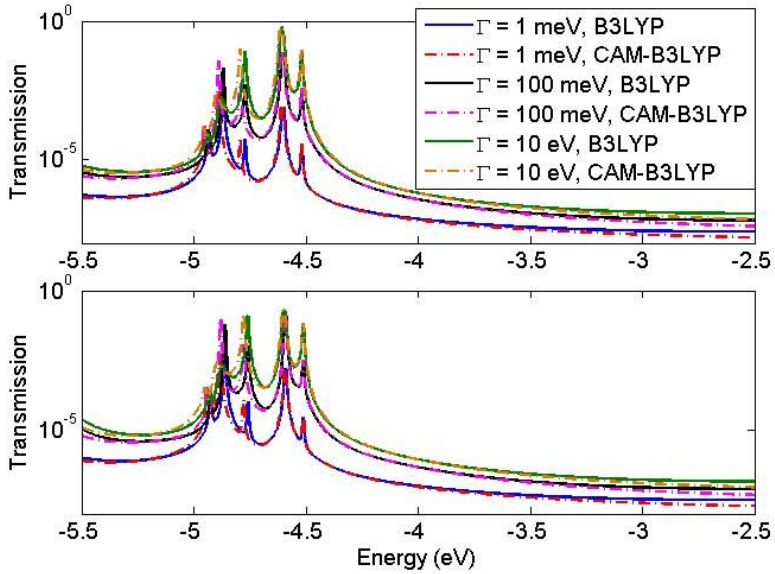


Figure 5.12: Transmission v.s. energy for $GC8$ (upper panel) and $GCm8$ (lower panel) calculated with B3LYP (solid curves) and the tuned CAM-B3LYP (dashed curves) functionals after including decoherence with a rate of 5 meV at $\Gamma = 1$ meV, $\Gamma = 100$ meV and $\Gamma = 10$ eV. The energy scale of transmission computed with CAM-B3LYP has been shifted by $|\mathrm{HOMO}_{\text{CAM-B3LYP}} - \mathrm{HOMO}_{\text{B3LYP}}|$.

is how large is this effect? and if it is observable in realistic systems. The other factor could be the inherent approximations within the Green's function-DFT approach itself. In order to quantify these effects, a systematic study of charge transport through well-characterized molecular junctions with different XC potentials and range-separation in the weak and strong coupling limits is needed.

5.5 Summary

In this chapter, we have studied the possibility of detecting a methylated DNA strand using its electronic properties from a theoretical perspective. We first focus on a single methylated cytosine base and base pair, and find that the methyl group can change the IP, hydrogen bonding and the HOMO distribution. We then explore the electronic structure and charge transport properties through an eight-base pair methylated DNA strand and its na-

tive counterpart used in the experiment of Reference [44]. We examine the TB parameters of the HOMO level for the two strands, and find that methylation has a large influence on the onsite energies of cytosine bases and hopping integrals between two nearest neighboring guanine and cytosine bases. The onsite energies of the methylated cytosine bases are approximately 200 meV higher than that of native cytosine bases. The intra-strand hopping integrals between two nearest neighboring guanine and cytosine bases from $5'$ to $3'$ ends for a methylated strand are about 30 meV stronger than that of native strand, and 15 meV stronger from $3'$ to $5'$ ends. We then carry out transport calculations in both phase-coherent and decoherent cases. Our calculations show that in phase-coherent transport, the transmission of electrons close to the HOMO band of the two strands are very similar. At energies away from HOMO, the transmission of $GCm8$ is larger than $GC8$. The linear response conductances of $GC8$ and $GCm8$ are also close to each other, with the conductance of $GCm8$ being slightly higher than $GC8$. The calculations also show that the conductance of these molecules can change by $5-10$ times if there is a small shift (50 meV) in the Fermi energy due to variations from experiment to experiment. In the phase-decoherent case, we find that the transmission and conductance trends occuring in the phase-coherent limit also hold in the HOMO vicinity if the decoherence rates for the two strands are the same. However, if the decoherence rate of the methylated strand is lower, we observe that the transmission and conductance of the methylated strand is also lower than the native strand, which is consistent with the experimental findings. Our results show that the effect of the two different functionals is to alter the onsite energies of the DNA bases at the HOMO level, while the transport properties do not depend much on the two functionals.

Chapter 6

CONFORMATIONAL GATING OF DNA CONDUCTANCE - ELECTRICAL TRANSPORT THROUGH A-FORM DNA

6.1 Introduction

In Chapters 3, 4 and 5, we have derived insight into electron flow in B-form DNA molecules by showing examples where engineered B-form DNA heterostructures behave as barriers and superlattices, developing a model including decoherence to explain the experimental data in four B-form DNA heterostructures, and examining the methylation effect on B-form DNA transport. We have seen that B-form DNA offers a unique framework for nanoscale electronic devices and detecting diseases by measuring electrical current in a single DNA molecule.

Besides B-form, DNA can also adopt other configurations in nature, such as A-form and Z-form. In Chapter 1, we have presented a description for the conformation of an A-form DNA. Here we briefly review the structural difference between a B-form DNA and an A-form DNA. A-form DNA is a more compact double-stranded helix and is also right-handed. In an A-form DNA, the distance between the two nearest stacking bases is 1 $\mathring{A}$ shorter than that in a B-form DNA (2.3 $\mathring{A}$ v.s. 3.4 $\mathring{A}$), and the twist angle is 33° instead of 36°. In addition, the inclination of base pair to axis in A-form is 19°, compared to $-1.2°$ in B-form. A-form DNA is found with low hydration conditions. Changes in the hydration level can induce conformational transitions between an A-form and B-form DNA [190]. This has been suggested as one of the possible reasons for the large variations in the conductance values in experimental measurements [6, 191, 192]. In biology, A-form DNA can serve as the prototypical structure for dsRNA and DNA:RNA hybrids, and is also very important in many biological processes, such as transcription [193].

During the last few decades, the charge transport through B-form DNA has been well

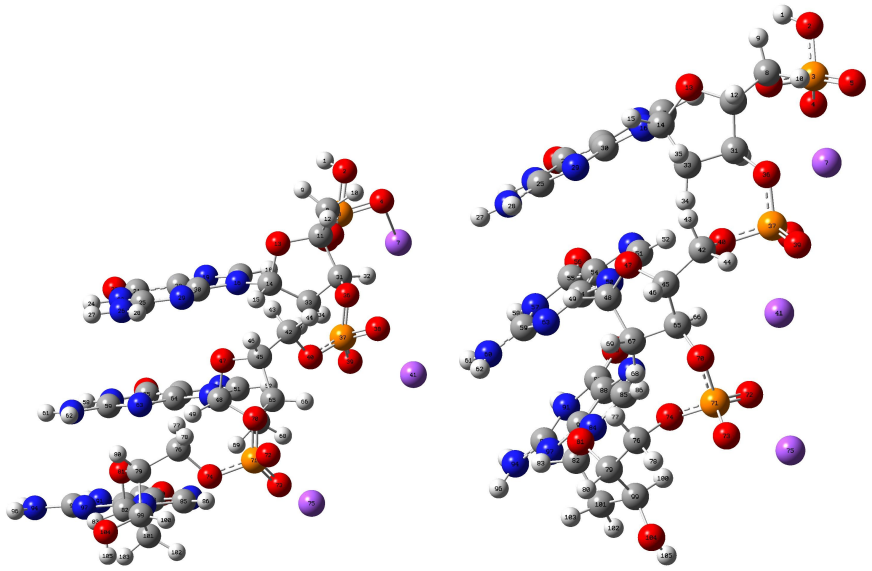
characterized from both experimental and theoretical perspectives. It has reached consensus that the electrical conduction in B-form DNA is caused by the π - π interaction between the neighboring bases, which is in turn determined by the distance and twist angles between the two bases [27]. In contrast to B-form DNA, the electrical conduction in A-form DNA is unclear. Studies suggest that A-form DNA has less effective orbital overlap due to the configuration of base stacking, and hence lower conductivity [6]. Besides this widely accepted framework, very little effort is devoted to the understanding of the charge migration mechanism in A-form DNA.

In this chapter, we report that the conductance of DNA can be changed greatly when the conformation switches from B-form to A-form. Our modeling shows counterintuitive results based on two sets of DNA strands with B-form and A-form configurations - A-form DNA can be more conducting than the B-form counterpart. Our results are consistent with the observations in the experimental measurements in our collaborators' group at University of California, Davis. We rationalize the role of backbone in the electrical transport through B-form and A-form DNA. The analysis of the transmission calculation, the HOMO pattern and the distributions of DOS for strands in B-form and A-form configurations all point to the fact that to get qualitatively correct information, including backbone is essential in the conductance modeling through an A-form DNA. We also study the effect of decoherence and find it is important in bringing the computed conductance values close to the experiments. The fact that we make similar conclusions with two different XC functionals and basis sets lends further support to our findings.

6.2 Method

We consider two sets of double-stranded structures for both B-form and A-form DNA, the sequences of which are: $3' - CCCGCGCCC - 5'$ $(B3/A3)$; $3' - CCCGCGCGCCC - 5'$ $(B5/A5)$; $3' - CCCGCGCGCGCCC - 5'$ $(B7/A7)$; $3' - CCCGCGCGCGCGCCC - 5'$ $(B9/A9)$. The ideal B-form and A-form structures are generated with the NAB software package. To identify the role of backbone in the electrical transport through B-form and

A-form DNA, we study the strands both with backbone and without backbone. We use sodium counterions to neutralize the negatively charged phosphate groups for the strands with backbone retained. The positions of the sodium atoms are determined by relaxing the sodium atoms in a single strand with three bases at the B3LYP/6-31G(d) level. The relaxation results for a B-form and A-form DNA are shown in Figure 6.1.

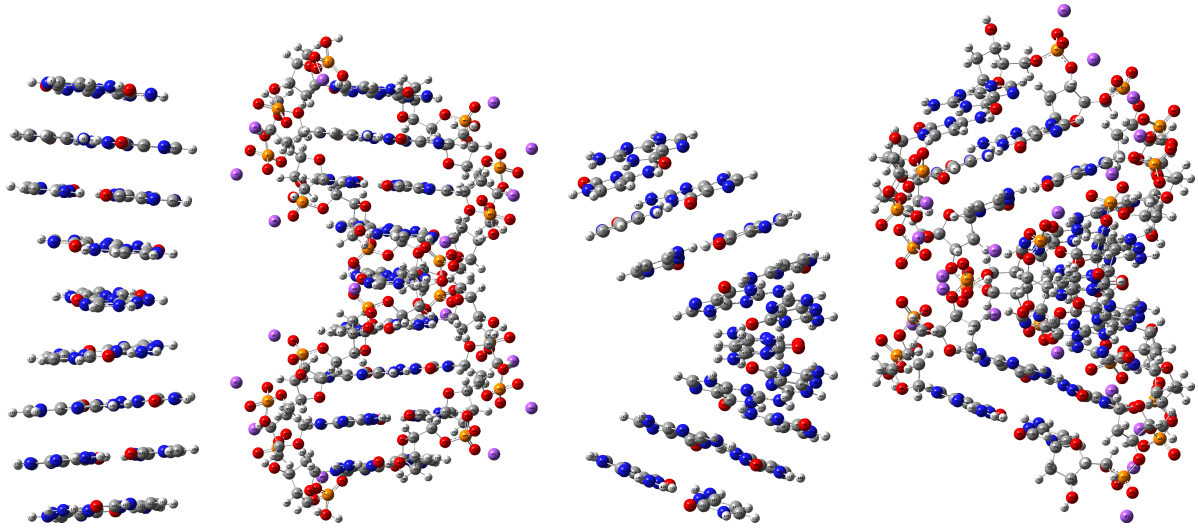


(a) (b)

Figure 6.1: Optimized positions of the sodium atoms in (a) B-form; and (b) A-form DNA. In both plots, sodium atoms are represented by purple balls. The sodium counterion stays close to the phosphate group for the B-form DNA. However, for an A-form DNA, it moves in between the two neighboring bases.

As shown in Figure 6.1(a), the sodium counterions placed around the B-form DNA stay close to the phosphate groups. In contrast, the sodium atoms placed around the A-form DNA move in between the two neighboring bases (Figure 6.1(b)). This interesting effect is due to structural difference between a B-form and an A-form DNA.

For the strands with backbone deleted, we simply terminate the bases using the hydrogen atoms. The side views of a 9-mer DNA with backbone included and deleted in both B-form and A-form configurations are shown in Figure 6.2.



(a) (b) (c) (d)

Figure 6.2: Side views of a 9-mer DNA with backbone deleted and included in both B-form and A-form conformations. (a) $B3$ without backbone; (b) $B3$ with backbone; (c) $A3$ without backbone; and (d) $A3$ with backbone.

In the next step, we use Gaussian 09 to carry out the DFT calculations. Two sets of XC functionals and basis sets, B3LYP/6-31G(d, p) and HF/6-31G, are used. Despite the difference, the results with these two sets of XC functionals and basis sets show similar results. We then calculate the transport using methods described in Chapter 2.

To obtain the weights of contributions to the HOMO orbital from each component of the DNA, we further compute the coefficients for the projected HOMO using the following method:

When solving the Schrödinger's equation, the ith molecular orbitals Ψ_i are expanded using a set of basis functions $\{\phi_\alpha\}$ [93]:

$$\Psi_i = \sum_{\alpha=1}^{K} c_\alpha^i \phi_\alpha \tag{6.1}$$

where K is the number of basis functions.

The eigen value equation that is solved is,

$$FC = SC\varepsilon \tag{6.2}$$

where F is the Fock matrix, $F_{\alpha\beta} = \langle\phi_\alpha|F|\phi_\beta\rangle$, and the S matrix represents the overlap between basis functions, $S_{\alpha\beta} = \langle\phi_\alpha|\phi_\beta\rangle$. C is a matrix of the expansion coefficient C^i_α. The ith molecular orbital Ψ_i is described by the ith column of C. Because the molecular orbitals $\{\Psi_i\}$ are orthonormal, $C^\dagger SC = 1$.

To identify the contribution from all the orbitals in a specific part within a molecule, we divide the system into N fragments (each fragment is represented by μ),

$$\Psi_i = \sum_{\mu=1}^{N}\sum_{\alpha\in\mu} c^i_\alpha \phi_\alpha \tag{6.3}$$

Correspondingly, the Fock matrix F, overlap matrix S, and the molecular orbital coefficient matrix C can also be transformed into the fragmental format. To find out the orbital contribution from a particular fragment at the ith energy level, we expand the following expression,

$$\begin{aligned} C^{i\dagger}SC^i &= \sum_{\mu=1}^{N}\sum_{p\in\mu} C_p^{i\dagger} S_{pp} C_p^i + \sum_{\mu\neq\nu}^{N}\sum_{p\in\mu;q\in\nu} C_p^{i\dagger} S_{pq} C_q^i \\ &= \sum_{\mu=1}^{N} f_\mu + \sum_{\mu\neq\nu}^{N} o_{\mu\nu} \end{aligned} \tag{6.4}$$

where f_μ is the contribution from the μth fragment with $f_\mu = \sum_{p\in\mu} C_p^{i\dagger} S_{pp} C_p^i$, and $o_{\mu\nu}$ is the overlap between the μth and νth fragments, $o_{\mu\nu} = \sum_{p\in\mu;q\in\ nu} C_p^{i\dagger} S_{pq} C_q^i$.

6.3 Results

In this section, we show the results from our calculations. We first present the results calculated with B3LYP/6-31G(d, p) and then show the results with HF/6-31G.

We are interested in the transport in the vicinity of the HOMO level. The HOMO energies for $B3-B9$ and $A3-A9$ with backbone deleted and included are summarized in Table 6.1

and 6.2, respectively.

Table 6.1: HOMO for Bn and An without backbone calculated with B3LYP/6-31G(d, p) (Unit, eV)

Strands	$B3$	$B5$	$B7$	$B9$
HOMO	-4.073	-4.049	-4.040	-4.036
Strands	$A3$	$A5$	$A7$	$A9$
HOMO	-3.927	-3.884	-3.882	-3.887

Table 6.2: HOMO for Bn and An with backbone calculated with B3LYP/6-31G(d, p) (Unit, eV)

Strands	$B3$	$B5$	$B7$	$B9$
HOMO	-3.079	-3.043	-3.028	-3.023
Strands	$A3$	$A5$	$A7$	$A9$
HOMO	-3.889	-3.948	-3.975	-3.987

The inclusion of the backbone can shift the HOMO by approximately 1 eV for B-form DNA. The difference in HOMO for strands with and without backbone are much smaller for A-form DNA.

6.3.1 Role of backbone on charge transport through A-form DNA

It is well-known that in a B-form DNA, the electrical conduction occurs through the orbital overlaps between the neighboring stacking bases, and the presence of the backbone in the B-form DNA is mainly to stabilize the structure. The role of the backbone in charge transport through an A-form DNA is rarely studied and uncertain. To this end, we first compare the role of the backbone in charge transport through a B-form DNA with that of an A-form

DNA. In Figure 6.3, we present the transmission for both B-form and A-form DNA with and without backbone.

We take the transmission of $B3$ and $A3$ as an example to analyze the role of backbone in the electrical transport through a B-form and an A-form DNA.The transmission of $B3$ without backbone (blue solid) is larger than $A3$ without backbone (red dashed) in a large range of energies - near the HOMO, it can be 10^{12} times larger and in the HOMO-LUMO gap, it can be 100 times larger. This suggests that the intrinsic arrangement of the B-form stacking bases is more effective than the A-form in electrical conduction. This finding is consistent with the fact that although the A-form DNA has a lower rise between the two nearest stacking bases, the orbital overlaps are poorer, because of the worse alignment between the bases. With the backbone included, the transmission of $B3$ (black solid) around the HOMO level stays similar in amplitude to the case with the backbone removed. The 1 eV shift in the energy axis agrees with our previous findings as discussed in Figure 4.3 of Chapter 4. In the HOMO-LUMO gap, the transmission of $B3$ with backbone is 10 times larger than $B3$ without backbone. Surprisingly, the backbone has a much more important impact on the conduction of A-form DNA. The transmission for $A3$ with backbone (magenta dashed) can be 15 orders of magnitude larger than the one without backbone in some energy windows. More importantly, after taking into account the backbone in the transport, the transmission of $A3$ is now larger than $B3$ in a large energy range. This is in agreement with the experimental observations. In the HOMO-LUMO gap, the transmission difference between $A3$ with backbone and $B3$ with backbone can be as much as 10^7 times. This observation suggests that backbone in A-form DNA plays a significant role in determining the electrical conduction. Our findings in $B3$ and $A3$ also hold in $B5$ and $A5$ (Figure 6.3(b)), $B7$ and $A7$ (Figure 6.3(c)), $B9$ and $A9$ (Figure 6.3(d)). With these observations, we are of the opinion that A-form DNA has been predicted to be less conductive in the previous reports [6] because of the omission of the backbone. Including backbone in the calculations induces a change in the conductance trend between a B-form and an A-form DNA with the same sequence.

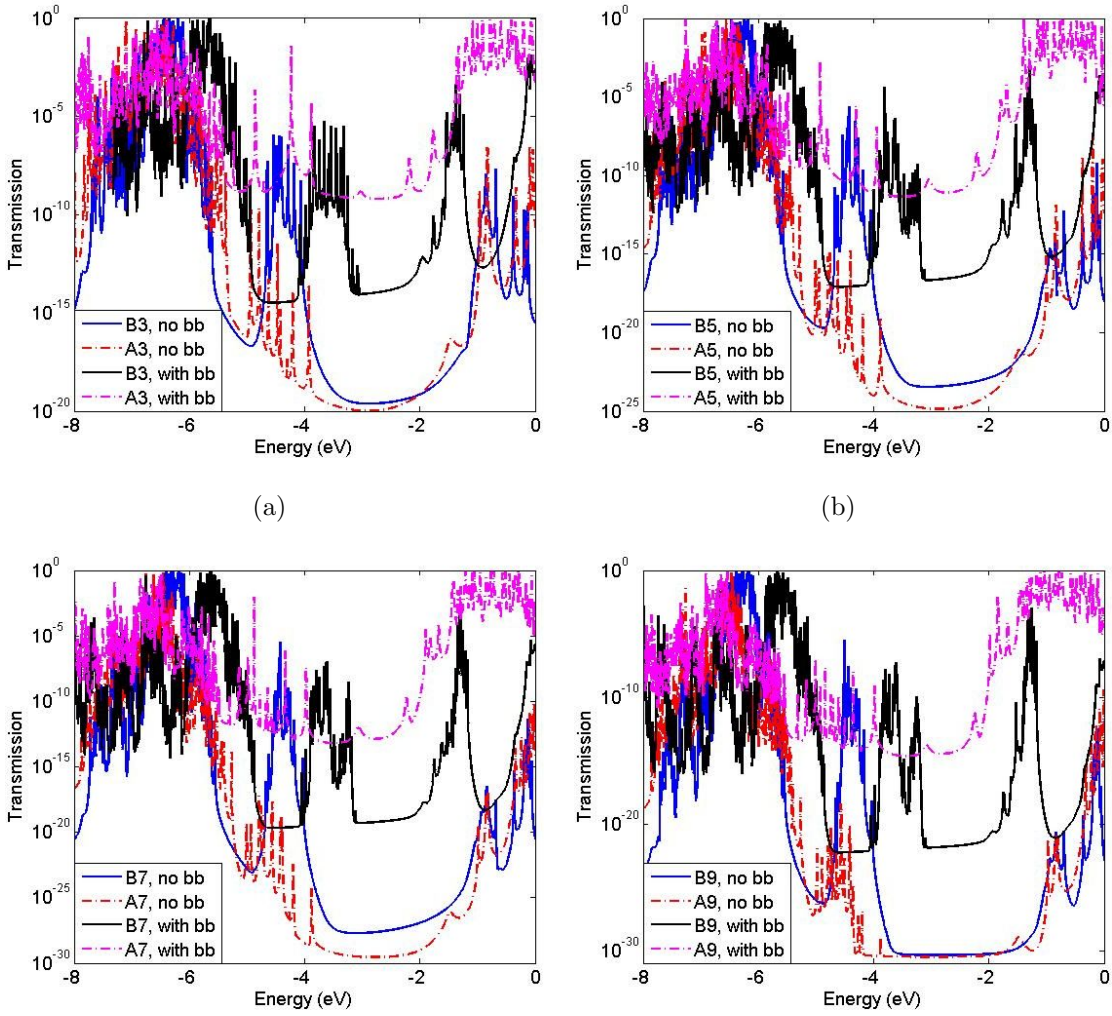
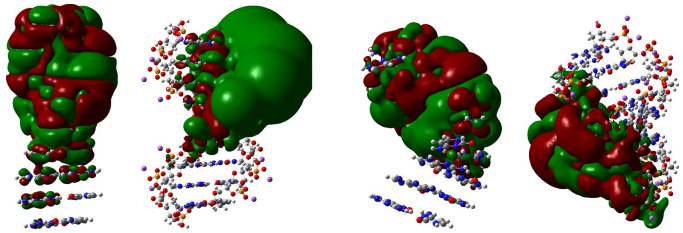


Figure 6.3: Transmission for (a) $B3$ and $A3$; (b) $B5$ and $A5$; (c) $B7$ and $A7$; and (d) $B9$ and $A9$ in the coherent transport limit calculated with B3LYP/6-31G(d, p). Both cases, where the backbone is deleted and included, are considered. Blue solid, B-form without backbone; red dashed, A-form without backbone; black solid, B-form with backbone; magenta dashed, A-form with backbone.

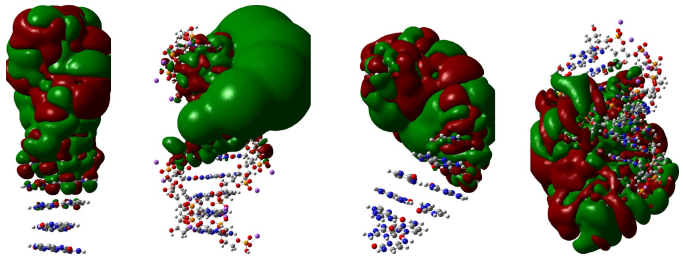
To provide additional insights to the transmission pattern, we further look at the HOMO iso-surface plots of the B-form and A-form DNA for both the cases with and without backbone. The plots are shown in Figure 6.4.

In Figure 6.4, it is clear that the remarkable change in the transmission of an A-form DNA upon the inclusion of the backbone is due to the significant modification in the charge density distribution. Take the HOMO distributions of $B3$ and $A3$ (Figure 6.4(a)) for example. The HOMO distribution of $B3$ with backbone follows a similar pattern to that of $B3$ without backbone. In both cases, the orbital mainly localizes on base pairs 1 to 6, with a little portion at base pairs 8 and 9. In contrast, the HOMO distribution of $A3$ with backbone is distinct from that of $A3$ without backbone. The HOMO orbital of $A3$ without backbone mainly resides on the first seven base pairs, following a similar pattern to that of $B3$ without backbone. However, with the introduction of the backbone, the main portion of the HOMO orbital for $A3$ shifts its location to the last eight base pairs. More interestingly, the overall HOMO orbital of $A3$ with backbone spreads out across the entire strand, making the charge transport through the strand more efficient. A similar trend is not observed in $B3$ with backbone, where the HOMO orbital is much less localized. We believe that this is the reason why $A3$ with backbone has a much larger transmission than $A3$ without backbone, and the similar trend is not observed in the transmission $B3$. A comparison between the HOMO orbitals of $B3$ and $A3$ also casts light on the transmission trend between $B3$ and $A3$ in Figure 6.3(a). In Figure 6.4(a), it is obvious that the HOMO orbital of $B3$ without backbone is more diffusive than $A3$ without backbone, resulting in a larger transmission. However, after the backbone is included, the HOMO distribution of $B3$ is less diffusive than $A3$ with backbone, which leads to a smaller transmission.

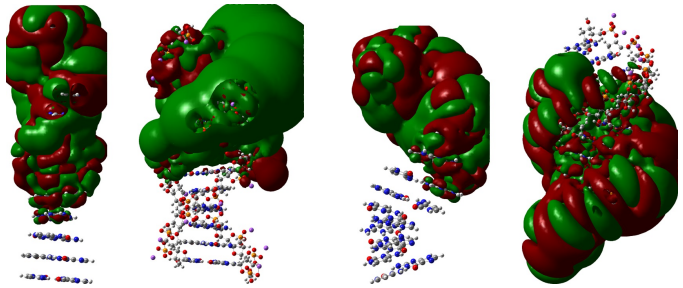
In order to obtain a more quantitative understanding of the role that the backbone plays in the electrical conduction through B-form and A-form DNA, we calculate the projection of the HOMO level onto the bases and backbone along each strand both with and without the backbone using Equation 6.4. The results are shown in Figure 6.5. These calculations show that the inclusion of the backbone has a dramatic impact on both the location and distribu-



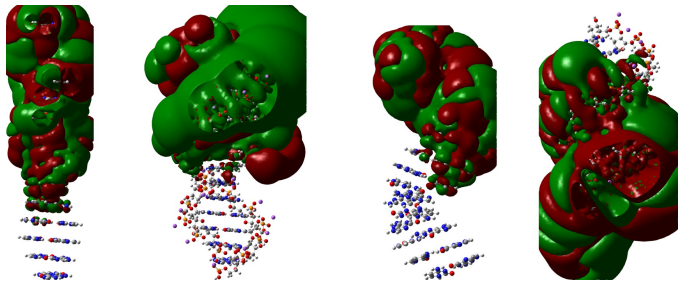
(a)



(b)



(c)



(d)

Figure 6.4: 3D iso-surface of the HOMO orbital for B-form and A-form DNA with and without the backbone calculated with B3LYP/6-31G(d, p). (a) $A3$ and $B3$, with iso-value $= 10^{-5}$; (b) $A5$ and $B5$, with iso-value $= 10^{-6}$; (c) $A7$ and $B7$, with iso-value $= 10^{-7}$; (d) $A9$ and $B9$, with iso-value $= 10^{-8}$. In each sub-figure, the HOMO orbitals are presented from left to right in the following order: B-form without backbone; B-form with backbone; A-form without backbone and A-form with backbone.

tion of the HOMO orbital in the A-form DNA, agreeing with the iso-surface plots in Figure 6.4. In the absence of backbone, the HOMO orbital of an A-form DNA is centered around the third base. When the backbone is included, it shifts to the other end of the molecule. Contrarily, the projected HOMO for B-form DNA with backbone stays similar to that without backbone. Thus, the backbone has a substantial impact on the energy-level distribution along the A-form DNA and an indirect (but significant) impact on the charge transport in these strands. As such, in order to obtain even a qualitative agreement between theory and experiments, one needs to include the backbone in the electronic/transport calculations through an A-form DNA.

We have also examined the orbital distribution of the HOMO $-$ 1 energy level, as shown in Figure 6.6. Our observation in the HOMO level also holds here. The contribution from each base pair of $B3$ without backbone is more uniform than that of $A3$ without backbone. The trend is reversed after including backbone. The finding at the HOMO $-$ 1 energy level lends more support to the important role of backbone in electrical conduction of the A-form DNA.

