UNIVERSITY OF CALIFORNIA,
IRVINE

Computational Studies of DNA Hybridization and Self-Assembly of DNA-Based Nanowires via Molecular Dynamics Simulations

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Chemical and Biochemical Engineering

by

Cade Bryant Markegard

Dissertation Committee:
Professor Hung D. Nguyen, Chair
Professor Elizabeth Read
Professor Ray Luo

2015

DEDICATION

To my parents,
for always supporting and encouraging me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Thank you to my advisor, Professor Hung Nguyen, for his support and guidance throughout my PhD. I greatly appreciate the knowledge I have gained in completing this work. Thank you to my qualifying and dissertation committee, Professor Elizabeth Read, and Professor Ray Luo, for giving their time and knowledge to further my work. Additional thanks to Professor Alon Gorodetsky for the guidance and feedback on the PTCDI nanowires project.

A huge thanks to my labmates, Iris & Quynh, who definitely help make every day in lab a lot better. I'm going to miss sending random gifs and memes and then to hearing a laugh on the other side of the room. Thanks to the Zen Pad + QLA + Subsessions crews for the fun times at the pads, station, or shows. I'm going to deeply miss the random pho trips. Huge thanks to Peter for the support and friendship. A big thanks for all my friends who have helped keep me sane and work through this project.

Audrey, thank you for always being there, being understanding, and helping me during this PhD adventure. I couldn't have done it without you.

Thank you to my family who supported me pursue this dream of mine.

Finally, thank you to the University of California, Irvine for the financial support with which this work would not be possible.

CURRICULUM VITAE

Cade Bryant Markegard

Education

University of California, Irvine
PhD Chemical and Biochemical Engineering 2015
MS Chemical and Biochemical Engineering 2012

University of California, Santa Barbara
BS Chemical Engineering 2010

Research Experience

Graduate Student Researcher at University of California, Irvine 2011 – 2015
Dissertation advisor: Professor Hung D. Nguyen

RISE Intern at University of California, Santa Barbara 2009 – 2010
Research advisor: Professor Michael Gordon

FASTNET/CISEI Intern at Tyndall National Institute, Cork, Ireland 2009
Research advisor: Professor Martyn Pemble

Undergraduate Researcher at University of California, Santa Barbara 2008 – 2009
Research advisor: Professor Michael Gordon

Publications

1. C.B. Markegard, D. Cheng, and H.D. Nguyen, "Molecular Dynamics Simulations of a DNA Computing Threshold" (in prep).
2. C. B. Markegard, and H. D. Nguyen, "The Role of PEGylated Tail on Stacking of Perylenediimide DNA Base Surrogates" (in prep).
3. C.B. Markegard, D. Cheng, and H.D. Nguyen, "Effect of Sequence, Temperature, and DNA Concentration on dsDNA Self-Assembly via Coarse-Grained Simulations" (in prep).
4. C.B. Markegard, A. Mazaheripour, J.M. Jocson, A.M. Burke, A.A. Gorodetsky, and H.D. Nguyen, "Molecular Dynamics Simulations of Perylenediimde DNA Base Surrogates", (submitted).
5. C.B. Markegard, I.W. Fu, K.A. Reddy, H.D. Nguyen, "Coarse-Grained Simulation Study of Sequence Effects of DNA Hybridization in a Concentrated Environment", *J. Phys. Chem. B.* (2015).

6. I.W. Fu, C.B. Markegard, and H.D. Nguyen, "Solvent Effects on Kinetic Mechanisms of Self-Assembly by Peptide Amphiphiles via Molecular Dynamics Simulations", *Langmuir,* 31(1): 315-324 (2015)
7. B.K. Chu, I.W. Fu, C.B. Markegard, S.E. Choi, and H.D. Nguyen, "A Tail of Two Peptide Amphiphiles: Effect of Conjugation with Hydrophobic Polymer on Folding of Peptide Sequences", *Biomacromolecules*, 15(9): 3313-3320 (2014).
8. I.W. Fu, C.B. Markegard, B.K. Chu, and H.D. Nguyen, "Role of Hydrophobicity on Self-Assembly by Peptide Amphiphiles via Molecular Dynamics Simulations", *Langmuir*, 30(26): 7745-7754 (2014).
9. I.W. Fu, C.B. Markegard, B.K. Chu, and H.D. Nguyen, "The role of electrostatics and temperature on morphological transitions of hydrogel nanostructures self-assembled by peptide amphiphiles via molecular dynamics simulations", *Adv. Healthc. Mater.*, 10(2): 1388-1400 (2013).

Contributed Conference Presentations

1. AIChE Annual Meeting; Atlanta, GA; Nov. 2014
   Talk: "Sequence Effect on Self-Assembled DNA Nanostructures via Molecular Dynamics Simulations"
2. Foundations of Nanoscience (FNANO) 14; Snowbird, UT; April 2014
   Poster: "Elucidation of DNA Self-Assembly Kinetics via Molecular Dynamics Simulations"
3. AIChE Annual Meeting; San Francisco, CA; Nov. 2013
   Talk: "Molecular Dynamics Simulations of Self-Assembled DNA Nanostructures"
4. AIChE Annual Meeting; Pittsburgh, PA; Nov. 2012
   Talk: "Self-Assembly of Complex DNA Architectures via Multiscale Simulations for Nanotechnology and Alternative Energy"
5. AIChE Annual Meeting; Pittsburgh, PA; Nov. 2012
   Talk: "Molecular Dynamics Simulations of Peptide Self-Assembly Under Confinement"
6. UC Bioengineering Symposium; Berkeley, CA; June 2012
   Poster: "Molecular Dynamics Simulations of Peptide Self-Assembly Under Confinement"
7. UC Irvine CCBS Retreat; Santa Monica, CA; March 2012
   Poster: "Molecular Dynamics Simulations of Peptide Self-Assembly Under Confinement"
8. GRC on Protein Folding Dynamics; Ventura, CA; Jan. 2012
   Poster: "Molecular Dynamics of Peptide Self-Assembly Under Confinement"
9. AIChE Annual Meeting; Minneapolis, MN; Oct. 2011
   Talk: "Molecular Dynamics Simulations of 2D Self-Assembled Peptide Monolayers"

ABSTRACT OF THE DISSERTATION

Computational Studies of DNA Hybridization and Self-Assembly of DNA-Based Nanowires via Molecular Dynamics Simulations

By

Cade Bryant Markegard

Doctor of Philosophy in Chemical and Biochemical Engineering

University of California, Irvine, 2015

Professor Hung D. Nguyen, Chair

DNA nanotechnology has been a rapidly growing field with many applications in drug delivery, energy, and molecular computing. First, a novel coarse-grained model of DNA is developed for elucidating the effects of sequence and environmental factors on DNA hybridization. This coarse-grained model has been integrated into our recently developed simulation package called BioModi (Biomolecular Multiscale Models at UC Irvine), enabling the study of the molecular interactions of large systems consisting of amino acids, nucleic acids, and polymers over long time scales. Second, for a design of organic nanowires, a DNA base surrogate, perylene-3,4,9,10-tetracarboxylic diimide, is parameterized and simulated via all-atom simulations to understand kinetic mechanisms of stacking and structural arrangement at equilibrium for optimized transfer of electrical charge.

DNA hybridization processes in a crowded environment are investigated using Biomodi. This model has been shown to capture the mechanical and thermal melting properties of DNA as compared to experimental data. Moreover, the effect of sequence, temperature and DNA concentration is elucidated in detail on the self-assembly process of

DNA, giving insight to choosing the right environmental conditions for the formation of double helices. Furthermore, DNA hybridization processes demonstrate many kinetic pathways that are dependent on the sequence.

The overlapping π system of stacked natural DNA bases can mediate the transfer of electrical charge over long distances. To enhance such electron transport property, novel nanowires can be designed by replacing the sugar group and natural DNA bases with perylene-3,4,9,10-tetracarboxylic diimides. Therefore, it is necessary to understand the stacking kinetics of such perylenediimide DNA base surrogates and their equilibrium structures. Simulations demonstrate the different kinetic pathways for stacking depending on the length of the nanowire even when the base is PEGylated. Equilibrium structures obtained are shown to be similar to expected structures found in experiments.

Implementation of our newly developed coarse-grained models, BioModi, and insight gained from our simulations will provide key parameters and understanding to advance computer-aided design and development of innovative smart biomaterials. Moreover, our parameterization, simulation and analysis methods developed for DNA base surrogates will enable the creation and design of organic nanowires with enhanced electron transport properties.

CHAPTER 1 Introduction

1.1 Motivation

Friedrich Miescher discovered deoxyribonucleic acid (DNA) in 1868 while determining the chemical composition of cells [1]. It was not until 1943 when Avery, MacLeod, and McCarty demonstrated that DNA was used as a hereditary material in cells rather than the originally thought proteins [2]. Watson and Crick proposed the double-helix structure of DNA in 1953 to provide an understanding on structural biology of DNA [3]. These discoveries have fueled a larger scientific community to further investigate DNA for genetics and medicine. Moreover, DNA nanotechnology has been a rapidly developing field with many potential applications for biofuels, biomaterials, and molecular computing.

DNA nanotechnology has been evolving quickly since Seeman first published a paper that proposed using Holliday junction structures, each of which is composed of four strands of DNA (**Figure 1.1A**), to build large 2D DNA lattices [6-7]. Moreover, various groups have developed methods for using the lattices for arranging metallic nanoparticles and detecting proteins [8-9]. Recently these methods have been further advanced to assemble 3D structures such as a tensegrity triangle [10]. Furthermore, Rothemund developed DNA origami via a bottom-up method of using scaffold and staple strands to self-assemble squares, rectangles, stars, smiley faces, and triangles [11]. This method has been widely popular in engineering a variety of structures and the software caDNAno developed by Douglas enables researchers to design a variety of structures by selecting the correct DNA sequences [12-13]. Through DNA origami, researchers have engineered DNA

nanopores in lipid membranes, molecular transporters, gears, and beams [13-18] **(Figure 1.1B)**. Although experimentalists can create almost any nanostructure they desire, the kinetic mechanisms and thermodynamics of such assembly processes are not well elucidated. Through the use of molecular dynamics simulations, these mechanisms can be investigated to reduce the kinetic trapping events of unwanted stable by-products.

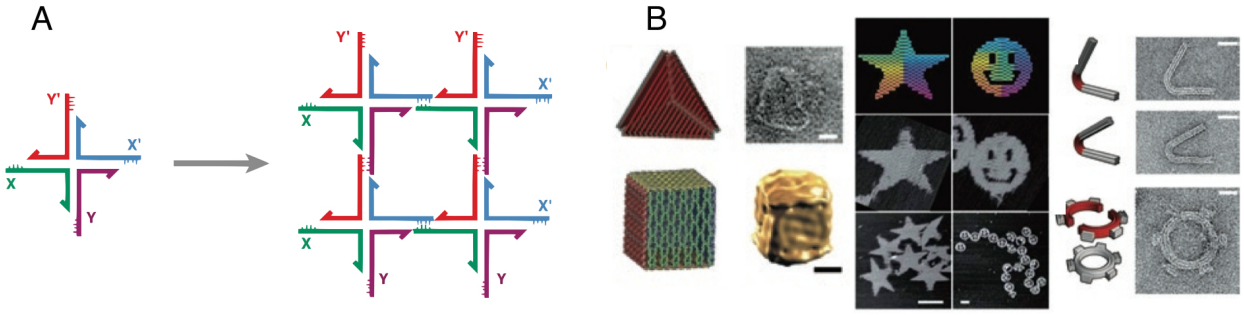

Figure 1.1: A. Holliday junction inspired 2D lattice developed by Seeman *et al*. Sticky ends enable the assembly of a 2D lattice. **B.** Various self-assembled DNA origami structures. DNA origami can be used to create complex structures such as gears, beams, stars, smiley faces, and boxes.

All-atom force fields such as AMBER or CHARMM offer full atomistic detail [19-20], though at a computational cost. These force fields have been employed in a large number of DNA related studies such as DNA transport through synthetic nanopores [21], DNA supercoils [22], DNA origami structure and dynamics [23], and conformational changes from B-form to A-form [24]. Currently, only short sequences (8 nucleotides) have been studied using advanced sampling techniques to measure thermal melting conditions [25]. It is not currently computationally feasible to use all-atom approaches to study the hybridization and self-assembly DNA due to the long time scales and rarity of events. Since the time scales and system sizes are incredibly large for self-assembly events, researchers have developed coarse-grained models to address these problems. Coarse-grained models have been used to study hybridization, self-assembly of double-stranded DNA and Holliday

junctions [26-29], DNA walkers [30], and crystallization of DNA modified nanoparticles lattices [31]. While these models have been used to study the aforementioned systems, the models have failed to (1) accurately represent the physical aspects of DNA, (2) be dependent on the environmental conditions (e.g. salt concentration), (3) accommodate large systems involving multiple types of strands, and (4) be compatible with coarse-grained models of proteins for studying protein-DNA interactions. Therefore, in this work a coarse-grained model is developed to closely represent the physical properties of DNA and accommodate large system sizes. Long time scales are accessed through the use of an event-driven molecular dynamics algorithm [32-33], which has been used to simulate peptide and peptide amphiphile self-assembly [34-35]. Using this nucleotide model, the hybridization and self-assembly of many strands chains of DNA are explored to further understand the thermodynamics and kinetics of the system.

Due to the ease of synthesizing DNA using solid-state supports, organic chemists have synthesized unnatural bases that can utilize the phosphate backbone of DNA [36]. For example, groups have developed libraries of artificial nucleosides as seen in **Figure 1.2** [37]. Recently, the Gorodetsky group at UC Irvine has synthesized perylene attached to the same phosphate backbone as DNA to create a perylene-based nanowire. Since DNA bases are similar in nature to the perylene π-conjugated base they are believed to stack as a single strand in a wire-like conformation. DNA bases have been shown to conduct electric charge [38-39], thus this wire formation offers a competitive edge in studying charge transport of other π-stacked organic molecules. In this work, a perylene conjugated to a DNA backbone is parameterized for all-atom molecular dynamics simulations through quantum mechanics calculations. Advanced sampling methods for molecular dynamics

simulations give insight into the atomic structures of these nanowires. The lowest free energy structures are taken and the molecular orbital structure is evaluated to find the location and energy of the molecular orbitals for electron transfer. Results from this work help determine the structure of the nanowires that are not easily observable to experimentalist.

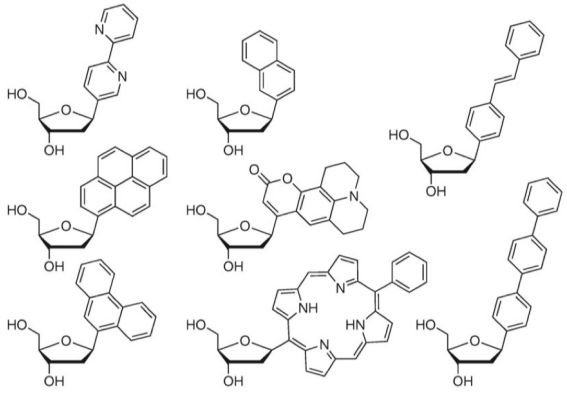

Figure 1.2: Synthetic nucleosides developed for insertion into DNA structures. Unnatural bases are able to stack within the double-helix structure [37].

1.2 DNA Nanotechnology: Structure and Applications

Seeman in 1982 developed a method of designing stable immobile junctions inspired by the four-armed Holliday junction [6-7]. By learning from genetic engineering at the time, he synthesized strands that had sticky ends to create a lattice structure using the immobile junctions. Fu and Seeman then developed double-crossover (DX) molecules in which strands of DNA crossover between nearby strands twice (**Figure 1.3A**) [40]. It was later found by Sa-Ardyen *et al*. that DX molecules have twice the persistence length of dsDNA [41], thus showing a path of creating very rigid DNA based structures. LaBean *et al*. developed a triple-crossover (TX) structure using four strands to form three double-helices [42] (**Figure 1.3B**). Lattice arrays can be formed from DX and TX structures, however TX

arrays have larger spacing between the molecules, thus enabling these sites to be used for tethering molecules. Additionally, it was found that TX structures could be used for molecular computation [42]. These crossovers and sticky ends have been applied to self-assemble 3D structures such as the tensegrity triangle that formed a crystal lattice [10] (**Figure 1.3C**). Furthermore, DNA has been applied to making non-array structures through DNA origami.

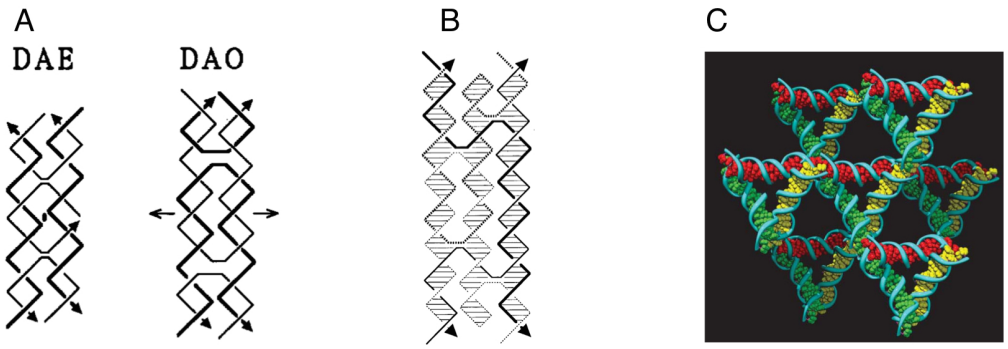

Figure 1.3: A. Two double crossover structures from Fu and Seeman. [40] **B.** Triple crossover structure from LaBean *et al*. [41] **C.** Tensegrity triangle lattice designed by Zheng *et al*. [10]

Rothemund developed DNA origami via a method using a long scaffold strand with smaller staple strands that help the scaffold fold into a desired structure [11]. Furthermore, this method has been applied in assembling structures such as smiley faces, gears, nanopores, triangles, and rods [11, 14, 43]. Researchers have developed methods for determining appropriate sequences to use at the thermodynamic level in order to reduce misfolded structures. Douglas has developed the program caDNAno to aid in the design of DNA origami structures. The program is freely available and is used as a plugin for Maya, the popular 3D CAD program [13]. Additionally, the CanDO (Computer-aided engineering for DNA origami) program was developed to give researchers feedback on the design of

their nanostructures made in caDNAno before experiments are performed [12, 44]. DNA origami has been applied to problems such molecular drug delivery systems and arrangement of molecules. For example, Andersen *et al*. designed a box, which had a lid capable of opening, and closing depending on the environment. This box (**Figure 1.4A**) is big enough to contain a ribosome or virus structure for delivery to a cell [43]. Similarly, Douglas *et al*. used a self-assembled molecular carrier (**Figure 1.4B**) to carry gold nanoparticles or antibody fragments into cells. A logic gate, using DNA aptamers, was used to control when the carrier would release the contents into the cell [16]. Moreover, Jiang *et al*. used 2D and 3D structures covalently attached to doxorubicin, a cancer medication (**Figure 1.4C**). The structures were found highly efficient in delivering the drug to cancer cells, even those that are resistant to the drug [45]. DNA origami has also been applied to the self-assembly of Holliday junction inspired 4x4 structure with sticky ends. The structures assemble into a lattice structure similar to those of Seeman (**Figure 1.4D**). The assembled structures have been used as a protein arrays and as a scaffold for creating silver nanowires [46]. Therefore, DNA origami offers scientists and engineers the ability of designing nanostructures for a wide range of applications.

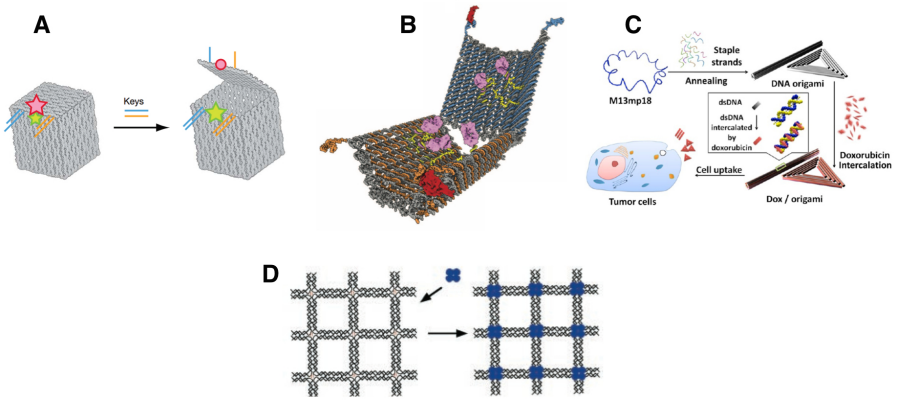

Figure 1.4: Applications of DNA origami structures. **A.** DNA box designed by Andersen *et al*. designed to open and close to hold small biological fragments [43]. **B.** Nanorobot used to carry molecules into cells designed by Douglas *et al*. [16]. The nanorobot has DNA aptamer ends to have controlled release of the molecules contained. **C.** Schematic of DNA origami coated with doxorubicin for cellular uptake into tumor cells designed by Jiang *et al*. [45] **D.** Protein array created by Yan *et al*. [46]. Array is similar in nature to Holliday junction lattices, however much larger using a scaffold strand and many staple strands. The arrays have also been used as templates for silver nanowires.

1.3 Coarse-Grained Nucleic Acid Models

Atomic-resolution simulations of DNA hybridization and assembly are computationally expensive due to the large system sizes required and the long timescales involved. As a consequence, researchers have employed coarse-grained models to study the kinetics and thermodynamics of nucleic acids. However, there is a large variety of coarse-grained models of nucleic acids ranging from one-bead models representing many nucleotides to more detailed models with six-beads representing one nucleotide [47-54]. Models in this field are designed to capture details such as the local dynamics of DNA, thermodynamics of melting, structural details such as persistence length, or hybridization kinetics. Below is an overview of some of the models that are capable of simulating DNA hybridization or assembly and their drawbacks.

Drukker *et al*. in 2001 developed a two-bead model, one representing the sugar and

phosphate sites, and one representing the base [47]. This model included important interactions such as hydrogen bonding, bending and torsion energies. Since this model implemented molecular dynamics, instead of the Monte Carlo method, it was the first to study the dynamics of the denaturation of DNA. The researchers were able to capture the sequence and chain length effects on thermal melting, and investigated the effect of mismatched base pairs on the melting. From their results they showed that the mismatches towards the end of strands does not affect the melting temperature as much as having mismatches in the middle of two strands. This indicates that melting is highly dependent on the core of the dsDNA being base paired. Another two-bead model is that from Ouldridge *et al*. to study DNA for nanotechnology purposes (**Figure 1.5A**) [51]. One drawback of the model is that it currently is parameterized for only 0.5 M NaCl, therefore investigating systems with different salt concentrations is not currently available. Also, the model is unable to capture the major and minor grooves of B-DNA due to the limitation of only two beads. This model has been very successful in exploring hybridization of dsDNA using forward flux sampling; researchers have found the rate constants from simulations compare favorably to ones obtained experimentally [27]. Ouldridge *et al*. have also investigated problems such as DNA tweezers, DNA walkers, and DNA stretching [30, 55, 56]. Other groups have used this model to investigate the self-assembly of dsDNA with comparisons of equilibrium states to liquid phase theories, and DNA nanostars [28, 57].

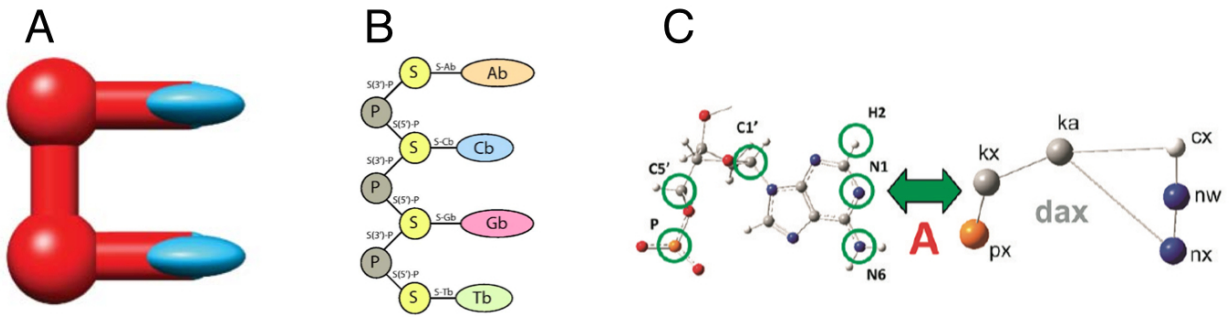

Figure 1.5: **A**. Two-bead model, oxDNA, used by Ouldridge *et al*. **B.** 3SPN model representation used by de Pablo and Dokholyan research groups. **C.** Direct mapping of coarse-grained atoms from all-atom simulations used by Dans *et al*. Enables the translation from coarse-grained model to all-atom representations.

3-site-per-nucleotide models have been developed by both the dePablo and Dokholyan research groups (**Figure 1.5B**). Knotts *et al*. first introduced the three-site model for DNA in which the phosphate, sugar, and base are each represented by one bead. [52] This initial model was able to accurately predict melting temperatures, however the persistence length was determined to be a factor of 2 smaller than experimental values. Sambriski *et al*. updated the model to more accurately capture the persistence length of dsDNA compared to experimental values and improved upon replicating experimental melting curves [53]. However, single-strand persistence length was greatly overestimated. Additionally, the hybridization of two single strands was investigated and the researchers determined that the two observed mechanisms were sequence dependent [26, 58]. This model has also been applied to studying the hybridization of strands in solution to strands bonded to surfaces. The researchers found that having one strand bound to the surface changes the hybridization kinetics by limiting how the strands can reorient themselves to successfully base pair [59-60]. Recently, Hinckley *et al*. updated the model even further being able to capture both dsDNA and ssDNA persistence lengths [54]. Additionally, using forward-flux sampling [61], the researchers were able to obtain similar rate constants for

hybridization as those found in experiments. Ding *et al*. from the Dokholyan group used a similar 3-site-per-nucleotide model with discontinuous molecular dynamics (DMD) to study the folding of RNA molecules. This implementation of a 3-bead model was able to predict the structure of many RNA molecules within 4 Å of the experimental crystal structure [49].

Other groups have developed more detailed coarse-grained models such as that from Dans *et al*. [50] This model uses six beads to represent a nucleotide and maps the beads directly from all-atom (AA) representation making it fairly simple to switch between coarse-grained and AA. Dans *et al*. applied the model to capture the conformational transition from A-form to B-form DNA, and the model is able to capture the melting temperature effect for ionic strength and sequence dependence.

1.3 Organic Semiconductors and Perylene Based Molecules

Organic semiconductors offer a significant advantage over crystalline semiconductors due to their ease of processing, flexibility, and low cost. They obtain semiconductor-like properties due to hybridized orbitals, sp^2, thus having both *s* and *p* character. Conjugated molecules are excellent examples of organic semiconductors, as their highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) act as a band gap for semiconductor behavior. In solid state physics terms, the HOMO is similar to the valence band and the LUMO is similar to the conduction band [62-63]. Charge transport in organic semiconductors is understood to occur through a hopping mechanism in which a charge hops from the HOMO of one molecule to the LUMO of another [64]. The effective charge transfer rate depends on the intermolecular distances of the π-

stacking, and the overlap of the molecular orbitals (greater overlap leads to better charge transfer) [62-63].

One molecule class is of great interest for this work is perylene tetracarboxylic diimides (PTCDI), due to their ability to conduct charge, and be functionalized (thus changing their optical and conducting properties) [65]. Klebe *et al.* investigated the packing of many perylene-based compounds (not containing a phosphate backbone) to understand how it affects the color of the dye. It was found that most have a stacking distance of ~3.35 Å, with a longitudinal shift of ~ 1.14 Å and no transverse shift [66]. Therefore perylene-based molecules like to stack with the π -conjugated segments on top of one another. An equation relating the maximum absorption (from UV/Vis) for the perylene molecules was determined and takes into account the transverse and longitudinal offsets. Since the absorption is dependent on the offsets, the packing of the molecules changes the electronic and optical properties of the material. Kazmaier *et al*. used semi-empirical quantum mechanics calculations to analyze the electronic properties of stacked perylene structures. They found a nodal structure of the valence and conduction bandwidth depending on the longitudinal and transverse offsets [67]. Delgado *et al*. then used density functional theory (DFT) to investigate the tuning of the charge-transport of PTCDIs by changing the functional groups. The researchers investigated single and packed substituted molecules; finding that intramolecular reorganization energy for electrons increases for both core and end substituted PTCDIs. An increase in the reorganization energy for holes can be achieved by substituting the perylene core with halogens. Additionally, transfer integrals were found to be nodal and depend on the transverse and longitudinal offsets [68]. Geng *et al*. used quantum calculations of many perylene compounds and found that imide substituents

could be used to change the intermolecular packing. This results with no change to the geometry and electrical properties of the actual conjugated core. Also found was that substituents at the core of the molecule changes the electronic and geometrical properties of the core and thus the intermolecular packing. Therefore, it is important to determine how the perylene core packs to determine its electronic properties. The properties can be tuned as shown by the previous quantum mechanics calculations by changing perylene core substituents [65].

1.4 Structure of Dissertation

The goal of this dissertation is to elucidate the kinetics and thermodynamics of DNA hybridization and self-assembly of DNA base-surrogate nanowires. In Chapter 2, a novel coarse-grained model is developed and shown to capture the mechanical and thermal melting properties of DNA. DNA hybridization and large scale self-assembly to double-helices are investigated at constant temperature. We show the presence of three hybridization mechanisms, zippering, slithering and strand displacement in the assembly towards double-helices. Additionally, it is found that DNA sequence affects yields of double-helices. In Chapter 3, the coarse-grained model is used to investigate the effects of sequence and environmental conditions on the kinetic mechanisms of double-helix assembly at high concentration. An analysis method is presented to delve deeper into those factors that affect aggregate formation and lifetimes. Kinetic modeling of the double-helix formation demonstrates activation energies similar to those found experimentally.

In Chapter 4, a DNA base-surrogate, perylene-3,4,9,10-tetracarboxylic diimides (PTCDIs), is constructed into a nanowire. Parameters for the PTCDI nanowires are

determined and simulated using all-atom molecular dynamics. The kinetic mechanisms of PTCDI stacking are investigated for PTCDI nanowires. Additionally, equilibrium structures are found to be similar to those expected from experimental evidence. In Chapter 5, the PTCDI is PEGylated and parameterized for all-atom simulations. Simulations are conducted to understand the effect of the PEG chain on the kinetic mechanisms of PTCDI stacking. It is found PTCDI stacking at equilibrium is not affected by the added PEG chain, however the PEG chain adds complexity in how PTCDIs stack. In Chapter 6, a summary of the dissertation is provided and some future directions for the projects presented.

1.5 References

1. R. Dahm. Friedrich Miescher and the Discovery of DNA. *Developmental Biology* 278, 274-288 (2005).
2. O. T. Avery, C. M. MacLeod, and M. McCarty. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by Desoxyribonucelic Acid Fraction Isolated from Pneumococcus Type III. *Journal of Experimental Medicine* 79, 137-158 (1943).
3. J. D. Watson, and F. H. C. Crick. Molecular Structure of Nucleic: A Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737-738 (1953).
4. N. C. Seeman. Structural DNA Nanotechnology: Growing Along with *Nano Letters*. *Nano Letters* 10, 1971-1978 (2010).
5. N. C. Seeman. Nanomaterials Based on DNA. *Annu. Rev. Biochem.* 79, 65-87 (2010).
6. N. C. Seeman. Nucleic Acid Junctions and Lattices. *J. Theor. Biol.* 99, 237-247 (1982).
7. N. C. Seeman and N. R. Kallenbach. Design of Immobile Nucleic Acid Junctions. *Biophys. J.* 44, 201- 209 (1983).
8. J. D. Carter and T. H. LaBean. Organization of Inorganic Nanomaterials via Programmable DNA Self-Assembly and Peptide Molecular Recognition. *ACS Nano* 5, 2200-2205 (2011).
9. C. Lin, E. Katilius, Y. Liu, J. Zhang, and H. Yan. Self-Assembled Signaling Aptamer DNA Arrays for Protein Detection. *Angew. Chem. Int. Ed.* 45, 5296-5301 (2006).
10. J. Zheng, J. J. Birktoft, Y. Chen, T. Wang, R. Sha, P. E. Constantinou, S. L. Ginell, C. Mao and N. C. Seeman. From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* 461, 74-77 (2009).
11. P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patters. *Nature* 440, 297-302 (2006).
12. C. E. Castro, F. Kilchherr, D-N. Kim, E. L. Shiao, T. Wauer, P. Wortmann, M. Bathe, and H. Dietz. A primer to scaffolded DNA origami. *Nature Methods* 8, 221-229 (2011).
13. S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, and W. M. Shih. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Research* 37, 5001-5006 (2009).
14. H. Dietz, S. M. Douglas, and W. M. Shih. Folding DNA into Twisted and Curved Nanoscale Shapes. *Science* 325, 725-730 (2009).
15. S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. H. Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* 459, 414-418 (2009).
16. S. M. Douglas, I. Bachelet, and G. M. Church. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* 335, 831-834 (2012).

17. J. R. Burns, E. Stulz, and S. Howorka. Self-Assembled DNA Nanopores That Span Lipid Bilayers. *Nano Letters* 13, 2351-2356 (2013).
18. J. R. Burns, K. Göpfrich, J. W. Wood, V. V. Thacker, E. Stulz, U. F. Keyser, and S. Howorka. Lipid- Bilayer-Spanning DNA Nanopores with a Bifunctional Porphyrin Anchor. *Angew. Chem. Int. Ed.* 52, 12069-12072 (2013).
19. R. Salomon-Ferrer, D. A. Case, and Ross C. Walker. An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* 3, 198-210 (2013).
20. B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Calfisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. IM, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. CHARMM: The Biomolecule Simulation Program. *J. Comput. Chem.* 30, 1545-1614 (2009).
21. A. Aksimentiev, J. B. Heng, G. Timp, and K. Schulten. Microscopic Kinetics of DNA Translocation through Synthetic Nanopores. *Biophysical Journal* 87, 2086-2097 (2004).
22. T. D. Lillian, M. Taranova, J. Wereszczynski, I. Andricioaei, and N. C. Perkins. A Multiscale Dynamic Model of DNA Supercoil Relaxation by Topoisomerase IB. *Biophysical Journal*, 100, 2016- 2023 (2011).
23. J. Yoo and A. Aksimentiev. In situ structure and dynamics of DNA origami determined through molecular dynamics simulations. *PNAS* 110, 20099-20104 (2013).
24. A. Noy, A. Pérez, C. A. Laughton, and M. Orozco. Theoretical study of large conformational transitions in DNA: the B <->A conformational change in water and ethanol/water. *Nucleic Acids Research* 35, 3330-3338 (2007).
25. S. Kannan, and M. Zacharias. Simulation of DNA double-strand dissociation and formation during replica-exchange molecular dynamics simulations. *Phys. Chem. Chem. Phys.* 11, 10589-10595 (2009).
26. E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo. Uncovering pathways in DNA oligonucleotide hybridization via transition state analysis. *PNAS* 106, 18125-18130 (2009).
27. T. E. Ouldridge, P. Sulc, F. Romano, J. P. K. Doye, DNA hybridization kinetics: zipping, internal displacement, and sequence dependence. *Nucleic Acids Research* 41, 8886-8895 (2013).
28. C. De Michele, L. Rovigatti, T. Bellini, and F. Sciortino. Self-assembly of short DNA duplexes: from a coarse-grained model to experiments through a theoretical link. *Soft Matter* 8, 8388-8398 (2012).
29. T. E. Ouldridge, I. G. Johnson, A. A. Louis, and J. P. K. Doye. The self-assembly of DNA Holliday junctions studied with a minimal model. *J. Chem. Phys.* 130, 065101 (2009).
30. T. E. Ouldridge, R. L. Hoare, A. A. Louis, J. P. K. Doye, J. Bath, A. J. Turberfield. Optimizing DNA Nanotechnology through Coarse-Grained Modeling: A Two-Footed DNA Walker. *ACS Nano* 7, 2479-2490 (2013).
31. T. I. N. G. Li, R. Sknepnek, R. J. Macfarlane, C. A. Mirkin, and M. Olvera de la Cruz. Modeling the Crystallization of Spherical Nucleic Acid Nanoparticle Conjugates with Molecular Dynamics Simulations. *Nano Letters* 12, 2509-2514 (2012).
32. B. J. Alder and T. E. Wainwright. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* 31, 459-466 (1959).
33. S. W. Smith, C. K. Hall and B. D. Freeman. Molecular Dynamics for Polymeric Fluids Using Discontinuous Potentials. *J. Comput. Phys.* 134
34. H. D. Nguyen, and C. K. Hall. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *PNAS* 101, 16180-16185 (2004).
35. I. W. Fu, C. B. Markegard, B. K. Chu, and H. D. Nguyen. The Role of Electrostatics and Temperature on Morphological Transitions of Hydrogel Nanostructures Self-Assembled by Peptide Amphiphiles via Molecular Dynamics Simulations. *Adv. Healthcare Mater*. 2, 1388-1400 (2013).
36. H-A. Wagenknecht. Synthetic Oligonucleotide Modifications for the Investigation of Charge Transfer and Migration Processes In DNA. *Current Organic Chemistry* 8, 251-266 (2004).
37. T. J. Bandy, A. Brewer, J. R. Burns, G. Marth, T. Nguyen, and E. Stulz. DNA as a supramolecular scaffold for functional molecules: progress in DNA nanotechnology. *Chem. Soc. Rev.* 40, 138-148 (2011).
38. D. D. Eley and D. I. Spivey. Semiconductivity of Organic Substances. Part 9—Nucleic Acid in the Dry State. *Trans. Faraday Soc.* 58, 411-415 (1961).
39. E. M. Boon and J. K. Barton. Charge transport in DNA. *Curr. Opin. Struct. Biol.* 12, 320-329 (2002).
40. T. Fu and N. C. Seeman. DNA Double-Crossover Molecules. *Biochemistry* 32, 3211-3220 (1993).

41. P. Sa-Ardyen, A. V. Vologodskii, N. C. Seeman. The flexibility of DNA double crossover molecules. *Biophysical Journal* 84, 3829-3837 (2003).
42. T. H. LaBean, H. Yan, J. Kopatsch, F. Liu, E. Winfree, J. H. Reif, and N. C. Seeman. Construction, Analysis, Ligation, and Self-Assembly of DNA Triple Crossover Complexes. *J. Am. Chem. Soc.* 122, 1848-1860 (2000).
43. E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, and J. Kjems. Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* 459, 73-77 (2009).
44. CanDO: http://cando-dna-origami.org.
45. Q. Jiang, C. Song, J. Nangreave, X. Liu, L. Lin, D. Qiu, Z-G. Wang, G. Zou, X. Liang, H. Yan, and B. Ding. DNA Origami as a Carrier for Circumvention of Drug Resistance. *J. Am. Chem. Soc.* 134, 13396-13403 (2012).
46. H. Yan, S. H. Park, G. Finkelstein, J. H Reif, and T. H. LaBean. DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires. *Science* 301, 1882-1884 (2003).
47. K. Drukker, G. Wu, G. C. Schatz. Model simulations of DNA denaturation dynamics. *J. Chem. Phys*. 114, 579-590 (2001).
48. A. Savelyev and G. A. Papoian. Chemically accurate coarse graining of double-stranded DNA. *PNAS* 107, 20340-20345 (2010).
49. F. Ding, S. Sharma, P. Chalasani, V. V. Demidov, N. E. Broude, and N. V. Dokholyan. Ab initio RNA folding by discrete molecular dynamics: From structure prediction to folding mechanisms. *RNA* 14: 1164-1173 (2008).
50. P. D. Dans, A. Zeida, M. R. Machado, and S. Pantano. A Coarse Grained Model for Atomic- Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.* 6, 1711-1725 (2010).
51. T. E. Ouldridge, A. A. Louis, and J. P. K. Doye. Structural, mechanical, and thermodynamics properties of a coarse-grained DNA model. *J. Chem. Phys*. 134, 085101 (2011).
52. T. A. Knotts, N. Rathore, D. C. Schwartz, and J. J. de Pablo. A coarse grain model for DNA. *J. Chem. Phys.* 126, 084901 (2007).
53. E. J. Sambriski, D. C. Schwartz, and J. J. de Pablo. A Mesoscale Model of DNA and Its Renaturation. *Biophysical Journal* 96, 1675-1690 (2009).
54. D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* 139, 144903 (2013).
55. T. E. Ouldridge, A.A. Louis, and J. P. K. Doye. DNA Nanotweezers Studied with a Coarse- Grained Model of DNA. *PRL* 104, 178101 (2010).
56. F. Romano, D. Chakraborty, J. P. K. Doye, T. E. Ouldridge, A. A. Louis. Coarse-grained simulations of DNA overstretching. *J. Chem. Phys*. 138, 085101 (2013).
57. L. Rovigatti, F. Bomboi, and F. Sciortino. Accurate phase diagram of tetravalent DNA nanostars. *J. Chem. Phys.* 140, 154903 (2014).
58. E. J. Sambriski, V. Ortiz, and J. J. de Pablo. Sequence effects in the melting and renaturation of short DNA oligonucleotides: structure and mechanistic pathways. *J. Phys.: Condens. Matter* 21, 034105 (2009).
59. T. J. Schmitt and T. A. Knotts. Thermodynamics of DNA hybridization on surfaces. *J. Chem. Phys*. 134, 2505105 (2011).
60. T. J. Schmitt, J. B. Rogers and T. A. Knotts. Exploring the mechanisms of DNA hybridization on a surface. *J. Chem. Phys.* 138, 035102 (2013).
61. F. A. Escobedo, E. E. Borrero, and J. C. Araque. Transition path sampling and forward flux sampling. Applications to biological systems. *J. Phys.: Condens. Matter.* 21, 333101 (2009).
62. *Organic Field-Effect Transistors*. Zhenan Bao and Jason Locklin. CRC Press 2007.
63. *Organic Electronics II: More Materials and Applications*. Chapter 3, R. Noriega, and A. Salleo. Edited by Hgen Klauk. Wiley-VCH Verlag GmbH & Co. KGaA. 2012.
64. J. Hirsch. Hopping transport in disordered aromatic solids: a re-interpretation of mobility of measurements on PNK and TNF. *J. Phys. C*: *Solid State Physics*. 12, 321-335 (1979).
65. Y. Geng, H-B. Li, S-X. Wu, and Z-M Su. The interplay of intermolecular interactions, packing motifs, and electron transport properties in perylene diimide related molecules: a theoretical prospective. *J. Mater. Chem*. 22, 20840-20851 (2012).

66. G. Klebe, F. Graser, E. Hädicke, and J. Berndt. Crystallochromy as a Solid-State Effect: Correlation of Molecular Conformation, Crystal Packing and Colour in Perylene-3,4:9,-bis(dicarboximide) Pigments. *Acta Cryst.* B45, 69-77 (1989).
67. P. M. Kazmaier and R. Hoffmann. A Theoretical Study of Crystallochromy. Quantum Interference Effects in the Spectra of Perylene Pigments. *J. Am. Chem. Soc*. 116, 9684-9691 (1994).
68. M. C. R. Delgado, E.-G. Kim, D. A. da Silva Filho, and J-L Bredas. *J. Am. Chem. Soc.* 132, 3375- 3387 (2010).

CHAPTER 2 Coarse-Grained Simulation Study of Sequence Effects of DNA Hybridization in a Concentrated Environment

2.1 Abstract

A novel coarse-grained model is developed to elucidate thermodynamics and kinetic mechanisms of DNA self-assembly. It accounts for sequence and solvent conditions to capture key experimental results such as sequence-dependent thermal property and salt-dependent persistence length of ssDNA and dsDNA. Moreover, constant-temperature simulations on two single strands of a homogeneous sequence show two main mechanisms of hybridization: a slow slithering mechanism and a one-order faster zippering mechanism. Furthermore, large-scale simulations at a high DNA strand concentration demonstrate that DNA self-assembly is a robust and enthalpically driven process in which the formation of double helices is deciphered to occur via multiple self-assembly pathways including the strand displacement mechanism. However, sequence plays an important role in shifting the majority of one pathway over the others and controlling size distribution of self-assembled aggregates. This study yields a complex picture on the role of sequence on programmable self-assembly and demonstrates a promising simulation tool that is suitable for studies in DNA nanotechnology.

2.2 Introduction

The field of DNA nanotechnology has been evolving quickly since Seeman [1] first proposed using Holliday junction structures composed of four strands of DNA to build large 2D lattices [2-3], which was later realized for arranging and detecting metallic

nanoparticles and proteins [4-5.] Further advances have been made to fabricate 3D structures such as tetrahedrons [6-7] and nanotubes [8]. Additionally, squares, stars, smiley faces [9] and a box with a controllable lid [10] can be assembled via DNA origami folding by using scaffold and staple strands. Current experimental approaches of programmable DNA self-assembly involves a slow annealing procedure in which all strands are mixed (1) altogether in 'one pot', (2) in hierarchical steps, or (3) with a long scaffolding strand to undergo origami folding. [11] However, these approaches often encounter multiple defects or mismatch steps, which greatly affect the degree of perfection of the final self-assembled nanostructures, possibly impairing morphology and function of the construct [12]. Therefore, an understanding on optimization of the physical conditions under which self-assembly occurs is valuable in minimizing errors that occur during the assembly process in order to improve the efficiency of nanostructure formation [11-12]. Through the use of molecular dynamics (MD), self-assembly kinetic mechanisms and thermodynamics can be investigated to reduce kinetic traps of unwanted by-products. This can potentially elucidate not only slow annealing process [5] but also isothermal self-assembly under diverse solvent conditions as demonstrated by recent studies [13].

MD simulations using atomistic force fields have been limited to relatively small systems to examine DNA transport through synthetic nanopores [14], DNA supercoiling [15], DNA origami structural dynamics [16], conformational changes from B-form to A-form [17], and thermal melting conditions of short sequences [18]. Currently, it is not computationally feasible to use all-atom approaches to study the hybridization and self-assembly of DNA at high concentrations due to the long time scales, large system sizes, and rarity of events. Instead, coarse-grained (CG) models [19-34] have been developed to

investigate hybridization and structural properties. Notable examples are the two-bead-per-nucleotide model called oxDNA [29, 35] and the three-bead model, called 3SPN [19,33].

The Doye group developed a two-bead model called oxDNA to study DNA for nanotechnology purposes [29, 35]. One drawback of the model is that it currently is parameterized for only 0.5 M NaCl, therefore investigating systems with different salt concentrations is not currently feasible. Also, the model is unable to capture the major and minor grooves of B-DNA due to the limitation of only two beads. However, this model has been successful in exploring hybridization of dsDNA using forward flux sampling, which showed that the rate constants from simulations compare favorably with experimental data [36]. Doye and co-workers have also investigated problems such as DNA tweezers [37], DNA walkers [38], DNA stretching [39], and toe-hold mediated DNA strand displacement [40]. Another research group led by Sciortino and co-wokers has used this model to examine an equilibrium bulk solution of 400 24-nucleotide strands designed to coaxially align on top of one another through base stacking [41] or 64 tetramers arranged as nanostars [42] for the formation of liquid crystal phases without investigating self-assembly kinetics.

The de Pablo research group developed the three-site-per-nucleotide model called developed by the 3SPN. The initial version, 3SPN.0, by Knotts *et al*. was able to predict melting temperatures; however, the persistence length was determined to be a factor of 2 smaller than experimental values [19]. Sambriski *et al*. updated this model to version 3SPN.1, which captures more accurately the persistence length of dsDNA and improves upon replicating experimental melting curves [43] however, single-strand persistence length was greatly overestimated. Additionally, the hybridization of two single strands was

investigated showing that different mechanisms were sequence dependent [44-45]. This model has also been applied to the hybridization by strands in solution to strands bonded to surfaces, which shows that having one strand bound to the surface changes the hybridization kinetics by limiting how the strands can reorient themselves to successfully base pair [46-47] Recently, Hinckley *et al*. updated the model even further to version 3SPN.2, which captures both dsDNA and ssDNA persistence lengths [48]. Using forward-flux sampling [49], they were able to obtain similar rate constants for hybridization as those found in experiments. Another version of the force field with explicit ions was introduced shortly after to study helix-coil transition, which requires local variations in ion concentrations [50].

These coarse-grained models use the traditional MD method with continuous potentials. In this work, we developed a CG model called BioModi (Biomolecular Multiscale Models at UC Irvine) to closely represent the structural properties of DNA (**Figure 2.1)** using an event-driven molecular dynamics algorithm [51] called Discontinuous Molecular Dynamics (DMD), wherein atoms are interacted with one another via discontinuous potentials. DMD has been shown to simulated large systems such as protein aggregation [52] and peptide-based self-assembly [53] on a single processor, which enables accessing to long time scales that have not been easily feasible by using the traditional MD method. Our DNA model, BioModi, is similar to the one developed by the Dokholyan group, which used an analogous 3-site-per-nucleotide model without explicit ions for use with DMD to study the folding of RNA molecules and was able to predict the structure of many RNA molecules within 4 Å of the experimental crystal structure [54]. Moreover, BioModi was built upon previous well-established models by incorporating some important features

from 3SPN and other models. For example, we have implemented from the 3SPN model an electrostatic potential that is dependent on ionic and thermal environment and a solvent induced potential that is dependent on the ionic strength which was originally present in the 3SPN.1 implementation.

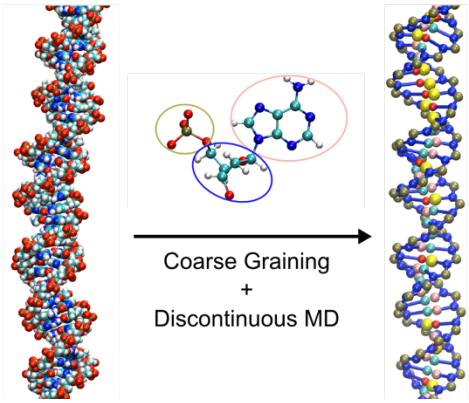

Figure 2.1: A coarse-grained model is derived from atomistic structure by representing each phosphate, sugar, and base group of a nucleotide as one bead particle.

BioModi employs a 3-site-per-nucleotide representation in which each phosphate, sugar, and base group of a nucleotide is represented as one bead particle [19, 54]. By using an implicit solvent, it reduces the number of degrees of freedom significantly and increases computational efficiency yet it takes the effects of the solvent into account. Using BioModi, we performed a set of replica-exchange and constant-temperature simulations to examine thermal and mechanical properties of DNA molecules as a mean to validate our model. Next, we conducted another set of constant-temperature simulations to elucidate different hybridization mechanisms by which a double helical structure is formed by two single strands of a homogenous sequence. Finally, we performed large-scale simulations on four systems of different sequences to understand sequence-dependent thermodynamics and kinetics of self-assembly at high strand concentrations.

2.3 Simulation Methods

2.3.1 Coarse-grained DNA Model

A coarse-grained model has been developed to represent DNA. This DNA model has been combined with our model ePRIME [53, 55, 56] for peptides and polymers, becoming an integrated simulation package called BioModi. For DNA, a 3-site-per-nucleotide model was chosen due to the success of other groups in accurately representing nucleic acids with the least number of particles [19, 54]. We simplified the DNA structural model by using a "bead-on-a-string" model polymer with three coarse-grained beads - phosphate (P), sugar (S), and base (B) - representing each nucleotide (**Figure 2.1**). The coordinates of these three chemical moieties are reduced to effective, united-atom interaction sites, the coordinates of which are derived from the structures generated by Nucleic Acid Builder [57]. Beads P and S are positioned at the center of mass of the corresponding phosphate group and the five-atom ring sugar. The bead representing each base is the N1 position for purine bases (adenine and guanine) and the N3 position for pyrimidine (thymine and cytosine). The bead sizes are given in **Table 2.1**.

Particle	**Bead Diameter (Å)**
Phosphate, P	5
Sugar, S	5
ADE, B	2.63
GUA, B	2.67
CYT, B	2.67
THY, B	2.63

Table 2.1: United atom sizes for nucleic acid beads in BioModi.

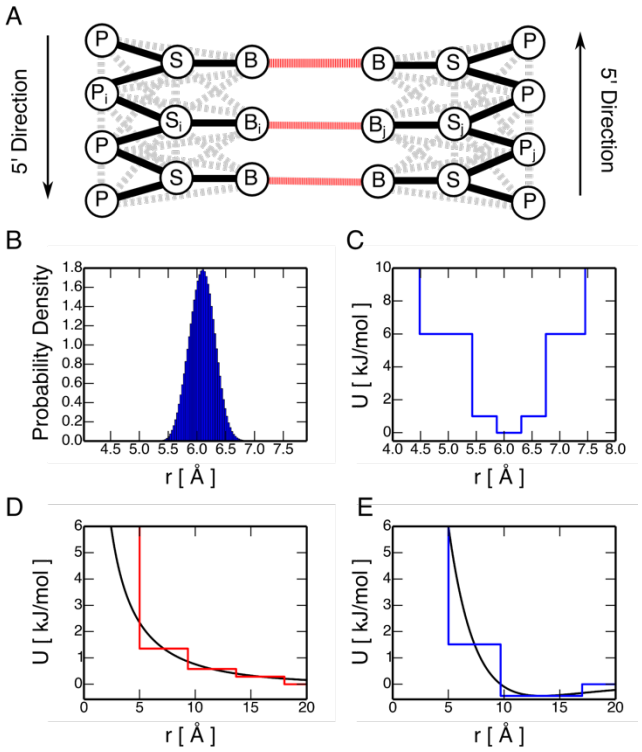

Figure 2.2: (A) Covalent bonds are schematically shown as dark black lines connecting united atoms. Pseudo-bonds are schematically shown as dashed blacks lines and are used to maintain bond angles and dihedrals. Base-pair interactions (red lines) are directional in orientation. (B) Probability distribution of a bond's distance extracted from all-atom MD simulations. The mean, minimum, maximum, and standard deviations are used to discretize a harmonic potential for bonding. (C) Example of the discretized potential for the bond in B. (D) Electrostatic potential for the phosphate groups with a monovalent salt concentration of 69 mM at 300 K. Continuous electrostatic potential is in black and three-step shoulder potential is in red. (E) Solvent induced potential for complementary sugar groups for a strand of DNA 30 nucleotides long at a salt concentration of 69 mM. Continuous Morse-potential is in black and the two-step discrete potential is in blue.

Besides covalent bonds, neighboring beads are constrained by imposing pseudo-bonds to mimic the chain connectivity and the local geometry such as bond angles and dihedrals (**Figure 2.2A**). The distances for 26 covalent bonds and pseudobonds (**Table 2.2**) were extracted from all-atom molecular dynamics simulations of dsDNA molecules with a random sequence of 352 nucleotides and a section of a nucleosome (pdb 1KX3). All-atom simulations were conducted using NAMD 2.8 with implicit solvent and the CHARMM 27 force field [58]. After a slow heating period to 300 K, the data was collected over a 40 ns equilibration run. Shown in **Figure 2.2B** is the probability density of a typical bond in the

system showing the Gaussian-like nature. Since covalent bonding is traditionally represented as a harmonic spring, this is discretized to a two-step potential as shown in **Figure 2.2C**. The first and second steps of the potential are placed at one and three standard deviations from the mean. Finally, the potential is set to infinity at the minimum and maximum of the bond distances found.

Bond Type	Mean (Å)	Standard Deviation	Bond Type	Mean (Å)	Standard Deviation
S_i-S_{i+1}	6.09	0.22	P_{i+1}-B_i (GUA)	9.25	0.41
P_i-P_{i+1}	6.98	0.22	P_{i+1}-B_i (CYT)	7.60	0.42
P_i-S_i	3.69	0.06	P_{i+1}-B_i (THY)	7.62	0.39
S_i-P_{i+1}	3.99	0.07	S_{i+1}-B_i (ADE)	8.34	0.45
S_i-B_i (ADE)	6.93	0.12	S_{i+1}-B_i (GUA)	8.61	0.47
S_i-B_i (GUA)	6.92	0.10	S_{i+1}-B_i (CYT)	7.10	0.38
S_i-B_i (CYT)	5.48	0.10	S_{i+1}-B_i (THY)	7.16	0.39
S_i-B_i (THY)	5.43	0.10	S_i-B_{i+1} (ADE)	7.52	0.50
P_i-B_i (ADE)	8.90	0.22	S_i-B_{i+1} (GUA)	7.37	0.54
P_i-B_i (GUA)	8.72	0.24	S_i-B_{i+1} (CYT)	7.20	0.51
P_i-B_i (CYT)	7.71	0.19	S_i-B_{i+1} (THY)	7.08	0.46
P_i-B_i (THY)	7.66	0.18	P_i-S_{i+1}	9.69	0.25
P_{i+1}-B_i (ADE)	9.03	0.41	S_i-P_{i+1}	9.87	0.34

Table 2.2: Mean distances and standard deviations of 26 covalent bond and pseudo-bonds shown in Figure 2.2A.