We further look at a more fundamental parameter, the DOS, to compare the electronic properties of a B-form DNA with that of an A-form DNA. By studying the DOS across the DNA molecule as a function of energy, we can investigate the transmission of the two conformations independently from any particular transport mechanism. Figure 6.7 shows the ratio of the DOS of the A-form DNA to the B-form DNA for each base pair through the entire sequence of the 9-mer (Figure 6.7(a)), 11-mer (Figure 6.7(b)), 13-mer (Figure 6.7(c)) and 15-mer (Figure 6.7(d)), for 1 eV energy range in the HOMO-LUMO gap starting from the HOMO level. The DOS of the base pairs at the two ends of a B-form DNA is similar to that of an A-form DNA. Depending on the energy we consider, sometimes the DOS of the B-form DNA is higher by a few times, and sometimes that of the A-form DNA is higher. However, in the central region with the alternating GC sequence, the difference between the DOS of a B-form DNA and an A-form DNA is remarkable. Here, the A-form DNA almost always has a larger DOS than the B-form DNA by as much as 4 orders of magnitude

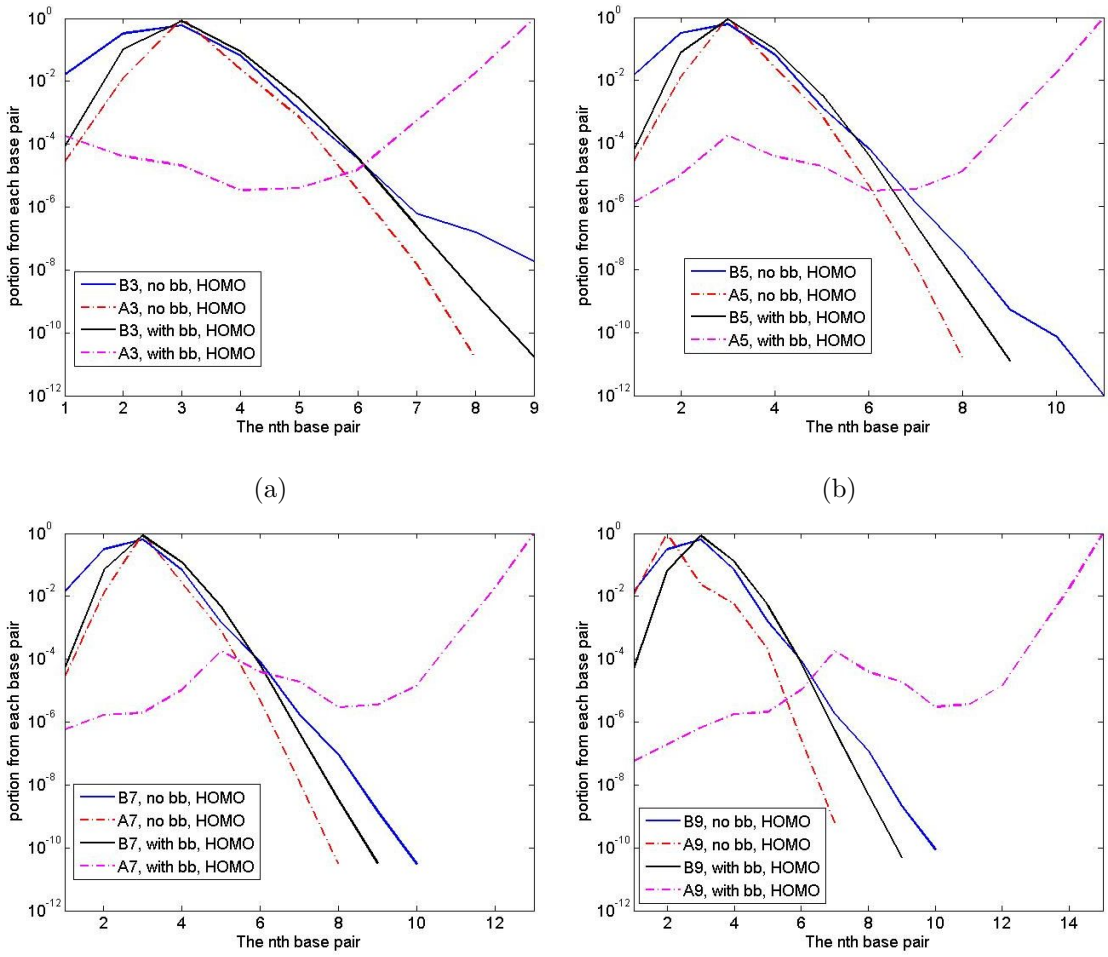
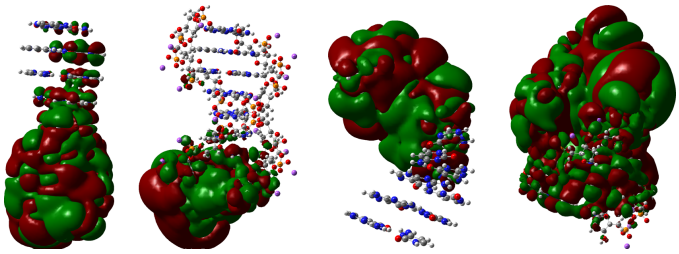
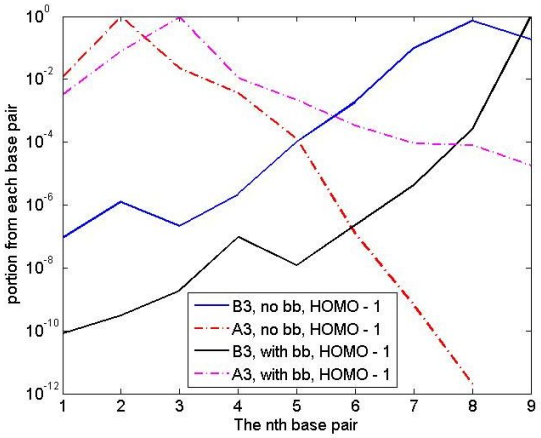


Figure 6.5: Projected HOMO of each base pair in the B-form and A-form DNA with and without backbone. Without backbone, the contribution from each base pair of B-form is more uniform than that of A-form DNA. After including backbone, the trend is reversed. Note that in these logarithm plots, the projected HOMOs do not have data for a few base pairs in some strands, meaning that the contributions from these base pairs are zero. The y-axis of these plots correspond to the quantity in Equation 6.4, where each fragment corresponds to a base pair without backbone (for the strands without backbone)/a base pair with backbone (for the strands with backbone).



(a)



(b)

Figure 6.6: (a) 3D iso-surface of the HOMO $-$ 1 orbital for $B3$ and $A3$ with and without the backbone calculated with B3LYP/6-31G(d, p), with iso-value $= 10^{-5}$. From left to right: $B3$ without backbone; $B3$ with backbone; $A3$ without backbone and $A3$ with backbone; (b) projected HOMO $-$ 1 of each base pair in the $B3$ and $A3$ with and without backbone. Without backbone, the contribution from each base pair of $B3$ is more uniform than that of $A3$. After including backbone, the trend is reversed. Note that in $A3$ without backbone, the contribution of the 9th base pair is zero.

times. For longer strands (11-mer, 13-mer and 15-mer), this trend is more evident. The DOS analysis provides more support that an A-form DNA should transport charge more effectively than the B-form with the same sequence, regardless of the specific mechanism.

6.3.2 Role of decoherence on charge transport through A-form DNA

In the previous chapters (Chapters 4 and 5), we have seen that decoherence is very important in the electrical conduction through a B-form DNA. In this section, we study the role that the decoherence plays in the electrical conduction through an A-form DNA. We first show the conductance of the B-form and A-form DNA with the backbone retained with the fully coherent transport model we used in Section 6.3.1. The results are presented in Figure 6.8. Except $B3$, the conductance of the strands have been shifted by $|IP - IP_{B3}|$ in the x-axis to account for the IP difference. We find that even taking into account the backbone, the conductance values of these strands are still several orders of magnitude below the experimental observations. The conductance values from the coherent model are only around $10^{-10}\ G_0$, while the ones in the measurements are about $10^{-4}\ G_0$. The low conductance from the fully-coherent model is consistent with the low β-values obtained in the experiments, which suggest that hopping, or partially decoherent processes dominates the transport.

To resolve the disagreement between the theory and the experiments, one may take into account the soft mattered nature of the DNA. Due to the flexibility of the DNA molecules, highly dynamic and structural fluctuations have been suggested to have an important influence on charge transport through DNA [192, 194, 195]. To include the effects of these factors phenomenologically, we again use the decoherent model by introducing a decoherence term to the Green's function. While the exact decoherence rate is difficult to determine in nanoscale systems, this model demonstrates a trend to provide explanations fro the experiments, as seen in Chapter 4. In this chapter, we set the decoherence rate to be 10 meV in the calculations. This value is a little larger than the ones used for the dry DNA experiment in Chapter 4, with the fact that the current experiment is done in solution, where more fluctuations are expected. The conductance of the B-form and A-form DNA with backbone

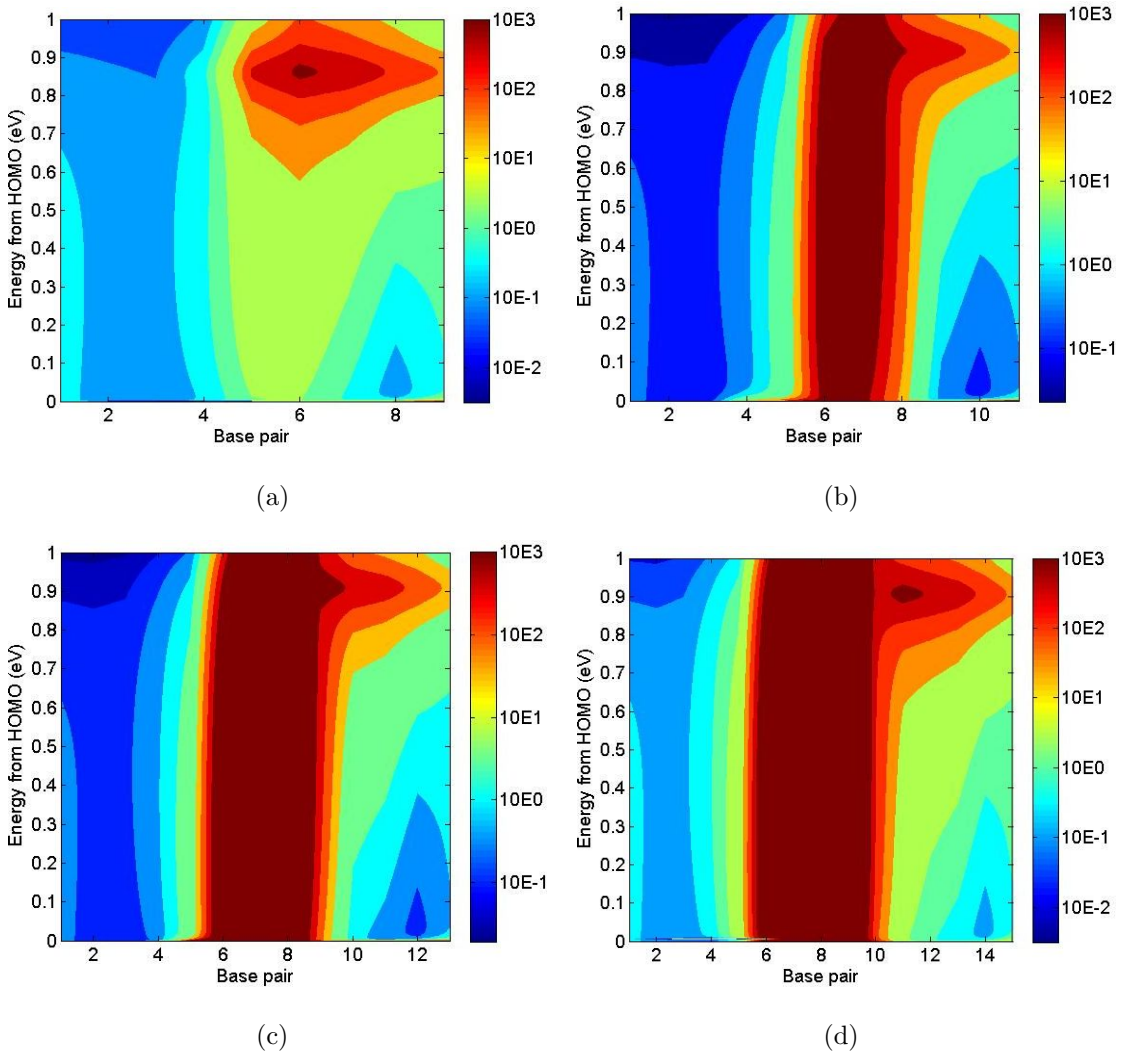


Figure 6.7: 2D representation for the DOS ratio of the A-form DNA to the B-form DNA for each base pair along the molecules for 1 eV energy range in the HOMO-LUMO gap starting from the HOMO level. (a) $B3$ and $A3$; (b) $B5$ and $A5$; (c) $B7$ and $A7$; (d) $B9$ and $A9$. In the central region of the molecule, DOS for the A-form DNA is several orders of magnitude larger than that of the B-form DNA. This observation holds true for all the sequences.

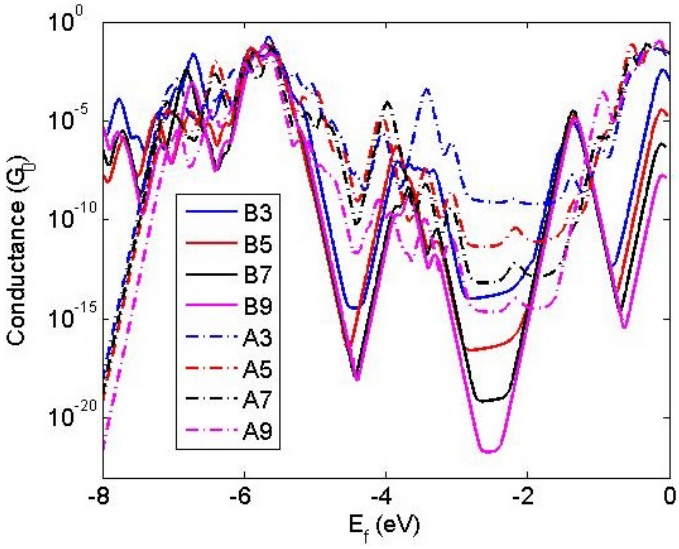


Figure 6.8: Conductance of B-form and A-form DNA with backbone in the coherent transport limit. Blue solid, $B3$ with backbone; red solid, $B5$ with backbone; black solid, $B7$ with backbone; magenta solid, $B9$ with backbone; Blue dashed, $A3$ with backbone; red dashed, $A5$ with backbone; black dashed, $A7$ with backbone; magenta dashed, $A9$ with backbone.

retained in the decoherent model is shown in Figure 6.9.

We again see that the decoherence term causes the conductance value of the B-form DNA to increase by several orders of magnitude, in agreement with the previous studies. We also find that the decoherence term has an important effect on the transport properties of the A-form DNA, by increasing the conductance values to the ones much closer to the experimental measurements. The conductance values of both B-form and A-form have been enhanced to 10^{-5} G_0 in the vicinity of the HOMO after the decoherence term with a rate of 10 meV is introduced. Note that even though the conductance values of the B-form and A-form DNA are similar once the decoherence is included, the A-form DNA is still more conducting than the B-form DNA over a large range of energies.

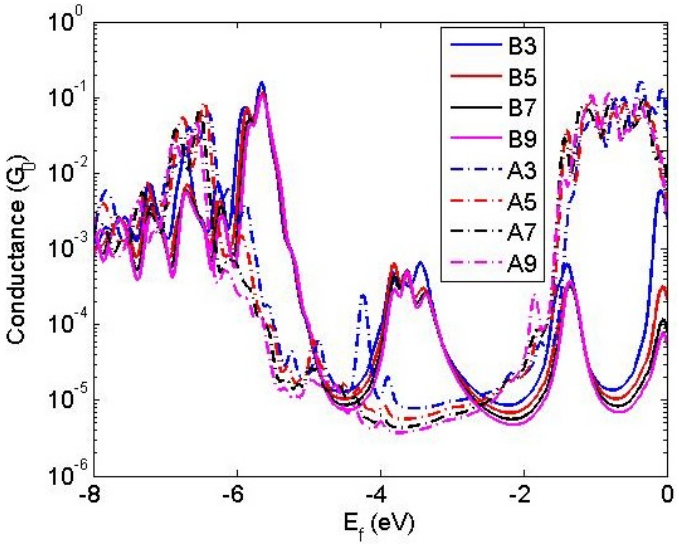


Figure 6.9: Conductance of B-form and A-form DNA with backbone retained in the decoherent transport model where 10 meV decoherence rate is considered. Blue solid, $B3$ with backbone; red solid, $B5$ with backbone; black solid, $B7$ with backbone; magenta solid, $B9$ with backbone; Blue dashed, $A3$ with backbone; red dashed, $A5$ with backbone; black dashed, $A7$ with backbone; magenta dashed, $A9$ with backbone.

6.3.3 Tight-binding model for B- and A-form DNA

In this section, we examine the inter-base pair electronic couplings in both B-form and A-form DNA, and consider the cases where the backbone is deleted and included. The motivation comes in two folds: (1) In Sections 6.3.1 and 6.3.2, we have seen that the structural difference between a B-form DNA and an A-form DNA results in different transmission patterns. To reveal the physics underneath this, it is helpful to examine the electronic coupling between the neighboring bases in the two conformations. (2) The size of the Hamiltonian matrix in the transport calculations depends on the size of the molecule and the choice of the basis set in the DFT calculations. The computational cost may be very high for a large molecule or when the DFT calculations involve a large basis set. In Figure 6.10, we present the size of Hamiltonian matrix for a set of DNA ($B3 - B9$) as a function of the number of base pairs. Two different basis sets, 6-31G and 6-31G(d, p), are considered. It clearly shows that

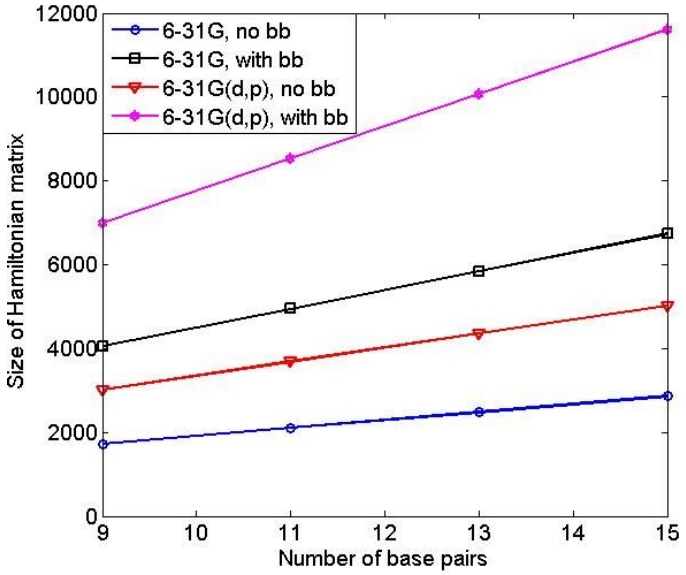
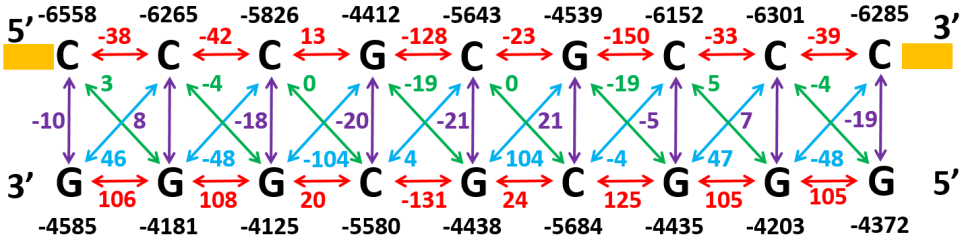


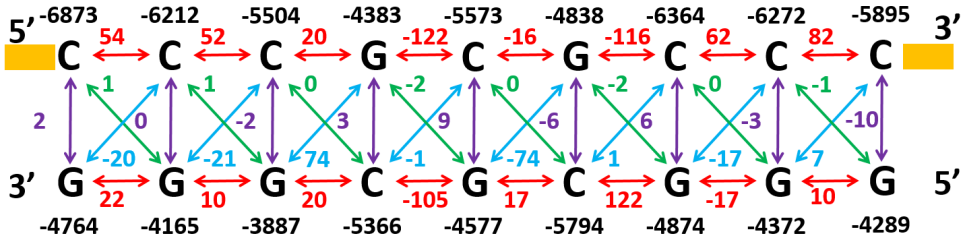
Figure 6.10: Size of the Hamiltonian matrix v.s. the number of base pairs for a set of DNA ($B3 - B9$) with two different basis sets. Blue dot, 6-31 G for the DNA without backbone; Black square, 6-31 G for the DNA with backbone; Red triangle, 6-31 G(d, p) for the DNA without backbone; magenta star, 6-31G(d, p) for the DNA with backbone. The size of Hamiltonian matrix grows linearly with the number of base pairs in the DNA strand.

the size of Hamiltonian matrix grows linearly with the number of base pairs in the DNA. Also, the polarization in the basis set affects the number of orbitals significantly. For a 15-mer DNA when the backbone is included, the size of the Hamiltonian can be as large as $12,000 \times 12,000$. Inverting a matrix is a very expensive computation, as the time complexity is typically $O(N^3)$ (N is the size of the matrix). Therefore, the transport calculation based on a full Hamiltonian for a large DNA molecule, say, a molecule that consists of hundreds of base pairs, is extremely difficult. In this case, a model Hamiltonian, which involves only a few energy levels to reduce the computational cost while still capturing the essential physics, is certainly useful.

Compared to a B-form DNA, an A-form DNA has a lower rise (2.3 $\mathring{A}$ v.s. 3.38 $\mathring{A}$) and a smaller angle (3.38° v.s. 3.6°). In addition to that, the two bases within a base pair in an A-form DNA do not lie in the same plane, making the orbital overlaps more complicated. In



(a)



(b)

Figure 6.11: TB parameters at the HOMO level obtained from DFT calculation with B3LYP/6-31G(d, p) for (a) $B3$ without backbone; and (b) $A3$ without backbone (Unit, meV). Numbers shown in black are the onsite energies of each base. Arrows and numbers shown in various colors follow the representations in Figure 2.3(e).

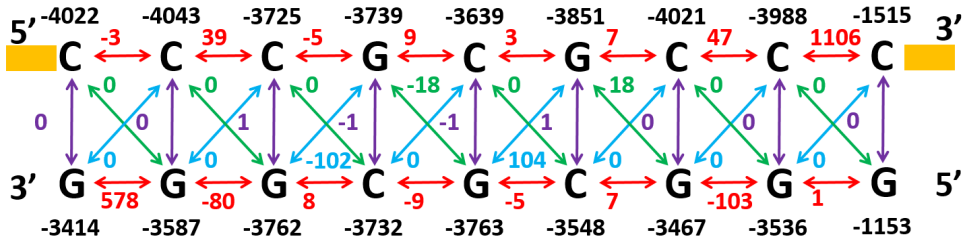
Figure 6.11, we show the inter- and intra- strand electronic couplings at the HOMO energy level for the ideal B-form and A-form 9-mer without backbone. Here, we find that both the intra- and inter- strand electronic couplings in the G-hopping paths in $A3$ are smaller than the corresponding ones in $B3$. For example, the intra-strand coupling between the two nearest neighboring Gs from $3'$ to $5'$ end is only $10 - 22$ meV in $A3$, compared to about 105 meV in $B3$. The inter-strand couplings between a $G : C$ base pair in $A3$ are also smaller than the ones in the corresponding positions in $B3$. The poorer electronic couplings in $A3$ are due to the less effective arrangement of the bases. Therefore, we conclude that the low rise in A-form DNA is not sufficient to explain the higher conductance observed in the experiment. This actually agrees with the transmission shown in Figure 6.3(a).

As the lower rise and smaller twist angles between the two neighboring stacking bases fail to explain the higher conductance in A-form DNA, we include the backbone to further

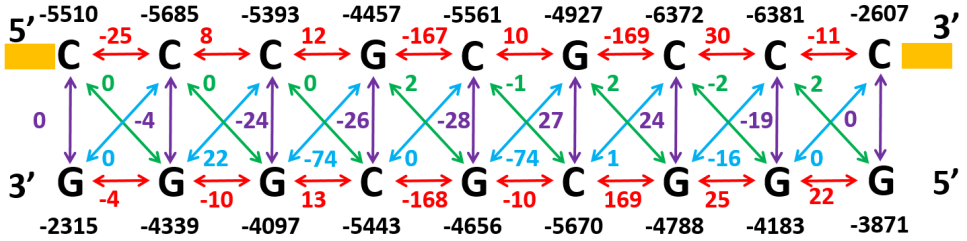
investigate the electronic structures of DNA in B-form and A-form configurations. In Figure 6.12, we show the inter- and intra- strand electronic couplings at the HOMO energy level for the ideal B-form and A-form 9-mer with the backbone included. With the inclusion of the backbone, the inter- and intra- strand coupling are rather irregular for both conformations. Overall, the inter-strand coupling within a base pair in $A3$ are larger than the ones in $B3$. However, it is difficult to draw a conclusive statement by only comparing the intra- strand couplings between the two molecules. Further inspection of the electronic couplings in the case without backbone (Figure 6.11) and with backbone (Figure 6.12) shows that for both $B3$ and $A3$, most of the coupling components along the G-hopping path are preserved while some of the components decrease after the backbone is included. This is contrary to the previously computed transmission that the backbone increases the electronic conduction for both B-form and A-form. Thus, we conclude that in the case where the backbone is included, energy levels besides HOMO may also play a role in the conduction, which is confirmed in our later calculations. This finding is consistent with the fact that the role of the backbone is important, as we have demonstrated above, and the issue of projecting the proper coupling and energy levels on to each base to achieve a proper hopping parameterization becomes non-trivial.

Although the analysis of the TB parameters at the HOMO level can provide an immediate intuition about the conduction, to ensure that the TB model can generate rigorous transport results, the comparison with the transmission based on the full Hamiltonian is needed. To this end, we show the transmission computed with the full Hamiltonian and the TB model for $B3$ and $A3$ without backbone in Figure 6.13, for $B3$ and $A3$ with backbone in Figure 6.14(b). To identify the influence of each energy level, we include the energy levels close to the HOMO and LUMO step by step in the TB model.

We first discuss the TB model for $B3$ without backbone (Figure 6.13(a)). By only including the HOMO level in the TB model (red dashed), we can get a qualitative match with the full Hamiltonian model near the HOMO level. However, the transmission information at the LUMO peak and in the HOMO-LUMO gap is completely missing. To evaluate the



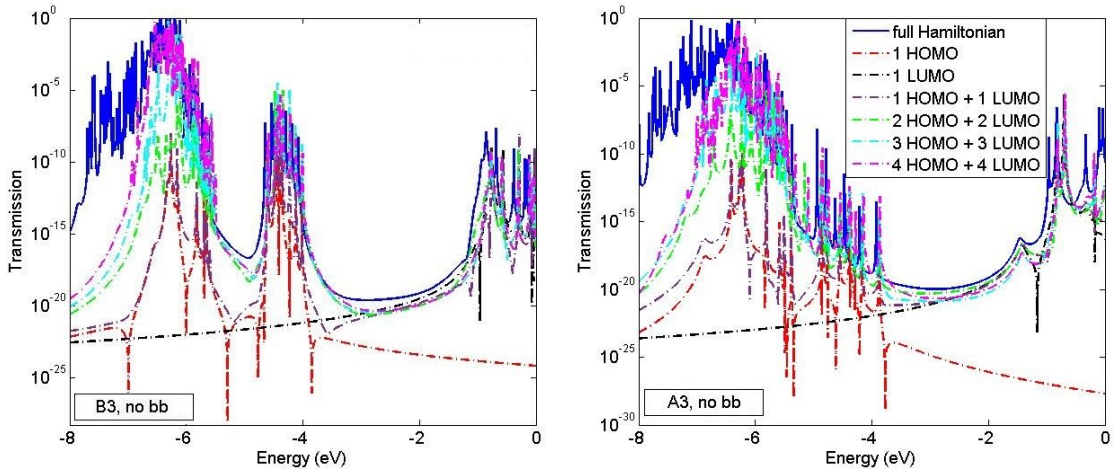
(a)



(b)

Figure 6.12: TB parameters at the HOMO level obtained from DFT calculation with B3LYP/6-31G(d, p) for (a) $B3$ with backbone; and (b) $A3$ with backbone (Unit, meV). Arrows and numbers shown in various colors follow the representations in Figure 2.3(e).

effect of the LUMO energy level, we then include only the LUMO energy level in the TB model (black dashed). Here, we find a strong match at the LUMO peak compared to the full Hamiltonian model, though the information at the HOMO peak is completely lost. In the next step, we improve the TB model by including both the HOMO and LUMO in the calculation (purple dashed). The transmission based on this model can qualitatively reproduce the full Hamiltonian result, with the physics at the HOMO peak, LUMO peak and the HOMO-LUMO gap all approximately captured. To get a more precise match with the full Hamiltonian, we sequentially include 2 HOMO + 2 LUMO levels (green dashed), 3 HOMO + 3 LUMO levels (cyan dashed), and 4 HOMO + 4 LUMO levels (magenta dashed) in the TB model. As expected, the more energy levels that are included, the better fitting with the full Hamiltonian model we obtain. With 2 HOMO + 2 LUMO levels (green dashed), the TB model can fix the narrower transmission width of the first HOMO window given by the 1 HOMO + 1 LUMO model, and increase the transmission values between the first and



(a) (b)

Figure 6.13: Transmission for (a) $B3$ without backbone; and (b) $A3$ without backbone based on the full Hamiltonian and the TB model. Blue solid, full Hamiltonian; red dashed, TB model with 1 HOMO level; black dashed, TB model with 1 LUMO level; purple dashed, TB model with 1 HOMO level + 1 LUMO level; green dashed, TB model with 2 HOMO levels + 2 LUMO levels; cyan dashed, TB model with 3 HOMO levels + 3 LUMO levels; magenta dashed, TB model with 4 HOMO levels + 4 LUMO levels.

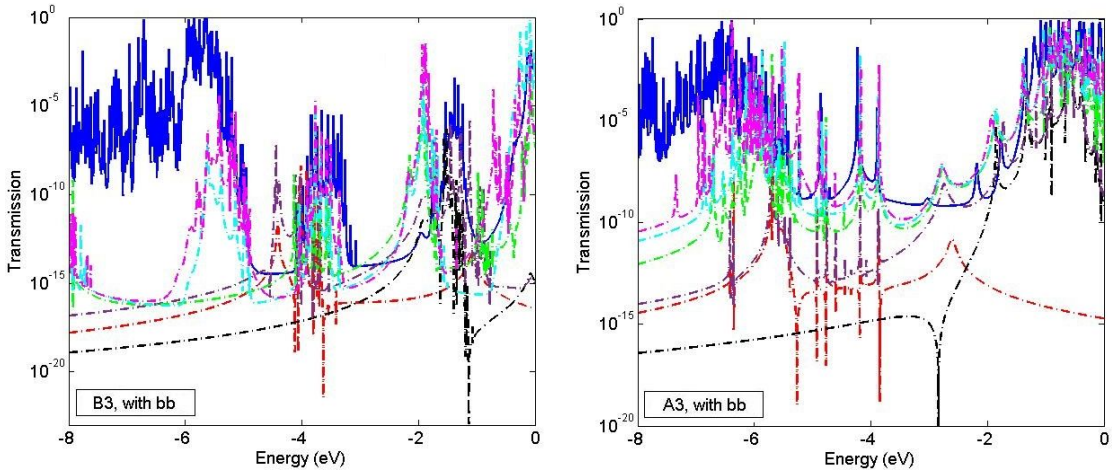
the second HOMO windows. Including 3 HOMO + 3 LUMO levels (cyan dashed), and 4 HOMO + 4 LUMO levels (magenta dashed) bring the transmission computed with the TB model even closer to the one given by the full Hamiltonian.