Base-pairing (A-T and G-C) interactions are crucial for modeling DNA hybridization, dynamics, and mechanical properties. The hydrogen bonds between complementary base pairs are strongly dependent upon orientation in the sense that the vectors connecting each base to the virtual hydrogen and oxygen (or nitrogen) atoms point toward each other within a relatively strict tolerance to create an obtuse angle. To enforce this angle constraint, we adopt an approach similar to that of Ding *et al*. [54] and require that the atoms of two pairs S_i-B_j and S_j-B_i shown in **Figure 2.2A** (hereafter referred to as auxiliary

pairs), be separated by predetermined distances that are chosen to maintain the hydrogen bond angle constraints. Upon the formation of a bond between two complementary nucleotides (e.g., A–T or G–C shown as B_i and B_j in **Figure 2.2A**), these auxiliary pairs temporarily interact via a single-step square-shoulder potential whose shoulder height is equal to the well depth of the hydrogen bond between B_i and B_j. In addition to preventing the hydrogen bond angles from straying outside of the desired values, the square-shoulder interaction within each auxiliary pair adds stability to the hydrogen bond by increasing the range of the repulsive forces between the beads that neighbor B_i and B_j. These auxiliary pairs return to their original interactions when the hydrogen bond is broken. The relative strength between complementary base-pair interactions CG and AT is 1.266 (**Table 2.3**) as previously adopted by SantaLucia [59] and De Pablo group [43].

Base Pair	Interaction Strength (kJ/mol)
ADE – THY	6.120
GUA – CYT	7.748

Table 2.3: Base pairing interaction strengths in BioModi

Base Stack Pair	kJ/mol	Base Stack Pair	kJ/mol
ADE ADE	2.7938	CYT ADE	3.0477
ADE GUA	3.5557	CYT GUA	3.0477
ADE CYT	3.0477	CYT CYT	2.7938
ADE THY	2.7938	CYT THY	2.7938
GUA ADE	3.5557	THY ADE	2.5398
GUA GUA	3.5557	THY GUA	3.3017
GUA CYT	3.0477	THY CYT	2.7938
GUA THY	3.3017	THY THY	2.0318

Table 2.4: Base stacking interaction strengths in BioModi

Also included in the model is base stacking, representing the attractive nature of the π-stacking in the bases. This attraction is only allowed to occur between adjacent bases on

the same strand of DNA. The interaction is also treated to be isotropic in nature, and the interaction strengths depend on the order of the stacking (5' → 3'). Stacking interaction strengths were obtained by Hobza and co-workers from quantum mechanics calculations and scaled appropriately via the hydrogen bond strength for the model [60] as shown in **Table 2.4**.

Electrostatic repulsions between phosphate groups are modeled through a Coulombic potential with a Debye-Huckel salt screening correction. This potential is dependent on the temperature and ionic environment as seen in **Equation 2.1**. Parameters for the relative permittivity (ϵ) follows the work of Sambriski et al. [43], which takes into account the temperature and monovalent salt concentration. Additionally, the Debye screening length, λ_D, determines how quickly the potential levels off. As one increases the amount of monovalent salt in the solution the electrostatic effect is reduced as the positively charged ions effectively screen the negative charge on the backbone.

$$U_{elect} = \sum_{i<j}^{n} \frac{q_i q_j Exp\left[{}^{-r_{ij}}/_{\lambda_D}\right]}{4\pi\epsilon_0\epsilon(T,C)r_{ij}} \qquad \text{Equation 2.1}$$

The electrostatic interaction is discretized as a three-step square shoulder as shown in **Figure 2.2D**. Cut-off lengths were tested to replicate the persistence length of double-stranded and single stranded DNA. The three shoulders were placed equidistant from the excluded volume distance (5 Å) and the cut off. The strength for each shoulder was set to be equal to the middle of each shoulder in order to accurately represent the potential with the least amount of detail. In order for the strands to hybridize together with this electrostatic potential it was necessary to also implement a solvent induced potential. The

potential is designed to overcome the entropic penalty of hybridization when water is ordered around the sugars [43].

The solvent induced potential, which is represented as a Morse-potential, with an attractive well at long distances and is repulsive at short distances (**Equation 2.2**). In Equation 2, α controls the range of the potential, r_s is the minimum of the potential (13.38 Å), and ε_s is the relative strength of the potential. The interaction is dependent on the ionic strength and number of nucleotides in the strand of DNA[43].

$$U_{Solv} = \varepsilon_s\left[1 - Exp\left(-\alpha(r_{ij} - r_s)\right]^2 - \varepsilon_s \right. \qquad \text{Equation 2.2}$$

The solvent induced potential occurs between sugars on complementary strands of DNA. A cutoff of 17 Å was chosen for the solvent induced interaction in order to still keep computational efficiency. The solvent induced interaction is discretized as a square shoulder and well potential as shown in **Figure 2.2E**.

2.3.2 Simulation and Analysis Procedures

Simulations were performed by using the DMD (discontinuous molecular dynamics) simulation algorithm [61], a variant on standard molecular dynamics that is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere, and square-well potentials. Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. DMD simulations proceed by locating the next collision, advancing the system to that collision, and then calculating the collision dynamics. Simulations are performed in the canonical ensemble with periodic boundary conditions imposed to eliminate artifacts due to box walls. Constant temperature

is achieved by implementing the Andersen thermostat method [62]. In this case, all beads are subjected to random, infrequent collisions with ghost particles whose velocities are chosen randomly from a Maxwell Boltzmann distribution centered at the system temperature. Simulation temperature is expressed in terms of the reduced temperature, $T^* = k_BT/\varepsilon_{HB}$, where k_B is Boltzmann's constant, T is the temperature, and ε_{HB} is the strength of one hydrogen bond (initially set at 1 kJ/mol as a reference so that other potential strengths can be accordingly scaled). Simulation time is defined in reduced units to be $t^*=t/\sigma(k_BT/m)^{1/2}$, where t is the simulation time, and σ and m are the average bead diameter and mass, respectively.

2.3.3 Thermal Melting

Thermal melting data of DNA was obtained using replica exchange molecular dynamics (REMD) [63]. The sequences used in the study are listed below in **Table 2.5**. A total of sixteen replicates were used over temperature ranges that capture fully hybridized and dissociated strands. Over 4,000 exchanges were attempted for each sequence with each attempt running five reduced time steps. Equilibrium is reached when energy (in term of the number of base-pair interactions) becomes constant. The equilibrium data reported in this manuscript is from the last 10% of simulation time.

Name	Sequence (5'→3')
GC 0%	AAAA AAAA AAAA AAAA AAAA AAAA AAAA AA
GC 40%	GTCA TACG ACTG AGTG CAAC ATTG TTCA AA

Table 2.5. Sequences for thermal melting simulations

2.3.4 Persistence Length

Calculations on the persistence length were performed on the data obtained from multiple constant-temperature simulations that were conducted for very long time scales at a constant temperature (T^*) of 0.5 (to ensure that all bases were hydrogen bonded together as it is below the melting temperature of DNA seen in **Figure 2.3**). Sixteen simulations for each salt concentration were conducted on a random sequence of dsDNA containing 352 nucleotides and a homogenous sequence of ssDNA containing 100 adenines (in order to keep intrastrand base pairing from occurring). By assuming that both dsDNA and ssDNA acted like worm-like chains, a method developed by Woodward and coworkers [64] was employed using the radius of gyration (R_g) to determine the orientational persistence length (P) per **Equation 2.3**.

$$R_g^2 = \frac{PL}{3} - P^2 + 2\frac{P^3}{L} - 2\frac{P^4}{L^2}\left(1 - Exp\left(\frac{-L}{P}\right)\right) \qquad \text{Equation 2.3}$$

In equation 2.3, L is the contour length of the polymer, which was assumed at 6.3 Å x (# nucleotides) for ssDNA [65] and 3.4 Å x (# nucleotides) for dsDNA. Specifically, the ensemble averages for each simulation were used to determine the radius of gyration, thus giving a value of the persistence length for each of the sixteen simulations conducted. Additionally, the green line in Figure 2.4A, was determined using Possion-Boltzmann theory [66] (**Equation 2.4**), and fitted using non-linear least squares regression to the mean of sixteen persistence length simulations at each salt concentration.

$$P = P_0 + \frac{1}{4\lambda_B/\lambda_D^2} \qquad \text{Equation 2.4}$$

In Equation 2.4, the baseline persistence length, P_0, is the persistence length without any electrostatic contributions, and the second term of the equation is the electrostatic contribution through the Bjerrum length (λ_B) and Debye length (λ_D).

2.3.5 Hybridization by Two Strands

Over 350 simulations were setup with one strand of A_{30}, and the other strand T_{30} with a random configuration. To further randomize the initial configuration, simulations were conducted at high temperature (T^* = 5.0) ensuring no base pairing could occur. Velocities of the particles were reinitialized to T^* = 0.7, just below the melting temperature and conducted with the NVT ensemble. A salt concentration of 0.5 M was used to determine initial contacts and 69 mM was used to determine the relative rate of hybridization. The time it takes for each mechanism to reach a given number of base pairs was determined by taking the given initial contact and running 64 additional simulations starting from that contact with different random number seeds. Data plotted only occurred from simulations that reached the maximum number of base pairs discussed in subsection 2.4.3.

2.3.6 Self-Assembly of Many Strands

Self-assembly simulations on four systems, each with 100 strands of DNA (50 sense strands and 50 anti-sense strands, sequences are listed in **Table 2.6**), were performed. These strands were placed randomly in the simulation box as an initial configuration. They were heated at high temperature (T^* = 5.0) for a period of time to ensure there is no bias in the initial randomization. Velocities of the particles were subsequently reinitialized to a

desired temperature and conducted with the NVT ensemble. Simulations were ran with a reduced temperature of T^* = 0.9, monovalent salt concentration of 69 mM, and DNA concentration of 1.45 mM. A total of four simulations at each condition were used in the analysis as it was found the standard deviations were sufficiently low. Moreover, additional simulations were conducted using 500 strands, which showed no significant differences in the final results compared to the 100-strand simulations, thus validating the use of the smaller system size. Each constant-temperature simulation was conducted for 1000 time units. This is enough for any simulation to reach equilibrium as indicated by the ensemble averages of the system's total potential energy which varied by no more than 2.5% toward the end of each simulation run. The simulation data analyzed for steady state is the last 10% of the simulation. In our analysis, an aggregate is defined as having at least two strands forming one base pair interaction at a minimum.

Name	**Sequence (5'→3')**
A_{32}	AAAA AAAA AAAA AAAA AAAA AAAA AAAA AAAA
RAND	CCAA TGCG GTAA GCCT GACA CCGA TCAA TCTT
$A_{16}T_{16}$	AAAA AAAA AAAA AAAA TTTT TTTT TTTT TTTT
$C_8T_8A_8G_8$	CCCC CCCC TTTT TTTT AAAA AAAA GGGG GGGG

Table 2.6. Sequences for self-assembly simulations

2.4 Results and Discussion

2.4.1 Thermal property as a function of sequence

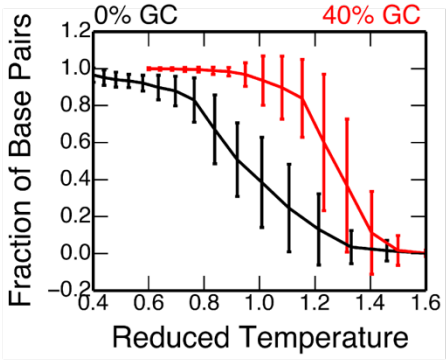

Figure 2.3: Melting curves of double-stranded DNA of a homogenous sequence (GC content 0%, black) and a random sequence of strand length 30 (GC content 40%, red).

Our model, BioModi, can capture the transition between the natured and denatured states of dsDNA as a function of temperature, similarly observed in UV/Vis experiments [67]. Such a two-state nature involving either strongly bound duplexes or widely separated strands [68] can bee seen in **Figure 2.3,** which shows the number of base-pair (bp) interactions from replica exchange molecular dynamics simulations over a wide range of temperatures for a homogeneous sequence containing adenines (**Table 2.5**). This hybridization transition gives a melting temperature of 0.9 in reduced temperature units (T^*). In comparison, the melting curve of a random sequence containing 40% guanine-cytosine (GC) nucleotides (**Table 2.5**) exhibits a sharper two-state transition and gives a melting temperature at approximately T^* = 1.25 (**Figure 2.3),** in qualitative agreement with experimental [69] and previous simulation [29, 43] studies.

2.4.2 Mechanical property as a function of ionic strength

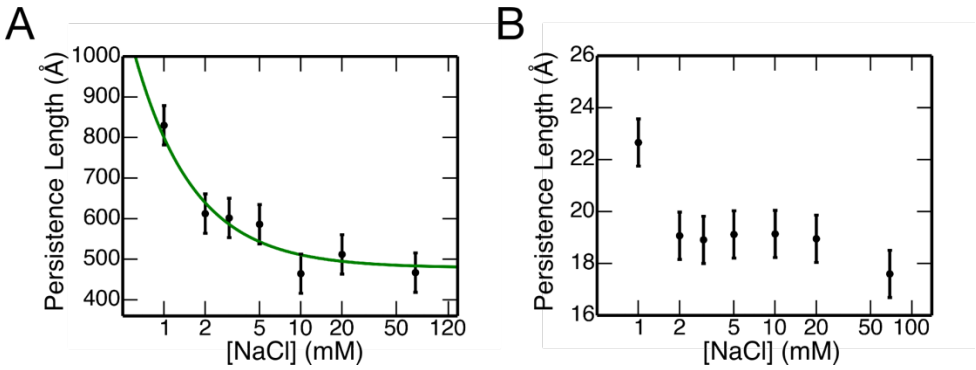

Figure 2.4: Simulation results of persistence length of (A) dsDNA and (B) ssDNA as a function of NaCl salt concentration. The green line is predicted from the Poisson-Boltzmann theory [66].

Our model can reproduce not only mechanical properties, such as molecular rigidity quantified as the persistence length of both dsDNA and ssDNA, but also the dependence of persistence length on salt concentration. As salt concentration decreases, the screening of the charges along the backbone of the DNA molecule decreases, and the phosphate groups seek to separate themselves from their neighbors. The result is a more elongated molecule at lower salt concentrations than in more screened environments. When the salt concentration is high, its effect on the persistence length fades away as the persistence length maintains a constant value. Our model yields a persistence length of ~50 nm at high salt concentrations and significantly higher values at low salt concentrations for dsDNA (**Figure 2.4A**). Similarly, the persistent length is close to 1.5 - 2.0 nm at high salt concentrations and higher values at low salt concentrations for ssDNA (**Figure 2.4B**). This relationship of the persistent length as a function of salt concentration for both dsDNA and ssDNA agrees with the Poisson-Boltzmann theoretical predictions [66] and experiments [65-66].

2.4.3 Kinetic mechanism of hybridization is determined by the first point of contact

Constant-temperature simulations on two complementary strands of a homogenous sequence A_{30} at $T^* = 0.7$, which is slightly below the melting point, demonstrate that hybridization of dsDNA from ssDNA occurs via two main hybridization mechanisms: zippering and slithering. Moreover, the choice of these aforementioned mechanisms is dictated by the first point of contact made by two approaching single strands (**Figure 2.5A)**. The nature of this first contact is relatively random as seen in **Figure 2.5B**, which shows probability (in non-normalized values) of making an initial contact between certain nucleotides of two strands from 350 independent simulations. However, closer inspection indicates that the end nucleotides of one strand are more likely to make an initial contact with any nucleotide on the other strand. Moreover, it demonstrates that the first contact is disproportionally biased toward non-native contacts compared to native contacts. By definition, base pairs that form when the sense and antisense strands are perfectly aligned are called native contacts.

The first contact transitions into a nucleus when at least another bp interaction between consecutive bases is swiftly formed, which creates a stable nucleating complex involving two or three base pairs. Due to thermal fluctuations, many first contacts are unable to become nuclei; therefore, two complementary strands break away from one another and approach each other again until a nucleus is established. In fact, each first contact has a 26% chance of unsuccessfully becoming a nucleus. The formation of such a nucleating complex has been formulated as the rate-limiting step of a second-order kinetic reaction based on the data extracted from experiments [70-71].

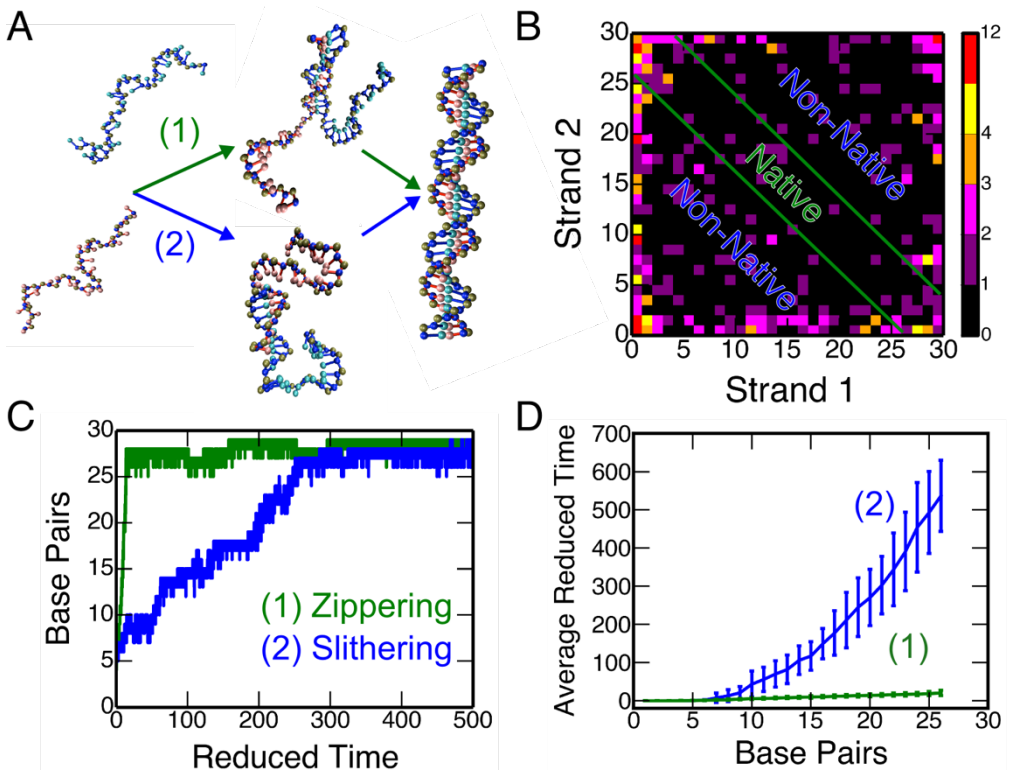

Figure 2.5: (A) Two representative mechanisms of hybridization by two strands of a homogenous sequence A_{30}: (1) zippering and (2) slithering. (B) A contact map showing probability (in non-normalized values) of making an initial contact between certain nucleotides of two strands from 350 simulations. (C) Hybridization kinetics in term of the number of bp interactions as a function of time (in reduced units) of two mechanisms. (D) Average hybridization time for each mechanism to reach a certain number of bp interactions.

When two strands initially make native contacts, hybridization is via the zippering mechanism, in which these strands quickly wrap around one another to form bp interactions reaching the fully hybridized state as seen in **Figure 2.5C** from a simulation in which the end nucleotides of one strand form initial contacts that are native with the end nucleotides of another strand. This mechanism is consistent with experimental data that indicated that hybridization is initiated by a nucleation event in which two or three base pairs of complementary strands are formed [70-71]. This is followed by a rapid process that accomplishes full complementarity by transitioning single stranded tails that are relatively flexible and unstructured (with a persistence length of ~1.5 – 2.0 nm as seen in

Figure 2.4B) into a rigidly double-helical duplex (with a persistence length of ~50 nm as seen in **Figure 2.4A**). This mechanism has been previously observed by both Doye [36] and de Pablo [33] groups.

In contrast, when two strands initially make non-native contacts, hybridization is via the slithering mechanism, in which the strands slowly reach the fully hybridized state as seen in **Figure 2.5C** from a simulation in which the end nucleotides of one strand form initial contacts that are non-native with the end nucleotides of another strand. The next step involves the nucleotides at the end of one strand breaking their existing bp interactions and moving up by a few nucleotides to form new bp interactions. This allows the end nucleotides of the other strands to form additional bp interactions. Such a process, which creates about 5 bp interactions, happens quickly with ease. However, the next sequence of steps faces some degree of difficulty. The nucleotides at the end of one strand break their existing bp interactions leaving the other bp interactions intact. Since the single stranded tail is highly flexible (with a persistence length of ~1.5 - 2 nm as seen in **Figure 2.4B**), they are able move up by a few nucleotides to form new bp interactions. This step creates a small bulge loop of one or two nucleotides between the new bp interactions and the intact bp interactions. Due to thermal fluctuations, the bulge breaks newly formed bp interactions or previously intact bp interactions. The latter results in a larger bulge; however, the nucleotides of the original bulge can form new bp interactions reducing the bulge loop's length. Gradually, the bulge moves down the strand to the other end. Overall, the whole process creates a few bp interactions. This process is repeated until the 5' end of one strand forms bp interactions with the 3' end of the other strand, which create a full double-helical structure.

The above-described slithering mechanism was similarly observed by Ouldridge *et al*., who referred to it as the inchworm displacement mechanism [36]. However, the bulge observed by Ouldridge *et al*. contains at least four nucleotides compared to just one or two nucleotides seen in our simulations. This difference is due to the sequences used for study; they examined a repetitive sequence of poly-AC whereas we examine a homogenous sequence of poly-A. Moreover, they observed that poly-AC also undergo another hybridization mechanism called pseudoknot in which the formation of the first non-native contacts precipitates the formation of a second segment of bp interactions a distance away elsewhere, which facilitates zippering into a full duplex. This mechanism is rarely seen in our simulations due to the homogenous sequence and strand length of 30 nucleotides in our simulations compared to 14 nucleotides in their study. In our simulations, after forming the first non-native contacts between the end nucleotides of two strands, the other two ends have a 32% probability to form additional bp interactions, which creates a circle. When thermal fluctuations are strong, the two strands in a circle can undergo the slithering mechanism until the two strands are perfectly aligned; at this point, the circle breaks out into a straight duplex.

The slithering mechanism requires overcoming many energetic barriers by constantly breaking existing bp and forming new bp interactions as indicated by multiple plateaus and fluctuations (**Figure 2.5C)**. Therefore, the slithering mechanism requires at least one order of magnitude longer than the zippering mechanism to reach the fully hybridized state starting from an initial bp interaction as seen in **Figure 2.5D,** which shows the average hybridization time for each mechanism to reach a certain number of bp interactions. The data plot for the zippering mechanism is a linear line indicating that the

amount of time to form one additional bp interaction is the same independent of the degree of hybridization. In contrast, the data plot for the slithering mechanism is an exponential growth curve indicating that more time is needed to form an additional bp interaction compared to the previous interaction. This means that the energetic barriers are higher as the duplex become more hybridized especially towards the end of the slithering process.

2.4.4 Self-assembly by homogeneous sequences produces polydisperse aggregate sizes

The homogeneous sequence, A_{32}, is the most susceptible to misaggregation as defined by forming aggregates larger than a dimer. This can be seen in **Figure 2.6A**, which shows the percentage of strands belonging to various aggregate sizes as a function of time for the A_{32} sequence. These results are from multiple independent simulations at a DNA strand concentration of 1.45 mM, temperature of T^* = 0.9, and a monovalent salt concentration of 69 mM. Each system contains 100 DNA strands of A_{32}; 50% sense strands and 50% anti-sense strands.

At the beginning of each simulation, there is a rapid decrease in monomers leading to the formation of dimers, then trimers, tetramers, pentamers and larger aggregates. As the simulation reaches steady state, the number of molecules in dimers is equivalent to that of the larger sized aggregates. The distribution of final aggregates for A_{32} at equilibrium is plotted in **Figure 2.6B**, ranging from dimers to large aggregates that contain up to 25 strands. This large distribution is primarily due to the high probability of hybridization occurring through the slithering mechanism that takes an order of magnitude longer than the zippering mechanism to reach the fully hybridized state (**Figure 2.5D**). Therefore, free strands are able to form bp interactions with partially hybridized strands and create larger

and larger aggregates (**Figure 2S2(A)**). Moreover, the dimers formed by A_{32} sequence are partially hybridized at steady state as shown by the distribution of dimers containing a broad number of bp interactions (**Figure 2.6C)**. The most probable state is at 26 out of 32 maximum bp interactions indicating that energetic barriers are more prohibitive as two strands become more hybridized as seen by increasingly high hybridization times are required as more bp interactions are formed in **Figure 2.5D**.

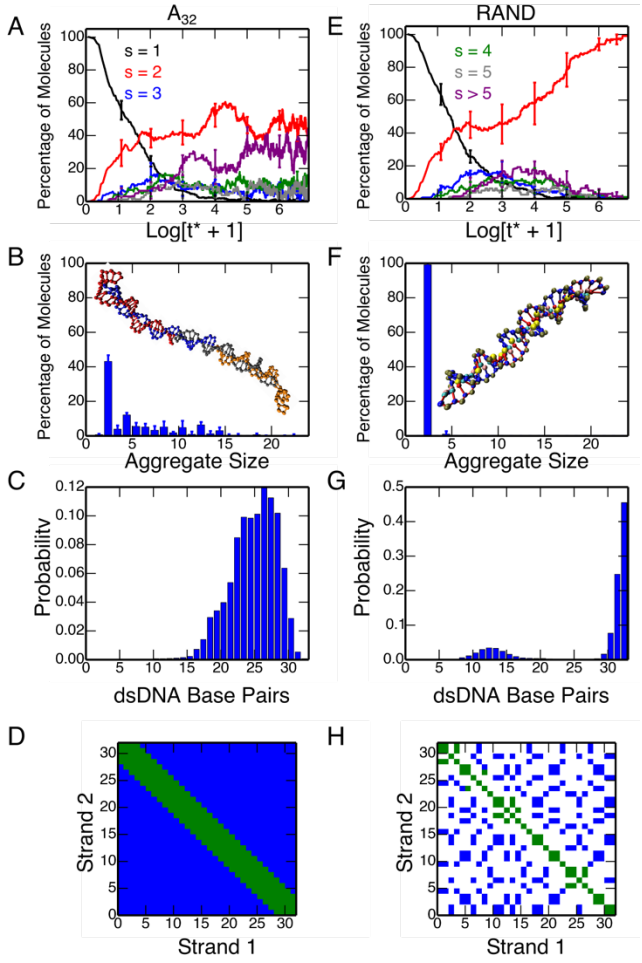

Figure 2.6. Percentage of molecules belonging to various aggregate sizes, *s*, as a function of time in natural logarithmic scale for (A) A_{32} and (E) RAND sequences. Distribution of aggregate size for (B) A_{32} and (F) RAND sequences at steady state. Snapshots showing a tetramer in (B) and dimer in (F). Distributions of dimers containing a certain number of base pair interactions for (C) A_{32} and (G) RAND sequences at steady state. Contact maps or possible first contact along each nucleotide by two strands of (D) A_{32} and (H) RAND sequences: non-native contacts in blue, native contacts in green, and no possible contacts in white. The native contacts should be a straight line; however, a small tolerance of ±3 is allowed.

2.4.5 Self-assembly by heterogeneous sequences produces monodisperse aggregate size

Unlike the homogeneous sequences, heterogeneous sequences in the form of a random sequence RAND (**Table 2.6**) produces no large aggregates at the end of simulations (**Figure 2.6E**). Indeed, all strands form dimers at steady state as seen in **Figure 2.6F**. Moreover, the RAND system has a very narrow distribution with the majority of

dimers contain 30 or more bp interactions at steady state (**Figure 2.6G).** This demonstrates that even though it is more entropically desirable to form a collection of aggregates of different sizes, DNA self-assembly by random sequences prefers to form a monodisperse distribution of dsDNA**.**

Large aggregates are also observed during self-assembly by a random sequence RAND; however, these aggregates disappear over time leading to formation of only dimers at steady state (**Figures 2.6E** and **2.6F**), in contrast to the A_{32} system (**Figures 2.6A** and **2.6B**). Even though hybridization by random sequences is through the zippering mechanism rather than the slithering mechanism, those strands can form non-native contacts (**Figure 2.6H)** albeit in small numbers that meet our definition of an aggregate (require at least one bp interaction between participating strands). Thus, those aggregates are metastable and likely to convert into more energetically favorable dimers. This is seen in **Figure 2S2(B)**, which shows a relatively high value of the number of bp interactions per strand by random sequences compared to homogenous sequences. The presence of large aggregates on route to the formation of dimers demonstrates the robust nature of DNA self-assembly, during which DNA strands get kinetically trapped in local minima yet they can overcome energetic barriers to reach the global minimum.

An example of large aggregates transitioning into dimers is shown in **Figure 2.7**. When two strands are misaggregated with another strand forming a trimer (**Figure 2.7A**), their total number of bp interactions is around 23 (**Figure 2.7B**). Then one strand (green) is displaced so that the other two strands can zip up via the zippering mechanism into a fully hybridized dimer (**Figure 2.7C-F**) as indicated by a jump in the number of bp interactions to 32 (**Figure 2.7B**). The displaced (green) strand (**Figure 2.7F**) can proceed

to attach to different strands, thus increase the number of bp interactions of the system. Such a process of forming metastable dimers and trimers before transforming into more energetically favorable dimers (**Figure 2.7**) indicates that DNA self-assembly is an enthalpically driven process. Moreover, it is similar to the strand displacement mechanism that has been observed in experiments [72] and playing an important role in designing a digital computation circuit using DNA strands [73].

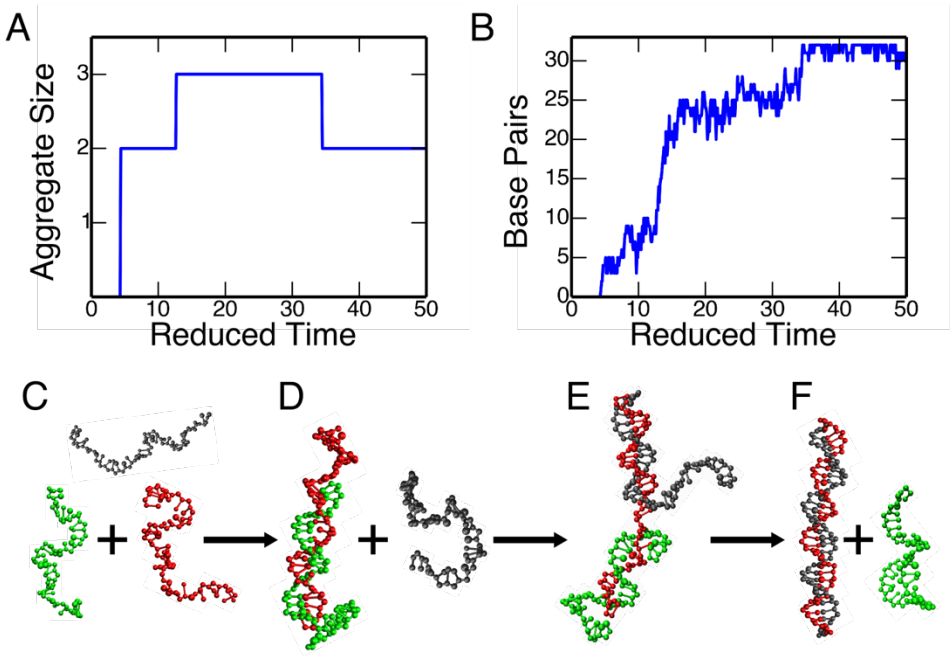

Figure 2.7. A typical trimer transitioning into a dimer by sequence RAND shown by (A) aggregate size and (B) number of bp interactions between three strands over time. Snapshots showing self-assembly to form a dsDNA starting from (C) free strands, which form (D) a partially hybridized dimer via the slithering mechanism, then (E) the free strand (gray) attaches to the unhybridized segment of the red strand to undergo hybridization via the zippering mechanism, finally (F) the green strand is displaced so that the red and gray strands can form a complete dsDNA.

The formation of double helices by random sequences is via multiple self-assembly pathways. On one hand, if two single strands approach each other and form initial contacts that are native, they create a dimer and undergo the zippering mechanism to form a complete double helix without interacting with other strands that are nearby. However, such a dimer during hybridization in a crowded environment is more likely to interact with

at least another nearby strand. That dimer can still reach full hybridization by discarding invading strands (similar to **Figure 2.5E** but the invading strand is green). In this case, the strands in the final double helix are the original strands that form the initial contacts. On the other hand, if two approaching strands make initial contacts that are non-native, they undergo the slithering mechanism, which is slower than the zippering mechanism. Therefore, they are more likely to undergo misaggregation, which creates a trimer or a larger aggregate. In this case, the invading strand can form initial contacts that are native with one of the original strands, zip up via the zippering mechanism and displace the other original strand in the process (**Figures 2.7C-F**). In this case, the formation of double helices is a product of a transition between metastable trimers or larger aggregates into more energetically favorable dimers through the strand displacement mechanism. Therefore, the choice of a self-assembly pathway to form a double helix is highly dependent upon the nature of the initial contacts.

2.4.6 Self-assembly by diblock sequences produces partially hybridized dimers

Reducing homogeneity in the self-assembling sequences by introducing patterned heterogeneity in the form of diblock $A_{16}T_{16}$ also produces monodisperse distribution of aggregates at the end of simulations (**Figure 2.8A**). Like the random sequence RAND, $A_{16}T_{16}$ quickly turn monomers into dimers and larger sized aggregates (**Figure 2.8A** and **2.8E**). However, the total number of strands in large aggregates at any time during self-assembly is lower than that in the RAND system. Therefore, this indicates that patterned heterogeneity reduces the probability of getting kinetically trapped in local minima of mis-aggregation and forming large aggregates.

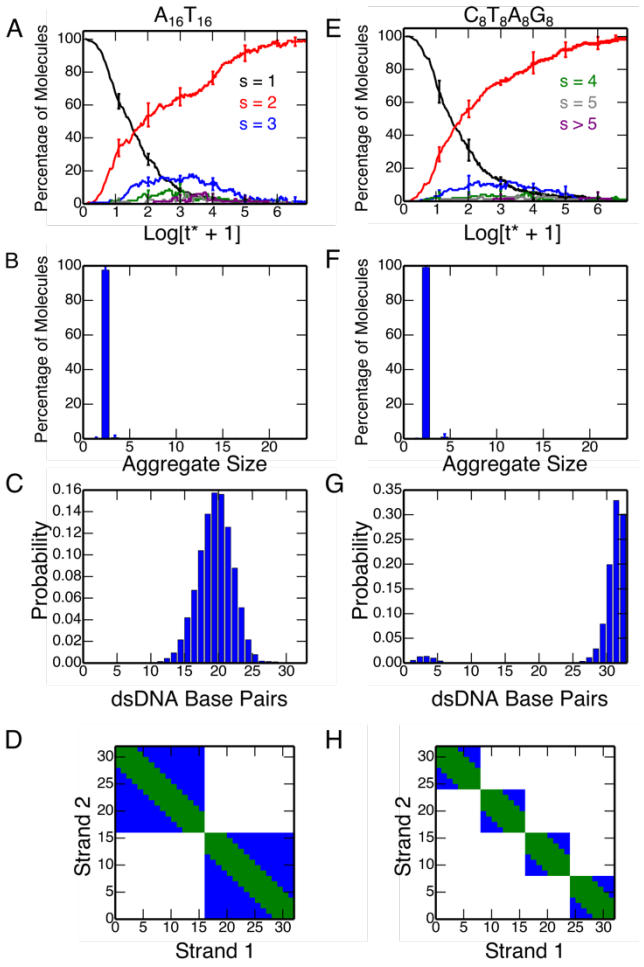

Figure 2.8: Percentage of molecules belonging to various aggregate sizes, *s*, as a function of time in logarithmic scale for (A) $A_{16}T_{16}$ and (E) $C_8T_8A_8G_8$ sequences. Aggregate size distribution for (B) $A_{16}T_{16}$ and (F) $C_8T_8A_8G_8$ sequences at steady state. Distributions of dimers containing a certain number of base pair interactions for (C) $A_{16}T_{16}$ and (G) $C_8T_8A_8G_8$ sequence at steady state. Contact maps or possible first contact along each nucleotide by two strands of (D) $A_{16}T_{16}$ and (H) $C_8T_8A_8G_8$ sequences: non-native contacts in blue, native contacts in green, and no possible contacts in white. The native contacts should be a straight line; however, a small tolerance of ±3 is allowed. Adding heterogeneity to the sequence reduces the possible number of non-native contacts thus shifting the mechanism from slithering toward zippering.

At steady state, dimers are the only species present in the $A_{16}T_{16}$ system (**Figure 2.8B**). However, the distribution of dimers at steady state in **Figure 2.8C** shows that a typical $A_{16}T_{16}$ dimer has 20 bp interactions out of 32 possible bp interactions. This is due to a relatively higher probability by two $A_{16}T_{16}$ strands to make initial non-native contacts instead of native contacts (**Figure 2.8D**); thus those $A_{16}T_{16}$ strands tend to undergo the

slithering mechanism. However, they become kinetically trapped in a conformation in which the middle region remains unhybridized while the two ends are constantly breaking and forming the same bp interactions (**Figure 2.9A)** without being able to slither into the fully hybridized state as shown in **Figure 2.9B-C**. We henceforth refer to this conformation as a kinetically trapped bubble structure.

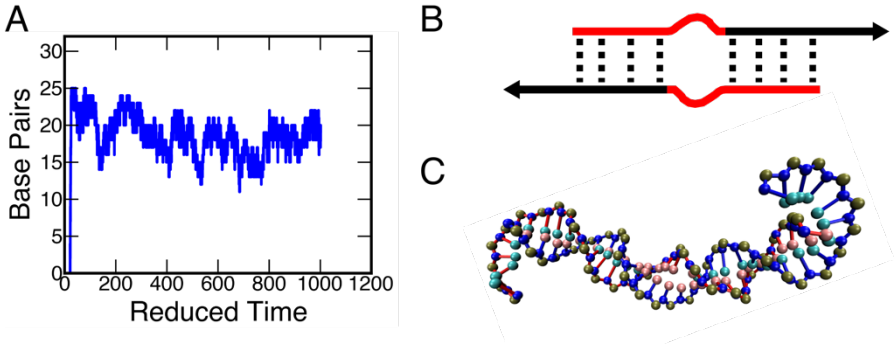

Figure 2.9: (A) Number of base pairs vs. time for a typical $A_{16}T_{16}$ dimer from the self-assembly simulations. (B) Schematic representation of a trapped kinetic structure called bubble dimer by $A_{16}T_{16}$. (C) A typical bubble dimer whose middle region containing the same nucleotides is unhybridized.

2.4.6 Self-assembly by tetrablock sequences produces fully hybridized dimers

Reducing homogeneity in the self-assembling sequences further by introducing patterned heterogeneity in the form of tetrablock $C_8T_8A_8G_8$ sequences also produces monodisperse distribution of aggregates at the end of simulations (**Figure 2.8E**).

Dimers are the only species at steady state in the $C_8T_8A_8G_8$ system (**Figure 2.8F**). In contrast to the $A_{16}T_{16}$ system, the distribution of dimers at steady state shows that a typical $C_8T_8A_8G_8$ dimer contains between 30-32 bp interactions in **Figure 2.8G**. This is due to a significantly reduced probability by two $C_8T_8A_8G_8$ strands to make initial non-native contacts compared to that of native contacts (**Figure 2.8H**); thus those $C_8T_8A_8G_8$ strands tend to undergo the zippering mechanism. However, the $C_8T_8A_8G_8$ system is still capable of

producing large aggregates in the similar amount as the $A_{16}T_{16}$ system (**Figure 2.8E**) but significantly less than the RAND system. This indicates that the formation of double helices in the $C_8T_8A_8G_8$ system is mainly through the zippering mechanism.

2.3.7 Self-assembly yield depends on sequence

Sequence	A_{32}	$A_{16}T_{16}$	$C_8T_8A_8G_8$	RAND
Yield, %	19.6 ± 3.2	0.6 ± 0.7	93.0 ± 1.0	79.1 ± 7.2

Table 2.7. Sequence effect on yield of dsDNA

The yield of dsDNA depends on the sequence as shown in **Table 2.7**. Using a requirement that dimers containing >80% of the maximum number of 32 bp interactions are classified as dsDNA, $C_8T_8A_8G_8$ and RAND were determined to form the largest yields of desired dsDNA over other sequences over a period of 1000 time units. However, RAND forms a slightly less amount of dsDNA than $C_8T_8A_8G_8$ since the probability of two RAND strands to make an initial contact that is non-native is higher than that by the $C_8T_8A_8G_8$ sequence (**Figures 2.6H** and **2.8H**) resulting in more kinetically trapped dimers that are partially hybridized (**Figures 2.6G** and **2.8G**). Moreover, A_{32} forms only ~20% dsDNA since its hybridization occurs mostly via the slithering mechanism resulting in a wide number of large aggregates. Surprisingly, $A_{16}T_{16}$ yields the lowest amount of highly hybridized dsDNA since its dimers tend to be in the kinetically trapped bubble state.

2.5 Conclusions

In summary, a novel coarse-grained model called BioModi was developed for nucleic acids to closely represent the structural properties of DNA strands and double helices. It is for use in conjunction with a very fast event-driven molecular dynamics algorithm, which

allows long time-scale simulations of large systems to explore hybridization and self-assembly kinetics. By taking into account sequence and solvent conditions, the model is able to capture key experimental results such as sequence-dependent thermal property from replica-exchange simulations and salt-dependent persistence length of ssDNA and dsDNA from constant-temperature simulations. Moreover, 350 constant-temperature simulations on two single strands of a homogeneous sequence were conducted to elucidate two main mechanisms of hybridization, the slow slithering mechanism and the much faster zippering mechanism.

We observe similar hybridization mechanisms as observed by both oxDNA [36] and 3SPN.2 [33]. Both models determined similar mechanisms such as the internal displacement (inchworm and psuedoknot) along with the zippering mechanism. Slithering mechanisms, defined by Sambriski *et al*., who used 3SPN.1 [43], is slightly different than the one observed by BioModi. Specifically, our slithering mechanism involves initial non-native contacts, then the breaking of these contacts in order to further hybridize the strands together into a double helix. In contrast, the description by Sambriski *et al*. involves the two strands forming initial non-native contacts then freely moving close to one another without overcoming energetic barriers in order to fully hybridize [43]. This is most likely due to their model which represents hydrogen bonding between base pairs as isotropic interactions; which were fixed in 3SPN.2 [48]. Since our hydrogen bond model is non-isotropic, direction dependent and allowing only one base to be paired to another base at a time, we observed a similar mechanism as Doye and coworkers using oxDNA [36].

All hybridization studies conducted by other groups thus far have focused on using advanced sampling techniques (forward flux sampling or transition path sampling) in

order to understand rare events and quantify the kinetics. Although these lay the foundation of the possibly mechanisms that can occur, they have not observed other complex mechanisms involved in the self-assembly of many strands at high DNA concentrations; for example, the trimer transition to dimers via the strand displacement mechanism [72-73] observed for heterogeneous sequences from our simulations. Simulations conducted in this work therefore give insights on how DNA self-assembly occurs and possible kinetic traps that exist. Therefore, our work builds on the foundation developed by oxDNA and 3SPN to further the understanding of how these complicated mechanisms affect the self-assembly of many strands.

Large-scale simulations at a high DNA strand concentration containing up to 500 strands demonstrate that DNA self-assembly is an extremely robust and enthalpically driven process. The formation of double helices is deciphered to occur via multiple self-assembly pathways: (1) by two single strands approaching each other and undergoing the zippering mechanism or possibly the slithering mechanism if the partially hybridized dimer does not form interactions with another nearby strands; (2) as a product of a transition between metastable trimers or larger aggregates into more energetically favorable dimers through the strand displacement mechanism as previously observed in experiments [72-73]. Moreover, sequence plays an important role in dictating hybridization mechanisms, which controls size distribution of self-assembled aggregates. Homogenous sequences, which hybridized through the slithering mechanism, yield kinetically trapped aggregates that are larger than dimers. When sequence heterogeneity is introduced as a diblock copolymer ($A_{16}T_{16}$), the slithering mechanism is significantly reduced yet its dsDNA yield is surprisingly low due to the formation of kinetically trapped

bubble structures. When more heterogeneity is introduced as a tetrablock copolymer ($C_8T_8A_8G_8$), the slithering mechanism is further reduced resulting the highest yield of dsDNA. Interestingly, a completely random sequence provides a high yet slightly lower yield of dsDNA compared to the tetrablock sequence.

The unpredictable simulation results presented in this work paint a complicated picture on the role of sequence on programmable self-assembly. Further studies on the effects of solvent conditions such as salt concentration and temperature on self-assembly by different sequences will provide valuable information on thermodynamics and kinetic mechanisms of hybridization in aiding the design and fabrication of complex DNA nanostructures.

It is important to note a number of limitations regarding our coarse-grained model, BioModi. Inherently, BioModi uses a simplified, united-atom representation and an implicit solvent model without physical representation of water and salt ions, which results in an approximation of the force field of the solvent. These simplifications exchange for the increase in computational tractability and performance, which allows us to conduct large-scale molecular-dynamics simulations to observe the effect of DNA sequence on hybridization in a concentrated environment. While a direct correspondence to experimental data is difficult, qualitative comparison can be established when examining the effects of temperature, salt concentration and DNA strand concentration on hybridization. For example, the strength of epsilon_HB was originally set as a reference at 1 kJ/mol, so the temperature can be scaled based on this reference point. If 1 kJ/mol was used as an absolute value, then our melting temperature T*=0.9 corresponds to 108.3 K, which is unrealistically low. In reality, the hydrogen bond strength is known ranging from 3

to 50 kJ/mol. If we used a value of 3.06 kJ/mol, which is half of a base-pair interaction representing two hydrogen bonds between A and T bases, our melting temperature of T*=0.9 at a GC content of 0% gives a temperature of 331.2 K, which is in a similar range as those observed by using the oxDNA [29, 35] or 3SPN models [19, 43, 48]. However, our melting temperature of T*=1.25 at a GC content of 40% gives a temperature of 460 K, which is higher than those observed by using the oxDNA [29, 35] or 3SPN models [19, 43, 48]. Therefore, our results are meant for qualitative rather than quantitative comparison and interpretation. We are currently working on the next version of our model focusing on re-parameterizing our potentials to match experimental values in order to make more genuine contributions to the field of DNA nanotechnology.

2.6 Supplementary Information

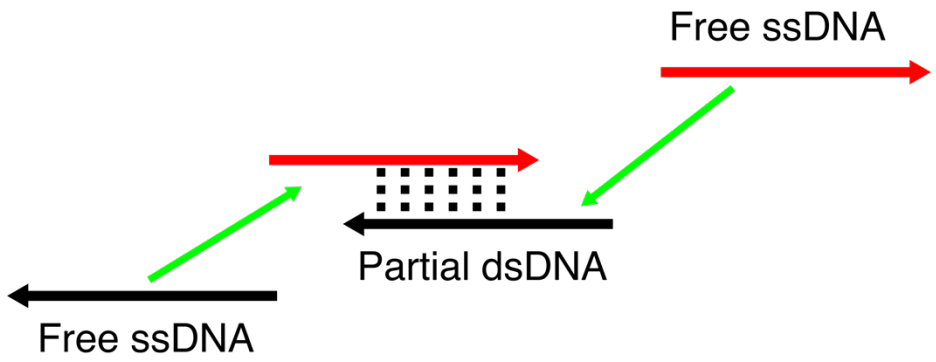

Figure S2.1: A schematic representation of a partially hybridized dimer undergoing a misaggregation process during which free strands are able to form bp interactions with unhybridized segments creating a larger aggregate.

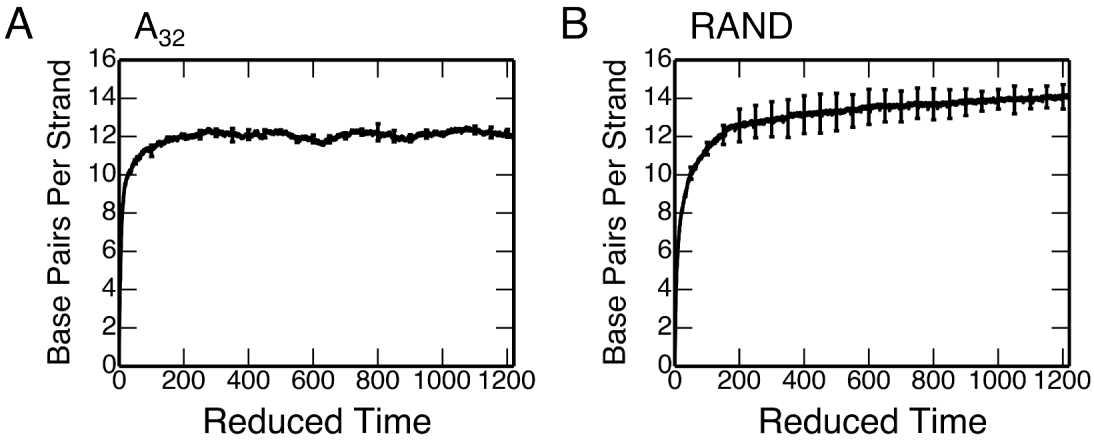

Figure S2.2: (A) Average number of base pairs per strand for A_{32}. (B) Average number of base pairs per strand for RAND. RAND has a higher number of base pairs per strand due to primarily forming dimers whereas A_{32} forms a large distribution of aggregates.

2.7 References

1. Seeman, N. C. Nucleic Acid Junctions and Lattices. *Journal of Theoretical Biology* **1982**, *99*, 237–247.
2. Rothemund, P. W. K.; Papadakis, N.; Winfree, E. PLOS Biology: Algorithmic Self-Assembly of DNA Sierpinski Triangles. *Plos Biol* **2004**, *2*, 2041-2053.
3. Malo, J.; Mitchell, J. C.; Vénien-Bryan, C.; Harris, J. R.; Wille, H.; Sherratt, D. J.; Turberfield, A. J. Engineering a 2D Protein-DNA Crystal. *Angew. Chem. Int. Ed.* **2005**, *44*, 3057–3061.
4. Lin, C.; Katilius, E.; Liu, Y.; Zhang, J.; Yan, H. Self-Assembled Signaling Aptamer DNA Arrays for Protein Detection. *Angew. Chem. Int. Ed.* **2006**, *45*, 5296–5301.
5. Carter, J. D.; LaBean, T. H. Organization of Inorganic Nanomaterials viaProgrammable DNA Self-Assembly and Peptide Molecular Recognition. *ACS Nano* **2011**, *5*, 2200–2205.
6. Goodman, R. P.; Schaap, I. A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science* **2005**, *310*, 1661–1665.
7. Zheng, J.; Birktoft, J. J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P. E.; Ginell, S. L.; Mao, C.; Seeman, N. C. From Molecular to Macroscopic via the Rational Design of a Self-Assembled 3D DNA Crystal. *Nature* **2009**, *461*, 74–77.
8. Ke, Y.; Liu, Y.; Zhang, J.; Yan, H. A Study of DNA Tube Formation Mechanisms Using 4-, 8-, and 12-Helix DNA Nanostructures. *J. Am. Chem. Soc.* **2006**, *128*, 4414–4421.
9. Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297–302.
10. Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Cristiano L. P. Oliveira; et al. Self-Assembly of a Nanoscale DNA Box with a Controllable Lid. *Nature* **2009**, *459*, 73–76.
11. Jaeger, L.; Chworos, A. The Architectonics of Programmable RNA and DNA Nanostructures. *Current Opinion in Structural Biology* **2006**, *16*, 531–543.
12. Zhang, G.; Surwade, S. P.; Zhou, F.; Liu, H. DNA Nanostructure Meets Nanofabrication. *Chem. Soc. Rev.* **2013**, *42*, 2488–2496.
13. Myhrvold, C.; Dai, M.; Silver, P. A.; Yin, P. Isothermal Self-Assembly of Complex DNA Structures Under Diverse and Biocompatible Conditions. *Nano Lett.* **2013**, *13*, 4242–4248.
14. Aksimentiev, A.; Heng, J. B.; Timp, G.; Schulten, K. Microscopic Kinetics of DNA Translocation Through Synthetic Nanopores. *Biophysical Journal* **2004**, *87*, 2086–2097.

15. Lillian, T. D.; Taranova, M.; Wereszczynski, J.; Andricioaei, I.; Perkins, N. C. A Multiscale Dynamic Model of DNA Supercoil Relaxation by Topoisomerase IB. *Biophysical Journal* **2011**, *100*, 2016–2023.
16. Yoo, J.; Aksimentiev, A. In Situ Structure and Dynamics of DNA Origami Determined Through Molecular Dynamics Simulations. *Proceedings of the National Academy of Sciences* **2013**, *110*, 20099–20104.
17. Noy, A.; Pérez, A.; Laughton, C. A.; Orozco, M. Theoretical Study of Large Conformational Transitions in DNA: the BA Conformational Change in Water and Ethanol/Water. *Nucleic Acids Research* **2007**, *35*, 3330–3338.
18. Kannan, S.; Zacharias, M. Simulation of DNA Double-Strand Dissociation and Formation During Replica-Exchange Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10589–10595.
19. Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. A Coarse Grain Model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901:1-12.
20. Ouldridge, T. E.; Johnston, I. G.; Louis, A. A.; Doye, J. P. K. The Self-Assembly of DNA Holliday Junctions Studied with a Minimal Model. *J. Chem. Phys.* **2009**, *130*, 065101:1-11.
21. Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. Coarse-Grained Model of Nucleic Acid Bases. *J. Comput. Chem.* **2010**, *31*, 1644–1655.
22. Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. PRIMO/PRIMONA: a Coarse-Grained Model for Proteins and Nucleic Acids That Preserves Near-Atomistic Accuracy. *Proteins* **2010**, *78*, 1266–1281.
23. Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711–1725.
24. Savelyev, A.; Papoian, G. A. Chemically Accurate Coarse Graining of Double-Stranded DNA. *Proceedings of the National Academy of Sciences* **2010**, *107*, 20340–20345.
25. Xia, Z.; Gardner, D. P.; Gutell, R. R.; Ren, P. Coarse-Grained Model for Simulation of RNA Three-Dimensional Structures. *J. Phys. Chem. B* **2010**, *114*, 13497–13506.
26. Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. A Systematically Coarse-Grained Model for DNA and Its Predictions for Persistence Length, Stacking, Twist, and Chirality. *J. Chem. Phys.* **2010**, *132*, 035105:1-18.
27. DeMille, R. C.; Cheatham, T. E., III; Molinero, V. A Coarse-Grained Model of DNA with Explicit Solvation by Water and Ions. *J. Phys. Chem. B* **2011**, *115*, 132–142.
28. Biyun, S.; Cho, S. S.; Thirumalai, D. Folding of Human Telomerase RNA Pseudoknot Using Ion-Jump and Temperature-Quench Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 20634–20643.
29. Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model. *J. Chem. Phys.* **2011**, *134*, 085101:1-22.
30. Hsu, C. W.; Fyta, M.; Lakatos, G.; Melchionna, S.; Kaxiras, E. Ab Initio Determination of Coarse-Grained Interactions in Double-Stranded DNA. *J. Chem. Phys.* **2012**, *137*, 105102:1-12.
31. Cragnolini, T.; Derreumaux, P.; Pasquali, S. Coarse-Grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047–8060.
32. He, Y.; Maciejczyk, M.; Ołdziej, S.; Scheraga, H.; Liwo, A. Mean-Field Interactions Between Nucleic-Acid-Base Dipoles Can Drive the Formation of a Double Helix. *Phys. Rev. Lett.* **2013**, *110*, 098101:1-5.
33. Hinckley, D. M.; Lequieu, J. P.; de Pablo, J. J. Coarse-Grained Modeling of DNA Oligomer Hybridization: Length, Sequence, and Salt Effects. *J. Chem. Phys.* **2014**, *141*, 035102:1-12.
34. Maffeo, C.; Ngo, T. T.; Ha, T.; Aksimentiev, A. A Coarse-Grained Model of Unstretched Single-Stranded DNA Derived From Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.* **2014**, 10, 2891-2896.
35. Šulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P. K.; Louis, A. A. Sequence-Dependent Thermodynamics of a Coarse-Grained DNA Model. *J. Chem. Phys.* **2012**, *137*, 135101:1-15.
36. Ouldridge, T. E.; Sulc, P.; Romano, F.; Doye, J. P. K.; Louis, A. A. DNA Hybridization Kinetics: Zippering, Internal Displacement and Sequence Dependence. *Nucleic Acids Research* **2013**, *41*, 8886–8895.
37. Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. DNA Nanotweezers Studied with a Coarse-Grained Model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101–178104.
38. Ouldridge, T. E.; Hoare, R. L.; Louis, A. A.; Doye, J. P. K.; BATH, J.; Turberfield, A. J. Optimizing DNA Nanotechnology Through Coarse-Grained Modeling: a Two-Footed DNA Walker. *ACS Nano* **2013**, *7*, 2479–2490.
39. Romano, F.; Chakraborty, D.; Doye, J. P.; Ouldridge, T. E.; Louis, A. A. Coarse-Grained Simulations of DNA Overstretching. *J. Chem. Phys.* **2013**, *138*, 085101:1-10.