We now discuss the TB model for $A3$ without backbone. When only including one HOMO (LUMO) level in the TB model, we have a qualitative match with the full Hamiltonian model near the HOMO (LUMO) level. However, the match in other energy levels and the HOMO-LUMO gap is very poor. This is similar to the observation in $B3$ without backbone (Figure 6.13(a)). To get a more comprehensive fitting, one needs to include both the HOMO and LUMO energy levels in the TB model. With 1 HOMO + 1 LUMO levels (purple dashed), we can get a good match at the first HOMO and LUMO peaks. With 2 HOMO + 2 LUMO (green dashed), 3 HOMO + 3 LUMO (cyan dashed) and 4 HOMO + 4 LUMO (magenta dashed) levels involved, we obtain more precise transmission results in the TB model, compared to

the full Hamiltonian model.

The transmission for $B3$ and $A3$ with backbone based on the full Hamiltonian and the TB model is also calculated, as shown in Figure 6.14. We first analyze the TB model for $B3$ with backbone (Figure 6.14(a)). For $B3$ with backbone, including only the HOMO level in the TB model is insufficient to even get a qualitative match with the full Hamiltonian model. This agrees with the findings in the TB parameters analysis in Figure 6.12(a) that energy levels besides HOMO are also important in the conduction through a B-form DNA with backbone. Including 1 HOMO + 1 LUMO (purple dashed) in the TB model can give a good match at the LUMO peak, while the HOMO peak is shifted by 0.5 eV in the energy axis. We attribute the shift to the missing of the electronic coupling between the HOMO energy level and the energy levels below the HOMO. We then include 2 HOMO + 2 LUMO energy levels in the TB model (green dashed). The position of the HOMO peak is improved. Next, we include 3 HOMO + 3 LUMO energy levels in the TB model (cyan dashed). The transmission values in the first HOMO window increase compared to the 2 HOMO + 2 LUMO model. In addition, the second HOMO window, which is completely missing in the 2 HOMO + 2 LUMO model, shows up. Including 3 HOMO + 3 LUMO energy levels in the TB calculations improves the transmission pattern even more. A problem with the TB model which has more than 1 LUMO levels (green dashed, cyan dashed and magenta dashed) is the occurrence of the fake LUMO window at -2 eV. We conclude that for $B3$ with backbone, the TB model including 4 HOMO + 1 LUMO energy levels is sufficient to capture the physics in the transmission computed with the full Hamiltonian model.

In Figure 6.14(b), we examine the TB model for $A3$ with backbone. Including only one HOMO level is difficult to get a match with the full Hamiltonian model, similar to the case in $B3$ with backbone. To ensure the TB model to produce a reasonable transport result, more energy levels are needed. At least 2 HOMO energy levels (green dashed) should be included in order to get a qualitative match with the full Hamiltonian at the HOMO window. And to match with full Hamiltonian quantitatively, 4 HOMO energy levels (magenta dashed) are required. For the LUMO window, the inclusion of more energy levels in the TB model



(a) (b)

Figure 6.14: Transmission for (a) $B3$ with backbone; and (b) $A3$ with backbone based on the full Hamiltonian and TB model. Blue solid, full Hamiltonian; red dashed, TB model with 1 HOMO level; black dashed, TB model with 1 LUMO level; purple dashed, TB model with 1 HOMO level + 1 LUMO level; green dashed, TB model with 2 HOMO levels + 2 LUMO levels; cyan dashed, TB model with 3 HOMO levels + 3 LUMO levels; magenta dashed, TB model with 4 HOMO levels + 4 LUMO levels.

contributes to a better transmission pattern.

6.3.4 Results with HF/6-31G

It is well-known that the results from DFT calculations may depend on the choice of XC functionals and basis sets. To investigate whether our conclusion above relies on the functionals and basis sets, we also perform electronic structure calculations with a low level of theory, HF/6-31G, where the electron correlation and polarization in the basis sets are ignored. The basis set in these new calculations is nearly 1.7 times smaller than the prior calculation with B3LYP/6-31G(d, p) in terms of the size of the Hamiltonian matrix. Despite the much simpler functional and basis set, our observation for the role of backbone in B-form and A-form DNA charge transport made with B3LYP/6-31G(d, p) is captured by HF/6-31G. That is, the backbone plays an important role in A-form DNA charge transport.

In Figure 6.15, we present the transmission results of B-form and A-form DNA in the coherent transport limit, where the *ab initio* part is calculated with HF/6-31G. Figure 6.15(a) confirms the different roles of the backbone in B-form and A-form DNA. Take the 9-mer (Figure 6.15(a)) for example. In the absence of the backbone, the overall transmission of $B3$ (blue solid) is larger than $A3$ (red dashed). In the HOMO-LUMO gap, it is approximately 15 times larger. After the backbone is included, the transmission curve of $B3$ is shifted by about 1 eV in the energy axis and the transmission value is increased by about 13 times in the HOMO-LUMO gap. For $A3$, the addition of backbone does not change the position of the HOMO level significantly. The main effect of adding the backbone in $A3$ is to increase the transmission by about 10^9 times. The observation here agrees qualitatively with the one calculated with B3LYP/6-31G(d, p), as shown in Figure 6.3. These findings also hold in the 11-mer (Figure 6.15(b)), 13-mer (Figure 6.15(c)) and 15-mer (Figure 6.15(d)) strands. Comparing with Figure 6.3, we conclude that the effect of HF/6-31G is to widen the HOMO-LUMO bandgap. The distinct roles that the backbone plays in charge transport through B-form and A-form are consistent in both methods.

In Figure 6.16, we show the orbital distributions at the HOMO level calculated with HF/6-31G for the 9-mer (Figure 6.16(a)), 11-mer (Figure 6.16(b)), 13-mer (Figure 6.16(c)) and 15-mer (Figure 6.16(d)) of the B-form and A-form configurations, with and without the backbone included. Take the 9-mer for example again. The HOMO orbital distributions of $B3$ without backbone and that with backbone follow a very similar pattern - the HOMO mainly resides on the first 6 base pairs. In contrast, the HOMO distribution of $A3$ without backbone and that with backbone are very different. The HOMO orbital is mainly on the first 6 base pairs in $A3$ without backbone, and after the backbone is included, it extends through the entire molecule. This further confirms the importance of backbone in charge transport through A-form DNA. By comparing Figures 6.4 and 6.16, we draw the conclusion that though the neglection of the correlation in the functional and the polarization functions in the basis set changes the detailed shape of the wavefunction, the HOMO orbital distributions of B-form and A-form DNA with HF/6-31G resemble the features in the ones with B3LYP/6-

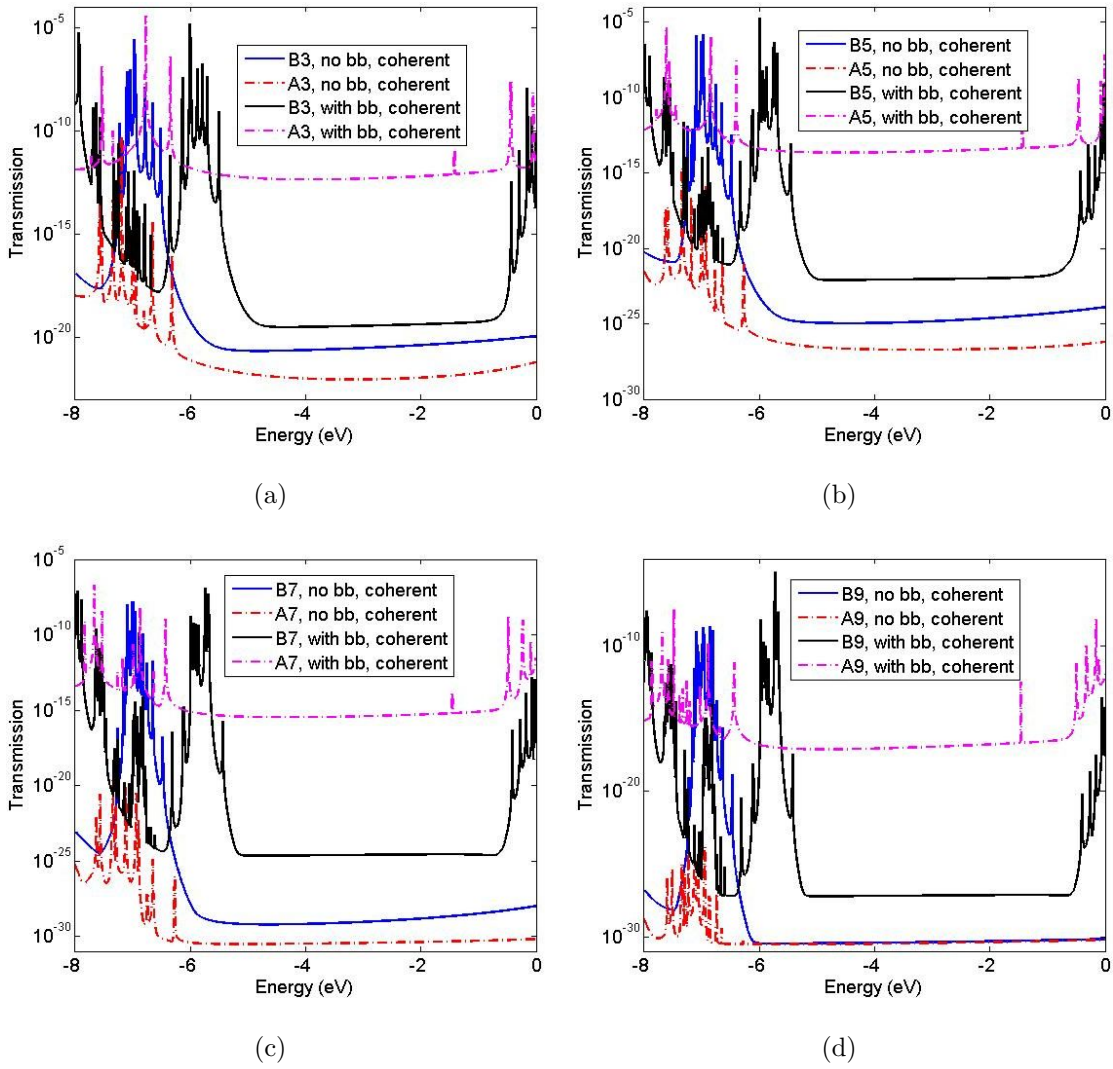


Figure 6.15: Transmission for (a) $B3$ and $A3$; (b) $B5$ and $A5$; (c) $B7$ and $A7$; and (d) $B9$ and $A9$ in the coherent transport limit calculated with HF/6-31G. Both cases, where the backbone is deleted and included, are considered. Blue solid, B-form without backbone; red dashed, A-form without backbone; black solid, B-form with backbone; magenta dashed, A-form with backbone.

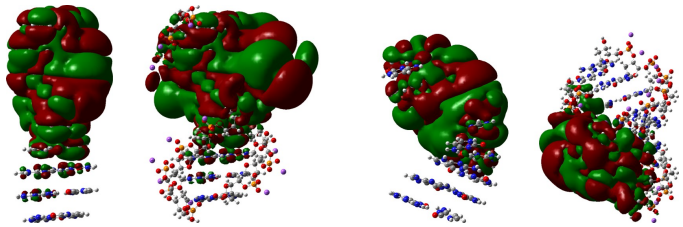
31G(d, p).

The similarities in both the transmission patterns and HOMO orbital distributions with two different methods, B3LYP/6-31G(d, p) and HF/6-31G, clearly demonstrate that the larger conductance in A-form DNA in the experiment is due to the backbone. In B-form DNA, the influence of the backbone is mainly to stabilize the stacking structures, while in A-form DNA, the backbone is involved in the electronic conduction. Including the backbone in the transport calculation of A-form DNA is important to achieve a qualitative agreement with the experimental results. The omission of the backbone in an A-form DNA may lead to incorrect results.

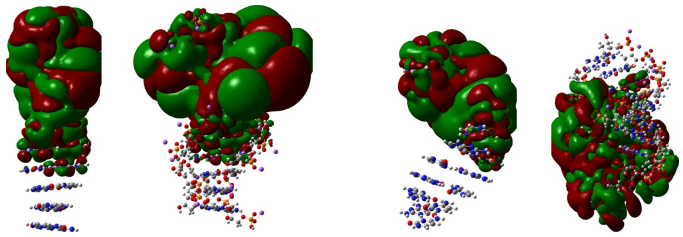
6.4 Discussion

The inter-base pair electronic couplings in a DNA molecule are sensitive to the structural details. In experiments, the structural details change as a result of temperature and time varying environmental effects. Consequently, both the energy levels and the coupling between the bases also change with time. This process is stochastic. The result of stochastic variations in the Hamiltonian is decoherence of an electron traveling in the quasi-1D system.

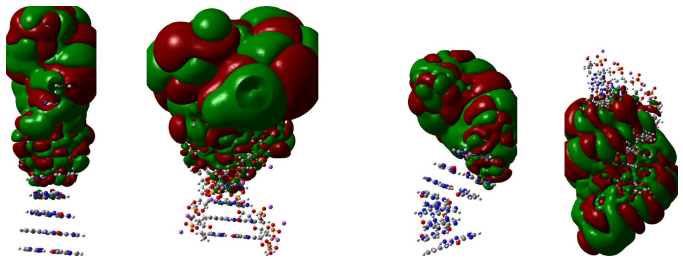
We are aware that it might be possible to do time dependent calculations with sub femtosecond resolution on a system with a Hamiltonian that varies due to fluctuations in environment (ions, water molecules, change in DNA conformation). These fluctuations are some of the reasons for decoherence. The work using this approach, while interesting, has so far involved very simplified Hamiltonians per base to represent the transport [196]. However, the method has not been applied to explain transport experiments in DNA strands, where a more complete Hamiltonian (including many energy levels on the bases) is necessary. The implementation of this approach to a realistic system with thousands of energy levels in the strand is not possible with current computational resources. We think that a method such as this, when parallelized over large computers with multiple energy levels from *ab initio* methods, coupling to the environment and including the geometry of the contacts, may be of use to the community.



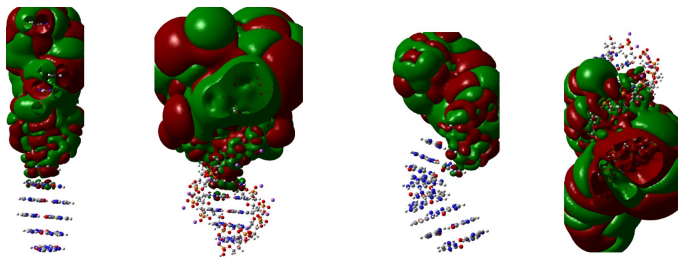
(a)



(b)



(c)



(d)

Figure 6.16: 3D iso-surface of the HOMO orbital for B-form and A-form DNA with and without the backbone calculated with HF/6-31G. (a) $A3$ and $B3$, with iso-value = 10^{-5} ; (b) $A5$ and $B5$, with iso-value = 10^{-6} ; (c) $A7$ and $B7$, with iso-value = 10^{-7} ; (d) $A9$ and $B9$, with iso-value = 10^{-8}. In each sub-figure, the HOMO orbitals are presented from left to right in the following order: B-form without backbone; B-form with backbone; A-form without backbone and A-form with backbone.

It is also worth pointing out that by looking at the electronic coupling in multiple DNA configurations, we may be able to further understand the charge transport mechanism through the two different structures. The calculations above are based on the two ideal structures, B-form and A-form which are generated from the standard software package, NAB. We are able to observe the distinct transport behaviors in these two ideal structures and explain the observation in experiment qualitatively. Moreover, by including the DOS in the argument, we are able to demonstrate that the A-form should be more conductive than B-form for these molecules, independent of the specific mechanism. Nevertheless, it is still helpful if we analyze the electronic couplings of multiple structures for both B-form and A-form DNA from dynamical trajectory.

For sufficiently long strands, we believe that the frozen atom approach from the trajectory is unlikely to unravel DNA conductance. Long strands tend to localize due to random changes in Hamiltonian. Decoherence makes the system more conducting. Decoherence can be included by time varying potentials and solving a time dependent equation [196], or in a phenomenological approach, like what has been done in our work. But the real test of these methods is in predicting experiments by comparing them with experimental data over many strands.

6.5 Conclusion

In this chapter, we study the configuration-dependent conduction through DNA by exploring the charge transport mechanism through a series of ideal B-form and A-form DNA. We show the transmission for B-form and A-form DNA with and without backbone to study the role of backbone in DNA conduction. Our main observation is that the addition of the backbone in A-form DNA is essential to get qualitative agreement with the experimental results where the A-form DNA is 10 times more conductive than B-form. Without the inclusion of the backbone in the modeling, qualitatively incorrect answers will result (there is proof of this from some recent experiments in Prof. Josh Hihath's group). The profound effects of the backbone on the transport properties of the system is due to the change in the charge

distribution when the structure changes from B-form to A-form, which have been confirmed by analyzing the HOMO distributions of DNA strands with and without backbone for both configurations. The HOMO level is distributed over the entire molecular length when the backbone is present in the A-form, but not without the backbone present, or in either case for the B-form. The findings in the HOMO pattern further demonstrate the importance of including the backbone in transport calculations on DNA. Moreover, these results allow us to qualitatively describe the observed conductance behavior. In the A-form, the energy levels are more distributed over the molecular structure resulting in a higher overall conductance.

We further provide DOS analysis for B-form and A-form DNA. In the central region, the DOS of the base pairs in A-form DNA can be larger than B-form by 1000 times. These findings hold true for the entire family of the molecules studied in this chapter. We note that with the fully-coherent model, the conductance from the modeling is several orders of magnitude smaller than the observation in the experiment for both B-form and A-form DNA, which is due to the lack of the fluctuations in the model. To model these effects phenomenologically, we employ the decoherent model developed in Chapter 2, which brings the conductance of both B-form and A-form DNA closer to the experiment, while maintaining the finding that A-form DNA is more conducting than B-form.

In order to further reveal the physics behind the conformation-dependent transport and reduce the expensive computational cost in the full Hamiltonian model, we examine the TB parameters and develop a model Hamiltonian for both B-form and A-form including and not including the backbone. For the strands without backbone, by including one HOMO and one LUMO orbitals in the TB model, one can get a qualitative agreement in the transmission with the full Hamiltonian model. After the backbone is included, the intrinsic transport mechanism becomes more complex. Four HOMO and one LUMO orbital are needed for a B-form DNA with backbone. In the case of an A-form DNA with backbone, four HOMO together with four LUMO orbitals are necessary to get a good match with the transmission calculated by the full Hamiltonian model.

To study the dependence of the result on the XC functionals and basis sets in the DFT

calculations, we re-do the DFT calculation with a low-level method, HF/6-31G, where the electron correlation in the XC functional and the diffusion in the basis sets are completely ignored. Despite the simplicity of this method, our conclusion in the results with B3LYP/6-31G(d, p) still holds. That is, the backbone plays an important role in A-form DNA charge transport. Both the transmission and HOMO plots show that although the electron correlation in the functional and the diffusion in the basis set can change the details of the pattern, the essential physics is consistent.

In summary, our current set of calculations based on the coherent and decoherent calculations of how electrons move along the DNA length, the HOMO distribution of the molecule (both qualitative visualization and quantitative coefficients calculation), the DOS across the length of the molecule, and the findings with two different DFT computational methods all give support to the experimental observation of higher conductance of A-form. This clearly suggests that the number of states available for transport is larger in the A-form than B-form, and as such, it is expected to display a larger conductance value regardless of whether the transport is dominated by tunneling, partial decoherence processes, hopping, or fluctuation dominated transport events, although continued studies of this system with each of these techniques may lead to a cohesive story for DNA charge transport. Work by different groups on the same sets of experimental data will go a long way into resolving the disagreements based on theory, and ultimately unravelling further details of transport in DNA.

Chapter 7

EFFECT OF WATER ON ELECTRICAL TRANSPORT THROUGH DNA

7.1 Introduction

Despite the recent progress in the conductance through DNA, many problems still remain unresolved. Currently, one of the controversial problems in this field is how the local environment of the DNA molecule affects its electrical conduction. The electronic coupling between the neighboring bases is an important factor in governing the charge transport in DNA, which subsequently depends on the conformation of the molecule [27, 197]. One manner that local environment influences DNA is via making changes to the structure of the molecule [62, 198, 199], and even switching the conformation of molecule by tuning the concentration of the counterions [200].

Water molecules in the local environment may have an impact on the electronic properties of DNA, because the dipole moments of water may change the local electric field through DNA. It has been found that the surrounding water can change the total energy and dipole moments of stacking bases [201], alter the IPs of single DNA bases [202], break some of the π-bonds [203], weaken and lengthen the hydrogen bonds [204]. The variation in the number of water molecules can also modify the HOMO and LUMO orbitals of the molecule [136], and even cause a conformational transition between B-form and A-form [6, 190]. These findings suggest that the presence of water molecules in the surrounding environment of a DNA has influence on the electronic properties of DNA. However, regarding the direct impact of water on the conductivity, results are controversial from both experiments and computations. Measurements on DNA with length of $0.5 - 2.9$ μm [205], bundles with $10 - 10^6$ molecules [59], individual strands with length of 140 nm [206], single-stranded and

double-stranded DNA with $120-3000$ nucleotides [207], and bundles consisting of 48500-base pair DNA [208] show that humidity affects the conductance of the DNA molecules strongly - an increase in humidity leads to a clear enhancement in DNA conductance. However, a more recent study [60] on a 21-mer oligonucleotide with the mechanically controlled break junction (MCBJ) technique shows that the conductance value of a DNA strand in solution is similar to that in vacuo. The main effect of the solution is to help stabilize the strand. Computational studies combining the large-scale QM/MD simulations with the Landauer theory on DNA undecamers [88] and Dickerson dodecamer [89] show that the transport through DNA is significantly affected by the dynamical disorder from the solvent. By adding the gold electrode explicitly, Reference [90] finds that conductance through two DNA strands with four base pairs can be enhanced by the humidity. In contrast, based on the comparison of the transmission between four 14-base pair DNA strands in dry and wet states, Reference [87] claims that water molecules and counterions do not create new transmission channels in addition to the intrinsic ones due to DNA itself. With these interesting findings, we note that a large number of studies [59, 88, 89, 90, 205, 206, 207, 208] have the problem of distinguishing the pure effect of water from that of counterions, which surround the negatively charged phosphate group of DNA backbone.

In this chapter, we simulate the effect of water on several DNA strands, including a four-base pair B-form strand with backbone (Section 7.3.1), a series of short oligonucleotides in both B-form and A-form configurations without backbone (Section 7.3.2), an eight-base pair [44] and a nine-base pair [54] strands of which the conductance values have been measured (Section 7.3.3). Specifically, we model the effect of water using three different models - we first add explicit water molecules around the DNA base pairs. We then use the polarizable continuum model (PCM) to simulate the effect of bulk water. We also employ a new model that combines the PCM and the explicit water molecules. The third model is inspired by the work in Reference [209] that studies the $fac-[Ru(CO)_3I_3]^-$. We find that the PCM is the best compromise in modeling the effect of water on electrical conduction through DNA. We also find that the effect of water on DNA conduction is conformation-dependent.

Transport calculations reveal that the peak value of the conductance for a DNA strand with experimental importance can be increased by $3-8$ times by water, depending on the sequence.

7.2 Strands and models

To get a comprehensive understanding of the effect of water, we consider the following *three sets* of DNA strands: (1) The *first set* is a B-form four-base pair sequence, $5'-GGGG-3'$ (with the complementary strand $3'-CCCC-5'$). In this structure, we keep the backbone which is neutralized by the Na^+ counterions. Using this strand, we obtain a fundamental understanding of the role of the water molecules on DNA by examining the structure and charge distribution of the DNA after the water molecules are introduced in different methods. To this end, we adopt four models. In the first model, we only consider the dry strand in the dry state to serve as a reference for the three solvation models. In the second model, we add three explicit water molecules around the bases. In the third model, we use the PCM, which models the solvation effect by viewing solvent as a polarizable continuum [210]. In the fourth model, we combine the second and third models by adding three explicit water molecules to the DNA molecule while implementing the PCM. In each of these models, we retain the backbone configuration from the NAB software package and relax the atoms in the bases, counterions and water molecules. (2) The *second set* of strands we have modeled is a series of DNA strands with both B-form and A-form stacking configurations. The sequences are $5'-G-3'$, $5'-GG-3'$, $5'-GGG-3'$ and $5'-GGGG-3'$ (with the complementary strands $3'-C-5'$, $3'-CC-5'$, $3'-CCC-5'$ and $3'-CCCC-5'$), which are abbreviated as $G1$, $G2$, $G3$ and $G4$, respectively. In this step, we use the structure from NAB, and simply terminate the backbone with a hydrogen as shown in Figure 1.2. (3) In the *third set* of strands, we consider an eight-base pair strand with sequence $5'-GCGCGCGC-3'$ (with complementary strand $3'-CGCGCGCG-5'$), labeled as $GC8$ (same as in Chapter 5), and a nine-base pair strand with sequence $5'-CCCGCGCCC-3'$ (with complementary strand $3'-GGGCGCGGG-5'$), labeled as $B3$ (same as in Chapter

6). $GC8$ is used in the experimental work of Reference [44] and $B3$ in used in Reference [211]. The third set consists only of B-form DNA. Identical to Set 2, the structure here is from NAB, with the bases terminated by hydrogen atoms as shown in Figure 1.2. The purpose of choosing the first and second sets of strands is to obtain a fundamental understanding of the effect of water on the electronic properties of DNA, while the choice of the third set is motivated by experiments.

All DFT calculations are performed using the Gaussian 09 software package with the B3LYP/6-31G(d) XC functional and basis set. The transport calculations are implemented with Green's function in the Landauer-Büttiker framework, with $\Gamma_L = \Gamma_R = 100\ meV$.

7.3 Results and discussion

In this section, we discuss the results from our calculations. The results are presented in the following order - we first compare the three solvation models described in Section 7.2, by inspecting the structure and the charge distribution of the four-base pair DNA optimized with the three models. We then study the roles of water molecules on DNA with different conformations by examining a series of short DNA strands in B-form and A-form configurations. Finally, we study the effect of water on conductance of DNA strands with experimental importance.

7.3.1 Models comparison (First set of DNA)

We first focus on the four-base pair B-form DNA strand, $5' - GGGG - 3'$. In order to obtain a quantitative comparison of the three solvation models, we first present the optimized structure of the strand from these models. For clarity, we focus on the structure of the second $G : C$ base pair ($GC2$) in the four-base pair DNA strand. As a reference, we first show the un-optimized structure of $GC2$ in the four-base pair DNA generated with NAB in Figure 7.1.

The optimized structures for DNA in the dry state and with the solvation model that includes three explicit water molecules are shown in Figures 7.2(a) and 7.2(b). The optimized

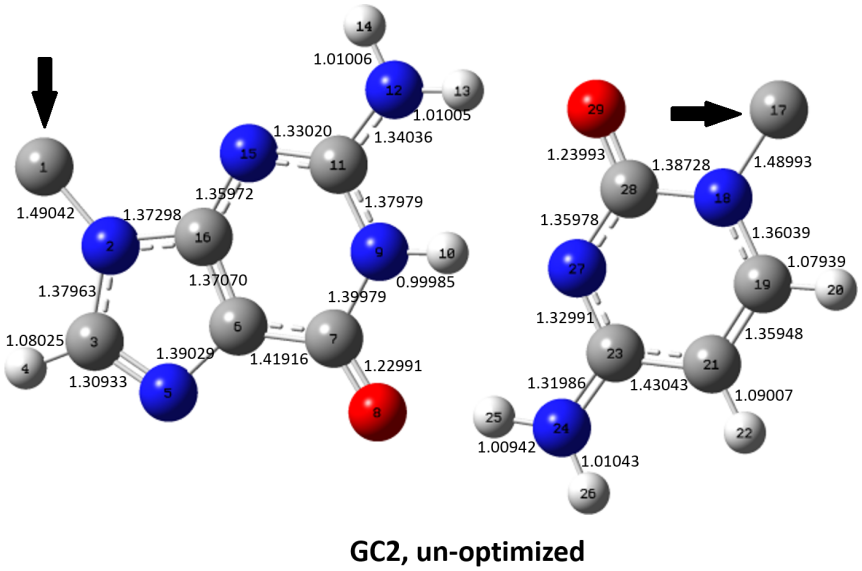
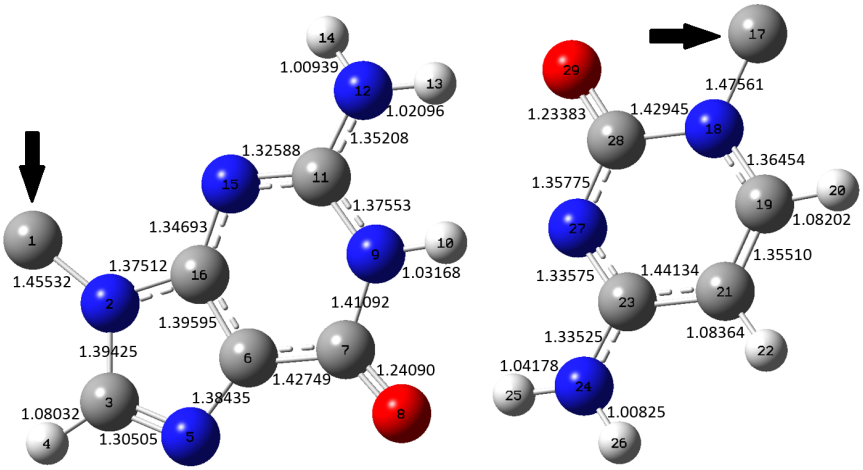


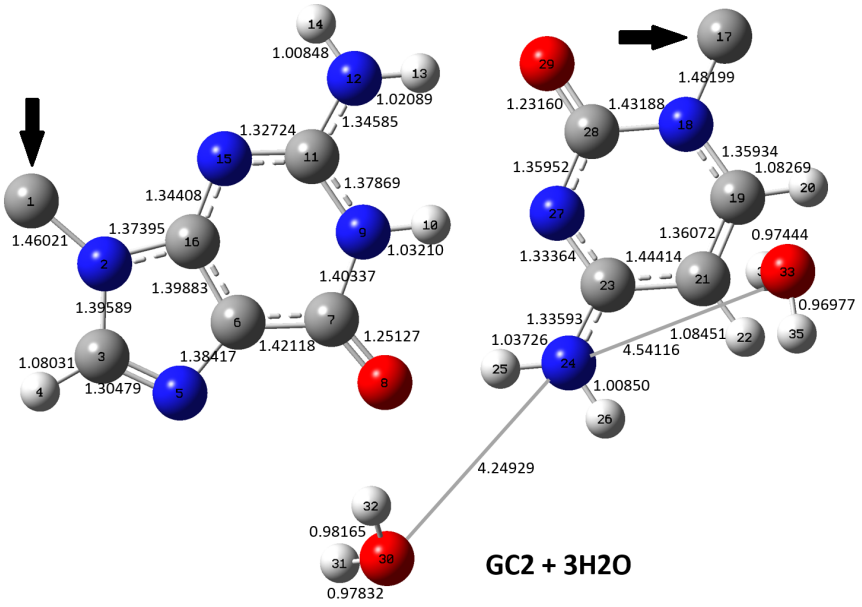
Figure 7.1: (First set of DNA) Un-optimized structure from NAB for $GC2$ in the B-form four-base pair sequence $5' - GGGG - 3'$ (with the complementary strand $3' - CCCC - 5'$), which is referred to as the first set in the previous page. Note that the calculations include the backbone and counterions.

structure of $GC2$ in the dry state has minor differences from the unoptimized one in Figure 7.1. For example, the bond length between C28 and O29 is 1.23383 $\mathring{A}$ in the optimized structure when the DNA is in the dry state and 1.23993 $\mathring{A}$ in the unoptimized one. The difference between the two structures is less than 0.5%. This suggests that the structure generated by the NAB software is reliable. In the solvation model where the three water molecules are added, the bond lengths for the atoms in the bases which are far away from the water molecules do not change very much, compared to those in the dry state. However, for the atoms close to the water molecules, the bond lengths are affected more. For example, in the dry state, the bond length between N15 and C16 is 1.34693 $\mathring{A}$, and after water molecules are included, it is 1.34408 $\mathring{A}$. The difference is only 0.2%. In contrast, the bond length between C7 and O8 increases from 1.24090 $\mathring{A}$ to 1.25127 $\mathring{A}$ ($\sim 0.8\%$) after the three water molecules are included, due to the water molecules nearby.