40. Srinivas, N.; Ouldridge, T. E.; Sulc, P.; Schaeffer, J. M.; Yurke, B.; Louis, A. A.; Doye, J. P. K.; Winfree, E. On the Biophysics and Kinetics of Toehold-Mediated DNA Strand Displacement. *Nucleic Acids Research* **2013**, *41*, 10641–10658.
41. De Michele, C.; Rovigatti, L.; Bellini, T.; Sciortino, F. Self-Assembly of Short DNA Duplexes: From a Coarse-Grained Model to Experiments Through a Theoretical Link. *Soft Matter* **2012**, *8*, 8388–8398.
42. Rovigatti, L.; Bomboi, F.; Sciortino, F. Accurate Phase Diagram of Tetravalent DNA Nanostars. *J. Chem. Phys.* **2014**, *140*, 154903:1-11.
43. Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. A Mesoscale Model of DNA and Its Renaturation. *Biophysical Journal* **2009**, *96*, 1675–1690.
44. Sambriski, E. J.; Ortiz, V.; de Pablo, J. J. Sequence Effects in the Melting and Renaturation of Short DNA Oligonucleotides: Structure and Mechanistic Pathways. *J. Phys.: Condens. Matter* **2008**, *21*, 034105:1-13.
45. Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. Uncovering Pathways in DNA Oligonucleotide Hybridization via Transition State Analysis. *Proceedings of the National Academy of Sciences* **2009**, *106*, 18125–18130.
46. Schmitt, T. J.; Knotts, T. A. Thermodynamics of DNA Hybridization on Surfaces. *The Journal of Chemical Physics* **2011**, *134*, 205105:1-9.
47. Schmitt, T. J.; Rogers, J. B.; Knotts, T. A., IV. Exploring the Mechanisms of DNA Hybridization on a Surface. *The Journal of Chemical Physics* **2013**, *138*, 035102:1-14.
48. Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. An Experimentally-Informed Coarse-Grained 3-Site-Per-Nucleotide Model of DNA: Structure, Thermodynamics, and Dynamics of Hybridization. *J. Chem. Phys.* **2013**, *139*, 144903:1-17.
49. Escobedo, F. A.; Borrero, E. E.; Araque, J. C. Transition Path Sampling and Forward Flux Sampling. Applications to Biological Systems. *J. Phys.: Condens. Matter* **2009**, *21*, 333101:1-23.
50. Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J. A Coarse-Grain Three-Site-Per-Nucleotide Model for DNA with Explicit Ions. *The Journal of Chemical Physics* **2011**, *135*, 165104:1-13.
51. Smith, S.; Hall, C.; Freeman, B. Molecular Dynamics for Polymeric Fluids Using Discontinuous Potentials. *Journal of Computational Physics* **1997**, *134*, 16–30.
52. Nguyen, H. D.; Hall, C. K. Molecular Dynamics Simulations of Spontaneous Fibril Formation by Random-Coil Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16180–16185.
53. Fu, I. W.; Markegard, C. B.; Chu, B. K.; Nguyen, H. D. The Role of Electrostatics and Temperature on Morphological Transitions of Hydrogel Nanostructures Self-Assembled by Peptide Amphiphiles via Molecular Dynamics Simulations. *Adv Healthc Mater* **2013**, *2*, 1388–1400.
54. Ding, F.; Sharma, S.; Chalasani, P.; Demidov, V. V.; Broude, N. E.; Dokholyan, N. V. Ab Initio RNA Folding by Discrete Molecular Dynamics: From Structure Prediction to Folding Mechanisms. *RNA* **2008**, *14*, 1164–1173.
55. Voegler Smith, A.; Hall, C. K. Alpha-Helix Formation: Discontinuous Molecular Dynamics on an Intermediate-Resolution Protein Model. *Proteins* **2001**, *44*, 344–360.
56. Nguyen, H. D.; Marchut, A. J.; Hall, C. K. Solvent Effects on the Conformational Transition of a Model Polyalanine Peptide. *Protein Science* **2004**, *13*, 2909–2924.
57. Macke, T. J.; Case, D. A. Modeling Unusual Nucleic Acid Structures. In *Molecular Modeling of Nucleic Acids*; Leontis, N. B.; SantaLucia, J., Eds.; American Chemical Society: Washington, DC, 1998; pp. 379–393.
58. Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: the Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
59. SantaLucia, J. A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460–1465.
60. Jure ka, P.; poner, J.; ern, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
61. Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31*, 459–466.
62. Andersen, H. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.

63. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chemical Physics Letters* **1999**, *314*, 141–151.
64. Ullner, M.; Woodward, C. Orientational Correlation Function and Persistence Lengths of Flexible Polyelectrolytes. *Macromolecules* **2002**, *35*, 1437–1445.
65. Murphy, M. C.; Rasnik, I.; Cheng, W.; Lohman, T. M.; Ha, T. Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy. *Biophysical Journal* **2004**, *86*, 2530–2537.
66. Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. Ionic Effects on the Elasticity of Single DNA Molecules. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6185–6190.
67. Aboul-Ela, F.; Koh, D.; Tinoco, I., Jr; Martin, F. H. Base-Base Mismatches. Thermodynamics of Double Helix Formation for dCA 3XA 3G + dCT 3YT 3G (X, Y = a,C,G,D. *Nucleic Acids Research* **1985**, *13*, 4811–4824.
68. SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu Rev Biophys Biomol Struct* **2004**, *33*, 415–440.
69. Owczarzy, R.; You, Y.; Moreira, B. G.; Manthey, J. A.; Huang, L.; Behlke, M. A.; Walder, J. A. Effects of Sodium Ions on DNA Duplex Oligomers: Improved Predictions of Melting Temperatures. *Biochemistry* **2004**, *43*, 3537–3554.
70. Wetmur, J. G.; Davidson, N. Kinetics of Renaturation of DNA. *Journal of Molecular Biology* **1968**, *31*, 349–370.
71. Wetmur, J. G. Hybridization and Renaturation Kinetics of Nucleic Acids. *Annual review of biophysics and bioengineering* **1976**, *5*, 337–361.
72. Turberfield, A. J.; Yurke, B.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. A DNA-Fuelled Molecular Machine Made of DNA : Abstract : Nature. *Nature* **2000**, *406*, 605–608.
73. Qian, L.; Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **2011**, *332*, 1196–1201.

Chapter 3 Effect of Sequence, Temperature, and DNA Concentration on dsDNA Self-Assembly via Coarse-Grained Simulations

3.1 Abstract

A novel coarse-grain model is utilized to elucidate the effects of sequence, temperature and concentration on DNA self-assembly. Large scale simulations demonstrate the presence of many kinetic pathways that depend on their environmental conditions. Trends in aggregate lifetimes, creation times, and base pairing are evaluated for each of the environmental conditions studied. Additionally, studies on equivalent G-C content sequences demonstrates the importance of DNA sequence design. Finally, from kinetic modeling of self-assembly, an experimentally accurate activation energy was found. The framework developed in this work can be applied to future investigations of DNA nanostructures.

3.2 Introduction

The study of DNA hybridization and renaturation has been extensively studied since the 1960's with Marmur and Doty [1]. A more comprehensive understanding of hybridization was the accomplished to understand the effects of sequence and environmental variables [2-3]. Inspired by the nature of DNA to renature and branch migration in genetics [4-5], Seeman developed an algorithmic method of having DNA strands hybridize into a 2D tile structure similar to the Holliday junction [6]. This innovation led to others utilizing DNA for developing complex nanostructures such as self-assembled DNA tetrahedrons, nanotubes, smiley faces, and squares [7-10]. Researchers have developed tools such as CadNano [11], and NuPACK [12] to help in the design of the

nanostructures. Additionally, the field of dynamic DNA nanostructures is very active with experimentalist and simulations investigating DNA nanomachines [13-15], DNA logic gates [16-18], and targeted delivery [19]. These dynamic DNA devices take advantage of the strand displacement mechanism [20]. The mechanism has been under investigation by experimentalist since the 1970's while exploring the DNA of *Escherichia coli* 15 under electron microscopy [4]**.** Strand displacement has recently been applied to the world of DNA nanotechnology, for example using DNA for computation. Aldeman was able to develop a method to obtain solutions to a Hamiltonian path problem using DNA and enzymes [21]. This has led to others to develop logic gates [18, 22], complex networks [23-24], and even programming languages using DNA strand displacement [25]. Microsoft Research labs has helped develop a program called DSD (DNA Strand Displacement tool) [26-27] for helping researchers design and test DNA computational devices. Both sides of the DNA nanotechnology field can benefit from the insights molecular dynamics can provide into the self-assembly processes and dynamic nature of DNA. This can be used to help improve assembly yields, reduce cross-talk in logic circuits, and being able to estimate reaction kinetics.

Molecular dynamics of nucleic acids has been evaluated by both all-atom [28-32] and coarse-grained force fields [33-48]. Although all-atom force fields give a high amount of detail, it is currently not computationally feasible to simulate the hybridization of DNA. Therefore, researchers have developed coarse-grained force fields to capture the physics at a larger scale. Coarse-grained models can vary from many nucleotides represented by one bead, which was used to investigate how DNA hybridization when covalently attached to

large metal particles [49], to a single nucleotide being represented by 3 beads, such as in the work of Dokholyan [50] and de Pablo [33, 47, 51].

In this work, we utilize BioModi, a coarse-grained simulation package developed in our own lab for studying self-assembly processes of peptides and/or nucleic acids (**Figure 3.1**). Biomodi combines our ePRIME model for peptides and polymers with a 3-site-per-nucleotide model for nucleic acids. The ePRIME model has been successful in modeling the aggregation of proteins [52] and peptide amphiphiles [53-55]. Our implementation of the nucleic acid model builds on 3-site-per-nucleotide models developed by the Dokholyan [50] and and de Pablo [33, 47, 51] groups includes the following interactions: psuedobonds, base stacking, electrostatic, solvent induced, and directional base pairing. We have previously shown that our DNA model captures the persistence length of ssDNA and dsDNA, and qualitatively captures the melting temperature of DNA [56]. Additionally, we have observed three kinetic mechanisms related to hybridization during a large scale self-assembly simulation: slithering, zippering, and strand displacement. Although we have qualitatively explained the differences of the mechanisms in our previous work, this was accomplished focusing on a single temperature, salt concentration and DNA concentration. In this work, we elucidate the full effects of those factor on the formation of double helices via traditional kinetic modeling of the self-assembly process. Previous works by the Doye [57] and de Pablo [47, 58] research groups have both used forward-flux sampling (FFS) in order to obtain the rate constant of the hybridization process of two strands. Although FFS is a very elegant method for obtaining rate constants, the drawback of these works is not obtaining the rate constants with more than two chains in the simulation. For our work, we

directly simulate the assembly of 100 strands (50 sense and 50 anti-sense), thus we are able to explore kinetic pathways not accessible in the aforementioned works.

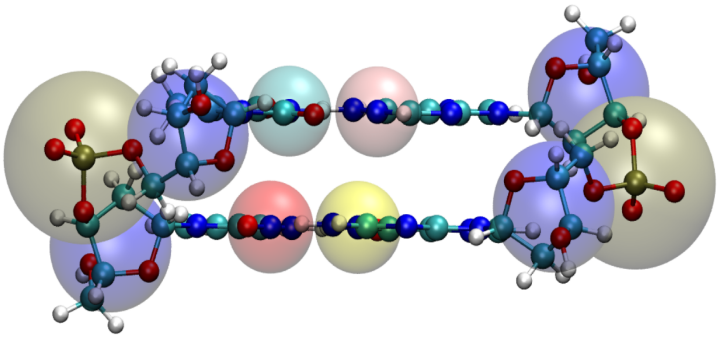

Figure 3.1: BioModi representation of each of the 4 nucleotides. Adenine in pink, thymine in cyan, cytosine in red, and guanine in yellow.

3.3 Methods

3.3.1 Simulation Methods

Simulations were performed by using the DMD (Discontinuous Molecular Dynamics) simulation algorithm [59], a variant on standard molecular dynamics that is applicable to systems of molecules interacting via discontinuous potentials, e.g., hard-sphere, and square-well potentials. Unlike soft potentials such as the Lennard-Jones potential, discontinuous potentials exert forces only when particles collide, enabling the exact (as opposed to numerical) solution of the collision dynamics. DMD simulations proceed by locating the next collision, advancing the system to that collision, and then

calculating the collision dynamics. Simulations are performed in the canonical ensemble with periodic boundary conditions imposed to eliminate artifacts due to box walls. Constant temperature is achieved by implementing the Andersen thermostat method [60]. In this case, all beads are subjected to random, infrequent collisions with ghost particles whose velocities are chosen randomly from a Maxwell–Boltzmann distribution centered at the system temperature. Simulation temperature is expressed in terms of the reduced temperature, $T^* = k_bT/\varepsilon_{HB}$, where k_b is Boltzmann's constant, T is the absolute temperature, and ε_{HB} is the strength of one hydrogen bond (initially set at 1 kJ/mol as a reference so that other potential strengths can be accordingly scaled). Simulation time is defined in reduced units to be $t^* = t/\sigma(k_bT/m)^{1/2}$, where t is the simulation time, and σ and m are the average bead diameter and mass, respectively.

3.3.1.1 Self-Assembly of Many Strands

Self-assembly simulations were performed for two sequences, each with 100 strands of DNA (50 sense strands and 50 antisense strands, sequences are listed in **Table 3.1**). These strands were placed randomly in the simulation box as an initial configuration. They were heated at high temperature ($T^* = 5.0$) for a period of time to ensure there is no bias in the initial randomization. Velocities of the particles were subsequently reinitialized to a desired temperature and conducted with the NVT ensemble.

Name	**Sequence (5' -> 3')**
C8T8A8G8	CCCC CCCC TTTT TTTT AAAA AAAA GGGG GGGG
RAND32	CCAA TGCG GTAA GCCT GACA CCGA TCAA TCTT

Table 3.1: Sequences used in this study C8T8A8G8 and RAND32 listed from 5' to 3'.

Our simulations are aimed to elucidate the effects of sequence, temperature and DNA concentration on the self-assembly of dsDNA. We evaluated temperatures from T* = 0.7 to T* = 1.3, and DNA concentrations from 1.45 mM to 0.09 mM. For the majority of this study we utilize 20 replicates for each condition, however at low concentrations, we only utilized 4 due to the large computational time it takes each replicate. Although these are much higher than experimental values (typically on the order of nM-μM), as exploring experimental concentrations is currently computationally unfeasible. A summary of the simulations used in this study is available in **Table 3.2**. Thus, we have simulated a total of 208 large scale simulations were ran for over a month on high performance computers on a single processor each. Each constant-temperature simulation was conducted for at least 600 time units. This was determined enough time for our simulations to reach equilibrium as indicated by the ensemble averages of the system's total potential energy which varied by no more than 2.5% toward the end of each simulation run.

Name	[DNA] (mM)	Temperatures (T*)	[NaCl] (mM)	Replicates
C8T8A8G8	1.45	0.7, 0.9, 1.0, 1.1, 1.3	69	20
RAND32	1.45	0.7, 0.9, 1.0, 1.1, 1.3	69	20
C8T8A8G8	0.36, 0.09	0.9, 1.1	69	4

Table 3.2: Summary of simulations used in this study. Total number of simulations is 208.

3.3.2 Analysis Method

Simulations were analyzed using an in-house analysis package developed for BioModi in FORTRAN. The code determines which strands are aggregated together via our criterion of having at least one base pair. Using this data, we are able to extract out the

unique aggregates and their properties using an in-house Python module. The module finds each unique aggregate by filtering through a Pandas data frame from the FORTRAN output. Each aggregate is stored as an Aggregate class with internal variables for the chains involved, the simulation frames it is present in, and the number of base pairs for each of the frames. Knowing this data for each aggregate we were able to determine the longest lifetime for each aggregate using a common longest increasing sequence algorithm. From the longest lifetime, we are able to determine the number of bp during that time period, and when that time period starts in our simulation (creation time). This data was then visualized in an iPython Notebook with matplotlib, and the results of this analysis can be seen in section 3.4.

Additionally, for our kinetic study we are interested in the aggregation of strands forming dimers and trimers only. Previously we have shown that there are three mechanisms for DNA hybridization that can occur during self-assembly: zippering, slithering, and strand displacement. The latter is a hybridization mechanism that has not been explored during a large scale self-assembly using molecular dynamics, but is responsible for the dynamic nature of DNA molecular machines and DNA computing devices. In order for our kinetic study to capture the all three mechanisms, we only consider the strands that are involved in dimer and trimer reactions. Since there are complex aggregates such as tetramers and above that could form we found it was necessary to filter these chains involved in other complex mechanisms due to their small percentage of the total number of chains for each aggregate. Filtering for only chains involved in dimer and trimer reactions we were able to obtain 60% of the chains per

simulation in our analysis. Therefore, the formation of dimers and trimers can be written with the following equations:

$$\text{1f. } \boldsymbol{M} + \boldsymbol{M} \xrightarrow{k_{1f}} \boldsymbol{D}$$
$$\text{1b. } \boldsymbol{D} \xrightarrow{k_{1b}} \boldsymbol{M} + \boldsymbol{M}$$
$$\text{2f. } \boldsymbol{M} + \boldsymbol{D} \xrightarrow{k_{2f}} \boldsymbol{T}$$
$$\text{2b. } \boldsymbol{T} \xrightarrow{k_{2b}} \boldsymbol{M} + \boldsymbol{D}$$

where M is monomer, D is dimer, and T is the trimer. Resulting is the set of differential equations are:

$$\frac{\boldsymbol{d}[\boldsymbol{M}]}{\boldsymbol{dt}} = -k_{1f}[\boldsymbol{M}]^2 + k_{1b}[\boldsymbol{D}] - k_{2f}[\boldsymbol{D}][\boldsymbol{M}] + k_{2b}[\boldsymbol{T}]$$

$$\frac{\boldsymbol{d}[\boldsymbol{D}]}{\boldsymbol{dt}} = k_{1f}[\boldsymbol{M}]^2 - k_{1b}[\boldsymbol{D}] - k_{2f}[\boldsymbol{D}][\boldsymbol{M}] + k_{2b}[\boldsymbol{T}]$$

$$\frac{\boldsymbol{d}[\boldsymbol{T}]}{\boldsymbol{dt}} = k_{2f}[\boldsymbol{D}][\boldsymbol{M}] - k_{2b}[\boldsymbol{T}]$$

Using this set of differential equations, we implemented a least squares regression on the simulation systems in order to find the rate constants ($k_{1f,b}$, $k_{2f,b}$). Specifically, we used the following set of python libraries: NumPy, SciPy and lm-fit to numerically solve the coupled differential equations and fitting of the rate constants. Lm-fit is an extension of Scipy's fitting method to put bounds on the rate constants, e.g. not allowing the solver to use values of $k_{1f,b} < 0$ or $k_{2f,b} < 0$.

3.4 Results and Discussion

3.4.1 Effect of Sequence on Kinetics of dsDNA Self-Assembly

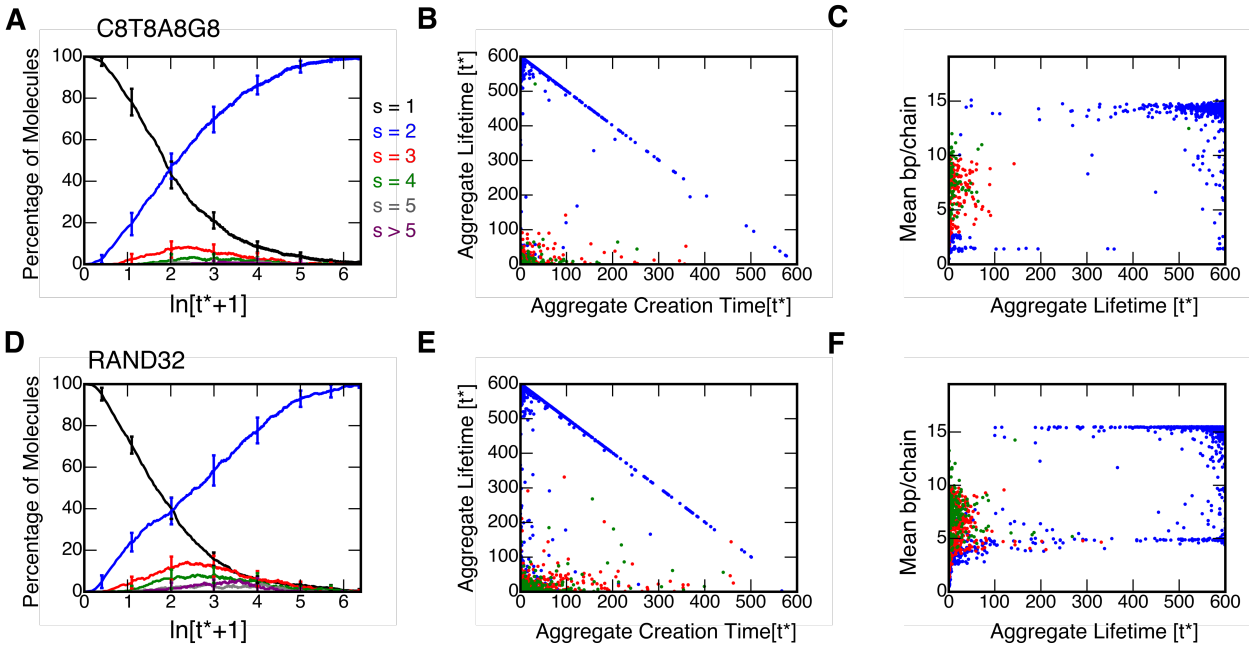

Figure 3.2: Effect of sequence on C8T8A8G8 and RAND32 self-assembly at [DNA] = 1.45mM, [NaCl] = 69 mM, T* = 1.0. Percentage of molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale for (A) C8T8A8G8, and (D) RAND32 sequences. Unique aggregate creation times and lifetimes plotted as a function of reduced time (t*) for (B) C8T8A8G8, and (E) RAND32. Unique aggregates lifetime and the mean number of base pairs per chain during the lifetime for (C) C8T8A8G8, and (F) RAND32.

Here we compare the two sequences of interest, C8T8A8G8 and RAND32 at T* = 1.0, a [DNA] = 1.45 mM, and [NaCl] = 0.069 M. Although both sequences form 100% dimers at the end of simulations as shown in **Figure 3.2A** and **3.2B**, there are noted differences of the population of intermediates during self-assembly. C8T8A8G8 initially forms a set of dimers from the free monomers in solution (**Figure 3.2A**). After a short delay, at ln[t* +1] = 0.7, trimers start forming and tetramers appear at ln[t* + 1] = 1.8. Both trimer and tetramer species reach a maximum around ln[t*+1] = 2.5, and then dissociate back into dimers and monomers. Free monomers in solution then form additional dimers. RAND32, similar to

the tetrablock, initially forms a set of dimers (**Figure 3.2D**), however trimers begin forming much quicker at ln[t* + 1] = 0.5. Then tetramers start forming at ln[t* + 1] = 1.0, followed by pentamers at ln[t* + 1] = 1.5, and larger aggregates at ln[t* + 1] = 3. Since these large aggregates are forming, the monomers for RAND32 are depleted quicker than the C8T8A8G8 system. All the large aggregates live past ln[t* + 1] = 5 for RAND32, whereas the trimers and tetramers of C8T8A8G8 are dissociated by ln[t* + 1] = 4.75. Therefore, both sequences are able to form dsDNA, but differ in their kinetic pathways to reach there. Additionally, RAND32 and C8T8A8G8 have the same G-C content so both are equally enthalpically driven to form dimers.

The differences of the two sequences kinetics can be further quantified by analyzing the trends in the aggregates formed throughout the simulation. In Figure **3.2B and 3.2E,** the aggregate creation time and lifetimes are plotted for all the aggregates for the 20 simulations conducted. From this it is evident that both sequences form a set of dimers that are long lived and created at the beginning of the simulation. Trimers and tetramers are formed during the beginning of the simulation (t* < 200) and are short lived. Dimers are also formed during the middle and ends of the simulation, indicating that some are formed from the dissociated trimers and tetramers. RAND32 forms more unique trimers and tetramers, 64.25 ± 3.98 and 18.65 ± 3.6, than that of the C8T8A8G8 sequence, 55.35 ± 2.35 and 9.3 ± 3.1, as observed in **Figure 3.2B,E** and plotted as a function of temperature for short and long lifetimes in **Figure S3.1**.

Next we can evaluate how the number of base pairs per chain affects the aggregate lifetimes. In **Figure 3.2C,F**, are the mean number of bp/chain and the aggregate lifetimes for dimers, trimers and tetramers of both sequences. For both C8T8A8G8 and RAND32, the

long lifetime dimers have approximately ~12-16 mean bp/chain, with 16 being the maximum for our sequences. It is apparent for trimers and tetramers that the reason they are short lived is that they have a low number of bp/chain, thus are not maximizing the number of base pairs. Both sequences have a metastable dimer that exists at low bp (< 5), however RAND32 forms significantly more. This group of aggregates will be further addressed in section 3.4.2.

A reason for RAND32 to form these misaggregates while forming the desired dsDNA, is due to the large number of possible base pair contacts compared to C8T8A8G8. The tetrablock sequence has a narrower path for possible contacts, therefore reducing the possible kinetic pathways observed in the RAND32 sequence. We have previously discussed these contact maps in our previous work [56].

3.4.2 Effect of Temperature on dsDNA Kinetic Self-Assembly

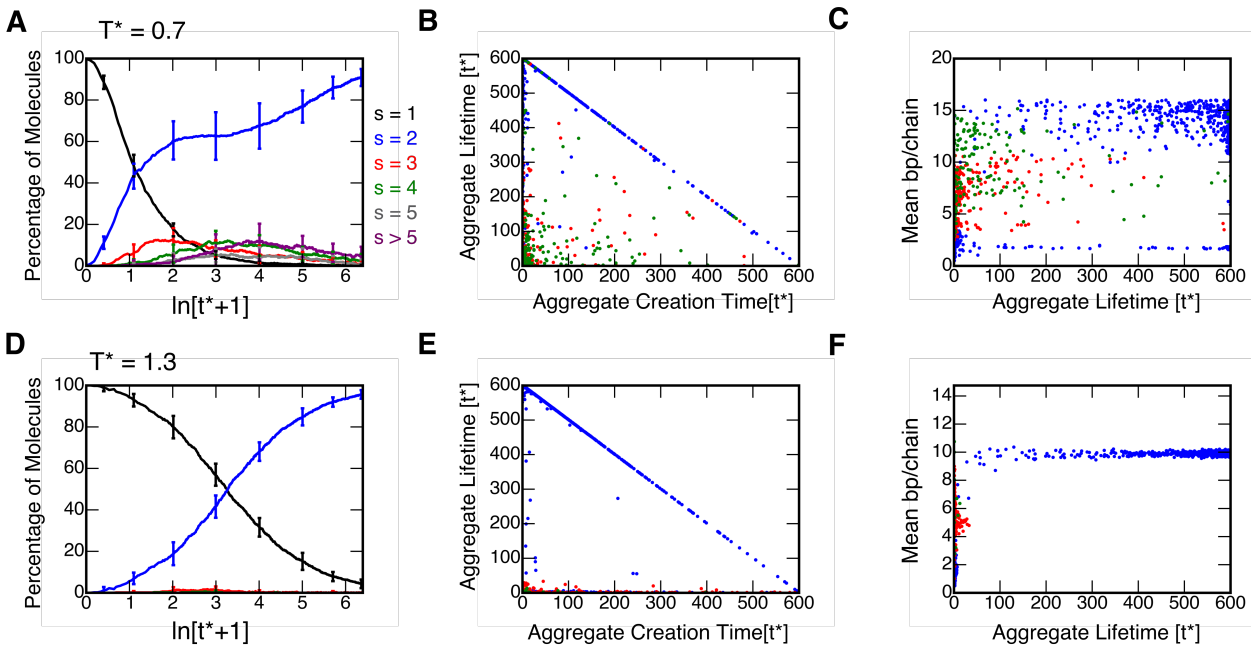

Figure 3.4: Effect of temperature on C8T8A8G8 self-assembly at [DNA] = 1.45mM and [NaCl] = 69 mM. Percentage of molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale for (A) T* = 0.7, and (D) T* = 1.3 reduced temperatures. Unique aggregate creation times and lifetimes plotted as a function of reduced time (t*) for (B) T* = 0.7, and (E) T* = 1.3 reduced temperatures. Unique aggregates lifetime and the mean number of base pairs per chain during the lifetime for (C) T = 0.7, and (F) T* = 1.3 reduced temperatures.

In addition to sequence, the effect of temperature on DNA self-assembly was also investigated. This analysis was performed for both sequences, but the following focuses on a discussion of the C8T8A8G8 results. The same analysis plots for RAND32 can be found in the Supplementary Information. In **Figure 3.4A**, T* = 0.7, dimers initially form, then trimers appear at ln[t* +1] = 0.5. There is then a flattening of the rate of formation of the dimers, while monomers and dimers are used to form trimers, tetramers and larger aggregates also appear. These aggregates maximize at ln[t* + 1] = 4, but then dissociate a slow rate compared to T* = 1.0 as discussed in the sequence section (3.4.1). At the end of the T* = 0.7 simulation there exists aggregates larger than dimers, unlike those at higher

temperatures. For self-assembly simulations conducted at T* = 1.3, dimers form very slowly due to the increased kinetic energy in the system. A very small amount of trimers do form, but constitute less than 5 % of the total ssDNA molecules. At the end of the simulation there exists both dimers and monomers in solution.

Aggregate lifetimes and creation times were evaluated for both temperatures in **Figure 3.4B,E.** At T* = 0.7, there exists many dimers, trimers, and tetramers with a large proportion having lifetimes under t* = 100 reduced units. At high temperatures, there are long lived dimers, but also many short lived dimers, trimers and tetramers. Trimers are created throughout the simulation but live very short lifetimes. Due to the availability of monomers in system at high temperature, these free chains can base pair shortly with other dimers. Additionally, due to the high kinetic energy, dimers are created throughout the simulation.

To evaluate a factor of aggregate lifetimes, one can obtain the mean number of base pairs per chain as seen in **Figure 3.4C,F**. At low T*, tetramers are able to form almost the maximum number of bp/chain, and generally have longer lifetimes than at moderate to high temperatures. Trimers also have a higher number of base pairs per chain than at higher temperatures. Dimers were found to have a larger spread for the mean number of bp than at high temperatures for aggregates with long lifetimes. This indicates that base pairing for our self-assembly simulations is highly dependent on temperature, as expected physically. Additionally, a set of meta-stable dimers exist with ~2 bp, but are not found at higher temperatures. At high temperatures there exists many dimers with short lifetimes, and low numbers of bp indicating they are not forming enough bp to overcome the kinetic energy of the system. We find trimers are generally short lived at high temperatures, with

most forming ~ 5 bp/chain. Long lived dimers at high temperatures exist with ~ 10 bp/chain, which is 6 bp/chain less than the maximum.

3.4.3 Effects of DNA Concentration on Self-Assembly

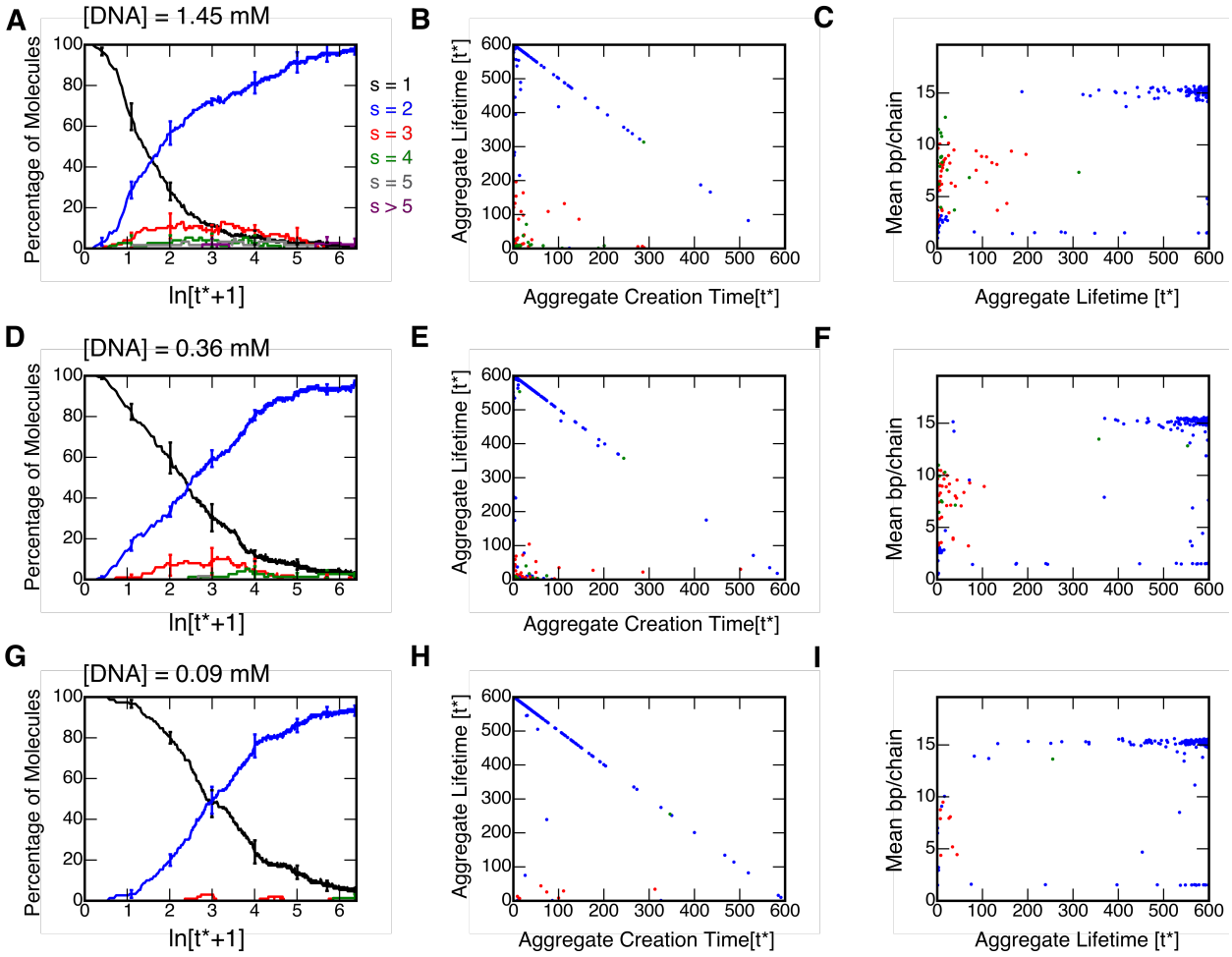

Figure 3.5: Effect of DNA concentration on C8T8A8G8 self-assembly at T* = 0.9. Percentage of molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale for (A) 1.45 mM, (D) 0.36mM, and (G) 0.09 mM. Unique aggregate creation times and lifetimes plotted as a function of reduced time (t*) for (B) 1.45 mM, (E) 0.36 mM, and (H) 0.09 mM. Unique aggregates lifetime and the mean number of base pairs per chain during the lifetime for (C) 1.45 mM, (F) 0.36 mM, and (I) 0.09 mM.

We evaluated the effect of concentration on DNA self-assembly through studying three different concentrations: 1.45, 0.36, and 0.09 mM at T* = 0.9 and 1.1. In **Figure 3.5** is a summary of the data for the three concentrations at T* = 0.9, and similar graphs can be

found in **Figure S3.3 for T*=1.1**. For a DNA concentration of 1.45 mM, dimers start forming by ln[t* +1] ~ 0.25, with the formation of trimers and tetramers nearly instantaneous after (**Figure 3.5A**). This is caused by the addition of free monomers in solution base pairing with partially hybridized dimers. This monomer addition to these aggregates leads to the formation of larger aggregates such as pentamers and hexamers. These large aggregates dissociate by the end of the simulation and form nearly 100% of the molecules are classified as dimers. At a lower concentration of 0.36 mM , the creation of dimers is delayed with respect to the higher concentration, as seen in **Figure 3.5D** occurring at ln[t* + 1] ~ 0.5. The formation of trimers starts ln[t* +1] ~ 1, and tetramers start forming at ln[t* +1] ~ 2.5. Both trimers and tetramers dissociate by the end of the simulation in favor for dimers. At our lowest concentration, 0.09 mM, the dimer lag time is further increased to ln[t* + 1] ~ 0.75 (**Figure 3.5G**). There exists some trimers and tetramers at this concentration however they account for less than 5% of the molecules in the simulation.

In Figure **3.5B, E,H** we plot the aggregate creation times and lifetimes. All three concentrations form a set of stable dimers initially, as seen in the upper left (0-100, 500-600) of each of the scatterplots. Lag times for dimers are also observed in the same region as the amount of points in that region decreases with decreasing concentration. Also trimers form near the beginning of the simulation but have short lifetimes for all concentrations. It is important to note the decreasing count of trimers and tetramers as the concentration decreases. Dimers are formed throughout the simulation due to the dissociation of the larger aggregates as seen along the y = x.

Base pairs play an important role for aggregate lifetimes as observed in **Figure 3.5C, F, I**. For all three concentrations, they form two distinct sets of stable dimers, those that have ~ 16bp/chain, and the meta-stable 2 bp/chain as seen previously at low and moderate temperatures. Additionally, at moderate to high concentrations, trimers have ~ 10 bp/chain, therefore having less than the theoretical maximum of 16 bp/chain. Additionally, the tetramers at these concentrations form 5 -10 bp. These low values of bp/chain result in the aggregates not being long lived and taking full enthalpic advantage of the available base pairing. It is therefore enthalpically favorable to form full duplexes as these are capable of forming 16 bp/chain.

3.4.4 Kinetic Modeling of dsDNA Self-Assembly

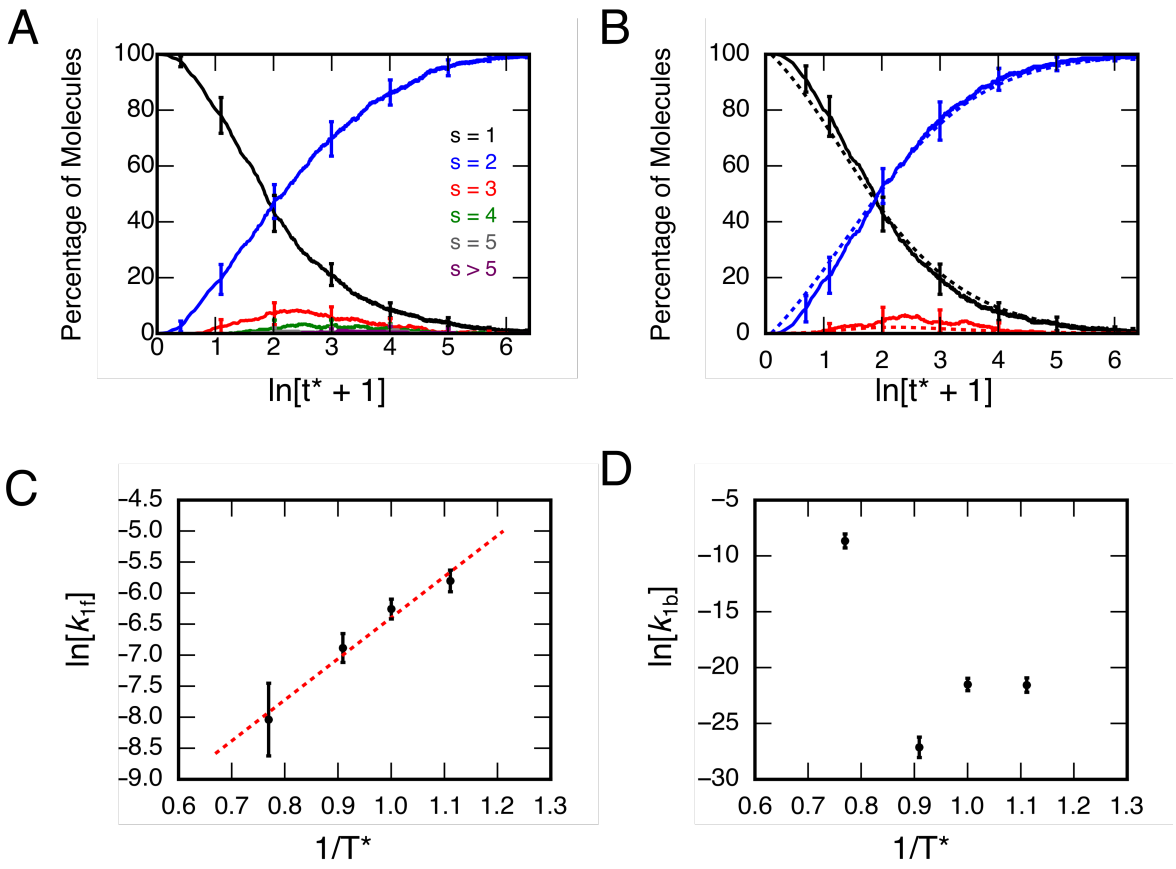

Figure 3.6: Kinetic Fits for C8T8A8G8 at 1.45 mM. (A) Percentage of molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale; (B) Percentage of filtered molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale; Arrhenius plots for (C) k1f and (D) k1b. Red line in (C) is the fitted line for the linearized Arrhenius equation.

In order to quantify the kinetics of the tetrablock according to the rate equations in the Analysis Methods, it as necessary to extract only relevant dimers, trimers, and monomers that are only involved in these aggregates. For each temperature T* = 0.9-1.3, we were able to extract the previously described chains. An example of the unfiltered kinetics is available in Figure 3.6A, and the filtered version in Figure 3.6B with the kinetic fit in the dashed lines. Using our rate equations and the data collected for each temperature

we were able to obtain values for the k-values. The fit is very close to our simulation values and stays within the error bars of the standard deviation of the simulation.

From the fitted k-values they can be plotted with 1/T* and ln(k), or Arrhenius style plots. The values for the the first rate equation, k_{1f} and k_{1b} are in Figure 3.6C and 3.6D. The values for k_{1f} have an increasing slope with regards to 1/T*, therefore having negative activation energy (E_A). This is expected as systems dependent on potential wells commonly have negative activation energy. The k-values were found for tetrablock at different concentrations of DNA and it was found that the average activation energy is -50.4 ± 4.03 kJ/mol, similar the the values obtained by experimentalist [61-62] and another coarse-grained model by Oulridge *et al*. [57] with activation energies ranging from ∼-32 to ∼-75 kJ/mol. Additionally the pre-exponential factor was found to be 1.63E-4 ± 4.15E-7 1/(t* ssDNA%). The values for k_{1b} follows non-Arrehenius kinetics also, with similar values close to zero for T* = 0.9 - 1.1 and at T* = 1.3 the value is no longer close to zero. This makes physical sense with our results of having monomers present at equilibrium.

3.5 Conclusion

In this work we use a novel coarse-grained model to study the effects of sequence, temperature and DNA concentration on the self-assembly process. Both sequences of interest, C8T8A8G8 and RAND32, have the same G-C content but the sequence pattern plays a large role in the kinetic pathways to form aggregates. We also explored the effect of temperature on C8T8A8G8 self-assembly. At low T* large aggregates form but have short lifetimes (t* < 200), however at high T* these aggregates are non-existent due to the high kinetic energy of system, thus destabilizing base pairing. It was determined the best

isothermal self-assembly temperature for both sequences is at T* = 1.0. This is due to the minimization of mis-aggregate formation and the large number of long lived dimers (Figure 3S1). Additionally, our model was able to capture the concentration effects on DNA self-assembly. It was found at high concentrations, monomers would form dimers then trimmers and larger aggregates during the early stages of the simulation (t* < 200). These aggregates would them dissociate into monomers and dimers to end forming the enthalpically favorable dimers. We applied kinetic modeling to mathematically capture the effects of temperature and concentration on the self-assembly process. Our system was found to have a negative activation energy similar to those found in experiments [61-62] and simulation [57].

Additionally, the aggregate analysis described in this work give us insights into the individual aggregates enabling us to identify the common themes of aggregates. This could be further expanded to include contact maps for each of the bases to more easily observe the bases involved in the hybridization process. Having the ability to do this analysis on individual aggregates is important in the field of DNA nanotechnology to further increase assembly yields and performance of dynamic DNA systems and the effect of environmental variables.

3.6 Supplementary Info

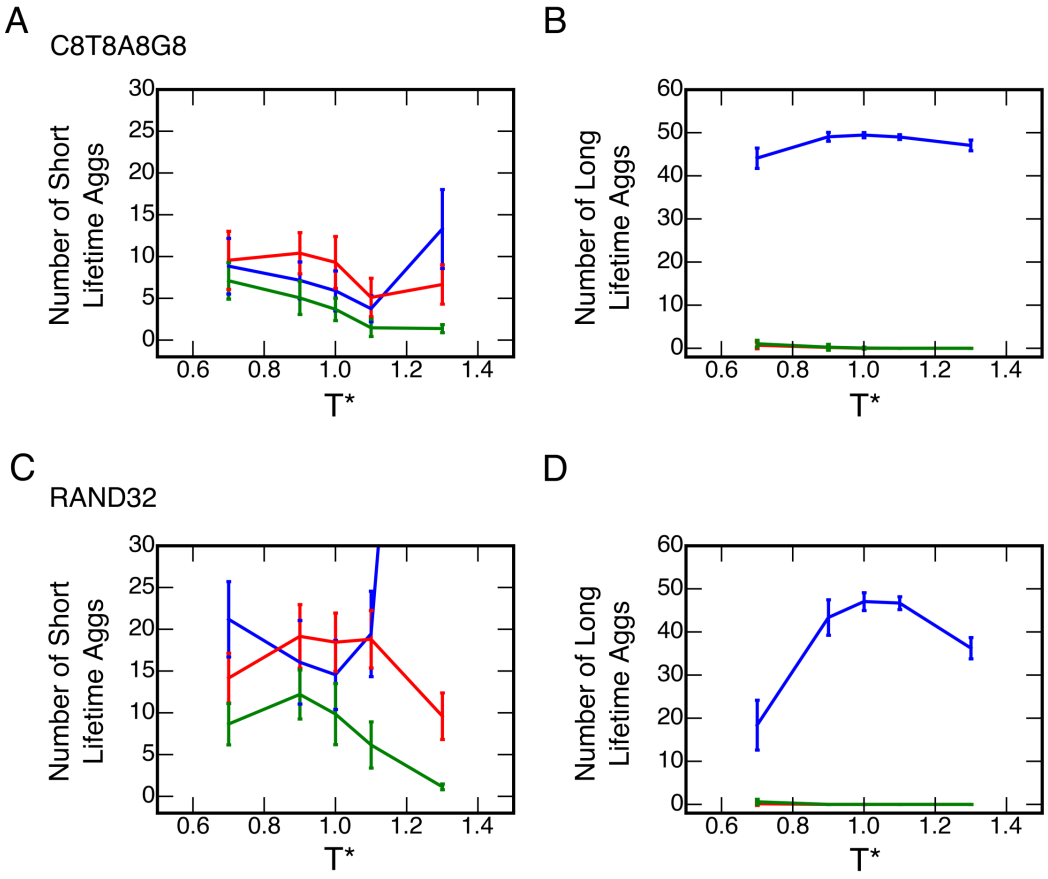

Figure S3.1: Effect of temperature on the number of aggregates with short lifetimes (t* < 200) for (A) C8T8A8G8 and (C) RAND32. At T* = 1.3, RAND32 (not in view in C) has a value of 121.70 ± 16.8. Effect of temperature on the number of aggregates with long lifetimes (t* > 200) for (B) C8T8A8G8 and (D) RAND32.

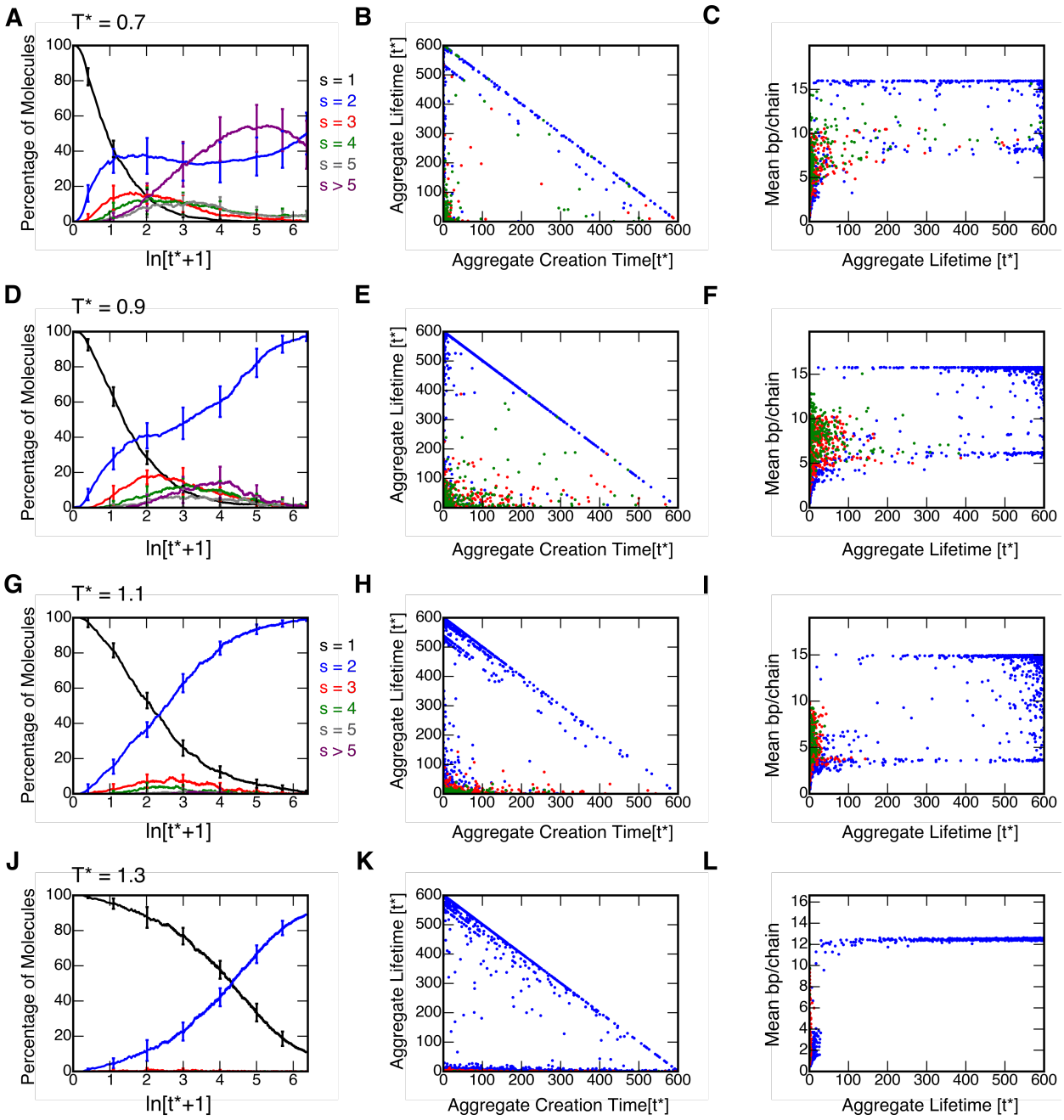

Figure S3.2: Effect of Temperature on RAND32. Percentage of molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale for (A) T* = 0.7, (D) T* = 0.9, (G) T* = 1.1, and (J) T* = 1.3. Unique aggregate creation times and lifetimes plotted as a function of reduced time (t*) for (B) T* = 0.7, , (E) T* = 0.9, (H) T* = 1.1, and (K) T* = 1.3 reduced temperatures. Unique aggregates lifetime and the mean number of base pairs per chain during the lifetime for (C) T = 0.7, (F) T* = 0.9, (I) T* = 1.1, (L) T* = 1.3 reduced temperatures.

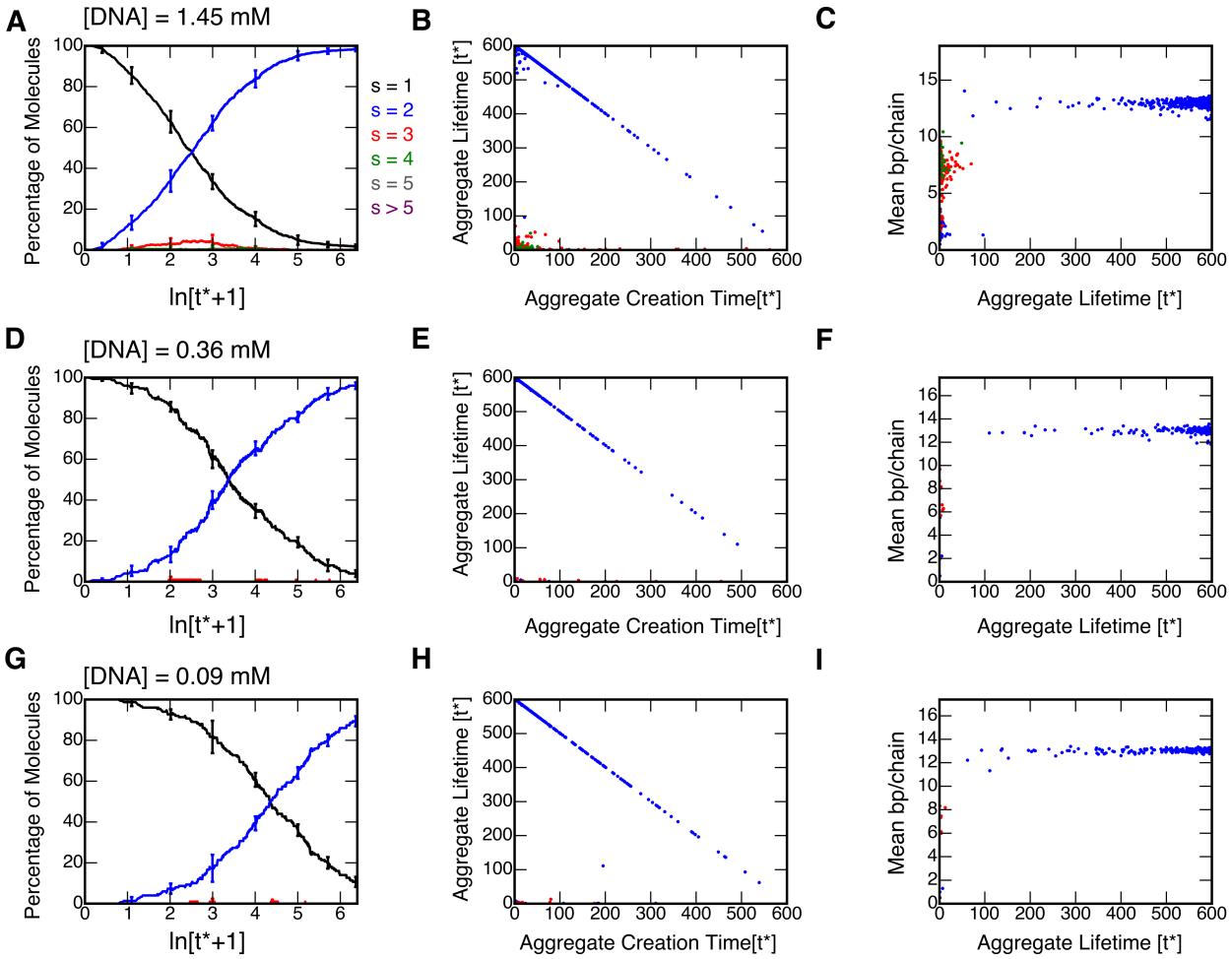

Figure S3.3: Effect of DNA concentration on C8T8A8G8 self-assembly at T* = 1.1. Percentage of molecules belonging to various aggregate sizes, s, as a function of time in natural logarithmic scale for (A) 1.45 mM, (D) 0.36mM, and (G) 0.09 mM. Unique aggregate creation times and lifetimes plotted as a function of reduced time (t*) for (B) 1.45 mM, (E) 0.36 mM, and (H) 0.09 mM. Unique aggregates lifetime and the mean number of base pairs per chain during the lifetime for (C) 1.45 mM, (F) 0.36 mM, and (I) 0.09 mM.