We then examine the optimized structures of DNA with the solvation models where the



(a)



(b)

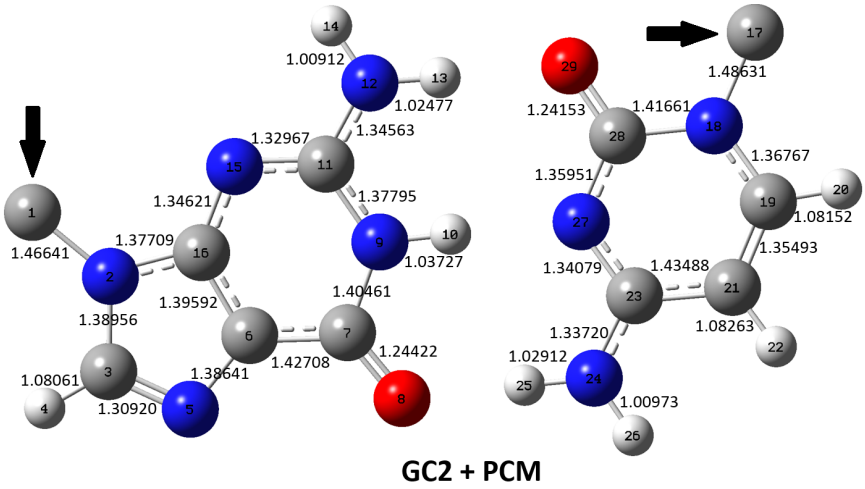
Figure 7.2: (First set of DNA) Structures of DNA in the dry state and with the solvation model which includes three explicit water molecules calculated by B3LYP/6-31G(d). (a) Optimized structure without water; (b) optimized structure with three explicit water molecules. Unit of bond length, $\mathring{A}$. For clarity, we only show the structure of the second base pair. The arrows show a carbon atom in the backbone. The calculations include the backbone and counterions.

PCM and PCM + three explicit water molecules are employed. The results are shown in Figures 7.3(a) and 7.3(b). Compared to the optimized structure of DNA in the dry state (Figure 7.2(a)), the optimized structure calculated with the PCM changes by a large amount, demonstrating that the bulk water molecules modeled with the PCM have a big impact on the structure of DNA. The optimized structure relaxed with the solvation model combining the PCM with three water molecules (Figure 7.3(b)) is very similar to that of the pure PCM model. Interestingly, the bulk water simulated with the PCM also has a big impact on the three explicit water molecules - both the locations with respect to the DNA base pairs and the structures of the water molecules change when the bulk water is involved (see the optimized structure of the water molecules in Figures 7.3(b) and 7.2(b)). The distances between water molecules and DNA base pairs in the model with PCM + three water molecules increase, compared to the one where only three water molecules are included.

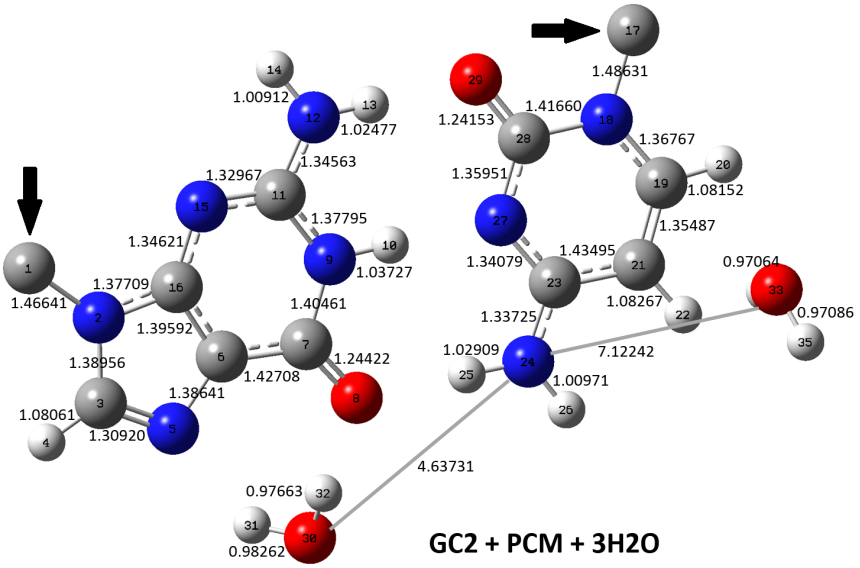
Next, we investigate the effect of the three solvation models on the charge distribution of the four-base pair DNA (first set of DNA). The results are shown in Figures 7.4 and 7.5. We find that as the solvation model changes, the charge distribution on atoms also changes. A comparison between Figures 7.4(a) and 7.4(b) shows that with three water molecules included, the charge re-distributes on each atom. After the bulk water is included (Figures 7.5(a) and 7.5(b)), the amount of charge on each atom changes even more.

In Figure 7.6, we show the energy levels of the four-base pair DNA strands (first set of DNA) between HOMO - 5 and LUMO + 5 calculated with the four different models. We make the following observations: (1) the inclusion of water can increase the band gap, irrespective of the solvation models. The addition of only three water molecules can enlarge the band gap from 1.251 eV to 1.432 eV, a change of 181 meV. With the inclusion of the PCM model, the band gap increases by 905 meV. We also find that the band gap given by the PCM model and *PCM with three explicit water molecules* are quite close, differing by only 37 meV; (2) with the addition of water, the energy levels of DNA become more dense close to the HOMO level.

Finally, we compute the transmission of the four strands with the four models, as shown in

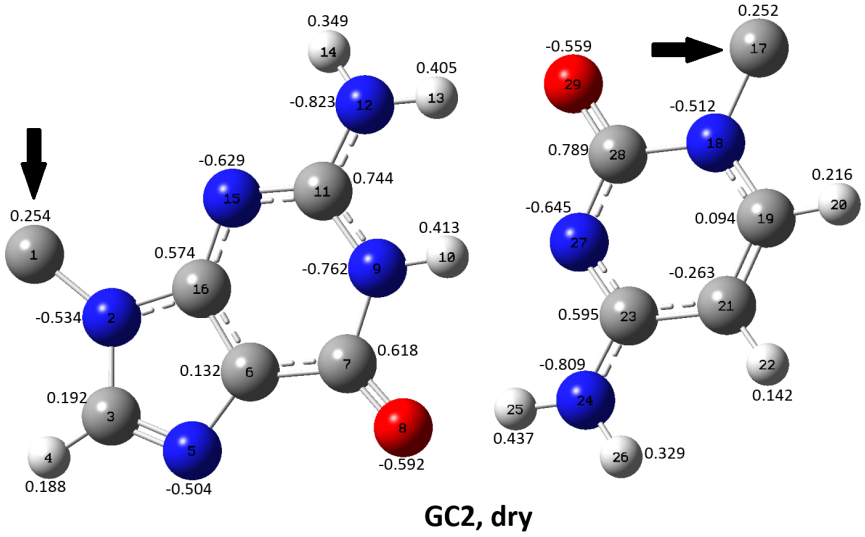


(a)

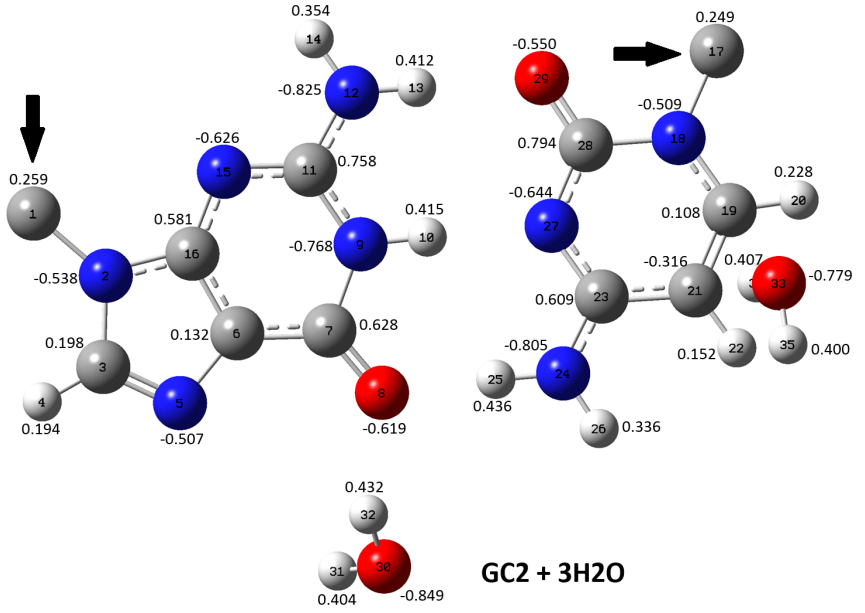


(b)

Figure 7.3: (First set of DNA) Structures of DNA with the solvation model where the PCM and PCM + three explicit water molecules are employed calculated by B3LYP/6-31G(d). (a) Optimized structure with the PCM model; (b) optimized structure with PCM + three explicit water molecules. Unit of bond length, $\mathring{A}$. For clarity, we only show the structure of the second base pair. The arrows show a carbon atom in the backbone. The calculations include the backbone and counterions.

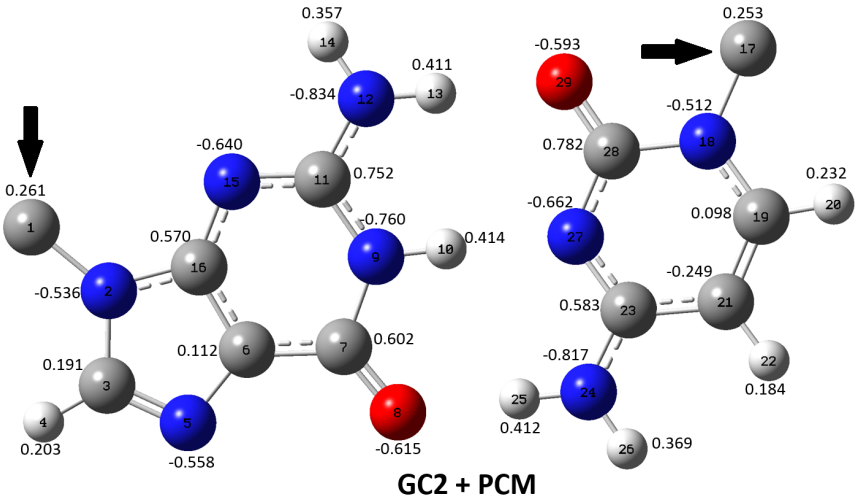


(a)

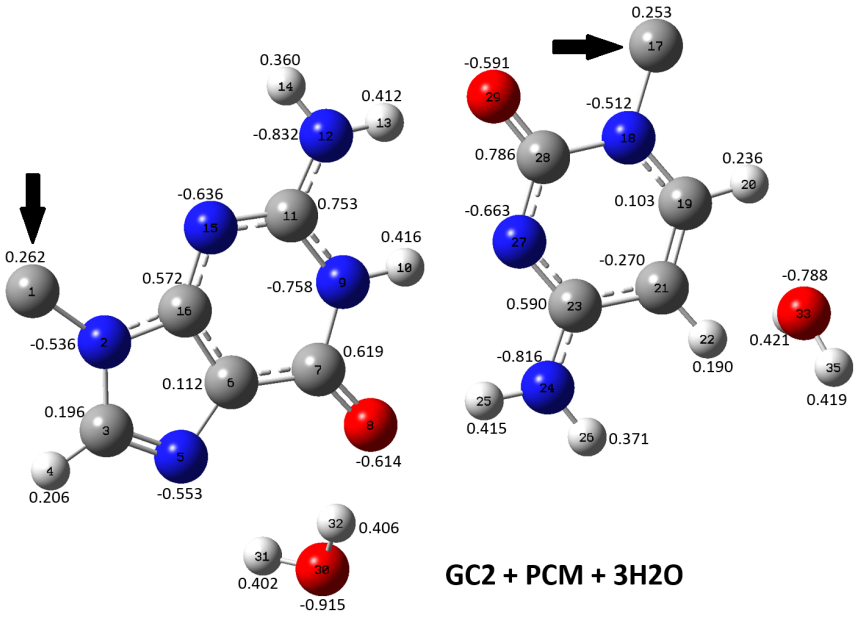


(b)

Figure 7.4: (First set of DNA) Charge distribution of DNA in the dry state and with the solvation model that includes three explicit water molecules calculated by B3LYP/6-31G(d). (a) Charge distribution in the model without water; (b) charge distribution in the model with three explicit water molecules. For clarity, we only show the structure of the second base pair. The arrows show a carbon atom in the backbone. The calculations include the backbone and counterions.



(a)



(b)

Figure 7.5: (First set of DNA) Charge distribution of DNA with the solvation model where the PCM and PCM + three explicit water molecules are employed. Both calculations use B3LYP/6-31G(d). (a) Charge distribution in the model with the PCM; (b) Charge distribution in the model with PCM + three explicit water molecules. For clarity, we only show the structure of the second base pair. The arrows show a carbon atom in the backbone. The calculations include the backbone and counterions.

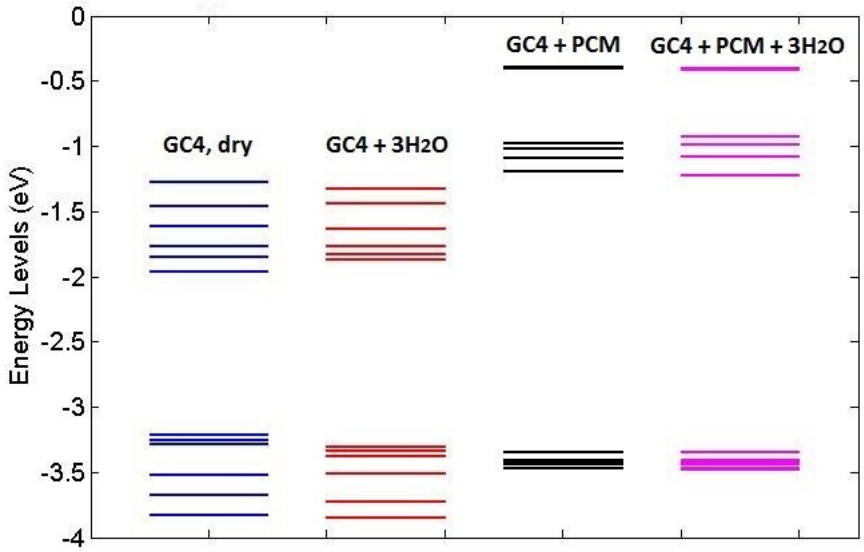


Figure 7.6: (First set of DNA) Energy levels of the four-base pair DNA strands between HOMO - 5 and LUMO + 5 calculated with the four different models. (a) The model without water; (b) the model with three explicit water molecules; (c) the model with PCM; and (d) the model with PCM + three explicit water molecules.

Figure 7.7. The transmission of the model that has three water molecules is shifted compared to the one that does not consider the solvation effect. This is consistent with the findings in the energy levels distribution (Figure 7.6). Deep in the HOMO band, the transmission for the model with the three water molecules included is different from the dry DNA. Sometimes the transmission values are larger and sometimes smaller, showing that the presence of the water molecule plays an important role in changing the transmission of the DNA. We then compare the transmission of the model with the PCM and the one with PCM + three water molecules. The transmission of these two models are quite similar in both the pattern and values, which agrees with the energy levels diagram (Figure 7.6). We now discuss the result with the PCM model (black curve) by comparing the transmission with the dry DNA model (blue curve). Beside the shift in the HOMO position, one can see that the PCM model can increase the transmission of DNA. The peak transmission value of the PCM model is about 5 times larger than the dry DNA model.

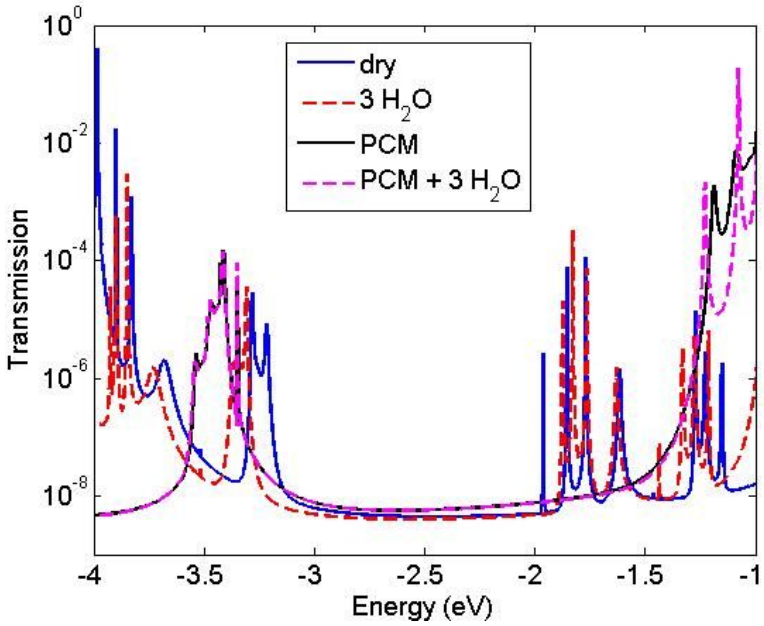


Figure 7.7: (First set of DNA) Transmission of the four-base pair DNA calculated with the four different models. Blue solid, without water; red dashed, with three explicit water molecules; black solid, with the PCM model; magenta dashed, with PCM + three explicit water molecules.

We also calculate the conductance of the DNA with the four models. The result is shown in Figure 7.8. Similar to the observation in transmission, we find that the presence of water molecules can shift the conductance in the Fermi energy axis and also change the conductance values. The addition of the three water molecules can increase the conductance values by 1.02 times and the PCM solvation model can increase the conductance by 9 times.

Based on the discussion of the four simulation models, we conclude that in order to model the effect of water on transport through DNA effectively, the PCM is a good compromise. The structure, charge distribution, energy levels, and transmission/conductance comparison between the PCM and PCM + three water molecules show a trend that the effect of the few explicit water molecules on the transport of DNA is immersed in the bulk water model. Thus, in the following part of this chapter, we use the PCM model to study the effect of water on DNA strands with different conformations, and the ones that were used in experiments.

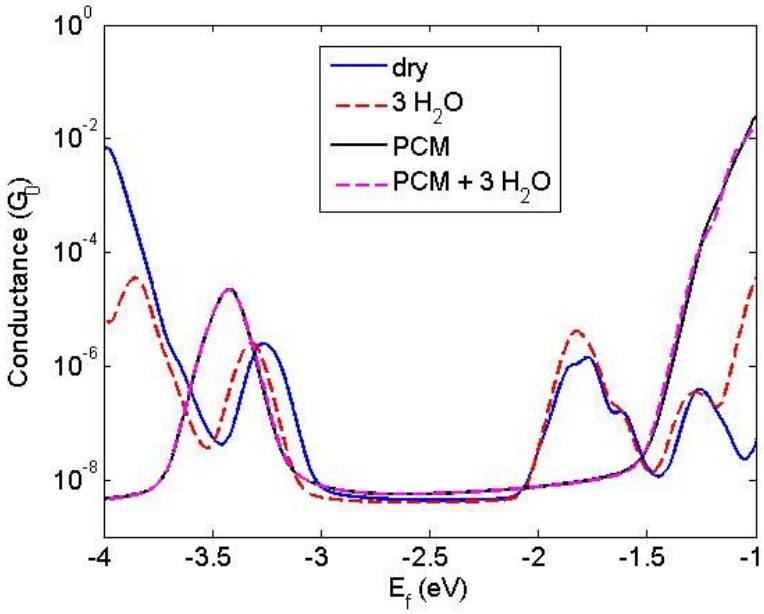


Figure 7.8: (First set of DNA) Conductance of the four-base pair DNA calculated with the four different models. Blue solid, without water; red dashed, with three explicit water molecules; black solid, with the PCM model; magenta dashed, with PCM and three explicit water molecules.

7.3.2 Effect of water on B- and A-form stacking bases (Second set of DNA)

In this section, we discuss the role of water molecules on DNA with different conformations by examining a series of short DNA strands in B-form and A-form configurations. We will see that the influence of water molecules on DNA is conformation-dependent.

7.3.2.1 Effect of water on short B-form DNA Oligomers

The HOMO orbitals of $G1 - G4$ with the B-form configuration in the dry and hydrated states are shown in Figure 7.9.

The HOMO orbitals tend to accumulate in G bases in both the dry and hydrated states, which is consistent with the fact that the IP of G is lower than C. As the number of base pairs increase, the distribution of HOMO orbitals for the dry DNA strands shows a larger nonuniformity - they are inclined to accumulate on the first few base pairs of the strand.

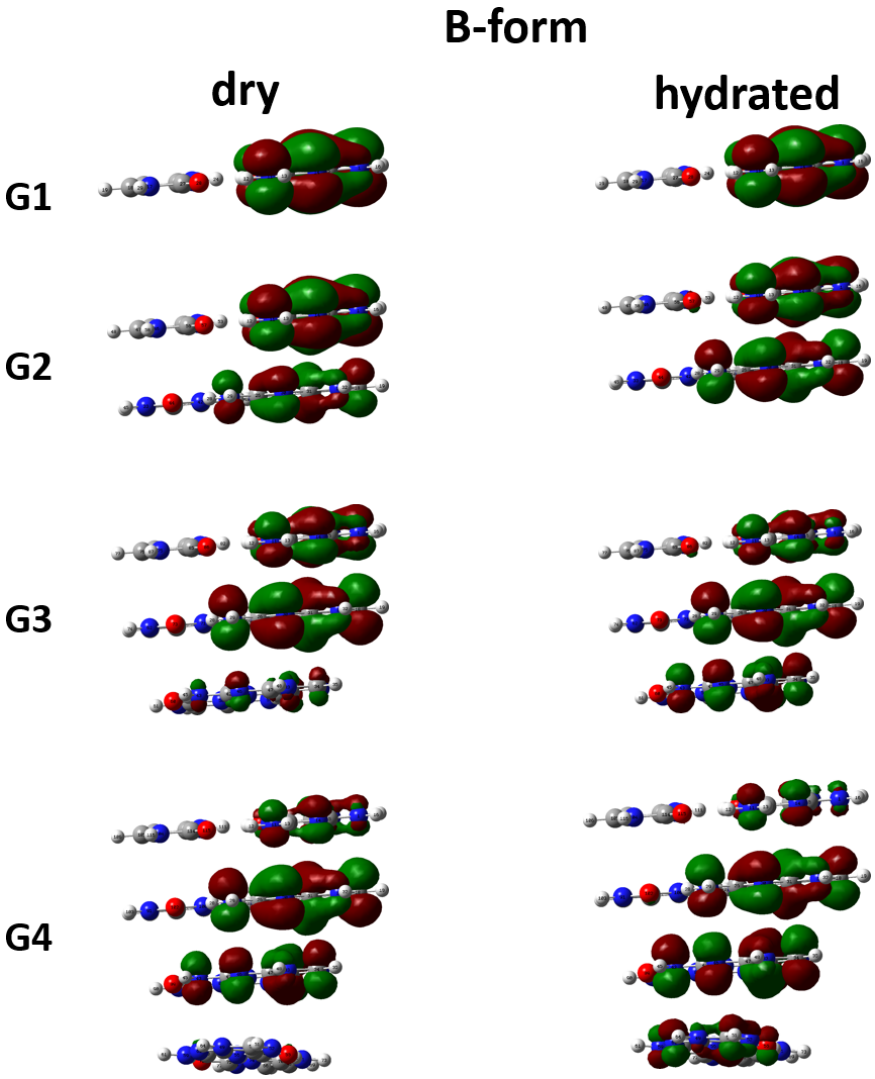


Figure 7.9: (Second set of DNA) HOMO distributions for $G1$, $G2$, $G3$ and $G4$ with the B-form configuration in the dry state (left panel) and hydrated state (right panel). Red and green colors represent the positive and negative components of the wavefunctions. Iso-value = 0.02.

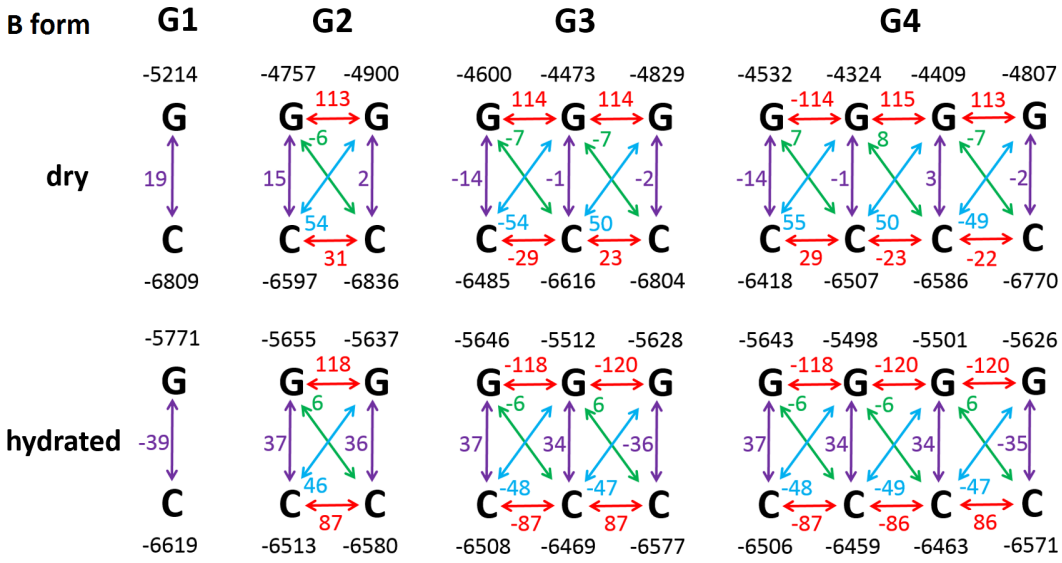


Figure 7.10: (Second set of DNA) TB parameters at the HOMO level for $G1$, $G2$, $G3$ and $G4$ with the B-form configuration in the dry state (upper panel) and hydrated state (lower panel). Arrows and numbers shown in various colors follow the representations in Figure 2.3(e).

However, after the water is included, the HOMO orbitals of these molecules tend to be more uniform, with a large portion on the center of the strands. This effect is clear for $G3$ and $G4$. In dry $G3$, the largest portion of the HOMO orbital is on the first two G bases. After including water, the orbital is more diffusive, with the second G having the largest contribution. Similar analysis applies to $G4$.

The above observation in HOMO orbital distribution of $G1 - G4$ in dry and hydrated states can be further understood by the TB parameters at the HOMO, as depicted in Figure 7.10.

It is evident from Figure 7.10 that the PCM model has a significant effect on the onsite energies of the DNA bases, especially the G bases. For $G1$ which has a single base pair, the onsite energy of G decreases by about 500 meV upon the inclusion of water, while that of C base increases by about 200 meV. As the length of the strand increases, the onsite energies of G and C bases undergo a non-trivial change. We also observe that in dry DNA strands, the

onsite energies of G have a non-uniform distribution even within the same strand, with the largest difference between the minimum and maximum values being as much as 483 meV in $G4$. Upon the inclusion of bulk water, the discrepancy in onsite energies reduces to 145 meV. In addition to the modification of the onsite energies of DNA bases, the presence of water around a DNA strand also changes some components of the hopping-integrals between the neighboring bases. The ones on which water has the largest influence are the inter-strand hopping integrals between the two nearest neighboring G and C bases from the two different complementary strands, and the intra-strand coupling hopping integrals between the two nearest neighboring C bases within the same single strand. The former not only undergoes a value change by an amount of $20 - 30$ meV after taking the bulk water into account, but also shows a more uniform distribution due to the smearing out of the edge effect, while the latter one mainly undergoes a value change by about 60 meV. It is also interesting to note that values of the inter-strand hopping integrals between the second nearest neighboring bases and the intra-strand hopping integrals between the two nearest neighboring G bases stay similar before and after the addition of water, indicating that these interactions mainly depend on the internal structure of the molecule.