3.7 References

1. J. Marmur, P. Doty. Thermal Renaturation of Deoxyribonucleic Acid. *J. Mol. Bio*., **1961,** 3, 585-594.
2. J. Wetmur and N. Davidson. Kinetics of Renaturation of DNA. *J. Mol. Bio*., **1968**, 31, 349-370.
3. B. J. McCarthy and R. B. Church. The Specificity of Molecular Hybridization Reactions. *Annu. Rev. Biochem.*, **1970**, 39: 131-150.
4. C. S. Lee, R. W. Davis, and N. Davidson. A Physical Study by Electron Microscopy of the Terminally Repetitious Circularly Permuted DNA from the Coliphage Particles of *Escherichia coli* 15. *J. Mol. Bio.*, **1970**, 48: 1-22.
5. M. S. Meselson and C. M. Radding. A General Model for Genetic Recombination. *Proc. Nat. Acad. Sci. USA*., **1975,** 72: 358-361.
6. N. C. Seeman, Nucleic Acid Junctions and Lattices. *Journal of Theoretical Biology* **1982**, *99*, 237–247.

7. Goodman, R. P.; Schaap, I. A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science* **2005**, *310*, 1661–1665.
8. Zheng, J.; Birktoft, J. J.; Chen, Y.; Wang, T.; Sha, R.; Constantinou, P. E.; Ginell, S. L.; Mao, C.; Seeman, N. C. From Molecular to Macroscopic via the Rational Design of a Self-Assembled 3D DNA Crystal. *Nature* **2009**, *461*, 74–77.
9. Ke, Y.; Liu, Y.; Zhang, J.; Yan, H. A Study of DNA Tube Formation Mechanisms Using 4-, 8-, and 12-Helix DNA Nanostructures. *J. Am. Chem. Soc.* **2006**, *128*, 4414–4421.
10. Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297–302.
11. S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, and W. M. Shih. Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Research*, **2009**, 37: 5001-5006.
12. J. N. Zadeh, B. R. Wolfe, and N. A. Pierce. Nucleic Acid Sequence Design via Efficient Ensemble Defect Optimization. *J. Comput. Chem.*, **2011**, 32: 439-452.
13. J-S. Shin, and N. A. Pierce. A Synthetic DNA Walker for Molecular Transport. *J. Am. Chem. Soc.*, **2004**, 126: 10834-10835.
14. J. Bath and A. J. Turberfield. DNA nanomachines. *Nature Nanotechnology,* **2007**, 2: 275-284.
15. T. E. Ouldridge, R. L. Hoare, A. A. Louis, J. P. K. Doye, J. Bath, A. J. Turberfield. Optimizing DNA Nanotechnology through Coarse-Grained Modeling: A Two-Footed DNA Walker. *ACS Nano*, **2013,** 7, 2479- 2490.
16. H. Lederman, J. Macdonald, D. Stefanovic, and M. N. Stojanovic. Deoxyribozyme-Based Three-Input Logic Gates and Construction of a Molecular Full Adder. *Biochemistry*, **2006**, 45: 1194-1199.
17. Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro. An autonomous molecular computer for logical control of gene expression. *Nature*, **2004,** 27: 423-429.
18. L. Qian and E. Winfree. A simple DNA gate motif for synthesizing large-scale circuits. *J. R. Interface*, **2011**, 8, 1281-1297.
19. S. M. Douglas, I. Bachelet, and G. M. Church. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science*, **2012**, 335, 831-834.
20. D. Y. Zhang and G. Seelig. Dynamic DNA Nanotechnology using strand displacement reactions. *Nature Chem*., **2011**, 3: 103-113.
21. L. M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science*, **1994**,266 :1021-1024.
22. G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree. Enzyme-Free Nucleic Acid Logic Circuits. *Science*, **2006**, 314: 1585-1588.
23. L. Qian, E. Winfree, J. Bruck. Neural network computation with DNA strand displacement cascades. *Nature*, **2011**, 475: 368-372.
24. C. C. Santini, J. Bath, A. J. Turberfield, A. M. Tyrrell. A DNA Network as an Information Processing System. *Int. J. Mol. Sci.*, **2012**, 13, 5125-5137.
25. Y-J. Chen, N. Dalchau, N. Srinivas, A. Phillips, L. Cardelli, D. Soloveichik, and G. Seelig. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, **2013**, 8, 755-762.
26. A. Phillips and L. Cardelli. A programming language for composable DNA circuits. *J. R. Soc. Interface*, **2009**,6: 419-436.
27. Microsoft Research DSD: http://research.microsoft.com/en-us/projects/dna/
28. A. Aksimentiev, J. B. Heng, G. Timp, K. Schulten. Microscopic Kinetics of DNA Translocation Through Synthetic Nanopores. *Biophysical Journal,* **2004**, *87*, 2086–2097.
29. T. D. Lillian, M. Taranova, J. Wereszczynski, I. Andricioaei, N. C. Perkins. A Multiscale Dynamic Model of DNA Supercoil Relaxation by Topoisomerase IB. *Biophysical Journal,* **2011**, *100*, 2016–2023.
30. J. Yoo, A. Aksimentiev. In Situ Structure and Dynamics of DNA Origami Determined Through Molecular Dynamics Simulations. *Proceedings of the National Academy of Sciences,* **2013**, *110*, 20099–20104.
31. A. Noy, A. Pérez, C. A. Laughton, M. Orozco. Theoretical Study of Large Conformational Transitions in DNA: the BA Conformational Change in Water and Ethanol/Water. *Nucleic Acids Research,* **2007**, *35*, 3330–3338.
32. S. Kannan, and M. Zacharias. Simulation of DNA Double-Strand Dissociation and Formation During Replica-Exchange Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10589–10595.

33. T. A. Knotts, N. Rathore; D. C. Schwartz, J. J. de Pablo. A Coarse Grain Model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901:1-12.
34. T. E. Ouldridge, I. G. Johnston; A. A. Louis; J. P. K. Doye. The Self-Assembly of DNA Holliday Junctions Studied with a Minimal Model. *J. Chem. Phys.,* **2009**, *130*, 065101:1-11.
35. M. Maciejczyk, A. Spasic, A. Liwo; H. A Scheraga. Coarse-Grained Model of Nucleic Acid Bases. *J. Comput. Chem.,* **2010**, *31*, 1644–1655.
36. S. M. Gopal; S. Mukherjee, Y.-M. Cheng, M. Feig, PRIMO/PRIMONA: a Coarse-Grained Model for Proteins and Nucleic Acids That Preserves Near-Atomistic Accuracy. *Proteins,* **2010**, *78*, 1266–1281.
37. P. D. Dans, A. Zeida, M. R. Machado, S. Pantano. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.,* **2010**, *6*, 1711–1725.
38. A. Savelyev, and G. A. Papoian. Chemically Accurate Coarse Graining of Double-Stranded DNA. *Proceedings of the National Academy of Sciences,* **2010**, *107*, 20340–20345.
39. Z. Xia, D. P. Gardner, R. R. Gutell, P. Ren. Coarse-Grained Model for Simulation of RNA Three-Dimensional Structures. *J. Phys. Chem. B,* **2010**, *114*, 13497–13506.
40. A. Morriss-Andrews, J. Rottler, S. S. Plotkin. A Systematically Coarse-Grained Model for DNA and Its Predictions for Persistence Length, Stacking, Twist, and Chirality. *J. Chem. Phys.,* **2010**, *132*, 035105:1-18.
41. R. C. DeMille, T. E. Cheatham III, V. Molinero. A Coarse-Grained Model of DNA with Explicit Solvation by Water and Ions. *J. Phys. Chem. B*, **2011**, *115*, 132–142.
42. S. Biyun, S. S. Cho, D. Thirumalai. Folding of Human Telomerase RNA Pseudoknot Using Ion-Jump and Temperature-Quench Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 20634–20643.
43. T. E. Ouldridge; A. A. Louis, J. P. K. Doye. Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model. *J. Chem. Phys.,* **2011**, *134*, 085101:1-22.
44. C. W. Hsu, M. Fyta, G. Lakatos, S. Melchionna, and E. Kaxiras. Ab Initio Determination of Coarse-Grained Interactions in Double-Stranded DNA. *J. Chem. Phys.* **2012**, *137*, 105102:1-12.
45. T. Cragnolini, P. Derreumaux, S. Pasquali. Coarse-Grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B,* **2013**, *117*, 8047–8060.
46. Y. He, M. Maciejczyk, S. Ołdziej, H. Scheraga, A. Liwo. Mean-Field Interactions Between Nucleic-Acid-Base Dipoles Can Drive the Formation of a Double Helix. *Phys. Rev. Lett.,* **2013**, *110*, 098101:1-5.
47. D. M. Hinckley, J. P. Lequieu, J. J. de Pablo. Coarse-Grained Modeling of DNA Oligomer Hybridization: Length, Sequence, and Salt Effects. *J. Chem. Phys.* **2014**, *141*, 035102:1-12.
48. C. Maffeo, T. T. Ngo, T. Ha, A. Aksimentiev. A Coarse-Grained Model of Unstretched Single-Stranded DNA Derived From Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.,* **2014**, 10, 2891-2896.
49. T. I. N. G. Li, R. Sknepnek, R. J. Macfarlane, C. A. Mirkin, and M. Olvera de la Cruz. Modeling the Crystallization of Spherical Nucleic Acid Nanoparticle Conjugates with Molecular Dynamics Simulations. *Nano Letters, **2012**,* 12, 2509-2514.
50. F. Ding, S. Sharma, P. Chalasani, V. V. Demidov, N. E. Broude, N. V. Dokholyan. Ab Initio RNA Folding by Discrete Molecular Dynamics: From Structure Prediction to Folding Mechanisms. *RNA,* **2008**, *14*, 1164–1173.
51. E. J. Sambriski, D. C. Schwartz; J. J. de Pablo. A Mesoscale Model of DNA and Its Renaturation. *Biophysical Journal* **2009**, *96*, 1675–1690.
52. H. D. Nguyen, C. K. Hall. Molecular Dynamics Simulations of Spontaneous Fibril Formation by Random-Coil Peptides. *Proc. Natl. Acad. Sci. U.S.A.,* **2004**, *101*, 16180–16185.
53. I. W. Fu, C. B. Markegard, B. K. Chu, and H. D. Nguyen. The Role of Electrostatics and Temperature on Morphological Transitions of Hydrogel Nanostructures Self-Assembled by Peptide Amphiphiles via Molecular Dynamics Simulations. *Adv . Healthc. Mater.,* **2013**, *2*, 1388–1400.
54. I. W. Fu, C. B. Markegard, and H. D. Nguyen. Solvent Effects on Kinetic Mechanisms of Self-Assembly by Peptide Amphiphiles via Molecular Dynamics Simulations. *Langmuir*, **2015**, *31*: 315–324.
55. I. W. Fu, C. B. Markegard, B. K. Chu, H. D. Nguyen. Role of Hydrophobicity on Self-Assembly by Peptide Amphiphiles via Molecular Dynamics Simulations. *Langmuir*, **2014**, *30*: 7745–7754
56. C. B. Markegard, I. W. Fu, K. Anki Reddy, H. D. Nguyen. Coarse-Grained Simulation Study of Sequence Effects on DNA Hybridization in a Concentrated Environment. *J. Phys. Chem. B*, **2015**, *119*: 1823–1834.

57. T. E. Ouldridge, P. Sulc, F. Romano, J. P. K. Doye. DNA hybridization kinetics: zipping, internal displacement, and sequence dependence. *Nucleic Acids Research*, **2013**, 41: 8886-8895.
58. D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. de Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.,* **2013**, 139: 144903.
59. B. J. Alder, T.E. Wainwright. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.,* **1959**, *31*, 459–466.
60. H. Andersen. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.,* **1980**, *72*, 2384–2393.
61. M. E. Craig, D. M. Crothers, and P. Doty. Relaxation Kinetics of Dimer Formation by Self Complementary Oligonucleotides. *J. Mol. Biol.*, **1971**, 62, 383-401.
62. D. Pörschke and M. Eigen. Co-operative Non-enzymic Base Recognition. III. Kinetics of Helix-Coil Transition of Oligoribouridylic- Oligoribouridylic Acid System and of Oligoriboadenylic Acid alone at Acidic pH. *J. Mol. Biol.*, **1971**, 62, 361-381.

CHAPTER 4 Molecular Dynamics Simulations of Perylenediimide DNA Base Surrogates

4.1 Abstract

Perylene-3,4,9,10-tetracarboxylic diimides (PTCDIs) are a well known class of materials. Recently, these molecules have been incorporated within DNA as base surrogates, finding ready applications as probes of DNA structure and function. However, the assembly dynamics and kinetics of these PTCDI base surrogates have received little attention to date. Herein, we employ constant temperature molecular dynamics simulations to gain an improved understanding of the assembly of PTCDI dimers and trimers. We also use replica-exchange molecular dynamics simulations to elucidate the energetic landscape dictating the formation of stacked PTCDI structures. Our studies provide insight into the equilibrium configurations of multimeric PTCDIs and hold implications for the construction of DNA-inspired systems from PTCDI-based organic semiconductor building blocks.

4.2 Introduction

Perylene-3,4,9,10-tetracarboxylic diimide (PTCDI) derivatives constitute a well known and extensively studied class of organic materials [1-8]. Due to their tunable coloration and excellent stability, these molecules have found ready applications as industrial dyes and pigments [1-2]. Moreover, PTCDIs' favorable electrochemical, photophysical, and self-assembly properties have facilitated not only the fundamental study of charge transport phenomena but also the development of various organic

electronic devices, such as transistors and solar cells [3-8]. Consequently, a number of synthetic methodologies have been developed for modulating the properties of PTCDIs [2, 7, 8]. For example, substitution of these molecules' aromatic core and imide positions provides a degree of control over their electronic properties and self-assembly behavior, respectively [8]. Thus, given the various advantageous features of PTCDIs, it is not surprising that these molecules have attracted much attention from both industry and academia for over 100 years [1-8].

Computational techniques have been used to investigate the electronic and self-assembly properties of PTCDI ensembles [5, 9-14]. For example, various studies have used density functional theory to establish relationships between the solid state packing and emergent electronic functionality of various substituted PTCDIs [5, 9-13]. Moreover, a handful of reports have used molecular dynamics simulations to understand the aggregation and assembly dynamics of PTCDI molecules both in solution and in the solid state [5, 11-14]. These efforts have afforded fundamental insight that is valuable for the design of improved PTCDI-based materials.

Recently, Wagenknecht and coworkers have developed a new class of PTCDI derivatives for use as artificial DNA base surrogates [15-20]. Within the context of oligonucleotide synthesis, such derivatives are advantageous because they can be incorporated in arbitrary positions within the DNA base pair stack in high yield *via* standard automated phosphoramidite chemistry [15-20]. To date, these molecules have been used for a number of applications, including the assembly of higher-order DNA ensembles [20-22], photophysical investigation of DNA structure/function [16, 17, 19], the study of charge transfer in DNA hairpins [23-25], and the electrochemical interrogation of

DNA monolayers [26]. However, the assembly dynamics and kinetics of PTCDI DNA base surrogates have not been extensively explored *via* computational techniques [21-22].

Herein, we present a molecular dynamics study of multimeric PTCDI DNA base surrogates. We first synthesize and characterize oligonucleotides featuring one, two, and three covalently attached PTCDI moieties. We then parametrize the oligonucleotides' PTCDI subunits and formulate an atomistic model of these compounds. We subsequently employ constant-temperature molecular dynamics simulations to develop an improved understanding of the assembly kinetics of PTCDI dimers and trimers. We in turn perform replica exchange molecular dynamics simulations to obtain the energetic landscape associated with ensembles of our stacked PTCDI structures at equilibrium. Altogether, our findings may hold implications for the design of DNA-inspired systems and materials from not only PTCDIs but also other organic semiconductor building blocks.

4.3 Methods

4.3.1 Preparation of the DNA Phosphoramidites

The phosphoramidites required for DNA synthesis were purchased from Glen Research, Inc. or Azco Biotech, Inc. and used as received. The perylenediimide phosphoramidites were synthesized and characterized according to established literature protocols [26]. The identity and purity of the phosphoramidites, as well as all intermediates required for their synthesis, were confirmed with ^{1}H NMR, ^{13}C NMR, ^{31}P NMR, and mass spectrometry.

4.3.2 Synthesis, Purification, and Characterization of the Oligonucleotides

The PTDCI-modified oligonucleotides (**Figure 4.1**) were synthesized according to standard commercial protocols recommended by Glen Research, Inc. for an Applied Biosystems (ABI) 394 DNA Synthesizer. Extended coupling times were used for the incorporation of perylenediimide phosphoramidites, as previously described [26]. After synthesis, the oligonucleotides were cleaved from the solid support by treatment with aqueous ammonium hydroxide and purified with high performance liquid chromatography (HPLC) on an Agilent 1260 Infinity system. The oligonucleotides were eluted with a gradient evolved from 95% solvent A and 5% solvent B to 0% solvent A and 100% solvent B over 30 min at a flow rate of 1 mL/min (solvent A, 50 mM ammonium acetate, pH = 6 buffer; solvent B, acetonitrile) on Agilent reverse phase C4 or C8 column (see Figure S4.1 for a typical chromatogram). The identity of the oligonucleotides was further confirmed via UV−visible spectra obtained during chromatographic purification with an Agilent 1260 Infinity Series Diode Array Detector and mass spectra obtained either with an Applied Biosystems Sciex MALDI-TOF/TOF instrument or a WATERS LCT Premier Electrospray Time of Flight instrument.

4.3.3 Parameterization of the Perylenediimide Base Surrogates

For the simulations, three separate residues (Figure S4.2) were designed according to established literature protocols [27]. Subsequently, the geometry of these structures was optimized by using the Hartree-Fock method with the 6-31G(d) basis set in Gaussian 09 [28]. After convergence of the structures, the electrostatic potentials were determined with

the Merz-Kollman scheme in Gaussian 09 [28]. In turn, the point charges necessary for molecular dynamics simulations were obtained via the two-step Restricted Electrostatic Potential (RESP) method [27]. The residue libraries necessary for construction of our oligonucleotides' PTCDI subunits (**Figure 4.3**) were generated in LEaP by combining the point charges with the Generalized AMBER Force Field (GAFF) [29-31].

4.3.4 Molecular Dynamics Simulations of Perylenediimide Base Surrogate Stacking Kinetics

Molecular dynamics simulations of PTCDI stacking kinetics (20 total) were performed with the Generalized AMBER Force Field (GAFF) in NAMD 2.9 [32-33]. The simulations employed the Generalized Born Implicit Solvent model (GBIS) and a monovalent salt concentration of 0.115 M [32-33]. For each simulation, the starting configuration was obtained by turning off the attractive van der Waals interactions in the force field and setting the temperature to 500 K, thereby ensuring that all PTCDI moieties were completely separated from one another in an unstacked random open configuration. To initiate the simulation, the attractive van der Waals interactions were turned on and the initial temperature was set to 300 K. All of the simulations were performed at a constant temperature of 300 K for 20 ns, ensuring that steady state was reached. The simulations were analyzed by monitoring the relative centers of mass (COM) distances and offset angles for every pair of PTCDIs. The COM distances were calculated from the atomic coordinates and atomic mass of the individual PTCDIs. The offset angles were calculated by constructing a vector from the nitrogen closest to the backbone to the nitrogen farthest

from the backbone for the individual PTCDIs (Figure S4.3). The dot product of these vectors for every pair of PTCDIs yielded their offset angles. The COM distances and offset angels indicated the relative separation and alignment of the PTCDIs, respectively. As an example, if the COM distance of two PTCDIs is ~ 3.4 Å and their offset angle is 0°, the two molecules are stacked and perfectly aligned on top of one another (Figure S4.4). The simulations also yielded the van der Waals and electrostatic interactions for each pair of PTCDIs. These interactions were monitored as a function of time to gain insight into the factors driving the self-assembly of the PTCDIs.

4.3.5 Replica-exchange Molecular Dynamics Simulations of Perylenediimide Base Surrogate Assembly

Replica exchange molecular dynamics (REMD) simulations were performed to explore the energetic landscape of the oligonucleotides' stacked PTCDI subunits [34]. The simulations were run for 16 replicates distributed over a temperature range of 290 to 700 K. The simulation time was 320 ns per replica, corresponding to 80,000 total exchanges, with an exchange attempt every 4 ps. The simulations generated the equilibrium atomic structures for an ensemble of stacked PTCDI subunits. These structures were analyzed via the MBAR method, enabling calculation of the potential of mean force (PMF) for the stacked PTCDIs as a function of their COM distances and offset angles [35]. This analysis yielded free energy landscape of our constructs at 300 K.

4.2 Results and Discussion

We began our experiments by preparing the series of model PTCDI-containing macromolecules shown in **Figure 4.1** (**Oligo1**, **Oligo2**, and **Oligo3**), via standard phosphoramidite chemistry [26]. The design of these macromolecules incorporated 1 to 3 PTCDI moeities, an alkanethiol functionality at their 5' end, and a polyadenine tract at their 3' end; the terminal modifications were included to mitigate intermolecular aggregation. Theses constructs were analogous to perylene-modified oligonucleotides previously investigated with photophysical techniques in solution [16, 17, 19] and electrochemical techniques at solid substrates [26].

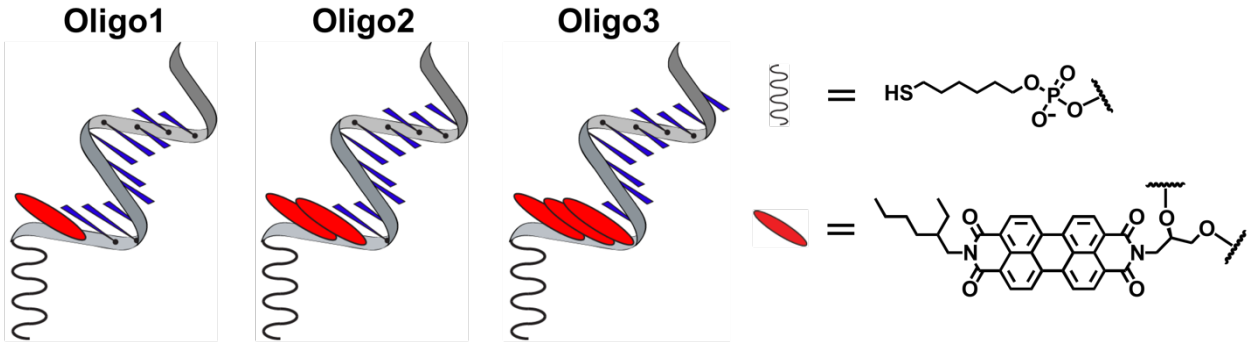

Figure 4.1: Illustration of oligonucleotides **Oligo1**, **Oligo2**, and **Oligo3** featuring 1, 2, and 3 PTCDI base surrogates (red ovals), respectively. The DNA sequences of these macromolecules were 3'-$(A)_{10}(P)_n$-S-5', where the A, P, and S indicate the locations of the adenines, PTCDIs, and thiols, respectively, and n corresponds to the number of PTCDIs.

We characterized **Oligo1**, **Oligo2**, and **Oligo3** with UV-Vis spectroscopy (Figure 4.2), observing a clear evolution in the molecules' absorbance spectra as the number of PTCDI moieties increased. For **Oligo1**, the spectrum of DNA1 was indicative of a single, isolated PTCDI, with the three characteristic absorbance peaks at 466 nm, 496 nm, and 534 nm

whose intensities increased at longer wavelengths (Figure 4.2). However, the situation was markedly different for **Oligo2** and **Oligo3**. Although the absorbance peaks maintained similar positions, the spectra were broadened, with the absorbance peaks at 498 nm for **Oligo2** and 497 nm for **Oligo3** now exhibiting the largest intensities. This type of evolution in our constructs' absorbance spectra was indicative of strong pi-pi stacking interactions between the adjacent PTCDI moieties [2-8, 36].

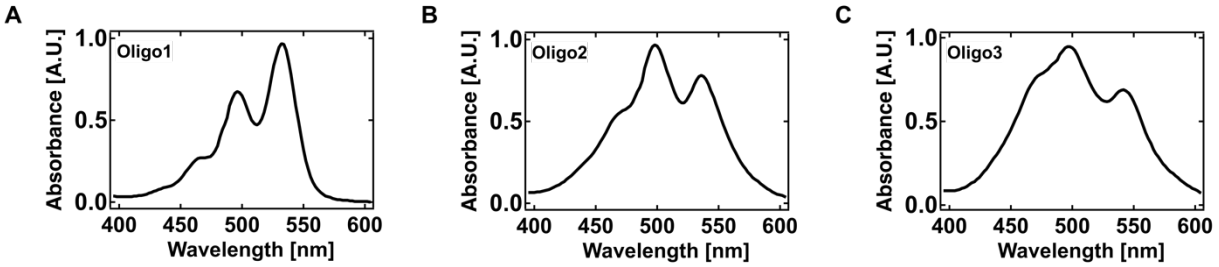

Figure 4.2: The UV-Vis absorbance spectra obtained for **Oligo1**, **Oligo2**, and **Oligo3** containing 1, 2, and 3 PTCDI moieties respectively. Note the change in the relative intensities of the absorbance peaks for **Oligo2** and **Oligo3** relative to **Oligo1**.

With our constructs in hand, we proceeded to perform constant temperature molecular dynamics simulations for the PTCDI subunits of **Oligo2** and **Oligo3** (Figure 4.3). For simplicity, we only considered the structures shown in Figure 4.3 (denoted as **P1**, **P2**, and **P3**), removing the alkyl tails on the PTCDIs, the 3' alkanethiol functionality, and the 5' polyadenine tract. We leveraged established literature protocols previously developed for oligonucleotides and DNA to parameterize the PTCDI moieties with the Generalized AMBER Force Field (GAFF), facilitating the computational analysis [29-31]. Here, through our simulations, we anticipated gaining insight into both the kinetics of self-assembly and the ultimate geometry of **P2** and **P3**.

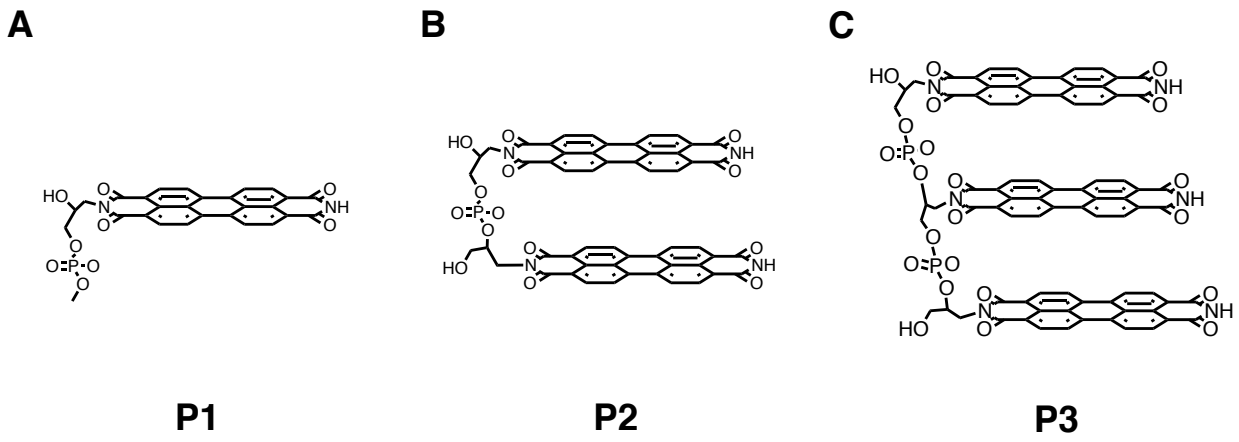

Figure 4.3: The chemical structures of the PTCDI subunits used for molecular dynamics simulations. The subunits are labeled as **P1**, **P2**, and **P3** corresponding to **Oligo1**, **Oligo2**, and **Oligo3**, respectively.