7.3.2.2 Effect of water on short A-form DNA Oligomers

In Figure 7.11, we show the HOMO orbitals distribution of $G1-G4$ with the A-form stacking configuration in both the dry and hydrated states. The configuration of the base stacking has a significant effect on the HOMO distribution of DNA. The distribution of the HOMO orbital for each strand with the B-form configuration is diffusive on the first few base pairs, as seen in Figure 7.9. In contrast, the HOMO orbital of the short A-form DNA mostly resides on the G base of the last base pair for all of the four strands. With the consideration of the effect of water, the HOMO of the A-form DNA tends to move to the middle region of the strand. For instance, the HOMO orbital of $G3$ shifts from the last G base to the middle G base after including water. Similarly, the HOMO orbital of $G4$ shifts from the last G base to the two bases in the middle region with the inclusion of water.

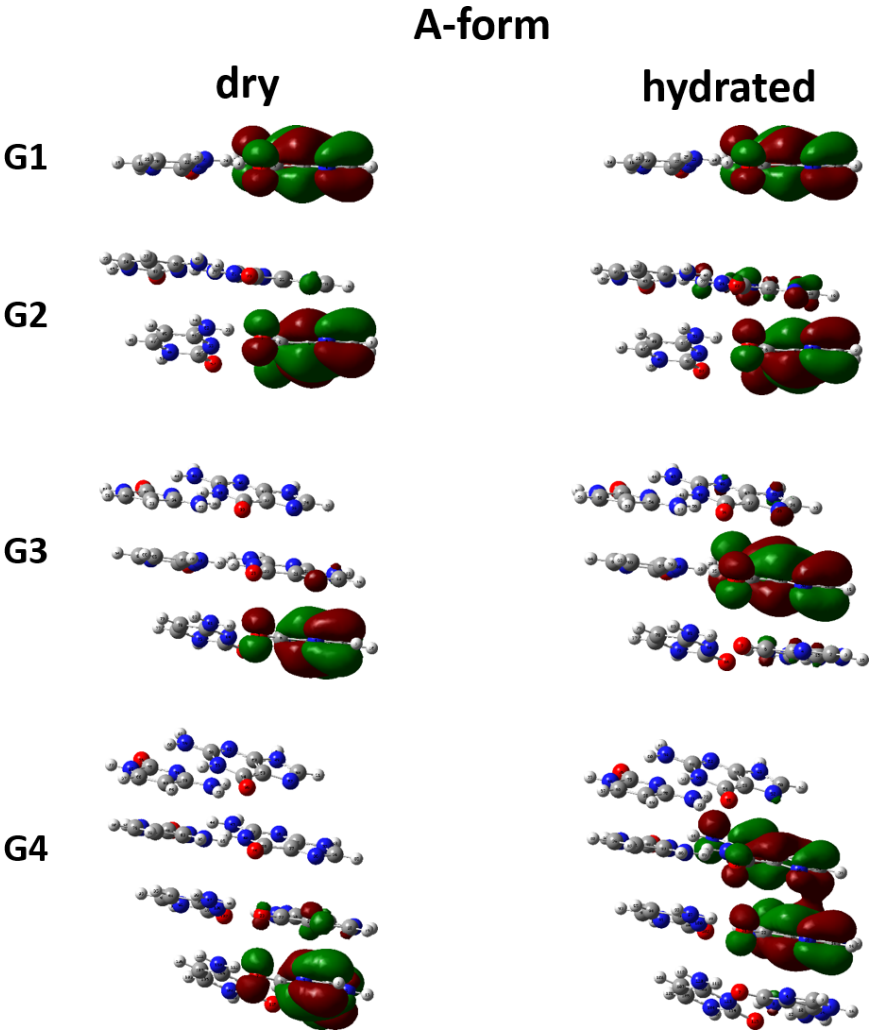


Figure 7.11: (Second set of DNA) HOMO distributions for $G1$, $G2$, $G3$ and $G4$ with the A-form configuration in the dry state (left panel) and hydrated state (right panel). Red and green colors represent the positive and negative components of the wavefunctions. Iso-value = 0.02.

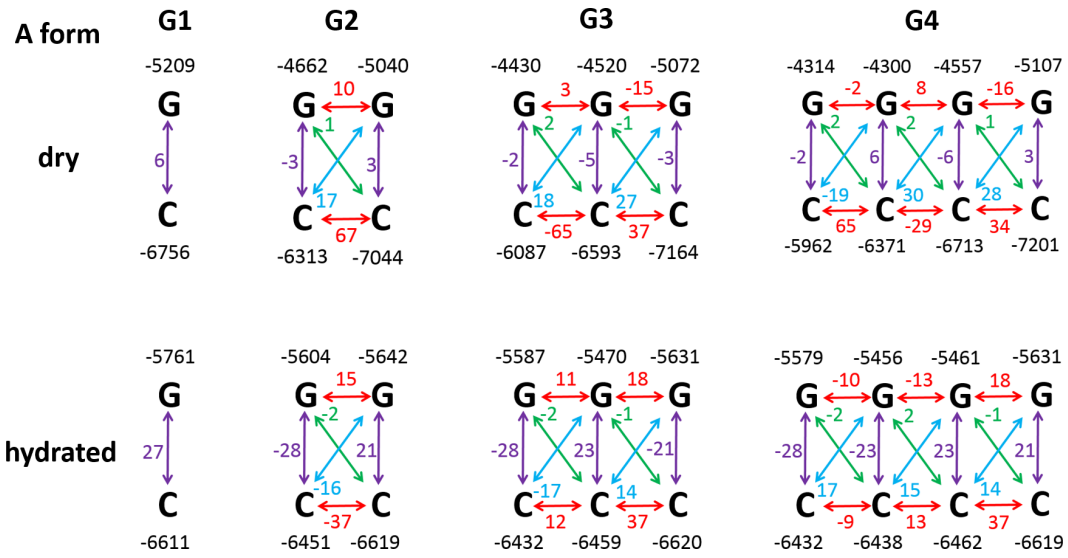


Figure 7.12: (Second set of DNA) TB parameters at the HOMO level for $G1$, $G2$, $G3$ and $G4$ with the A-form configuration in the dry state (upper panel) and hydrated state (lower panel). Arrows and numbers shown in various colors follow the representations in Figure 2.3(e).

In Figure 7.12, we show the TB parameters of the A-form short DNA strands at the HOMO level. The bulk water also plays an important role in the TB parameters for A-form DNA, including the onsite energies of the bases and some components of the hopping integrals. The hopping integrals that are affected most are the inter-strand hopping integrals between the two nearest neighboring G and C from the complementary strands, and the intra-strand hopping integrals between the two nearest neighboring C bases within the same strand. The onsite energy of G base increases by up to 1000 meV. The nearest neighboring inter-strand hopping between G and C increases by $10-20$ meV, and some components of the nearest neighboring intra-strand hopping between two C bases changes by $10-60$ meV. An interesting observation is that the intra-strand hopping between two consecutive C decreases by a different amount after the introduction of water.

To summarize this section, our analysis of the orbital visualizations and TB parameters at the HOMO level for a series of short DNA oligomers with the B-form and A-form con-

figurations shows the effect of water on electrical conduction through DNA is conformation-dependent. The inclusion of water in B-form DNA makes the HOMO orbital more uniform through the entire strands (Figure 7.9), while in A-form DNA the HOMO shifts to the middle of the strands (Figure 7.11). For both B-form and A-form DNA, water can make the onsite energies of G and C bases more smooth and change some components of the hopping integrals. The inclusion of water increases the hopping between G and C within a base pair by about $20-30\ meV$ for B-form, and $10-20\ meV$ for A-form DNA. In addition, it increases the hopping between two consecutive C bases by about $60\ meV$ for B-form DNA. However, it can decrease the same quantity for A-form DNA by as much as $\sim 60\ meV$.

7.3.3 Effect of water on experimental DNA strands (Third set of DNA)

So far we have only considered the ideal short $G:C$ duplexes. In this section, we study the effect of water on two longer strands, $GC8$ and $B3$, which were used in two recent experiments [44, 211]. The HOMO orbital distributions for $GC8$ and $B3$ in the dry and hydrated states are shown in Figures 7.13 and 7.14.

Similar to the findings in the short $G1-G4$ strands with the B-form configuration, the HOMO orbitals for $GC8$ and $B3$ in hydrated state are more diffusive than the ones in the dry state. The main portion of the HOMO orbital for $GC8$ without water resides on the two G bases in the middle of the strand. After including the effect of water, the HOMO orbital extends to four more bases in the central region. We can make a similar statement for $B3$, but the shape of the orbital is more asymmetric. We infer that the symmetric distribution of the HOMO orbital for $GC8$ is due to their regular sequence.

To further our understanding, we extract the TB parameters at the HOMO level for $GC8$ and $B3$, as shown in Figures 7.15 and 7.16, with the gold electrodes depicted in yellow.

As seen in Figures 7.15 and 7.16, in both $GC8$ and $B3$, upon including the water, the onsite energies of G and C bases, and the inter-strand hopping integrals between the two nearest neighboring G and C bases have an appreciable change, but the inter-strand hopping integrals between the second nearest neighboring bases only undergo minor modifications.

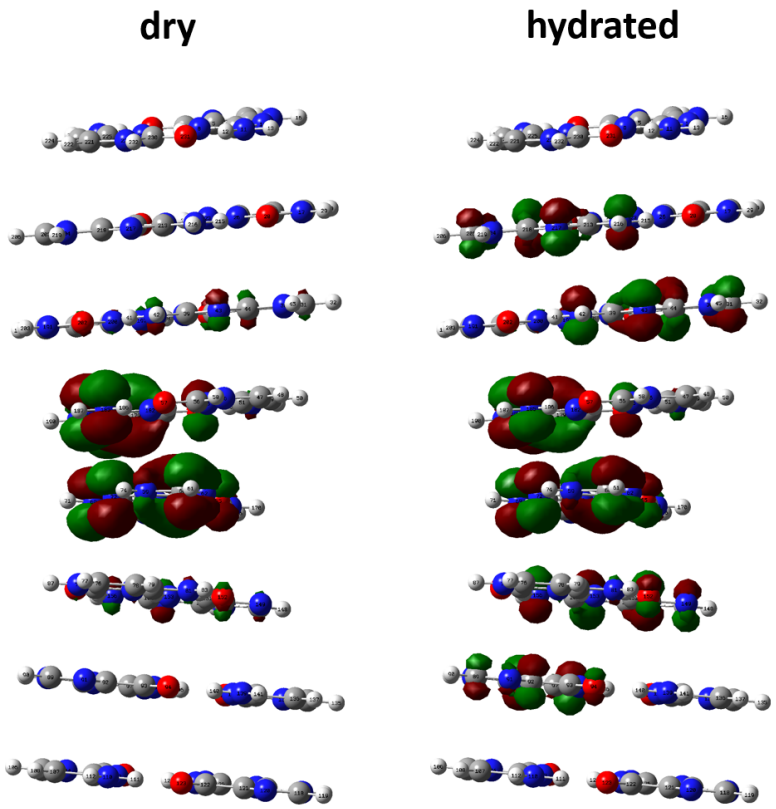


Figure 7.13: (Third set of DNA) HOMO distribution for *GC*8 in the dry state (left panel) and hydrated state (right panel). Red and green colors represent the positive and negative components of the wavefunctions. Iso-value = 0.02.

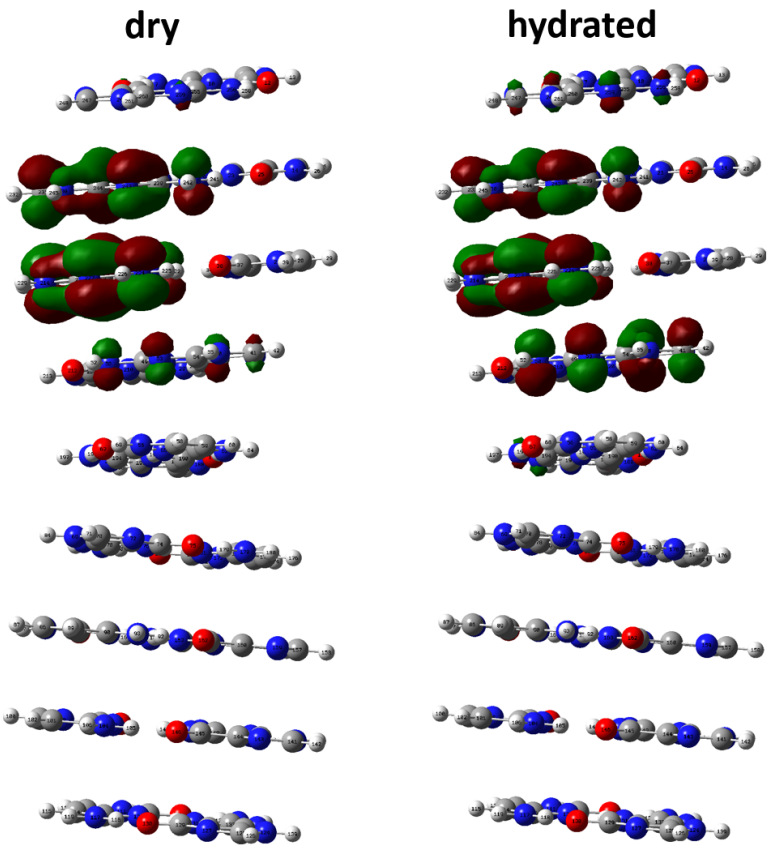


Figure 7.14: (Third set of DNA) HOMO distribution for $B3$ in the dry state (left panel) and hydrated state (right panel). Red and green colors represent the positive and negative components of the wavefunctions. Iso-value = 0.02.

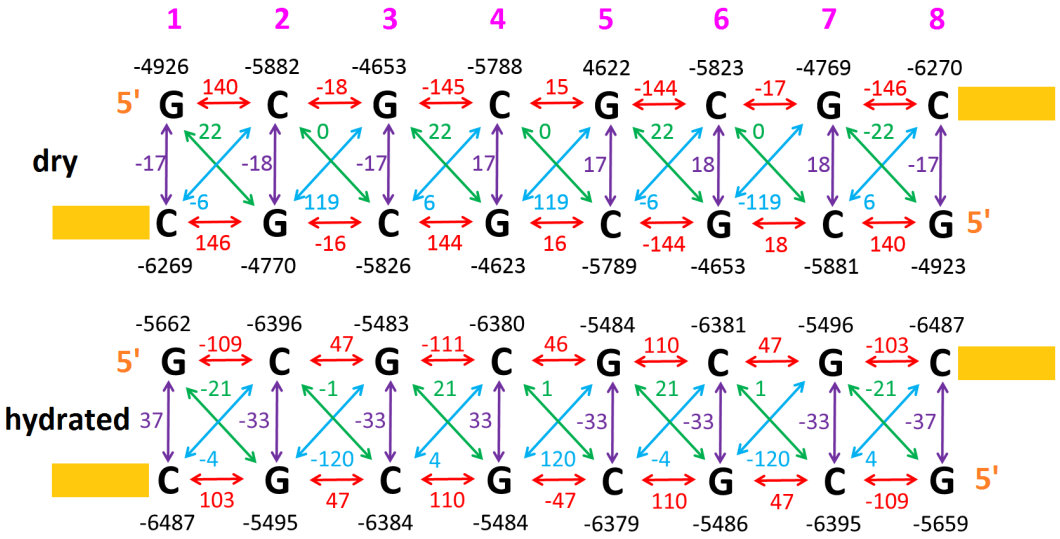


Figure 7.15: (Third set of DNA) TB parameters at the HOMO level for $GC8$ in the dry state (upper panel) and hydrated state (lower panel). The gold electrodes are depicted in yellow.

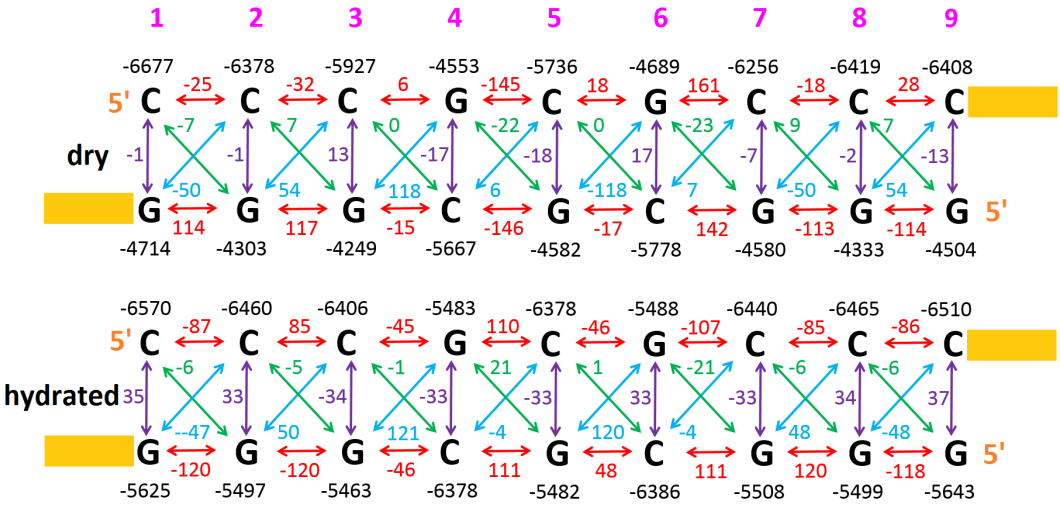


Figure 7.16: (Third set of DNA) TB parameters at the HOMO level for $B3$ in the dry state (upper panel) and hydrated state (lower panel). The gold electrodes are depicted in yellow.

This is similar to the observation in the TB models for $G1 - G4$ short oligomers. The responses of the intra-strand hopping integrals between the nearest neighboring bases to water in $GC8$ and $B3$ depends on the type of the bases and the orientation of the strand. In $GC8$, the intra-strand hopping integrals between the nearest neighboring bases G and C from the $5'$ end to $3'$ end decrease by about 30 meV after including water (the same holds true for the intra-strand hopping from the $3'$ end to $5'$ end). In $B3$, in the presence of water, the strengths of the intra-strand hopping integrals between two G bases stay similar and that for the two consecutive C bases increases by 60 meV. The intra-strand hopping integrals between the two nearest neighboring bases G and C changes by 30 meV, same as the observation in $GC8$.

Coupled with the findings in $G1 - G4$, we conclude that for a strand consisting of $G : C$ base pairs, the inclusion of water around the molecule can change the onsite energies of DNA bases and the inter-strand hopping integrals between the two nearest neighboring G and C bases by a large amount. The inter-strand hopping integral between the second nearest neighboring bases are not affected too much by the bulk water. Other components of the TB parameters depend on the type of the neighboring bases and the orientation of the strand. The overall tendency in TB parameters is that the variations in both the onsite energies and hopping-integrals are reduced with the addition of water.

The onsite energies of DNA bases and hopping integrals between the neighboring bases in the TB model are important quantities in determining the conducting properties of a DNA strand. In the hopping model for G-rich DNA strands, the G bases serve as the stepping stones. Therefore, the change in onsite energies of G bases and G-relevant hopping integrals due to the bulk water may be related to the charge transport properties modeled with the hopping model. We then calculate the transmission and conductance for $GC8$ and $B3$ in both dry and hydrated states. The results are shown in Figures 7.17 and 7.18.

Inspection of the transmission schemes for $GC8$ and $B3$ reveals that the addition of the water can enlarge the HOMO-LUMO gap of the molecule. Indeed, based on the DFT calculations, the HOMO-LUMO gap increases by 0.74 eV and 1.36 eV for $GC8$ and $B3$,

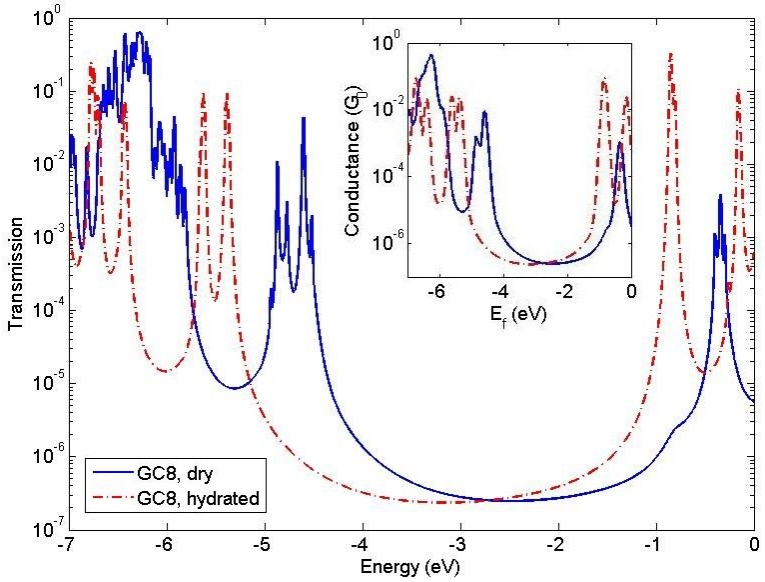


Figure 7.17: (Third set of DNA) Transmission for $GC8$ in the dry state (blue solid) and hydrated state (red dash). Inset, conductance for $GC8$ in the dry state (blue solid) and hydrated state (red dash).

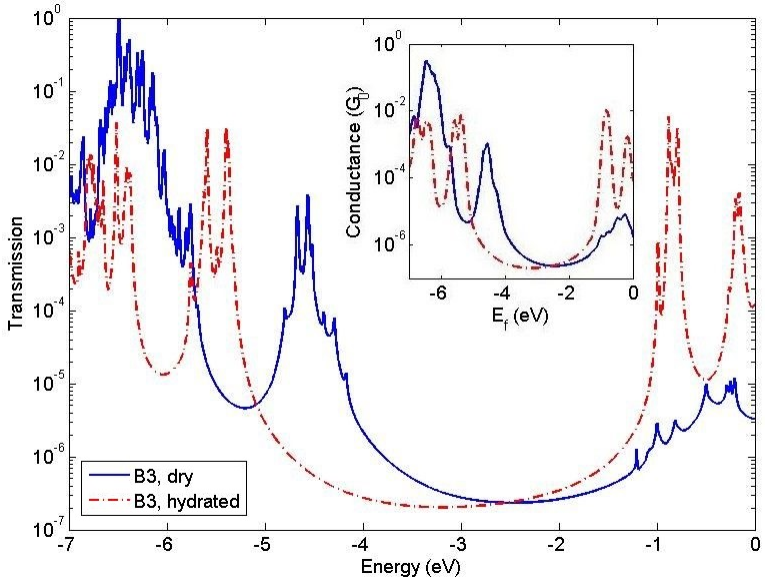


Figure 7.18: (Third set of DNA) Transmission for $B3$ in the dry state (blue solid) and hydrated state (red dash). Inset, conductance for $B3$ in the dry state (blue solid) and hydrated state (red dash).

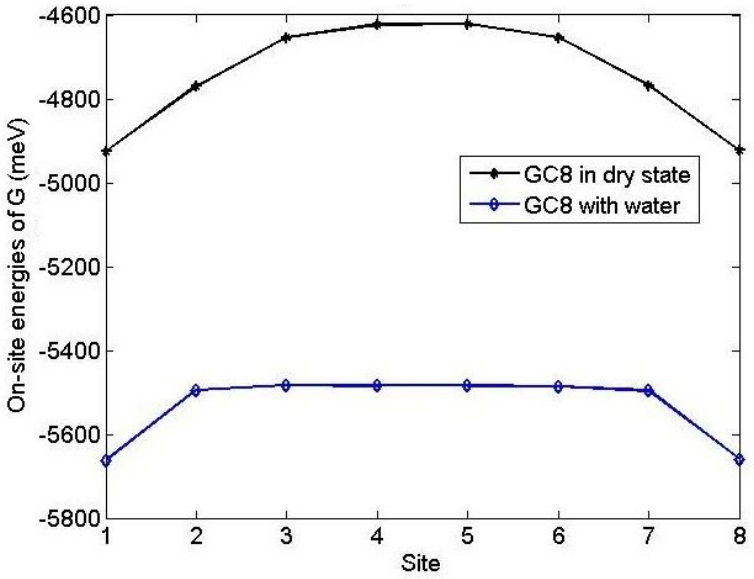


Figure 7.19: (Third set of DNA) Onsite energies of G bases in $GC8$ in the dry state (black) and hydrated state (blue).

respectively. The peak value of the transmission around HOMO levels for the strand in hydrated state increases by 2 (8) times for $GC8$ ($B3$), compared to the one in the dry state. Correspondingly, the peak value of the conductance around HOMO for the hydrated strands increases by 3 (8) times for $GC8$ ($B3$), with respect to the dry one, as shown in the inset of Figure 7.17 (Figure 7.18). We attribute the enhancement in the transmission and conductance of a hydrated strand to the more diffusive orbital overlap as observed previously.

The change in conducting properties of $GC8$ and $B3$ by bulk water can be understood better by a closer inspection of the TB parameters at the HOMO level, as shown in Figures 7.15 and 7.16. The TB parameters show that the strengths of the hopping integrals between two G bases are quite close before and after adding water, while the variation in the onsite energies is considerably reduced by the water. In order to quantitatively study the effect of the continuum solvent model, we plot the onsite energies of G bases in each base pair for $GC8$ and $B3$ in the dry state and with bulk water, as shown in Figures 7.19 and 7.20.

The lower values of at the two ends in both the dry and hydrated states for $GC8$ and $B3$

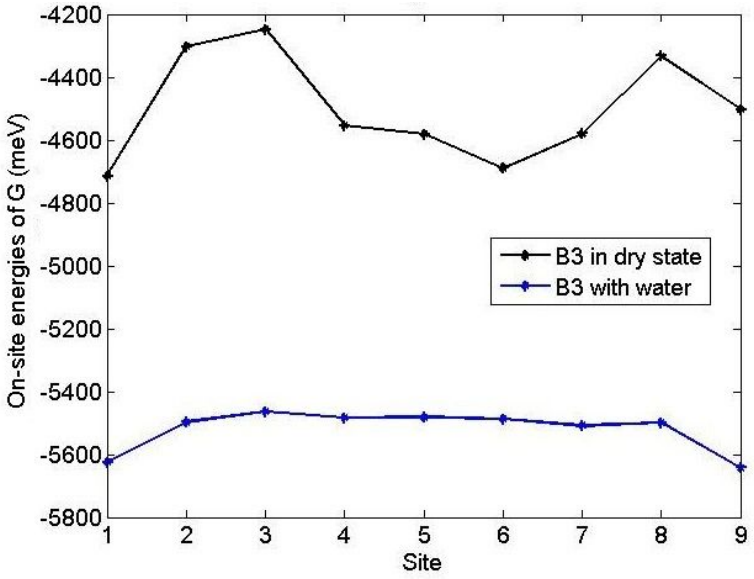


Figure 7.20: (Third set of DNA) Onsite energies of G bases in $B3$ in the dry state (black) and hydrated state (blue).

are due to the edge effect. It is evident that the onsite energies of each G base for $GC8$ and $B3$ in hydrated state are significantly smaller than the corresponding ones without water, as described before. More importantly, in the strand without water, the distribution of the onsite energies of G bases are over a wider range of energies for both $GC8$ and $B3$. After including water, the distribution is smoother. The standard derivation of onsite energies of G bases for $GC8$ without water is 127 meV, compared to 80 meV in the case in the hydrated state. Similarly, the standard derivation of onsite energies of G bases for $B3$ without water is as much as 169 meV, while with the presence of water, the standard derivation is only 60 meV. Using a much more complicated method, the work in Cuniberti's group draw a similar observation [89]. The larger change in the standard derivation of the G bases for $B3$ after including water corresponds to the larger increase (8 times) in conductance.

7.4 Summary

In this chapter, we numerically study the effect of water on the electronic structure and electrical conduction of DNA. We consider three sets of DNA strands with the purpose of studying the effect of water on transport in a somewhat detailed manner. To simulate the the effect of water, we employ three solvation models - adding three explicit water molecules around DNA, using the pure PCM, and combining the PCM with three explicit water molecules. We first compare three solvation models using a four-base pair DNA strand with backbone (first set of DNA). We find that the PCM model is the best compromise. We also examine the effect of water on different DNA configurations by comparing the HOMO orbital distribution and TB parameters for short DNA oligomers with the B-form and A-form configurations (second set of DNA). For both configurations, we are able to demonstrate that the presence of water makes the HOMO orbitals more diffusive (for A-form DNA, it is more centered on the middle of the strands) and also changes the onsite energies of the bases and some component of the hopping integrals in TB parameters at the HOMO level. The inclusion of water not only makes the onsite energies of the bases more smooth, but also increases the inter-strand hopping between G and C in complementary strands by $20-30\ meV$ for B-form and $10-20\ meV$ for A-form DNA. Depending on the conformation, it has different effects on the intra-strand hopping between consecutive C bases. The intra-strand hopping between two consecutive C bases for a B-form DNA increase by $60\ meV$. The same quantity can decrease by $\sim 60\ meV$ for an A-form DNA. We also study the influence of water on two longer strands which were used in two recent experiments (third set of DNA), and find that the addition of bulk water can facilitate the charge transport through DNA. In fact, water can increase the peak values of conductance by 3 to 8 times, depending on the sequence of the strand. This enhancement can be rationalized by the reduction in variation of the onsite energies for the G bases with the addition of water.

Chapter 8

CHARGE TRANSPORT PROPERTIES OF RNA:DNA HYBRIDS

8.1 Introduction

As our understanding of the electrical transport in DNA is getting a little more mature, research exploring the electronic properties of RNA and PNA (peptide nucleic acid) is also beginning to be pursued [212, 213, 214].

The study of RNA has been an active topic during the last few decades mainly because of its essential importance in many biological processes. RNA can be classified into a few types based on its functions, such as the messenger RNA (mRNA), the transfer RNA (tRNA), the ribosomal RNA (rRNA) and the transfer-messenger RNA (tmRNA). RNA plays a vital role in many activities, including gene transportation, gene expression, regulating chemical reactions in cells and immunological responses of various organisms [215, 216].

The structure of RNA has been described in Chapter 1. Here, we review a new structure - the RNA:DNA hybrid, which is formed by a single RNA strand with a single DNA strand. The hybrid structure we study in this chapter is an A-form right-handed helix, similar to an A-form dsDNA. The RNA:DNA hybrids have been found to be important in many biological activities and systems, including DNA replication, transcription during gene expression and reverse transcription [217, 218, 219, 220, 221].

In contrast to the intensive study of the electrical conduction in DNA, we have little knowledge about the electronic properties and the potential application in electronics made of RNA and RNA:DNA hybrids. Given the structural similarities of these molecules to that of DNA, it is worth devoting effort to the electronic properties of RNA and RNA:DNA hybrids.

Recently, there has been a proposal to detect disease by observing the change in mRNA's electrical properties [212]. Since mRNA plays the role of carrying information required to synthesize protein from DNA to the ribosome, the change in mRNA can be important biomarkers for genetic diseases such as cancer [212]. Thus, the detection of RNA may reveal a wealth of information for disease diagnosis. Currently, the detecting methods have focused on PCR and fluorometric detection [214]. Meanwhile, the use of the electrical and electrochemical readouts is also attracting more and more attention from researchers, because they have the features of high selectivity, high sensitivity, high speed, low detection limit and low cost [213, 214]. The direct conductance measurements through single nucleotides show that the conductances of RNA nucleotides are higher than the corresponding DNA nucleotides [36], indicating RNA may be used in designing molecular devices that require higher conductivity. This reveals that RNA may be relevant to bio-electronics.

Motivated by these findings, we study the charge transport through a series of RNA:DNA hybrids in this chapter. We find that the conductance of each RNA:DNA strand can be higher than that of the B-form dsDNA counterpart in certain energy windows. We extract the length-dependent exponential decay constants β and observe that the decay constant of RNA:DNA is higher than that of DNA. The different β values between RNA:DNA hybrids and DNA are rationalized by the binding energy difference between these two types of molecules.

8.2 Strands

The four RNA:DNA hybrid strands we study consist of only $G:C$ base pairs, with the following sequences on the RNA side: $5'-GGGCGCGGG-3'$ (referred to as $R3$, with the DNA complement $3'-CCCGCGCCC-5'$); $5'-GGGCGCGCGGG-3'$ (referred to as $R5$, with the DNA complement $3'-CCCGCGCGCCC-5'$); $5'-GGGCGCGCGCGGG-3'$ (referred to as $R7$, with the DNA complement $3'-CCCGCGCGCGCCC-5$') and $5'-GGGCGCGCGCGCGGG-3'$ (referred to as $R9$, with the DNA complement $3'-CCCGCGCGCGCGCCC-5'$). In the experiment, the $3'$ and $5'$ ends of the DNA strands

Figure 8.1: Sketch of the 11-mer RNA:DNA hybrid, $R5$, with sequence on the RNA side, $5' - GGGCGCGCGGG - 3'$ (the sequence of the DNA complement is $3' - CCCGCGCGCCC - 5'$). Also shown are the diamine linkers at the $3'$ and $5'$ ends of the DNA complement.

are modified with the diamine linkers to build contact between the hybrids and the gold electrodes in the STM junction. The sketch of the 11-mer RNA:DNA hybrid, $R5$, with the diamine linkers, is shown in Figure 8.1.

The initial coordinates of the atoms are produced with the NAB software package. To construct the RNA:DNA hybrid, we first generate an ideal A-form DNA structure and replace the hydrogen atom in the 2' position of the sugar ring with an OH group in one strand. We then relax the hydroxyl groups at the B3LYP/6-31G(d) level to optimize the sugar ribose ring. To neutralize the negatively charged backbone, we add the sodium counterions around the phosphate groups. The positions of the sodium counterions are determined by optimizing the sodium counterions in a 3-base single strand at the B3LYP/6-31G(d) level. In Figure 8.2, we show the structure of the 11-mer RNA:DNA hybrid, $R5$, with the RNA and DNA strands marked in red and blue, respectively.

In the next step, we perform the DFT calculations with the B3LYP/6-31G(d,p) XC functional and basis set. Transport calculations for both the coherent and decoherent cases follow the methods in Chapter 2. We also study the B-form dsDNA, $B3$, $B5$, $B7$ and $B9$, with the same sequences as that of $R3$, $R5$, $R7$ and $R9$. The results of the B-form DNA are used to compare with the findings in the RNA:DNA hybrids.

8.3 Result

The HOMOs, LUMOs and gaps for the B-form dsDNA and A-form ds-RNA:DNA hybrids are shown in Tables 8.1 and 8.2. For both the dsDNA and the hybrids, the value of the

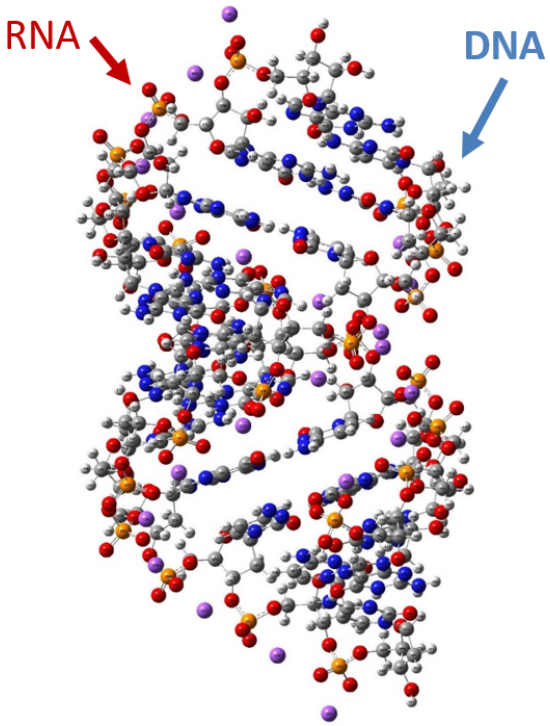


Figure 8.2: Structure of the 11-mer RNA:DNA hybrid, $R5$, with sequence on the RNA side $5' - GGGCGCGCGGG - 3'$ (the sequence of the DNA complement is $3' - CCCGCGCGCCC - 5'$). The RNA and DNA strands are marked in red and blue, respectively.

HOMOs, LULMOs and gaps are similar with different lengths.

Table 8.1: HOMO, LUMO and gap for the B-form dsDNA computed with B3LYP/6-31G(d,p) (Unit, eV).

Quantities	$B3$	$B5$	$B7$	$B9$
HOMO	-3.079	-3.043	-3.028	-3.023
LUMO	-1.956	-1.953	-1.944	-1.936
Gap	1.123	1.090	1.085	1.088

Table 8.2: HOMO, LUMO and gap for the A-form RNA:DNA computed with B3LYP/6-31G(d,p) (Unit, eV).

Quantities	$R3$	$R5$	$R7$	$R9$
HOMO	-4.163	-4.233	-4.267	-4.286
LUMO	-3.111	-3.135	-3.161	-3.180
Gap	1.053	1.098	1.106	1.106

8.3.1 Importance of contact coupling

In Chapters 3 and 5, we have seen the importance of the contact coupling in the charge transport through DNA. In this chapter, we investigate the significance of the contact coupling in charge transport through the RNA:DNA hybrids. The conductance value v.s. the Fermi level of $R3$ within a broad range of Γ value is shown in Figure 8.3.

When the Γ value is less than 5 eV, the conductance values depend on the contact coupling strength sensitively - a small change in Γ can cause the conductance to change significantly. However, as the Γ value is approaching a value that is larger than 5 eV, the conductance is less sensitive to the change in Γ value. To best fit the experiment while rationalizing the physics of the gold electrode, we choose the Γ value to be 5 eV.

8.3.2 Binding energy of dsDNA and RNA:DNA hybrids

We then calculate the hydrogen binding energies for both dsDNA and the RNA:DNA hybrids with various lengths, as shown in Figure 8.4.

As the length of the strand increases, the binding energy also increases. This is true for both dsDNA and the RNA:DNA hybrids, indicating that the longer strands are more stable than the short ones, and have less fluctuations when placed in solution. Also, the binding energy of the dsDNA is larger than that of the DNA:RNA hybrid with the same sequence by about 35 meV. This suggests that each dsDNA is more stable than the corresponding DNA:RNA hybrid counterpart. The observation here agrees with a recent experiment which

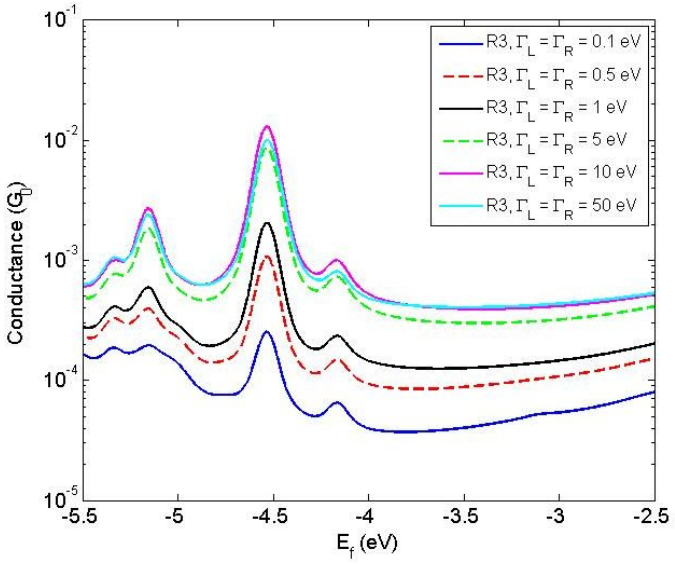


Figure 8.3: Conductance v.s. Fermi level of $R3$ with different Γ values. Decoherence rate is set to be 30 meV for all the Γs.

shows that DNA has a higher melting temperature than the RNA:DNA hybrid [222]. This information is useful for choosing the decoherence rate, because decoherence has a contribution from vibrations. The more stable the molecule is, the less vibrations it suffers from. With these benchmarks, we make the assumption that a longer strand has a smaller decoherence rate, and DNA has a smaller decoherence rate than RNA with the same sequence. These statements seem plausible but a more rigorous calculation to verify them would be useful.

8.3.3 Charge transport in RNA:DNA hybrids

In this section, we look at the transport properties of the RNA:DNA hybrids and compare them with that of dsDNA. In the previous study of DNA, we have demonstrated the importance of including decoherence in the calculation to rationalize the experimental observations. While decoherence plays an important role in charge transport through molecules, the precise value of the decoherence rate is difficult to choose. Previously (Chapters 4, 5

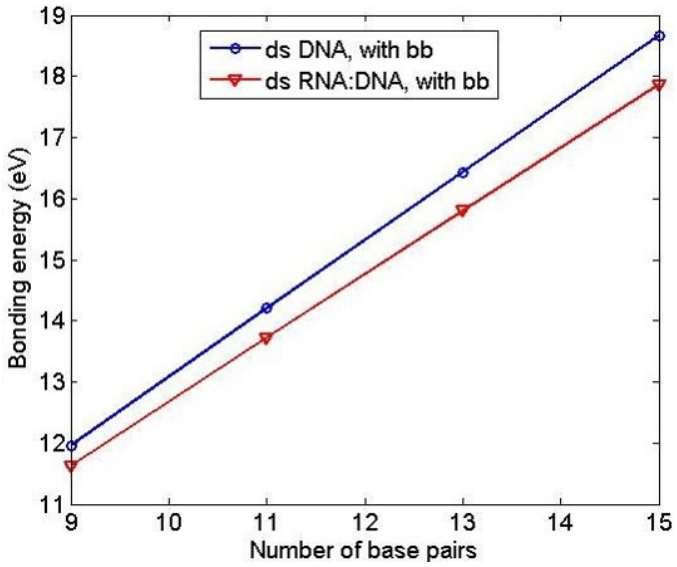
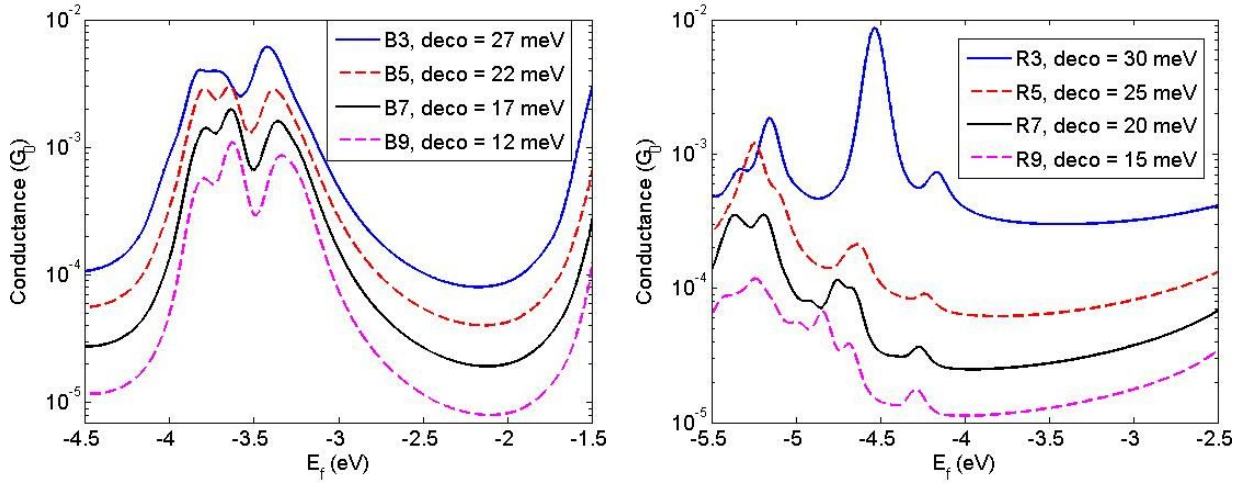


Figure 8.4: Binding energy for dsDNA, $B3$, $B5$, $B7$, $B9$, and the RNA:DNA hybrids, $R3$, $R5$, $R7$, $R9$. The binding energy increases with the increasing length. Also, the binding energy of each DNA is larger than that of the RNA:DNA hybrid with the same sequence.

and 6), we chose the decoherence rates in dsDNA to have values comparable to other low dimensional solid state systems, such as carbon nanotubes, so as to get a qualitative match with experiments. Here, we find that to obtain a qualitative match with experiments, the choice of the decoherence rate on the G and C bases have to be binding energy dependent (a plausibility argument for this was given earlier).

We further find that in current set of calculations, we need to use the decoherence rates that are higher than what we used to model the experiment discussed in Chapter 4. This is because the experiment we modeled in Chapter 4 is based on the MCBJ technique. During the measurement, the DNA molecule sits on a substrate, which provides interactions to make them more stable. However, in the STM-break junction technique, the molecule is pulled out from the solution with the tip. In this process, most of the molecules are suspended between the tip and the substrate. Thus, we expect the decoherence rates in this experiment to be larger. We assume the values to be $2-5$ times higher than we previously used for dry DNA in Chapter 4.



(a) (b)

Figure 8.5: Conductance for (a) B-form dsDNA; and (b) A-form DNA:RNA hybirds with different decoherence rates. $\Gamma = 5\ eV$.

Also, in order to match the experiments, we assume that the decoherence rate depends on the length of strands. The motivation for this is that the longer strands are more stable, as discussed in Section 8.3.2. In addition, the decoherence rates for DNA are larger than that of RNA:DNA with the same sequence, because DNA are more stable than the corresponding RNA:DNA hybrids, as shown in the binding energy plot Figure 8.4.

With the above considerations, we set decoherence rates to be 27 meV for $B3$, 22 meV for $B5$, 17 meV for $B7$, 12 meV for $B9$, 30 meV for $R3$, 25 meV for $R5$, 20 meV for $R7$, and 15 meV for $R9$. The conductance results are shown in Figure 8.5.

The trend in the conductance from our modeling matches the experiment nicely. In the HOMO-LUMO gap, the conductance of RNA:DNA is larger than that of dsDNA hybrid in a certain energy range, which is consistent with the experimental measurement. The values of the decay constant β for DNA and the RNA:DNA hybrid are 0.105 $\mathring{A}^{-1}$ and 0.251 $\mathring{A}^{-1}$ in our modeling. This is in qualitative agreement with the experiments (0.196 $\mathring{A}^{-1}$ for DNA and 0.314 $\mathring{A}^{-1}$ for the RNA:DNA hybrid). The comparison between the theoretical model and the experimental observations suggests the importance of choosing different decoherence

rate for strands with different lengths and structures.

8.4 Conclusion

In this chapter, we study the electrical conduction through a series of DNA:RNA hybrids made of RNA strands with DNA strands. The choice of the contact coupling is very important to rationalize the experimental result. We also calculate the binding energy for the dsDNA and RNA:DNA hybrids and find that as the length of the strand increases, the binding energy also increases. In addition, the DNA has a larger binding energy than the RNA:DNA hybrid with the same sequence. These observations help in choosing the decoherence rates for strands with different lengths and structures. By considering the details of the experiment involving the STM-break junction, we set the decoherence rates to be 27 meV for $B3$, 22 meV for $B5$, 17 meV for $B7$, 12 meV for $B9$, 30 meV for $R3$, 25 meV for $R5$, 20 meV for $R7$ and 15 meV for $R9$. With this sets of decoherence rates, the β value is 0.105 $\mathring{A}^{-1}$ for dsDNA and 0.251 $\mathring{A}^{-1}$ for RNA:DNA hybrid, qualitatively agreeing with the experiments. Note that conductance values for the RNA:DNA hybrids from our simulation have difficulty in getting matches with the experiment, suggesting that other mechanisms may be involved in the electrical conduction through these molecules.

Chapter 9

SUMMARY AND FUTURE WORK

9.1 Summary

In this thesis, we systematically model the electrical transport through a series of nucleic acids using DFT and the Landauer-Büttiker formalism. The topics covered in this thesis are summarized below.

In Chapter 1, we review the physical structure of DNA and RNA, and the motivation for the study of their electronic properties. The current research status in this field, and the challenges for electrical transport modeling of DNA electronics are given.

In Chapter 2, we present a detailed description of the methods and models that we use. We review a few key concepts in *ab initio* calculations. We also establish the scheme of extracting the TB parameters (onsite energies and hopping integrals) for a model Hamiltonian. We then review the Landauer-Büttiker formalism, which is used for the transmission and conductance computations. Green's function approach is also briefly reviewed, based on a two-probe system. Specifically, a decoherent transport model, which involves the implementation of Büttiker probes based on the D'Amato-Pastawski model, is presented in detail. Matlab code for the computation of transmission using both coherent and decoherent models is provided in Appendix A.

In Chapter 3, we explore the possibility of using DNA as simple device building blocks by studying the transmission and conductance through a set of B-form DNA heterostructures and superlattices in both single-stranded and double-stranded configurations. We observe that T base acts as a barrier in single-stranded DNA heterostructures $poly\{G\}(T)_N poly\{G\}$. As the number of T increases, the transmission in a superlattice made of $poly\{G\}(T)_N poly\{G\}$ decreases exponentially. One more T causes the transmission of the heterostructure to drop

by 72 times. Minibands with a width of 40 meV and 3 μeV are found in $poly\{GT\}$ and $poly\{GGTT\}$. In the double-stranded structures $poly\{(G:C)\}(A:T)_N poly\{(G:C)\}$, $A:T$ base pair is found to be a barrier. Miniband with a width of 10 meV is observed in the supperlattice consisting of $poly\{(G:C)(A:T)\}$.

In Chapter 4, we develop a unified model based on Green's function picture that can rationalize the measured conductance of four DNA strands consisting of $G:C$-rich base pairs and a few $A:T$ base pairs in the central region. We find that decoherence on $G:C$ base pairs is crucial in getting agreement with the experiments. However, the decoherence on $G:C$ base pairs alone does not explain the experimentally determined dependence of conductance in strands containing a number of $A:T$ base pairs. Including decoherence on $A:T$ base pairs is also essential. By fitting the experimental magnitudes of the conductance for the four DNA strands, we estimate that the deocherence rate is 6 meV for $G:C$ and 1.5 meV for $A:T$ base pairs.

In Chapter 5, we investigate the charge transport through an eight-base pair methylated DNA strand and its native counterpart. Our results show that with the same choice of contact coupling, in the phase-coherent limit, the transmission of the methylated strand is smaller in the bandgap at energies close to the HOMO, while inside the HOMO band, the transmission is oscillatory and the methylated DNA may have a larger transmission in certain energy windows. The trend in transmission also holds in the presence of the decoherence though there is a crossover in the transmission of the native and methylated strands away from the HOMO level. We then re-do the DFT calculations using the CAM-B3LYP based range-separated functional, and find that our transport results are insensitive to the choice of XC funtionals. The introduction of the methyl groups can stabilize the molecular structure of DNA, which leads to smaller decoherence rates. This may be the reason for the lower conductance of the methylated DNA in the experiment.

In Chapter 6, we study the conformation-dependent conductance of DNA. We rationalize the conductance difference between a series of B-form DNA and A-form DNA by studying the role of backbone. The HOMO orbital distribution of an A-form DNA is more diffusive

than that of a corresponding B-form DNA with the same sequence, making it easier for the electrons to flow. The inclusion of backbone in the transmission calculation for an A-form DNA can cause the transmission to increase by a few orders of magnitude, which is not seen in the B-form DNA. The comparison of the DOS and the projected HOMO along the strand between a B-form DNA and A-form DNA additionally point to the fact that the backbone plays a crucial role in the electrical transport through an A-form DNA. The TB Hamiltonians for both the B-form and A-form DNA are also developed in this chapter.

In Chapter 7, we analyze the effect of water on the electrical conduction through DNA. Three different solvation models - adding a few explicit water molecules around DNA, employing the PCM to simulate the bulk water, and combining the explicit water molecules with the PCM - are used to model the effect of water. We compare the three solvation models using a four-base pair B-form DNA strand with backbone and find that water has influence on the structure, charge distribution, energy levels and transmission of DNA. To model the effect of water, the PCM is the best comprise. We then use the PCM to study the effect of water on a series of short DNA strands without backbone in B-form and A-form configurations. We compare the HOMO orbital distribution and the TB parameters for short B-form and A-form stacking bases in both gas phase and bulk water, and find that the effect of water on DNA depends on the conformation. We also study the impact that water has on DNA strands that were used in the experiments, and observe that water can increase the conductance of these strands by $3 - 8$ times, depending on the sequence.

In Chapter 8, we study the charge transport through a series of RNA:DNA hybrids. We find that the binding energy of the strands increases with the increasing length. Also, the binding energy of the dsDNA is higher than that of the RNA:DNA hybrid with the same sequence. By choosing proper values of decoherence rates for strands with different lengths and structures, we are able to obtain a decay constant whose value qualitatively matches with the experiment.

9.2 Future work

In this thesis, we mainly use two models - the coherent transport model and the decoherent transport model, to study electrical conduction in nucleic acids. There are three open variables in the models we use - the Fermi energy, the contact coupling and the decoherence rate in the decoherent model. The determination of the Fermi energy is difficult, as the Fermi energy can easily change with small modifications of the surface conditions near the molecule-electrode interface. The modeling of the contact coupling and the precise choice of decoherence rate are also very challenging problems. However, there is room to make improvement in the current model.

In order to model the contact effect of the gold electrode used in the conductance measurement, we use the WBL by assuming the self-energies are constants. To improve this, one can either use an extended molecular system which includes a few layers of electrode atoms. This is a very popular method in the coherent model, and can be extended into the decoherent model. Alternatively, we can do *ab intio* calculations on a short electrode layer and get the TB parameters for the electrode material. With the TB parameters, one can fold the effect of the semi-infinite electrodes into the two sides of the molecule using surface Green's function.

In this thesis, we have demonstrated that the inclusion of decoherence is essential to bring the computed conductance values close to the experiments. A purely coherent model yields much lower transmission when compared to the measurements. To choose the value of the decoherence rate, we refer to values used in solid state systems such as carbon nanotubes, and compare the resulting conductance with experiment. Since the location of the Fermi level is difficult to determine, there is a certain degree of freedom when choosing the value of decoherence rate. Methods that theoretically determine the decoherence rate would be a beneficial direction to further unravel the mechanisms behind the electrical conduction in nucleic acids.

BIBLIOGRAPHY

[1] *https* : *//en.wikipedia.org/wiki/Nucleic_acid.*

[2] *http* : *//www.scienceprofonline.com/images/DNA_chemical_structure.jpg.*

[3] J. D. Watson, F. H. Crick, *et al.*, "Molecular structure of nucleic acids," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[4] C. Dekker and M. A. Ratner, "Electronic properties of DNA," *Physics World*, vol. 14, no. 8, pp. 29–33, 2001.

[5] D. Porath, N. Lapidot, and J. Gomez-Herrero, "Charge transport in DNA-based devices," in *Lecture Notes in Physics, Berlin Springer Verlag*, vol. 680, pp. 411–444, 2005.

[6] R. Endres, D. Cox, and R. Singh, "Colloquium: The quest for high-conductance DNA," *Reviews of Modern Physics*, vol. 76, no. 1, p. 195, 2004.

[7] J. M. Tour, "Molecular electronics. synthesis and testing of components," *Accounts of Chemical Research*, vol. 33, no. 11, pp. 791–804, 2000.

[8] A. Aviram and M. A. Ratner, "Molecular rectifiers," *Chemical Physics Letters*, vol. 29, no. 2, pp. 277–283, 1974.

[9] N. Zhitenev, H. Meng, and Z. Bao, "Conductance of small molecular junctions," *Physical Review Letters*, vol. 88, no. 22, p. 226801, 2002.

[10] F. Chen, J. Hihath, Z. Huang, X. Li, and N. Tao, "Measurement of single-molecule conductance," *Annu. Rev. Phys. Chem.*, vol. 58, pp. 535–564, 2007.

[11] Y. Kim, H. Song, F. Strigl, H.-F. Pernau, T. Lee, and E. Scheer, "Conductance and vibrational states of single-molecule junctions controlled by mechanical stretching and material variation," *Physical Review Letters*, vol. 106, no. 19, p. 196804, 2011.

[12] S. Yasutomi, T. Morita, Y. Imanishi, and S. Kimura, "A molecular photodiode system that can switch photo-current direction," *Science*, vol. 304, no. 5679, pp. 1944–1947, 2004.

[13] J. H. Schön, H. Meng, and Z. Bao, "Self-assembled monolayer organic field-effect transistors," *Nature*, vol. 413, no. 6857, pp. 713–716, 2001.

[14] M.-K. Ng, D.-C. Lee, and L. Yu, "Molecular diodes based on conjugated diblock cooligomers," *Journal of the American Chemical Society*, vol. 124, no. 40, pp. 11862–11863, 2002.

[15] Y. Chen, D. A. Ohlberg, X. Li, D. R. Stewart, R. S. Williams, J. O. Jeppesen, K. A. Nielsen, J. F. Stoddart, D. L. Olynick, and E. Anderson, "Nanoscale molecular-switch devices fabricated by imprint lithography," *Applied Physics Letters*, vol. 82, no. 10, pp. 1610–1612, 2003.

[16] D. Eley and D. Spivey, "Semiconductivity of organic substances. Part 9.–Nucleic acid in the dry state," *Transactions of the Faraday Society*, vol. 58, pp. 411–415, 1962.

[17] I. Saito, M. Takayama, H. Sugiyama, K. Nakatani, A. Tsuchida, and M. Yamamoto, "Photoinduced DNA cleavage via electron transfer: demonstration that guanine residues located 5' to guanine are the most electron-donating sites," *Journal of the American Chemical Society*, vol. 117, no. 23, pp. 6406–6407, 1995.

[18] S. O. Kelley and J. K. Barton, "Radical migration through the DNA helix: chemistry at a distance.," *Metal Ions in Biological Systems*, vol. 36, pp. 211–249, 1999.

[19] B. Giese, "Long-distance charge transport in DNA: the hopping mechanism," *Accounts of Chemical Research*, vol. 33, no. 9, pp. 631–636, 2000.

[20] F. D. Lewis, X. Liu, S. E. Miller, R. T. Hayes, and M. R. Wasielewski, "Dynamics of electron injection in DNA hairpins," *Journal of the American Chemical Society*, vol. 124, no. 38, pp. 11280–11281, 2002.

[21] G. B. Schuster, "Long-range charge transfer in DNA: transient structural distortions control the distance dependence," *Accounts of Chemical Research*, vol. 33, no. 4, pp. 253–260, 2000.

[22] B. Giese, J. Amaudrut, A.-K. Köhler, M. Spormann, and S. Wessely, "Direct observation of hole transfer through DNA by hopping between adenine bases and by tunnelling," *Nature*, vol. 412, no. 6844, pp. 318–320, 2001.

[23] M. A. O'Neill and J. K. Barton, "Effects of strand and directional asymmetry on base–base coupling and charge transfer in double-helical DNA," *Proceedings of the National Academy of Sciences*, vol. 99, no. 26, pp. 16543–16550, 2002.

[24] F. D. Lewis, H. Zhu, P. Daublain, T. Fiebig, M. Raytchev, Q. Wang, and V. Shafirovich, "Crossover from superexchange to hopping as the mechanism for photoinduced charge transfer in DNA hairpin conjugates," *Journal of the American Chemical Society*, vol. 128, no. 3, pp. 791–800, 2006.

[25] F. D. Lewis, H. Zhu, P. Daublain, B. Cohen, and M. R. Wasielewski, "Hole mobility in DNA a tracts," *Angewandte Chemie International Edition*, vol. 45, no. 47, pp. 7882–7985, 2006.

[26] J. D. Slinker, N. B. Muren, S. E. Renfrew, and J. K. Barton, "DNA charge transport over 34 nm," *Nature Chemistry*, vol. 3, no. 3, pp. 228–233, 2011.

[27] C. Adessi, S. Walch, and M. Anantram, "Environment and structure influence on DNA conduction," *Physical Review B*, vol. 67, no. 8, p. 081405, 2003.

[28] B. Göhler, V. Hamelbeck, T. Markus, M. Kettner, G. Hanne, Z. Vager, R. Naaman, and H. Zacharias, "Spin selectivity in electron transmission through self-assembled monolayers of double-stranded DNA," *Science*, vol. 331, no. 6019, pp. 894–897, 2011.

[29] Z. Xie, T. Z. Markus, S. R. Cohen, Z. Vager, R. Gutierrez, and R. Naaman, "Spin specific electron conduction through DNA oligomers," *Nano Letters*, vol. 11, no. 11, pp. 4652–4655, 2011.

[30] N. Matsuo, S. Takagi, K. Yamana, A. Heya, T. Takada, and S. Yokoyama, "Electrical property of DNA field-effect transistor: Charge retention property," *Japanese Journal of Applied Physics*, vol. 51, no. 4S, p. 04DD13, 2012.

[31] B. Chakraborty, R. Sha, and N. C. Seeman, "A DNA-based nanomechanical device with three robust states," *Proceedings of the National Academy of Sciences*, vol. 105, no. 45, pp. 17245–17249, 2008.

[32] M. McCullagh, I. Franco, M. A. Ratner, and G. C. Schatz, "Defects in DNA: Lessons from molecular motor design," *The Journal of Physical Chemistry Letters*, vol. 3, no. 6, pp. 689–693, 2012.

[33] N. Kang, A. Erbe, and E. Scheer, "Observation of negative differential resistance in DNA molecular junctions," *Applied Physics Letters*, vol. 96, no. 2, p. 023701, 2010.

[34] P. A. Sontz, T. P. Mui, J. O. Fuss, J. A. Tainer, and J. K. Barton, "DNA charge transport as a first step in coordinating the detection of lesions by repair proteins," *Proceedings of the National Academy of Sciences*, vol. 109, no. 6, pp. 1856–1861, 2012.

[35] K. Kim, M. Yoon, J. Koo, and Y. Roh, "Electrical characteristics of oxygen doped DNA molecules," *Thin Solid Films*, vol. 519, no. 20, pp. 7057–7059, 2011.

[36] T. Ohshiro, K. Matsubara, M. Tsutsui, M. Furuhashi, M. Taniguchi, and T. Kawai, "Single-molecule electrical random resequencing of DNA and RNA," *Scientific Reports*, vol. 2, p. 501, 2012.

[37] H. W. C. Postma, "Rapid sequencing of individual DNA molecules in graphene nanogaps," *Nano Letters*, vol. 10, no. 2, pp. 420–425, 2010.

[38] P. Xie, Q. Xiong, Y. Fang, Q. Qing, and C. M. Lieber, "Local electrical potential detection of DNA by nanowire-nanopore sensors," *Nature Nanotechnology*, vol. 7, no. 2, pp. 119–125, 2012.

[39] E. A. Manrao, I. M. Derrington, M. Pavlenok, M. Niederweis, and J. H. Gundlach, "Nucleotide discrimination with DNA immobilized in the MspA nanopore," *PLoS One*, vol. 6, no. 10, p. e25723, 2011.

[40] E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, and J. H. Gundlach, "Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase," *Nature Biotechnology*, vol. 30, no. 4, pp. 349–353, 2012.

[41] A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, and J. H. Gundlach, "Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA," *Proceedings of the National Academy of Sciences*, vol. 110, no. 47, pp. 18904–18909, 2013.

[42] J. Shim, G. Humphreys, B. Venkatesan, J. Munz, X. Zou, C. Sathe, K. Schulten, F. Kosari, A. Nardulli, G. Vasmatzis, *et al.*, "Detection and quantification of methylation in DNA using solid-state nanopores.," *Scientific Reports*, vol. 3, p. 1389, 2013.

[43] M. Tsutsui, K. Matsubara, T. Ohshiro, M. Furuhashi, M. Taniguchi, and T. Kawai, "Electrical detection of single methylcytosines in a DNA oligomer," *Journal of the American Chemical Society*, vol. 133, no. 23, pp. 9124–9128, 2011.

[44] J. Hihath, S. Guo, P. Zhang, and N. Tao, "Effects of cytosine methylation on DNA charge transport," *Journal of Physics: Condensed Matter*, vol. 24, no. 16, p. 164204, 2012.

[45] P. Tan Bui, T. Nishino, H. Shiigi, and T. Nagaoka, "One-by-one single-molecule detection of mutated nucleobases by monitoring tunneling current using a DNA tip," *Chemical Communications*, vol. 51, no. 9, pp. 1666–1669, 2015.

[46] H. Mehrez and M. P. Anantram, "Interbase electronic coupling for transport through DNA," *Physical Review B*, vol. 71, no. 11, p. 115405, 2005.

[47] T. Liu and J. K. Barton, "DNA electrochemistry through the base pairs not the sugar-phosphate backbone," *Journal of the American Chemical Society*, vol. 127, no. 29, pp. 10160–10161, 2005.

[48] A. A. Gorodetsky, O. Green, E. Yavin, and J. K. Barton, "Coupling into the base pair stack is necessary for DNA-mediated electrochemistry," *Bioconjugate Chemistry*, vol. 18, no. 5, pp. 1434–1441, 2007.

[49] E. Meggers, M. E. Michel-Beyerle, and B. Giese, "Sequence dependent long range hole transport in DNA," *Journal of the American Chemical Society*, vol. 120, no. 49, pp. 12950–12955, 1998.

[50] J. Jortner, M. Bixon, T. Langenbacher, and M. E. Michel-Beyerle, "Charge transfer and transport in DNA," *Proceedings of the National Academy of Sciences*, vol. 95, no. 22, pp. 12759–12765, 1998.

[51] A. A. Voityuk, N. Rösch, M. Bixon, and J. Jortner, "Electronic coupling for charge transfer and transport in DNA," *The Journal of Physical Chemistry B*, vol. 104, no. 41, pp. 9740–9745, 2000.

[52] M. Bixon and J. Jortner, "Incoherent charge hopping and conduction in DNA and long molecular chains," *Chemical Physics*, vol. 319, no. 1, pp. 273–282, 2005.

[53] E. B. Brauns, C. J. Murphy, and M. A. Berg, "Local dynamics in DNA by temperature-dependent stokes shifts of an intercalated dye," *Journal of the American Chemical Society*, vol. 120, no. 10, pp. 2449–2456, 1998.

[54] A. K. Mahapatro, K. J. Jeong, G. U. Lee, and D. B. Janes, "Sequence specific electronic conduction through polyion-stabilized double-stranded DNA in nanoscale break junctions," *Nanotechnology*, vol. 18, no. 19, p. 195202, 2007.

[55] K.-H. Yoo, D. Ha, J.-O. Lee, J. Park, J. Kim, J. Kim, H.-Y. Lee, T. Kawai, and H. Y. Choi, "Electrical conduction through poly (dA)-poly (dT) and poly (dG)-poly (dC) DNA molecules," *Physical Review Letters*, vol. 87, no. 19, p. 198102, 2001.

[56] S. P. Walch, "Model calculations of the electron affinities and ionization potentials of DNA," *Chemical Physics Letters*, vol. 374, no. 5, pp. 496–500, 2003.

[57] B. Xu, P. Zhang, X. Li, and N. Tao, "Direct conductance measurement of single DNA molecules in aqueous solution," *Nano Letters*, vol. 4, no. 6, pp. 1105–1108, 2004.

[58] E. Maciá, F. Triozon, and S. Roche, "Contact-dependent effects and tunneling currents in DNA molecules," *Physical Review B*, vol. 71, no. 11, p. 113106, 2005.

[59] T. Heim, D. Deresmes, and D. Vuillaume, "Conductivity of DNA probed by conducting–atomic force microscopy: Effects of contact electrode, DNA structure, and surface interactions," *Journal of Applied Physics*, vol. 96, no. 5, pp. 2927–2936, 2004.

[60] S.-P. Liu, J. Artois, D. Schmid, M. Wieser, B. Bornemann, S. Weisbrod, A. Marx, E. Scheer, and A. Erbe, "Electronic transport through short dsDNA measured with mechanically controlled break junctions: New thiol–gold binding protocol improves conductance," *Physica Status Solidi (b)*, vol. 250, no. 11, pp. 2342–2348, 2013.

[61] F. Zahid, M. Paulsson, and S. Datta, "Electrical conduction through molecules," *Advanced Semiconductors and Organic Nano-Techniques*, vol. 3, 2003.

[62] R. N. Barnett, C. L. Cleveland, A. Joy, U. Landman, and G. B. Schuster, "Charge migration in DNA: ion-gated transport," *Science*, vol. 294, no. 5542, pp. 567–571, 2001.

[63] C. Adessi and M. Anantram, "Influence of counter-ion-induced disorder in DNA conduction," *Applied physics letters*, vol. 82, no. 14, pp. 2353–2355, 2003.

[64] M. Xu, R. Endres, S. Tsukamoto, M. Kitamura, S. Ishida, and Y. Arakawa, "Conformation and local environment dependent conductance of DNA molecules," *Small*, vol. 1, no. 12, pp. 1168–1172, 2005.

[65] P. De Pablo, F. Moreno-Herrero, J. Colchero, J. G. Herrero, P. Herrero, A. Baró, P. Ordejón, J. M. Soler, and E. Artacho, "Absence of dc-conductivity in λ-DNA," *Physical Review Letters*, vol. 85, no. 23, pp. 4992–4995, 2000.

[66] D. Porath, A. Bezryadin, S. De Vries, and C. Dekker, "Direct measurement of electrical transport through DNA molecules," *Nature*, vol. 403, no. 6770, pp. 635–638, 2000.

[67] S. Takagi, T. Takada, N. Matsuo, S. Yokoyama, M. Nakamura, and K. Yamana, "Gating electrical transport through DNA molecules that bridge between silicon nanogaps," *Nanoscale*, vol. 4, no. 6, pp. 1975–1977, 2012.

[68] H.-W. Fink and C. Schönenberger, "Electrical conduction through DNA molecules," *Nature*, vol. 398, no. 6726, pp. 407–410, 1999.

[69] H. Zhang, W. Xu, X. Liu, F. Stellacci, and J. T. Thong, "Capturing a DNA duplex under near-physiological conditions," *Applied Physics Letters*, vol. 97, no. 16, p. 163702, 2010.

[70] A. Y. Kasumov, M. Kociak, S. Gueron, B. Reulet, V. Volkov, D. Klinov, and H. Bouchiat, "Proximity-induced superconductivity in DNA," *Science*, vol. 291, no. 5502, pp. 280–282, 2001.

[71] A. Chepelianskii, D. Klinov, A. Kasumov, S. Guéron, O. Pietrement, S. Lyonnais, and H. Bouchiat, "Conduction of DNA molecules attached to a disconnected array of metallic Ga nanoparticles," *New Journal of Physics*, vol. 13, no. 6, p. 063046, 2011.

[72] R. G. Lerner and G. L. Trigg, *Quantum transport: atom to transistor*. VHC publishers, 1991.

[73] S. Datta, *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[74] J. Patterson and B. Bailey, *Solid-State Physics*. Springer-Verlag Berlin Heidelberg, 2010.

[75] S. Selberherr, *Analysis and simulation of semiconductor devices*. Springer-Verlag Wien, 1984.

[76] U. Dersch, B. Pohlmann, and P. Thomas, "The rate equation for hopping transport of interacting electrons," *Journal of Physics C: Solid State Physics*, vol. 16, no. 19, pp. 3725–3737, 1983.

[77] K. Senthilkumar, F. C. Grozema, C. F. Guerra, F. M. Bickelhaupt, F. D. Lewis, Y. A. Berlin, M. A. Ratner, and L. D. Siebbeles, "Absolute rates of hole transfer in DNA," *Journal of the American Chemical Society*, vol. 127, no. 42, pp. 14894–14903, 2005.

[78] A. A. Voityuk, "Can charge transfer in DNA significantly be modulated by varying the π stack conformation?," *The Journal of Physical Chemistry B*, vol. 113, no. 43, pp. 14365–14368, 2009.

[79] J. Yi, "Conduction of DNA molecules: A charge-ladder model," *Physical Review B*, vol. 68, no. 19, p. 193103, 2003.

[80] R. Gutiérrez, S. Mandal, and G. Cuniberti, "Quantum transport through a DNA wire in a dissipative environment," *Nano Letters*, vol. 5, no. 6, pp. 1093–1097, 2005.

[81] R. Gutiérrez, S. Mandal, and G. Cuniberti, "Dissipative effects in the electronic transport through DNA molecular wires," *Physical Review B*, vol. 71, no. 23, p. 235116, 2005.

[82] R. Gutiérrez, S. Mohapatra, H. Cohen, D. Porath, and G. Cuniberti, "Inelastic quantum transport in a ladder model: Implications for DNA conduction and comparison to experiments on suspended DNA oligomers," *Physical Review B*, vol. 74, no. 23, p. 235105, 2006.

[83] A. V. Malyshev, "DNA double helices for single molecule electronics," *Physical Review Letters*, vol. 98, no. 9, p. 096801, 2007.

[84] S. S. Mallajosyula, J. C. Lin, D. L. Cox, S. K. Pati, and R. R. P. Singh, "Sequence dependent electron transport in wet DNA: Ab initio and molecular dynamics studies," *Physical Review Letters*, vol. 101, no. 17, p. 176805, 2008.

[85] A. Malyshev, E. Díaz, F. Dominguez-Adame, and V. Malyshev, "Effects of the environment on the electric conductivity of double-stranded DNA molecules," *Journal of Physics: Condensed Matter*, vol. 21, no. 33, p. 335105, 2009.

[86] X.-F. Wang, T. Chakraborty, and J. Berashevich, "Quantum transport anomalies in DNA containing mispairs," *Nanotechnology*, vol. 21, no. 48, p. 485101, 2010.

[87] E. B. Starikov, C. Nganou, K. H. Lee, G. Cuniberti, and W. Wenzel, "Single-molecule DNA conductance in water solutions: Role of explicit water-counterion sheath and chemical modification of nucleobases," *Biophysical Reviews and Letters*, vol. 4, no. 3, pp. 231–243, 2009.

[88] P. B. Woiczikowski, T. Kubař, R. Gutiérrez, R. A. Caetano, G. Cuniberti, and M. Elstner, "Combined density functional theory and Landauer approach for hole transfer in DNA along classical molecular dynamics trajectories," *The Journal of chemical physics*, vol. 130, no. 21, p. 215104, 2009.

[89] R. Gutiérrez, R. Caetano, P. Woiczikowski, T. Kubař, M. Elstner, and G. Cuniberti, "Structural fluctuations and quantum transport through DNA molecular wires: A combined molecular dynamics and model Hamiltonian approach," *New Journal of Physics*, vol. 12, no. 2, p. 023022, 2010.

[90] Y. Maeda, A. Okamoto, Y. Hoshiba, T. Tsukamoto, Y. Ishikawa, and N. Kurita, "Effect of hydration on electrical conductivity of DNA duplex: Green's function study combined with DFT," *Computational Materials Science*, vol. 53, no. 1, pp. 314–320, 2012.

[91] S. S. Mallajosyula and S. K. Pati, "Toward DNA conductivity: A theoretical perspective," *The Journal of Physical Chemistry Letters*, vol. 1, no. 12, pp. 1881–1894, 2010.

[92] H. Mehrez, S. Walch, and M. Anantram, "Electronic properties of O_2-doped DNA," *Physical Review B*, vol. 72, no. 3, p. 035441, 2005.

[93] F. Jensen, *Introduction to computational chemistry*. John Wiley & Sons, 2007.

[94] R. M. Martin, "Electronic structure: Basic theory and practical methods," 2008.

[95] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, *et al.*, "General atomic and molecular electronic structure system," *Journal of computational chemistry*, vol. 14, no. 11, pp. 1347–1363, 1993.

[96] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical Review B*, vol. 54, no. 16, p. 11169, 1996.

[97] J. Kong, C. A. White, A. I. Krylov, D. Sherrill, R. D. Adamson, T. R. Furlani, M. S. Lee, A. M. Lee, S. R. Gwaltney, T. R. Adams, *et al.*, "Q-Chem 2.0: A high-performance ab initio electronic structure program package," *Journal of Computational Chemistry*, vol. 21, no. 16, pp. 1532–1548, 2000.

[98] G. Te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. van Gisbergen, J. G. Snijders, and T. Ziegler, "Chemistry with ADF," *Journal of Computational Chemistry*, vol. 22, no. 9, pp. 931–967, 2001.

[99] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, "The SIESTA method for ab initio order-N materials simulation," *Journal of Physics: Condensed Matter*, vol. 14, no. 11, p. 2745, 2002.

[100] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. E. P. J. A. Montgomery, Jr., F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin,

R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, *Gaussian 09*. Gaussian, Inc., Wallingford, CT, 2009.

[101] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, *et al.*, "NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations," *Computer Physics Communications*, vol. 181, no. 9, pp. 1477–1489, 2010.

[102] T. Macke and D. Case, *Modeling unusual nucleic acid structures*, pp. 379–393. Washington, DC: American Chemical Society, 1998.

[103] P.-O. Löwdin, "On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals," *The Journal of Chemical Physics*, vol. 18, no. 3, pp. 365–375, 1950.

[104] B. Gazdy, M. Seel, and J. Ladik, "The role of self-consistency in quantum-mechanical studies of disordered quasi-one-dimensional systems," *Chemical Physics*, vol. 86, no. 1, pp. 41–48, 1984.

[105] Y.-J. Ye and Y. Jiang, "Electronic structures and long-range electron transfer through DNA molecules," *International Journal of Quantum Chemistry*, vol. 78, no. 2, pp. 112–130, 2000.

[106] G. Cuniberti, L. Craco, D. Porath, and C. Dekker, "Backbone-induced semiconducting behavior in short DNA wires," *Physical Review B*, vol. 65, no. 24, p. 241314, 2002.

[107] G. Cuniberti, E. Maciá, A. Rodriguez, and R. Römer, "Tight-binding modeling of charge migration in dna devices," in *Charge Migration in DNA*, pp. 1–20, Springer, 2007.

[108] H. Yamada and K. Iguchi, "Some effective tight-binding models for electrons in dna conduction: A review," *Advances in Condensed Matter Physics*, vol. 2010, 2010.

[109] R. Landauer, "Spatial variation of currents and fields due to localized scatterers in metallic conduction," *IBM Journal of Research and Development*, vol. 1, no. 3, pp. 223–231, 1957.

[110] M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, "Generalized many-channel conductance formula with application to small rings," *Physical Review B*, vol. 31, no. 10, p. 6207, 1985.

[111] S. Datta, *Electronic transport in mesoscopic systems*. Cambridge university press, 1995.

[112] D. Ryndyk, R. Gutiérrez, B. Song, and G. Cuniberti, "Green function techniques in the treatment of quantum transport at the molecular scale," in *Energy Transfer Dynamics in Biomaterial Systems*, pp. 213–335, Springer Berlin Heidelberg, 2009.

[113] M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, "Modeling of nanoscale devices," *Proceedings of the IEEE*, vol. 96, no. 9, pp. 1511–1550, 2008.

[114] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *Journal of Applied Physics*, vol. 91, no. 4, pp. 2343–2354, 2002.

[115] S. Hershfield, J. H. Davies, and J. W. Wilkins, "Probing the kondo resonance by resonant tunneling through an Anderson impurity," *Physical Review Letters*, vol. 67, no. 26, p. 3720, 1991.

[116] A.-P. Jauho, N. S. Wingreen, and Y. Meir, "Time-dependent transport in interacting and noninteracting resonant-tunneling systems.," *Physical review. B*, vol. 50, no. 8, pp. 5528–5544, 1994.

[117] A. Prociuk, B. Van Kuiken, and B. D. Dunietz, "Benchmarking the performance of density functional theory based Green's function formalism utilizing different self-energy models in calculating electronic transmission through molecular systems," *The Journal of Chemical Physics*, vol. 125, no. 20, p. 204717, 2006.

[118] X. Zheng, F. Wang, C. Y. Yam, Y. Mo, and G. Chen, "Time-dependent density-functional theory for open systems," *Physical Review B*, vol. 75, no. 19, p. 195127, 2007.

[119] C. Herrmann, G. C. Solomon, and M. A. Ratner, "Local pathways in coherent electron transport through iron porphyrin complexes: A challenge for first-principles transport calculations," *The Journal of Physical Chemistry C*, vol. 114, no. 48, pp. 20813–20820, 2010.

[120] D. Papaconstantopoulos, "Handbook of the band structure of elemental solids," 1986.

[121] M. Zilly, O. Ujsághy, and D. E. Wolf, "Conductance of DNA molecules: Effects of decoherence and bonding," *Physical Review B*, vol. 82, no. 12, p. 125125, 2010.

[122] M. Lee, S. Avdoshenko, R. Gutierrez, and G. Cuniberti, "Charge migration through DNA molecules in the presence of mismatches," *Physical Review B*, vol. 82, no. 15, p. 155455, 2010.

[123] M. Büttiker, "Four-terminal phase-coherent conductance," *Physical Review Letters*, vol. 57, no. 14, pp. 1761–1764, 1986.

[124] M. Büttiker, "Symmetry of electrical conduction," *IBM Journal of Research and Development*, vol. 32, no. 3, pp. 317–334, 1988.

[125] T. Kubař and M. Elstner, "What governs the charge transfer in DNA? The role of DNA conformation and environment," *The Journal of Physical Chemistry B*, vol. 112, no. 29, pp. 8788–8798, 2008.

[126] X.-Q. Li and Y. Yan, "Electrical transport through individual DNA molecules," *Applied Physics Letters*, vol. 79, no. 14, pp. 2190–2192, 2001.

[127] J. L. D'Amato and H. M. Pastawski, "Conductance of a disordered linear chain including inelastic scattering events," *Physical Review B*, vol. 41, no. 11, pp. 7411–7420, 1990.

[128] D. Nozaki, C. G. Da Rocha, H. M. Pastawski, and G. Cuniberti, "Disorder and dephasing effects on electron transport through conjugated molecular wires in molecular junctions," *Physical Review B*, vol. 85, no. 15, p. 155327, 2012.

[129] L. C. Gosule and J. A. Schellman, "Dna condensation with polyamines: I. Spectroscopic studies," *Journal of Molecular Biology*, vol. 121, no. 3, pp. 311–326, 1978.

[130] H. Deng, V. A. Bloomfield, J. M. Benevides, and G. J. Thomas Jr, "Structural basis of polyamine–DNA recognition: spermidine and spermine interactions with genomic B-DNAs of different GC content probed by Raman spectroscopy," *Nucleic Acids Research*, vol. 28, no. 17, pp. 3379–3385, 2000.

[131] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *The Journal of Chemical Physics*, vol. 98, no. 7, pp. 5648–5652, 1993.

[132] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Physical Review B*, vol. 37, no. 2, pp. 785–789, 1988.

[133] J. Qi, M. G. Rabbani, S. Edirisinghe, and M. Anantram, "Transport of charge in DNA heterostructures," in *Nanotechnology (IEEE-NANO), 2011 11th IEEE Conference on*, pp. 487–491, IEEE, 2011.

[134] C. Weisbuch and B. Vinter, *Quantum semiconductor structures: Fundamentals and applications*. Academic press, 1993.

[135] R. Dennington, T. Keith, J. Millam, *et al.*, "Gaussview, version 5," *Semichem Inc., Shawnee Mission, KS*, 2009.

[136] R. N. Barnett, C. L. Cleveland, U. Landman, E. Boone, S. Kanvah, and G. B. Schuster, "Effect of base sequence and hydration on the electronic and hole transport properties of duplex DNA: Theory and experiment," *The Journal of Physical Chemistry A*, vol. 107, no. 18, pp. 3525–3537, 2003.

[137] P. B. Woiczikowski, T. Kubař, R. Gutiérrez, R. A. Caetano, G. Cuniberti, and M. Elstner, "Combined density functional theory and Landauer approach for hole transfer in DNA along classical molecular dynamics trajectories," *The Journal of Chemical Physics*, vol. 130, no. 21, p. 215104, 2009.

[138] J. Qi, N. Edirisinghe, M. G. Rabbani, and M. Anantram, "Unified model for conductance through DNA with the Landauer-Büttiker formalism," *Physical Review B*, vol. 87, no. 8, p. 085404, 2013.

[139] K. D. Robertson, "DNA methylation, methyltransferases, and cancer.," *Oncogene*, vol. 20, no. 24, pp. 3139–3155, 2001.

[140] J. E. Gill, J. A. Mazrimas, and C. C. Bishop, "Physical studies on synthetic DNAs containing 5-methylcytosine," *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis*, vol. 335, no. 3, pp. 330–348, 1974.

[141] L. C. Sowers, B. R. Shaw, and W. D. Sedwick, "Base stacking and molecular polarizability: Effect of a methyl group in the 5-position of pyrimidines," *Biochemical and biophysical research communications*, vol. 148, no. 2, pp. 790–794, 1987.

[142] P. M. Severin, X. Zou, H. E. Gaub, and K. Schulten, "Cytosine methylation alters DNA mechanical properties," *Nucleic Acids Research*, vol. 39, no. 20, pp. 8740–8751, 2011.

[143] W. Doerfler, "DNA methylation and gene gctivity," *Annual Review of Biochemistry*, vol. 52, no. 1, pp. 93–124, 1983.

[144] K. J. Vining, K. R. Pomraning, L. J. Wilhelm, H. D. Priest, M. Pellegrini, T. C. Mockler, M. Freitag, and S. H. Strauss, "Dynamic DNA cytosine methylation in the Populus trichocarpa genome: Tissue-level variation and relationship to gene expression," *BMC Genomics*, vol. 13, no. 1, p. 27, 2012.

[145] D. A. Khavari, G. L. Sen, and J. L. Rinn, "DNA methylation and epigenetic control of cellular differentiation," *Cell Cycle*, vol. 9, no. 19, pp. 3880–3883, 2010.

[146] T. Goto and M. Monk, "Regulation of X-chromosome inactivation in development in mice and humans," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 2, pp. 362–378, 1998.

[147] E. Li, T. H. Bestor, and R. Jaenisch, "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality," *Cell*, vol. 69, no. 6, pp. 915–926, 1992.

[148] H. Lei, S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li, "De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells," *Development*, vol. 122, no. 10, pp. 3195–3205, 1996.

[149] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, *et al.*, "Human DNA methylomes at base resolution show widespread epigenomic differences," *Nature*, vol. 462, no. 7271, pp. 315–322, 2009.

[150] H. Guo, P. Zhu, L. Yan, R. Li, B. Hu, Y. Lian, J. Yan, X. Ren, S. Lin, J. Li, *et al.*, "The DNA methylation landscape of human early embryos," *Nature*, vol. 511, no. 7511, pp. 606–610, 2014.

[151] Z. D. Smith, M. M. Chan, K. C. Humm, R. Karnik, S. Mekhoubad, A. Regev, K. Eggan, and A. Meissner, "DNA methylation dynamics of the human preimplantation embryo," *Nature*, vol. 511, no. 7511, pp. 611–615, 2014.

[152] T. A. Hore, R. W. Rapkins, and J. A. M. Graves, "Construction and evolution of imprinted loci in mammals," *TRENDS in Genetics*, vol. 23, no. 9, pp. 440–448, 2007.

[153] K. D. Robertson, "DNA methylation and human disease," *Nature Reviews Genetics*, vol. 6, no. 8, pp. 597–610, 2005.

[154] P. M. Das and R. Singal, "DNA methylation and cancer," *Journal of Clinical Oncology*, vol. 22, no. 22, pp. 4632–4642, 2004.

[155] S. B. Baylin, "DNA methylation and gene silencing in cancer," *Nature Clinical Practice Oncology*, vol. 2, pp. S4–S11, 2005.

[156] L. Shen and R. A. Waterland, "Methods of DNA methylation analysis," *Current Opinion in Clinical Nutrition & Metabolic Care*, vol. 10, no. 5, pp. 576–581, 2007.

[157] S. J. Clark, J. Harrison, C. L. Paul, and M. Frommer, "High sensitivity mapping of methylated cytosines," *Nucleic Acids Research*, vol. 22, no. 15, pp. 2990–2997, 1994.

[158] C. Bock, E. M. Tomazou, A. B. Brinkman, F. Müller, F. Simmer, H. Gu, N. Jäger, A. Gnirke, H. G. Stunnenberg, and A. Meissner, "Quantitative comparison of genome-wide DNA methylation mapping technologies," *Nature Biotechnology*, vol. 28, no. 10, pp. 1106–1114, 2010.

[159] M. Tsutsui, M. Taniguchi, K. Yokota, and T. Kawai, "Identifying single nucleotides by tunnelling current," *Nature Nanotechnology*, vol. 5, no. 4, pp. 286–290, 2010.

[160] J. Clarke, H.-C. Wu, L. Jayasinghe, A. Patel, S. Reid, and H. Bayley, "Continuous base identification for single-molecule nanopore DNA sequencing," *Nature Nanotechnology*, vol. 4, no. 4, pp. 265–270, 2009.

[161] E. V. Wallace, D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, and H. Bayley, "Identification of epigenetic DNA modifications with a protein nanopore," *Chemical Communications*, vol. 46, no. 43, pp. 8195–8197, 2010.

[162] C. Toher, A. Filippetti, S. Sanvito, and K. Burke, "Self-interaction errors in density-functional calculations of electronic transport," *Physical Review Letters*, vol. 95, no. 14, p. 146402, 2005.

[163] N. A. Temiz, D. E. Donohue, A. Bacolla, B. T. Luke, and J. R. Collins, "The role of methylation in the intrinsic dynamics of B-and Z-DNA," *PLoS One*, vol. 7, no. 4, p. e35558, 2012.

[164] P. Stephens, F. Devlin, C. Chabalowski, and M. J. Frisch, "Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields," *The Journal of Physical Chemistry*, vol. 98, no. 45, pp. 11623–11627, 1994.

[165] W. J. Hehre, R. Ditchfield, and J. A. Pople, "Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules," *The Journal of Chemical Physics*, vol. 56, no. 5, pp. 2257–2261, 1972.

[166] J. Olofsson and S. Larsson, "Electron hole transport in DNA," *The Journal of Physical Chemistry B*, vol. 105, no. 42, pp. 10398–10406, 2001.

[167] A. A. Voityuk, J. Jortner, M. Bixon, and N. Rösch, "Electronic coupling between Watson–Crick pairs for hole transfer and transport in desoxyribonucleic acid," *The Journal of Chemical Physics*, vol. 114, no. 13, pp. 5614–5620, 2001.

[168] A. A. Voityuk and N. Rösch, "Fragment charge difference method for estimating donor–acceptor electronic coupling: Application to DNA π-stacks," *The Journal of Chemical Physics*, vol. 117, no. 12, pp. 5607–5616, 2002.

[169] H. Ishii, N. Kobayashi, and K. Hirose, "Quantum electron transport through carbon nanotubes with electron-phonon coupling," *Journal of Vacuum Science & Technology B*, vol. 27, no. 2, pp. 882–886, 2009.

[170] C. Köhler, T. Watermann, and E. Malic, "Relaxation dynamics via acoustic phonons in carbon nanotubes," *Physica Status Solidi (b)*, vol. 249, no. 12, pp. 2483–2486, 2012.

[171] C. Köhler, T. Watermann, and E. Malic, "Time-and momentum-resolved phonon-induced relaxation dynamics in carbon nanotubes," *Journal of Physics: Condensed Matter*, vol. 25, no. 10, p. 105301, 2013.

[172] S. Wang and E. T. Kool, "Origins of the large differences in stability of DNA and RNA helixes: C-5 methyl and 2'-hydroxyl effects," *Biochemistry*, vol. 34, no. 12, pp. 4125–4132, 1995.

[173] J. Sponer, H. Gabb, J. Leszczynski, and P. Hobza, "Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: A quantum-chemical study.," *Biophysical Journal*, vol. 73, no. 1, pp. 76–87, 1997.

[174] N. Sugimoto, M. Nakano, and S.-i. Nakano, "Thermodynamics-structure relationship of single mismatches in RNA/DNA duplexes," *Biochemistry*, vol. 39, no. 37, pp. 11270–11281, 2000.

[175] S. Derreumaux, M. Chaoui, G. Tevanian, and S. Fermandjian, "Impact of CpG methylation on structure, dynamics and solvation of cAMP DNA responsive element," *Nucleic Acids Research*, vol. 29, no. 11, pp. 2314–2326, 2001.

[176] J. Norberg and M. Vihinen, "Molecular dynamics simulation of the effects of cytosine methylation on structure of oligonucleotides," *Journal of Molecular Structure*, vol. 546, no. 1, pp. 51–62, 2001.

[177] L. R. Rutledge, H. F. Durst, and S. D. Wetmore, "Computational comparison of the stacking interactions between the aromatic amino acids and the natural or (cationic) methylated nucleobases," *Physical Chemistry Chemical Physics*, vol. 10, no. 19, pp. 2801–2812, 2008.

[178] L. Fojt, V. Vetterl, and T. Doneux, "Adsorption and two-dimensional condensation of 5-methylcytosine," *Bioelectrochemistry*, vol. 75, no. 2, pp. 89–94, 2009.

[179] A. Moser, R. Guza, N. Tretyakova, and D. M. York, "Density functional study of the influence of C5 cytosine substitution in base pairs with guanine," *Theoretical Chemistry Accounts*, vol. 122, no. 3-4, pp. 179–188, 2009.

[180] C. Acosta-Silva, V. Branchadell, J. Bertran, and A. Oliva, "Mutual relationship between stacking and hydrogen bonding in DNA. Theoretical study of guanine-cytosine, guanine-5-methylcytosine, and their dimers," *The Journal of Physical Chemistry B*, vol. 114, no. 31, pp. 10217–10227, 2010.

[181] T. Yanai, D. P. Tew, and N. C. Handy, "A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)," *Chemical Physics Letters*, vol. 393, no. 1, pp. 51–57, 2004.

[182] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into current limitations of density functional theory," *Science*, vol. 321, no. 5890, pp. 792–794, 2008.

[183] R. Baer, E. Livshits, and U. Salzner, "Tuned range-separated hybrids in density functional theory," *Annual Review of Physical Chemistry*, vol. 61, pp. 85–109, 2010.

[184] L. Kronik, T. Stein, S. Refaely-Abramson, and R. Baer, "Excitation gaps of finite-sized systems from optimally tuned range-separated hybrid functionals," *Journal of Chemical Theory and Computation*, vol. 8, no. 5, pp. 1515–1531, 2012.

[185] M. Srebro and J. Autschbach, "Tuned range-separated time-dependent density functional theory applied to optical rotation," *Journal of Chemical Theory and Computation*, vol. 8, no. 1, pp. 245–256, 2011.

[186] S. Refaely-Abramson, S. Sharifzadeh, N. Govind, J. Autschbach, J. B. Neaton, R. Baer, and L. Kronik, "Quasiparticle spectra from a nonempirical optimally tuned range-separated hybrid density functional," *Physical Review Letters*, vol. 109, no. 22, p. 226405, 2012.

[187] K. Lopata and N. Govind, "Near and above ionization electronic excitations with non-hermitian real-time time-dependent density functional theory," *Journal of Chemical Theory and Computation*, vol. 9, no. 11, pp. 4939–4946, 2013.

[188] T. Stein, L. Kronik, and R. Baer, "Reliable prediction of charge transfer excitations in molecular complexes using time-dependent density functional theory," *Journal of the American Chemical Society*, vol. 131, no. 8, pp. 2818–2820, 2009.

[189] M. A. Rohrdanz and J. M. Herbert, "Simultaneous benchmarking of ground-and excited-state properties with long-range-corrected density functional theory," *The Journal of Chemical Physics*, vol. 129, no. 3, p. 034107, 2008.

[190] B. Gu, F. Zhang, Z. Wang, and H. Zhou, "Solvent-induced DNA conformational transition," *Physical Review Letters*, vol. 100, no. 8, p. 088104, 2008.

[191] R. Venkatramani, S. Keinan, A. Balaeff, and D. N. Beratan, "Nucleic acid charge transfer: Black, white and gray," *Coordination Chemistry Reviews*, vol. 255, no. 7, pp. 635–648, 2011.

[192] G. Brancolini, A. Migliore, S. Corni, M. Fuentes-Cabrera, F. J. Luque, and R. Di Felice, "Dynamical treatment of charge transfer through duplex nucleic acids containing modified adenines," *ACS nano*, vol. 7, no. 10, pp. 9396–9406, 2013.

[193] J. Kypr, I. Kejnovská, D. Renčiuk, and M. Vorlíčková, "Circular dichroism and conformational polymorphism of DNA," *Nucleic Acids Research*, vol. 37, no. 6, pp. 1713–1725, 2009.

[194] Y. A. Berlin, A. A. Voityuk, and M. A. Ratner, "DNA base pair stacks with high electric conductance: A systematic structural search," *ACS nano*, vol. 6, no. 9, pp. 8216–8225, 2012.

[195] Y. Zhang, C. Liu, A. Balaeff, S. S. Skourtis, and D. N. Beratan, "Biological charge transfer via flickering resonance," *Proceedings of the National Academy of Sciences*, vol. 111, no. 28, pp. 10049–10054, 2014.

[196] B. Popescu, P. B. Woiczikowski, M. Elstner, and U. Kleinekathöfer, "Time-dependent view of sequential transport through molecules with rapidly fluctuating bridges," *Physical Review Letters*, vol. 109, no. 17, p. 176802, 2012.

[197] F. C. Grozema, L. D. Siebbeles, Y. A. Berlin, and M. A. Ratner, "Hole mobility in DNA: Effects of static and dynamic structural fluctuations," *ChemPhysChem*, vol. 3, no. 6, pp. 536–539, 2002.

[198] P. Várnai and K. Zakrzewska, "DNA and its counterions: A molecular dynamics study," *Nucleic Acids Research*, vol. 32, no. 14, pp. 4269–4280, 2004.

[199] Y. A. Mantz, F. L. Gervasio, T. Laino, and M. Parrinello, "Solvent effects on charge spatial extent in DNA and implications for transfer," *Physical Review Letters*, vol. 99, no. 5, p. 058104, 2007.

[200] K. Wang, J. M. Hamill, B. Wang, C. Guo, S. Jiang, Z. Huang, and B. Xu, "Structure determined charge transport in single DNA molecule break junctions," *Chemical Science*, vol. 5, no. 9, pp. 3425–3431, 2014.

[201] D. Sivanesan, V. Subramanian, and B. Nair, "Solvent effect on DNA base stacked dimers: An isodensity polarizable continuum model approach," *International Journal of Quantum Chemistry*, vol. 84, no. 6, pp. 750–758, 2001.

[202] C. E. Crespo-Hernández, R. Arce, Y. Ishikawa, L. Gorb, J. Leszczynski, and D. M. Close, "Ab initio ionization energy thresholds of DNA and RNA bases in gas phase and in aqueous solution," *The Journal of Physical Chemistry A*, vol. 108, no. 30, pp. 6373–6377, 2004.

[203] J. Berashevich and T. Chakraborty, "How the surrounding water changes the electronic and magnetic properties of DNA," *The Journal of Physical Chemistry B*, vol. 112, no. 44, pp. 14083–14089, 2008.

[204] J. Poater, M. Swart, C. F. Guerra, and F. M. Bickelhaupt, "Solvent effects on hydrogen bonds in Watson–Crick, mismatched, and modified DNA base pairs," *Computational and Theoretical Chemistry*, vol. 998, pp. 57–63, 2012.

[205] D. H. Ha, H. Nham, K.-H. Yoo, H.-m. So, H.-Y. Lee, and T. Kawai, "Humidity effects on the conductance of the assembly of DNA molecules," *Chemical Physics Letters*, vol. 355, no. 5, pp. 405–409, 2002.

[206] S. Tuukkanen, A. Kuzyk, J. Toppari, V. Hytönen, T. Ihalainen, and P. Törmä, "Di-electrophoresis of nanoscale double-stranded DNA and humidity effects on its electrical conductivity," *Applied Physics Letters*, vol. 87, no. 18, p. 183102, 2005.

[207] T. Kleine-Ostmann, C. Jördens, K. Baaske, T. Weimann, M. H. de Angelis, and M. Koch, "Conductivity of single-stranded and double-stranded deoxyribose nucleic acid under ambient conditions: The dominance of water," *Applied Physics Letters*, vol. 88, no. 10, p. 102102, 2006.

[208] C. Yamahata, D. Collard, T. Takekawa, M. Kumemura, G. Hashiguchi, and H. Fujita, "Humidity dependence of charge transport through DNA revealed by silicon-based nanotweezers manipulation," *Biophysical Journal*, vol. 94, no. 1, pp. 63–70, 2008.

[209] J. Autschbach and S. Zheng, "Density functional computations of ^{99}ru chemical shifts: Relativistic effects, influence of the density functional, and study of solvent effects on fac[Ru $(CO)_3I_3$]-," *Magnetic Resonance in Chemistry*, vol. 44, no. 11, pp. 989–1007, 2006.

[210] J. Tomasi, B. Mennucci, and R. Cammi, "Quantum mechanical continuum solvation models," *Chemical Reviews*, vol. 105, no. 8, pp. 2999–3094, 2005.

[211] J. M. Artés, Y. Li, J. Qi, M. P. Anantram, and J. Hihath, "Conformational gating of dna conductance," *submitted*.

[212] S. R. Morris and L. A. Carey, "Gene expression profiling in breast cancer," *Current Opinion in Oncology*, vol. 19, no. 6, pp. 547–551, 2007.

[213] Z. Fang and S. O. Kelley, "Direct electrocatalytic mRNA detection using PNA-nanowire sensors," *Analytical Chemistry*, vol. 81, no. 2, pp. 612–617, 2009.

[214] X. Chen, S. Roy, Y. Peng, and Z. Gao, "Electrical sensor array for polymerase chain reaction-free messenger RNA expression profiling," *Analytical Chemistry*, vol. 82, no. 14, pp. 5958–5964, 2010.

[215] L. Bessman, M. Simms, and E. Kornberg, "Enzymatic synthesis of deoxyribonucleic acid. Preparation of substrates and partial purification of an enzyme from Escherichia. coli," *The Journal of Biological Chemistry*, vol. 233, no. 1, pp. 163–170, 1985.

[216] F. Crick, L. Barnett, S. Brenner, and R. Watts-Tobin, "General nature of the genetic code for proteins," *Nature*, vol. 192, pp. 1227–1232, 1961.

[217] M. J. Bibb, R. A. Van Etten, C. T. Wright, M. W. Walberg, and D. A. Clayton, "Sequence and gene organization of mouse mitochondrial DNA," *Cell*, vol. 26, no. 2, pp. 167–180, 1981.

[218] J. Summers and W. S. Mason, "Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate," *Cell*, vol. 29, no. 2, pp. 403–415, 1982.

[219] A. Kornberg, "DNA replication.," *The Journal of Biological Chemistry*, vol. 263, no. 1, pp. 1–4, 1988.

[220] P. F. Agris, "Decoding the genome: A modified view," *Nucleic Acids Research*, vol. 32, no. 1, pp. 223–238, 2004.

[221] R. Barrangou, "Cas9 targeting and the CRISPR revolution," *Science*, vol. 344, no. 6185, pp. 707–708, 2014.

[222] J. C. Zhou, B. Feller, B. Hinsberg, G. Sethi, P. Feldstein, J. Hihath, E. Seker, M. Marco, A. Knoesen, and R. Miller, "Immobilization-mediated reduction in melting temperatures of DNA-DNA and DNA-RNA hybrids: Immobilized DNA probe hybridization studied by SPR," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 481, pp. 72–79, 2015.

[223] C. Herrmann, G. C. Solomon, J. E. Subotnik, V. Mujica, and M. A. Ratner, "Ghost transmission: How large basis sets can make electron transport calculations worse?," *The Journal of Chemical Physics*, vol. 132, no. 2, p. 024103, 2010.

[224] G. Sedghi, V. M. García-Suárez, L. J. Esdaile, H. L. Anderson, C. J. Lambert, S. Martín, D. Bethell, S. J. Higgins, M. Elliott, N. Bennett, *et al.*, "Long-range electron tunnelling in oligo-porphyrin molecular wires," *Nature Nanotechnology*, vol. 6, no. 8, pp. 517–523, 2011.

[225] W. Fichtner, *Quantum Transport for Nanostructures*. Swiss Federal Institute of Technology Zurich.

[226] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. Lundstrom, "A simple quantum mechanical treatment of scattering in nanoscale transistors," *Journal of Applied Physics*, vol. 93, no. 9, pp. 5613–5625, 2003.

[227] N. Russo, M. Toscano, and A. Grand, "Theoretical determination of electron affinity and ionization potential of DNA and RNA bases," *Journal of Computational Chemistry*, vol. 21, no. 14, pp. 1243–1250, 2000.

Appendix A

CODE DEVELOPMENT

The Matlab codes developed in this work are provided here.

A.1 Code for coherent transport

```
% Transmission in coherent model for DNA strands.
% Developer:        Jianqing Qi (University of Washington, Seattle)
% Input parameters: strand_name; left and right contact coupling; sites
%                   for electron to be injected; sites for electron to
%                   be extraced
% Output:           transmission

function  DNATransmission_Coherent(workpath, filepath, strand_name,...
    gammaL, gammaR, InjectSite, ExtractSite)
format long
workdir = workpath
filedir = filepath
strand = char(strand_name)
if (ischar(gammaL)), gammaL = str2num(gammaL), end;
if (ischar(gammaR)), gammaR = str2num(gammaR), end;
if (ischar(InjectSite)), Lsite = str2num(InjectSite), end;
if (ischar(ExtractSite)), Rsite = str2num(ExtractSite), end;
addpath(workdir)
%%%%%%%%%%%%%Load Matrices%%%%%%%%%%%%%%%%%%%%%%%%%%%
Fm = ['load ', filedir, strand, '.mat'];
Bm = ['load ', filedir, 'Base_', strand, '.mat'];
```

```
Em = ['load ', filedir, 'Energy_', strand, '.mat'];
eval(Fm)
eval(Bm)
eval(Em)
H0 = eval([strand]);
Base = eval(['Base_', strand]);
Energy = eval(['Energy_', strand]);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
eta = 0;
sizeH = size(H0, 1)  % size of Hamiltonian
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
NBase = length(Base)
Norb = sum(Base)
if(Norb ~= sizeH)
    return;
end
%%%%%%%%%%%Initialize gamma %%%%%%%%%%%%%%
sumSig = zeros(sizeH,sizeH);
GammaL = zeros(Base(Lsite));
GammaR = zeros(Base(Rsite));
%%%%%%%%%set sumSig %%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%% Injection %%%%%%%%%%%%%%%%
TempLenL1 = sum(Base(1 : Lsite)) - Base(Lsite);
TempLenL2 = sum(Base(1 : Lsite));
GammaL = gammaL * eye(Base(Lsite));
sumSig(TempLenL1 + 1 : TempLenL2, TempLenL1 + 1 : TempLenL2) = GammaL;
%%%%%%%%%%% Extraction %%%%%%%%%%%%%%%%
TempLenR1 = sum(Base(1 : Rsite))-Base(Rsite);
TempLenR2 = sum(Base(1 : Rsite));
GammaR = gammaR * eye(Base(Rsite));
sumSig(TempLenR1 + 1 : TempLenR2, TempLenR1 + 1 : TempLenR2) = GammaR;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
sumSig = -1i * sumSig / 2;
```

```
%%%%%%%%%%%% Transmission & DOS %%%%%%%%%%%%%%%%%%%%
NE = length(Energy);
DOS = zeros(1, NE);
DOSBase = zeros(NBase, NE);
%%%%%%%%%%%%%%%%%%%%%%%%%%
for nE = 1 : NE
    nE
    E = Energy(nE)
    Gr=((E + 1i * eta) * eye(sizeH) - H0 - sumSig) \ eye(sizeH);
    Ga = Gr';
%%%%%%%%%%%%%%%%%%% DOS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
     tempM = imag(diag(Gr));
     DOS(nE)=-sum(tempM) / pi;

     t1 = 1;
     t2 = Base(1);

     for nBase = 1 : NBase
         DOSBase(nBase, nE) = -sum(tempM(t1 : t2)) / pi;
         if nBase < NBase
            t1 = t1 + Base(nBase);
            t2 = t2+Base(nBase + 1);
         end
     end
%%%%%%%%%% Transmission %%%%%%%%%%%%
Tran(nE) = real(trace(GammaL * Gr(TempLenL1 + 1 : TempLenL2, ...
    TempLenR1 + 1 : TempLenR2) * GammaR * Ga(TempLenR1 + 1 : ...
    TempLenR2, TempLenL1 + 1 : TempLenL2)));
T = Tran(nE)
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Tname = strcat(workdir, 'Tran_',strand, '_gammaL_', num2str(gammaL), ...
    '_gammaR_', num2str(gammaR), '.mat');
```

```
88 Dname = strcat(workdir, 'DOS_',strand, '_gammaL_', num2str(gammaL), ...
89     '_gammaR_', num2str(gammaR), '.mat');
90 save(Tname, 'Energy', 'Tran')
91 save(Dname, 'Energy', 'DOS', 'DOSBase')
92 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

A.2 Code for decoherent transport

```
1  % Transmission in decoherent model for DNA strands.
2  % Developer:        Jianqing Qi (University of Washington, Seattle)
3  % Input parameters: strand_name; left and right contact coupling;
4  %                   Buttiker probes coupling; sites for electron
5  %                   to be injected; sites for electron to be extraced
6  % Output:           transmission
7  % Reference:        J. L. D'Amato, H. M. Pastawski, Phys. Rev. B 41, ...
8  %                   7411 (1990)
9  function  DNATransmission_Decoherent(workpath, filepath, strand_name, ...
10     gammaL, gammaR, bprobe, InjectSite, ExtractSite)
11 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
12 format long
13 workdir = workpath
14 filedir = filepath
15 strand = char(strand_name)
16 if (ischar(gammaL)), gammaL = str2num(gammaL), end;
17 if (ischar(gammaR)), gammaR = str2num(gammaR), end;
18 if (ischar(bprobe)), bprobe = str2num(bprobe), end;
19 if (ischar(InjectSite)), Lsite = str2num(InjectSite), end;
20 if (ischar(ExtractSite)), Rsite = str2num(ExtractSite), end;
21 addpath(workdir)
22 %%%%%%%%%%%%%Load Matrices%%%%%%%%%%%%%%%%%%%%%%%%%%%
23 Fm = ['load ', filedir, strand, '.mat'];
24 Bm = ['load ', filedir, 'Base_', strand, '.mat'];
```

```
25 Em = ['load ', filedir, 'Energy_', strand, '.mat'];
26 Dm = ['load ', filedir, 'Dsites_', strand, '.mat'];
27 eval(Fm)
28 eval(Bm)
29 eval(Em)
30 eval(Dm)
31 H0 = eval([strand]);
32 Base = eval(['Base_', strand]);
33 Energy = eval(['Energy_', strand]);
34 Dsites = eval(['Dsites_', strand]);
35 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
36 eta = 0;
37 sizeH = size(H0, 1)  % size of Hamiltonian
38 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
39 NBase = length(Base)
40 Norb = sum(Base)
41 if(Norb ~= sizeH)
42     return;
43 end
44 Nb = length(Dsites);
45 Nsite = Nb + 2
46 %%%%%%%%%%Initialize gamma %%%%%%%%%%%%%%
47 sites = zeros(1, Nsite);
48 gamma = zeros(1, Nsite);
49 sites = [Lsite Rsite Dsites];
50 gamma = [gammaL gammaR bprobe * ones(1, Nb)];
51 %%%%%%%%%%set sumSig %%%%%%%%%%%%%%%%%%%%%
52 sumSig = zeros(sizeH, sizeH);
53 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
54 for ii = 1 : Nsite
55     isite = sites(ii);
56     TempLen1 = sum(Base(1 : isite)) - Base(isite);
57     TempLen2 = sum(Base(1 : isite));
```

```
58      Len = TempLen2 - TempLen1;
59      sumSig(TempLen1 + 1 : TempLen2, TempLen1 + 1 : TempLen2) = ...
60          gamma(ii) * eye(Len);
61  end
62  %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
63  sumSig = -1i * sumSig / 2;
64  %%%%%%%%%%%% Transmission & DOS %%%%%%%%%%%%%%%%%%%%%
65  NE = length(Energy);
66  Teff = zeros(1, NE);
67  DOS = zeros(1, NE);
68  DOSBase = zeros(NBase, NE);
69  %%%%%%%%%%%%%%%%%%%%%%%%%%
70  for nE = 1 : NE
71      nE
72      E = Energy(nE)
73      Gr = ((E + 1i * eta) * eye(sizeH) - H0 - sumSig) \ eye(sizeH);
74      Ga = Gr';
75  %%%%%%%%%%%%%%%%%%%% DOS %%%%%%%%%%%%%%%%%%%%%%%%%%%%%
76       tempM = imag(diag(Gr));
77       DOS(nE) = -sum(tempM) / pi;
78
79       t1 = 1;
80       t2 = Base(1);
81
82       for nBase = 1 : NBase
83           DOSBase(nBase, nE) = -sum(tempM(t1 : t2)) / pi;
84           if nBase < NBase
85              t1 = t1 + Base(nBase);
86              t2 = t2 + Base(nBase + 1);
87           end
88       end
89  %%%%%%%%%% Transmission %%%%%%%%%%%%%
90  %%%%%%%%%% transmission between every 2 probes %%%%%%%%%%%%%
```

```
Tmat = zeros(Nsite, Nsite);
for ii = 2 : Nsite
    isite = sites(ii);
    TempLeni1 = sum(Base(1 : isite)) - Base(isite);
    TempLeni2 = sum(Base(1 : isite));
    Leni = TempLeni2 - TempLeni1;
    Gammai = gamma(ii) * eye(Leni);

    for jj = 1 : (ii - 1)
        jsite = sites(jj);
        TempLenj1 = sum(Base(1 : jsite)) - Base(jsite);
        TempLenj2 = sum(Base(1 : jsite));
        Lenj = TempLenj2 - TempLenj1;
        Gammaj = gamma(jj) * eye(Lenj);

        Tmat(ii, jj) = real(trace(Gammai * Gr(TempLeni1 + 1 : ...
            TempLeni2, TempLenj1 + 1 : TempLenj2) * Gammaj * ...
            Ga(TempLenj1 + 1 : TempLenj2, TempLeni1 + 1 : TempLeni2)));
        Tmat(jj, ii) = Tmat(ii, jj);
    end
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
TL = [];
TR = [];
for ii = 3 : Nsite
    TL = [TL Tmat(1, ii)];
    TR = [TR; Tmat(ii, 2)];
end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Wmat = zeros(Nb);

for ii = 1 : Nb
    for kk = 1 : Nsite
```

```
124             Wmat(ii, ii) = Wmat(ii, ii) + Tmat(ii + 2, kk);
125         end
126 %        Wmat(ii, ii) = Wmat(ii, ii) - Tmat(ii + 2, ii + 2);
127         for jj = ii + 1 : Nb
128             Wmat(ii, jj) = -Tmat(ii + 2, jj + 2);
129         end
130 end
131
132 Wmat = Wmat' + Wmat;
133 for ii = 1 : Nb
134     Wmat(ii, ii) = Wmat(ii, ii) / 2;
135 end
136 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
137 Td = Tmat(1, 2) + TL * (Wmat \ eye(Nb)) * TR
138 Teff(nE) = Td;
139 end
140 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
141 Tname = strcat(workdir, 'Teff_', strand, '_gammaL_', num2str(gammaL), ...
142     '_gammaR_', num2str(gammaR), '_bprobe_', num2str(bprobe), '.mat');
143 Dname = strcat(workdir, 'DOS_', strand, '_gammaL_', num2str(gammaL), ...
144     '_gammaR_', num2str(gammaR), '_bprobe_', num2str(bprobe), '.mat');
145 save(Tname, 'Energy', 'Teff')
146 save(Dname,'Energy','DOS','DOSBase')
147 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

Appendix B

ATOMIC COORDINATES FOR CYTOSINE AND METHYLATED CYTOSINE

The atomic coordinates for C, Cm, $G:C$ and $G:Cm$ are provided here.

B.1 Atomic coordinates for a single C nucleobase

1	N	-3.95100000	-2.40500000	26.64800000
2	C	-4.81500000	-1.35500000	26.60500000
3	H	-5.87600000	-1.53200000	26.51400000
4	C	-4.36400000	-0.07400000	26.67500000
5	H	-5.07000000	0.75500000	26.63900000
6	C	-2.95000000	0.10900000	26.79500000
7	N	-2.44400000	1.32500000	26.86800000
8	H	-1.44100000	1.41200000	26.95300000
9	H	-3.05100000	2.13200000	26.84000000
10	N	-2.11300000	-0.92400000	26.83700000
11	C	-2.58400000	-2.19800000	26.76500000
12	O	-1.83600000	-3.18600000	26.80000000
13	H	-4.28126762	-3.34747102	26.59631177

B.2 Atomic coordinates for a single Cm nucleobase

1	N	-3.95100000	-2.40500000	26.64800000
2	C	-4.81500000	-1.35500000	26.60500000
3	H	-5.87600000	-1.53200000	26.51400000

4	C	-4.36400000	-0.07400000	26.67500000
5	C	-5.31077921	1.10030563	26.62767464
6	H	-5.24924873	1.71389660	27.53693365
7	H	-5.10547767	1.75831156	25.77220787
8	H	-6.34630982	0.76083323	26.53497799
9	C	-2.95000000	0.10900000	26.79500000
10	N	-2.44400000	1.32500000	26.86800000
11	H	-1.44100000	1.41200000	26.95300000
12	H	-3.05100000	2.13200000	26.84000000
13	N	-2.11300000	-0.92400000	26.83700000
14	C	-2.58400000	-2.19800000	26.76500000
15	O	-1.83600000	-3.18600000	26.80000000
16	H	-4.28518790	-3.35865815	26.59569823

B.3 Atomic coordinates for a single $G:C$ base pair

1	N	-3.95100000	-2.40500000	26.64800000
2	C	-4.81500000	-1.35500000	26.60500000
3	H	-5.87600000	-1.53200000	26.51400000
4	C	-4.36400000	-0.07400000	26.67500000
5	H	-5.07000000	0.75500000	26.63900000
6	C	-2.95000000	0.10900000	26.79500000
7	N	-2.44400000	1.32500000	26.86800000
8	H	-1.44100000	1.41200000	26.95300000
9	H	-3.05100000	2.13200000	26.84000000
10	N	-2.11300000	-0.92400000	26.83700000
11	C	-2.58400000	-2.19800000	26.76500000
12	O	-1.83600000	-3.18600000	26.80000000
13	H	-4.28126762	-3.34747102	26.59631177
14	N	4.61700000	0.38600000	27.42600000
15	C	4.51500000	1.76200000	27.45200000
16	H	5.40500000	2.36700000	27.54300000

17	N	3.28600000	2.20700000	27.36000000
18	C	2.51900000	1.05200000	27.26800000
19	C	1.11300000	0.88800000	27.14600000
20	O	0.23900000	1.75100000	27.09500000
21	N	0.76100000	-0.46500000	27.08300000
22	H	-0.21100000	-0.68400000	26.99800000
23	C	1.64700000	-1.52200000	27.12900000
24	N	1.10700000	-2.74600000	27.05400000
25	H	0.12600000	-2.97400000	26.98000000
26	H	1.74800000	-3.52500000	27.09600000
27	N	2.96400000	-1.36700000	27.24100000
28	C	3.32300000	-0.05700000	27.30300000
29	H	5.43366560	-0.18841722	27.48169700

B.4 Atomic coordinates for a single $G:Cm$ base pair

1	N	-3.95100000	-2.40500000	26.64800000
2	C	-4.81500000	-1.35500000	26.60500000
3	H	-5.87600000	-1.53200000	26.51400000
4	C	-4.36400000	-0.07400000	26.67500000
5	C	-5.31077921	1.10030563	26.62767464
6	H	-5.24924873	1.71389660	27.53693365
7	H	-5.10547767	1.75831156	25.77220787
8	H	-6.34630982	0.76083323	26.53497799
9	C	-2.95000000	0.10900000	26.79500000
10	N	-2.44400000	1.32500000	26.86800000
11	H	-1.44100000	1.41200000	26.95300000
12	H	-3.05100000	2.13200000	26.84000000
13	N	-2.11300000	-0.92400000	26.83700000
14	C	-2.58400000	-2.19800000	26.76500000
15	O	-1.83600000	-3.18600000	26.80000000
16	H	-4.28518790	-3.35865815	26.59569823

17	N	4.61700000	0.38600000	27.42600000
18	C	4.51500000	1.76200000	27.45200000
19	H	5.40500000	2.36700000	27.54300000
20	N	3.28600000	2.20700000	27.36000000
21	C	2.51900000	1.05200000	27.26800000
22	C	1.11300000	0.88800000	27.14600000
23	O	0.23900000	1.75100000	27.09500000
24	N	0.76100000	-0.46500000	27.08300000
25	H	-0.21100000	-0.68400000	26.99800000
26	C	1.64700000	-1.52200000	27.12900000
27	N	1.10700000	-2.74600000	27.05400000
28	H	0.12600000	-2.97400000	26.98000000
29	H	1.74800000	-3.52500000	27.09600000
30	N	2.96400000	-1.36700000	27.24100000
31	C	3.32300000	-0.05700000	27.30300000
32	H	5.44242842	-0.19458072	27.48229463

VITA

Jianqing Qi was born in Jiangsu, China. She graduated from Nanjing University in 2007 with a Bachelor's degree in Physics. She then went to the Institute of Physics, Chinese Academy of Sciences for graduate studies and earned a Master's degree in 2010. She started pursuing her Ph.D in Department of Electrical Engineering at University of Washington, Seattle in Fall 2010, and obtained her Ph.D in 2015.