We initially investigated the kinetics of self-assembly for the two PTCDI moieties of **P2**, as illustrated for a typical simulation in Figure 4.4. At t = 0.051 ns, the two PTCDIs maintained a random open configuration (Figure 4.4A), with a COM distance of 18.34 Å (Figure 4.4E) and an offset angle of 125.7 ° (Figure 4.4F). After 1.718 ns, the top PTCDI had started to flip (Figure 4.4B), leading to a slight decrease in the COM distance to 14.34 Å (Figure 4.4E) and a reduction in the offset angle to 68.28 ° (Figure 4.4F). After 1.912 ns, the top PTCDI had begun to orient itself towards the bottom PTCDI (Figure 4.4C), leading to a further decrease in the COM distance (Figure 4.3E) and an additional reduction in the offset angle (Figure 4.4F). Finally, after 2.412 ns, the two PTCDIs had collapsed to a stacked state (Figure 4.4D), with an average COM distance of 4.6 Å (Figure 4.4E) and an average offset angle of 15.51° (Figure 4.4F). However, although the two PTCDIs remain stacked for the remainder of the ~ 20 ns simulation, their COM distances and offset angles varied by ~ 0.44 Å and ~ 7.83°, respectively. Such geometric variability likely resulted from thermal fluctuations, hinting at some dynamic character for the final configuration. Nonetheless, the final stacked dimer was highly reproducible, with nearly identical results obtained for ten

independent repetitions.

To gain additional insight into the kinetics of stacking for the two PTCDI moieties of **P2**, we analyzed the energetics driving their self-assembly. When the two PTCDIs collapsed to a stacked configuration between ~ 1.9 ns and ~ 2.4 ns (Figure 4.4E and 4.4F), there was a sharp decrease in the van der Waals energy of the system, due to pi-pi stacking between the molecules' aromatic cores, (Figure 4.4G) and a sharp increase in the electrostatic energy of the system, due to repulsion between the PTCDIs' carbonyl groups located distal to the alkane phosphate backbone (Figure 4.4H). The competition between these two sets of interactions appeared to dictate the final arrangement of the system, where there was substantial overlap between the perylene cores but a misalignment of the carbonyl groups (Figure 4.4D).

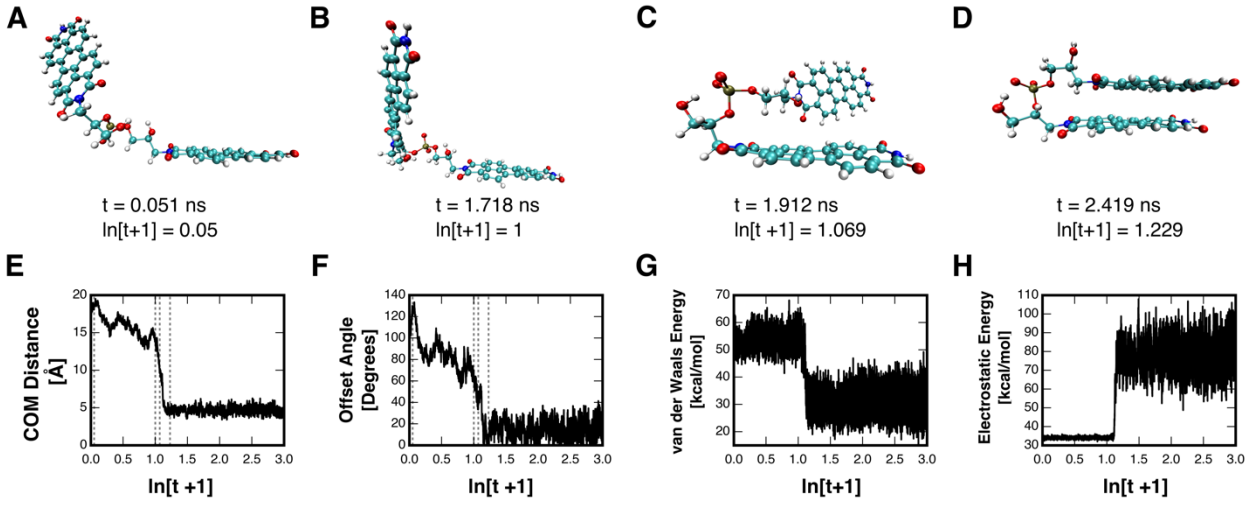

Figure 4.4: Snapshots of **P2** during a molecular dynamics simulation at times of A) t = 0.051 ns, B) t = 1.718 ns, C) t = 1.912 ns, and D) t = 2.419 ns. The sequence demonstrates the transition of **P2** from an open random to a stacked configuration. E) The evolution of the centers of mass (COM) distance between the two PTCDIs of **P2** as a function of time. (F) The evolution of the offset angle between the two PTCDIs of **P2** as a function of time. G) The evolution of the van der Waals energy of **P2** as a function of time. H) The evolution of the electrostatic energy of the PTCDIs' carbonyl oxygens distal to the alkane phosphate backbone as a function of time. The four vertical dashed lines in (E) and (F) correspond to the times used for the snapshots in A), B), C), and D). The simulations were performed at a constant temperature of 300 K.

We next investigated the stacking kinetics of the three PTCDI moieties of **P3**. In our experiments, we observed that self-assembly of **P3** into the final stacked arrangement necessitated a two-step mechanism. First, two of the PTCDIs formed a dimer analogous to the one found for the final configuration of **P2**. Second, the remaining third PTCDI stacked with this dimer to form a trimer. Although our simulations revealed multiple possibilities for the PTCDIs' stacking order (Figure 4.5 and Figures S4.5 and S4.6), this general mechanism was highly reproducible, with similar results obtained for ten independent simulations.

As an example, Figure 4.5 illustrates a typical **P3** stacking kinetics simulation, which demonstrates the aforementioned multi-step mechanism and the most likely assembly pathway. At t = 0.051 ns, the three constituent PTCDIs (denoted as **1**, **2**, and **3**) were unstacked with a random open configuration (Figure 4.5A). Subsequently, during the first step of the stacking mechanism at t = 0.649 ns (Figure 4.5B), **2** and **3** formed a dimer, with a corresponding sharp drop from 13.34 Å to 5.02 Å for their COM distance (Figure 4.5E) and from 100.61° to 1.30° for their offset angle (Figure 4.5F). In turn, during the second step of the stacking mechanism at t = 2.935 ns, **1** stacked on top of the dimer to form a trimer (Figure 4.5C). For this step, there was a sharp drop from 9.09 Å to 5.61 Å in the COM distance (Figure 4.5E) and from 63.26° to 13.76° in the offset angle (Figure 4.5F) for **1** and **2**. There also was an accompanying drop from 10.92 Å to 9.29 Å in the COM distance (Figure 4.5E) and from 73.07° to 30.45° in the offset angle (Figure 4.5F) between **1** and **3**. Although the three PTCDIs remained in a stacked arrangement for the remainder of the ~ 20 ns simulation, we again observed variability in their relative COM distances and offset angles (Figure 4.5E and 4.5F), indicating some dynamic character for the final

configuration.

To further understand the stacking kinetics of the three PTCDI moieties of **P3**, we analyzed the energetics driving their self-assembly. In the first step of the mechanism, the formation of the dimer was driven by a sharp decrease in the van der Waals energy, due to pi-pi stacking between the molecules' aromatic cores (Figure 4.5G), and a sharp increase in the electrostatic energy, due to repulsion between the PTCDIs' carbonyl groups located distal to the alkane phosphate backbone (Figure 4.5H), as also observed for **P2** above (Figure 4.4G and 4.4H). In the second step of the mechanism, a similar interplay of favorable and unfavorable interactions appeared to drive formation of the trimer, as indicated by similar changes in the van der Waals and electrostatic energies (Figure 4.5G and 4.5H). In their final equilibrium structure, **P3**'s three constituent PTCDIs adopted a twisted arrangement, where they were stacked but offset with respect to one another (Figure 4.4D).

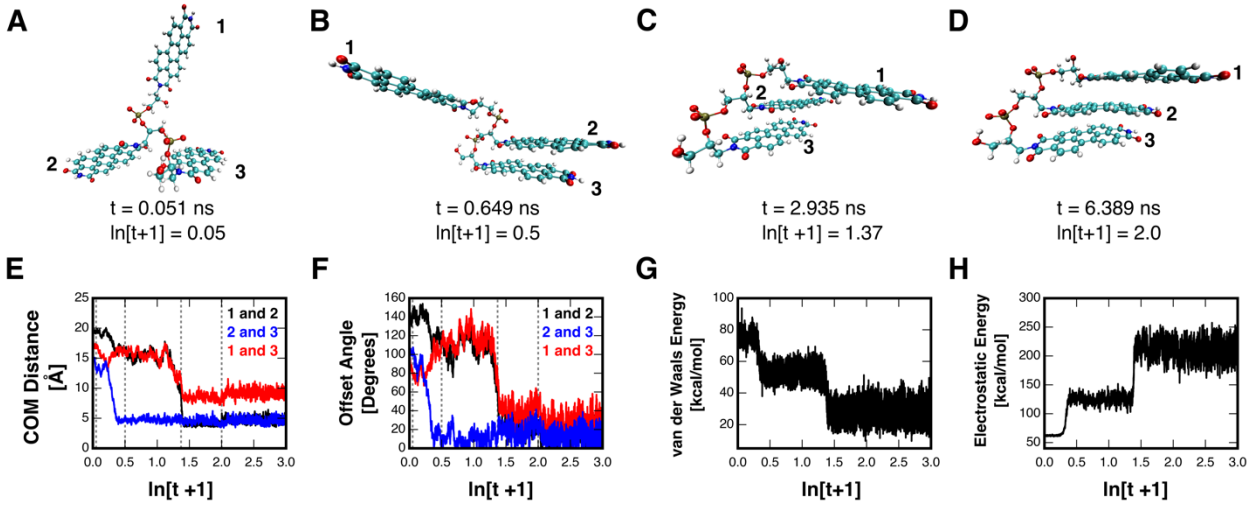

Figure 4.5: Snapshots of **P3** during a molecular dynamics simulation at times of A) t = 0.051 ns, B) t = 0.649 ns, C) t = 2.935 ns, and D) t = 6.389 ns. The constituent PTCDIs of **P3** are labeled as **1**, **2**, and **3**. The sequence demonstrates the transition of **P3** from an open random to a stacked configuration. E) The evolution of the centers of mass (COM) distance between PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. (F) The evolution of the offset angle between PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. G) The evolution of the van der Waals energy of **P3** as a function of time. H) The evolution of the electrostatic energy of the PTCDIs' carbonyl oxygens distal to the alkane phosphate backbone as a function of time. The four vertical dashed lines in (E) and (F) correspond to the times used for the snapshots in A), B), C), and D). The simulations were performed at a constant temperature of 300 K.

We proceeded to perform REMD simulations for **P2** and **P3**. Relative to constant temperature simulations, REMD simulations are advantageous because they minimize the possibility of kinetic traps [34]. Moreover, replica exchange simulations enable direct calculation of the potential of mean force (PMF), facilitating comparisons between the relative free energies of different equilibrium structures [35]. Here, through our simulations, we anticipated gaining insight into the distinct configurations possible for **P2** and **P3** over a broad temperature range.

We initially performed REMD simulations for **P2**, obtaining equilibrium structures such as the one illustrated in Figure 4.6A and 4.6B. These configurations spanned a range of COM distance and average offset angle combinations, corresponding to various PMF values

(Figure 4.6C). As a general rule, the lowest free energy structures found for **P2** featured COM distances between 4.2 Å and 5.2 Å and offset angles between 3° and 22°. For example, the lowest energy equilibrium structure shown in Figure 4.6A and 4.6B featured a COM distance of 4.53 Å and an offset angle of 14.61°, in excellent agreement with the kinetic simulations above. Notably, configurations with COM distances of > ~ 7 Å possessed large free energies, underscoring the PTCDIs known propensity for stacking [2-8]. However, configurations with COM distances of ~ 3.4 Å and offset angles close to 0° still possessed large free energies, indicating that perfect overlap of the two PTCDIs was not favored. Overall, these simulations provided additional insight into the preferred equilibrium arrangement of the PTCDI moieties of **P2**.

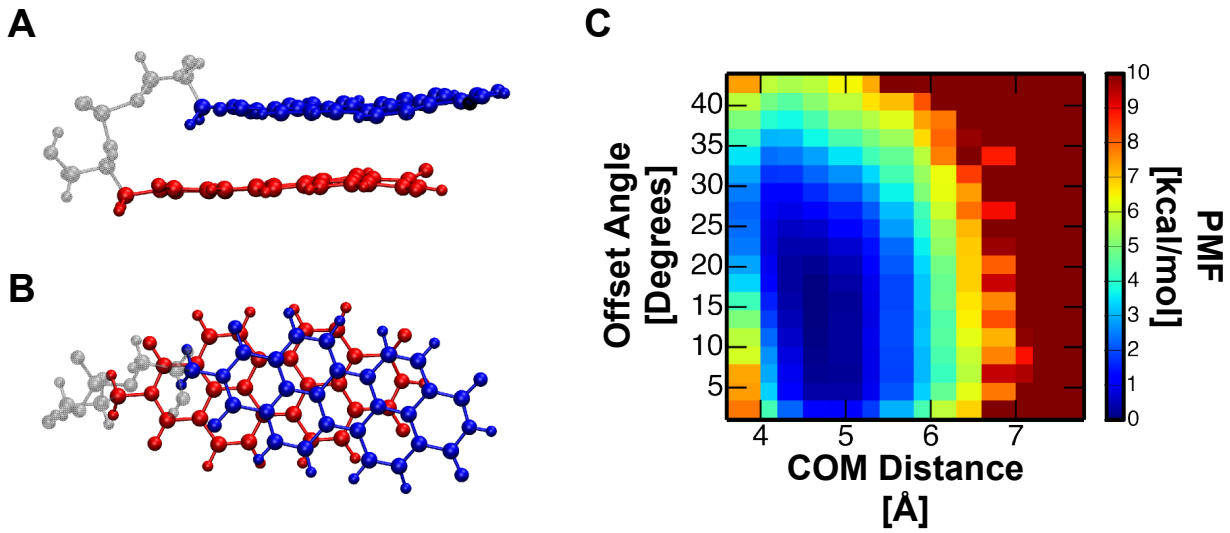

Figure 4.6: Snapshots of the lowest energy equilibrium structure observed for **P2** from A) a side view and B) a top view. The backbone is colored gray, and two PTCDIs are colored red and blue. C) The potential of mean free force (PMF) in kcal/mole as a function of the centers of mass (COM) distance and the offset angle between the two PTCDIs of **P2**, as obtained from a replica exchange simulation at 300 K.

We in turn proceeded to perform REMD simulations for **P3**, obtaining equilibrium structures such as the one illustrated in Figure 5.7A and 5.7B. These simulations again yielded an ensemble of possible **P3** configurations, which we analyzed by comparing the

relative orientations of the three constituent PTCDIs (denoted as **1**, **2**, and **3**). The possible COM distance and offset angle combinations obtained for **1** and **2**, along with their corresponding PMF values, are illustrated in Figure 5.7A, and the possible COM distance and offset angle combinations obtained for **2** and **3**, along with their corresponding PMF values, are illustrated in Figure 5.7B. As observed for **P2**, the equilibrium structures with the lowest free energies featured COM distances between 4.2 Å and 5.2 Å and offset angles between 3° and 22°. For example, the lowest energy equilibrium structure shown in Figure 5.7A and 5.7B, featured a COM distance of 4.86 Å and an offset angle of 6.01° for **1** and **2**, as well as a COM distance of 5.24 Å and an offset angle of 4.5° for **2** and **3**. Although the three PTCDIs comprising this structure adopted a stacked arrangement, they did not overlap perfectly and were slightly offset with respect to one another. These findings were again in agreement with the kinetics simulations and afforded an improved understanding of the energetic landscape driving the self-assembly of **P3**.

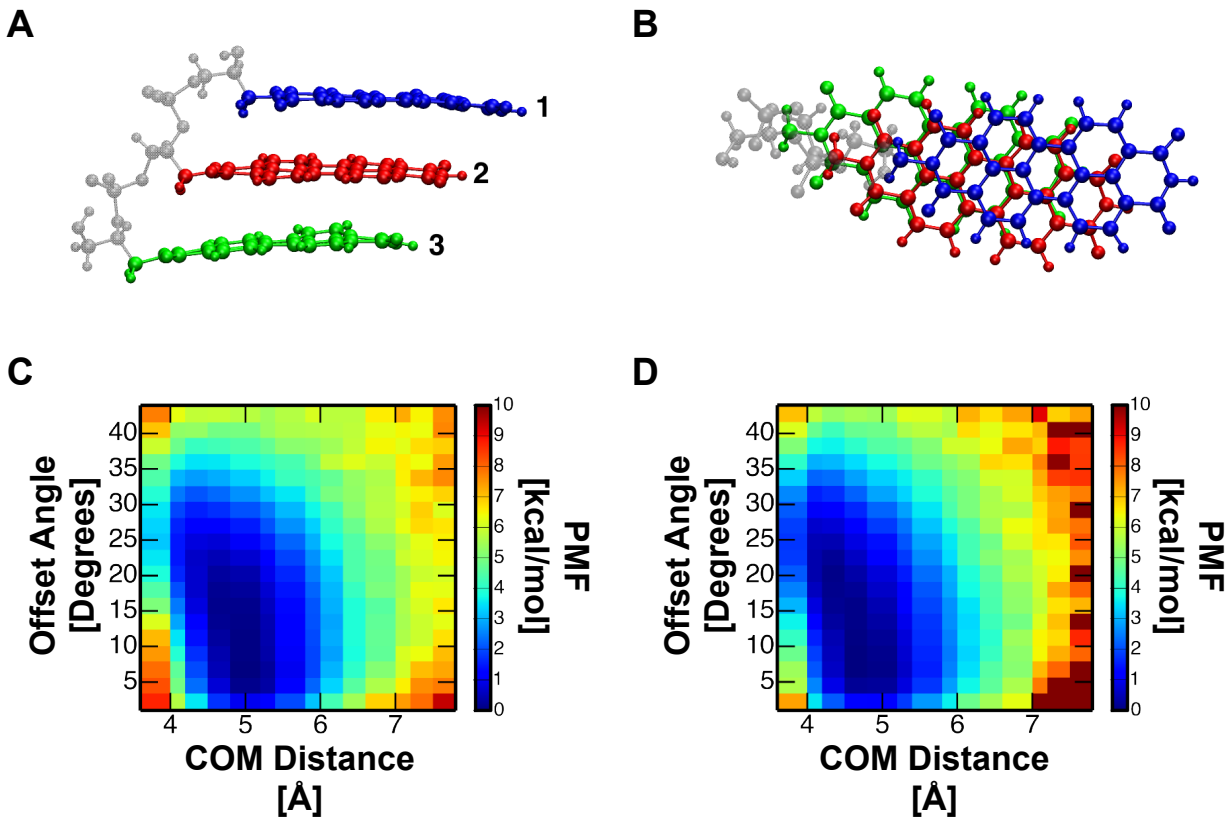

Figure 4.7: Snapshots of the lowest energy equilibrium structure observed for **P3** from A) a side view and B) a top view. The backbone is colored gray, and PTCDIs **1**, **2**, and **3** are colored red, blue, and green, respectively. C) The potential of mean free force (PMF) in kcal/mole as a function of the centers of mass (COM) distance and the offset angle between the PTCDIs **1** and **2** of **P3**, as obtained from a replica exchange simulation at 300 K. D) The potential of mean free force (PMF) in kcal/mole as a function of the centers of mass (COM) distance and the offset angle between the PTCDIs **2** and **3** of **P3**, as obtained from a replica exchange simulation at 300 K.

Finally, we note that our simulations indicated that the PTCDI moieties comprising the **P2** and **P3** equilibrium structures were non-planar on short time scales (Figure 4.6A 4.6B and Figure 4.7A and 4.7B). To gain insight into the origin of this effect, we compared the distortion from planarity found for the stacked, interacting PTCDIs of **P2** and **P3** with the distortion of planarity found for the independent PTCDI of **P1**. For this comparison, we defined a bend angle between two vectors from the center to the edge of the PTCDIs (Figure S4.7). The PTCDI of **P1** featured a range of bend angles between 153.27° and

179.92°, with an average value of 174.66 ± 3.67° (Figure S4.7). Interestingly, we found that the PTCDIs of **P2** and **P3** all featured nearly identical bending angle distributions (Figures S4.8 and S4.9). Given that the average distortions were small and not dependent on stacking interactions, we postulated that the observed deviations from planarity were likely due to thermal fluctuations. These findings further underscored the rich dynamics of stacked PTCDIs, which warrant additional exploration.

4.4 Conclusion

In summary, we have used constant temperature and replica exchange molecular dynamics simulations to investigate the self-assembly of covalently-linked perylene-3,4,9,10-tetracarboxylic diimide DNA base surrogates. Together, our computational analyses yielded insight not be readily apparent from experimental methodologies and were significant for several reasons. First, to the best of our knowledge, the stacking kinetics of PTCDI DNA base surrogates have received little attention from a computational perspective, especially with regard to investigating the process in its entirety. Second, our simulations have allowed for observation of the assembly kinetics in atomistic detail, revealing that the underlying mechanism is quite complex and may encompass multiple distinct pathways. Third, our studies have unveiled the energetics of multimeric PTCDI ensembles, indicating that a complex interplay of attractive van der Waals and repulsive electrostatic interactions dictates their assembly and final structures. Fourth, the simulations demonstrate that the observed equilibrium structures are not fully static and possess some dynamic character, as evidenced by fluctuations in the relative positions and planarity of the stacked PTCDIs. Fifth, the presented computational framework is quite

general and can be readily extended to the study of more complex sequence-and length-variable systems consisting of covalently linked pi-conjugated organic semiconductor building blocks. Overall, our studies constitute a foundation for the rational design and construction of precisely-defined one-dimensional nanowires that draw inspiration from the structure of DNA.

4.5 Supplementary Info

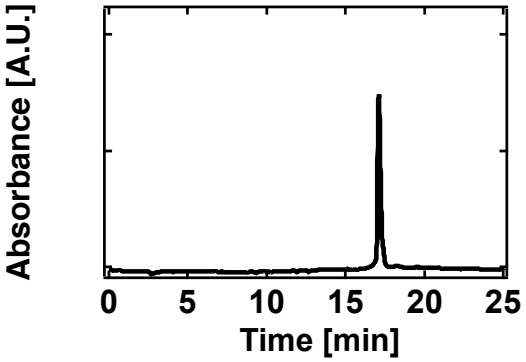

Figure S4.1: A typical HPLC chromatogram corresponding to the purification of **Oligo2**. The DNA sequence was 3'-$(A)_{10}(P)_2$-*S*-5', where the *A*, *P*, and *S* indicate the locations of the adenines, PTCDIs, and thiol, respectively.

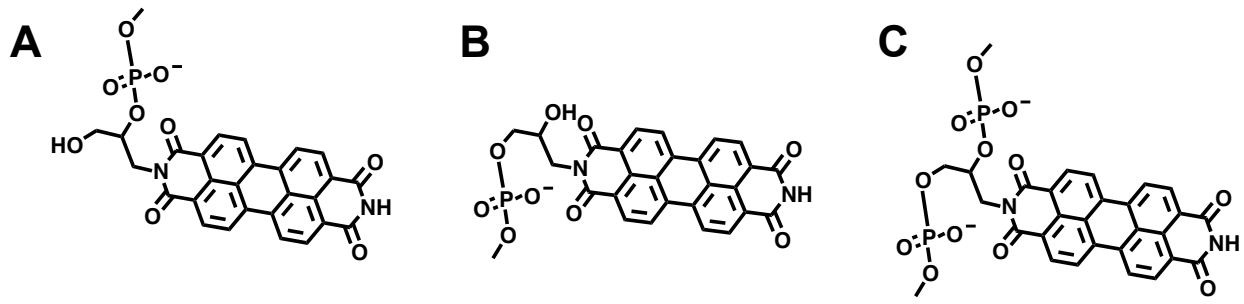

Figure S4.2: The chemical structure of the three PTCDI residues (along with the corresponding phosphate groups) that were designed and parameterized for the molecular dynamics simulations. The terminal residues of **P2** and **P3** are illustrated in A) and B), and the middle residue is illustrated in C). When two residues are joined, one of the redundant phosphate groups will be removed to leave a single phosphate between the joined PTCDIs.

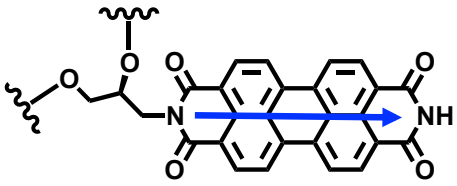

Figure S4.3: Illustration of the PTCDI DNA base surrogate. The blue line indicates the vector connecting the nitrogen closest to the backbone to the nitrogen furthest away from the backbone. The vector was used for analysis of the stacking of adjacent PTCDIs.

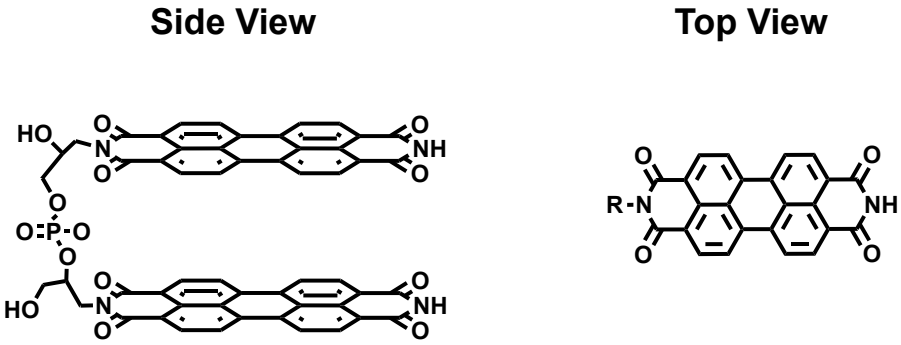

Figure S4.4: The side view (left) and top view (right) of the chemical structure of **P2**, where the two constituent PTCDIs feature a COM distance of ~ 3.4 Å and an offset angle of 0°. The top view, where the backbone has been removed for clarity, shows that the PTCDI moieties perfectly overlap in this scenario.

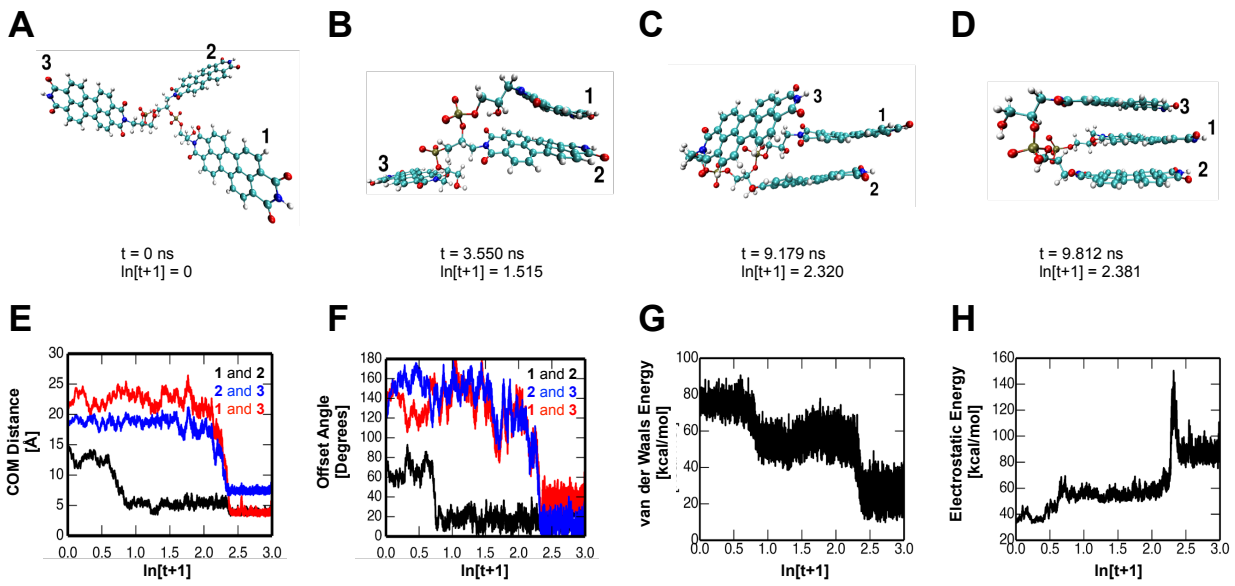

Figure S4.5: Assembly of **P3** into a stacked configuration through a possible alternative pathway, as observed during molecular dynamics simulations. Note that this pathway is not physically likely when the PTCDIs are iteratively incorporated in an oligonucleotide. Snapshots of **P3** at times of A) t = 0 ns, B) t = 3.550 ns, C) t = 9.179 ns, and D) t = 9.812 ns. The constituent PTCDIs of **P3** are labeled as **1**, **2**, and **3**. The sequence demonstrates the transition of **P3** from an open random to a stacked configuration. E) The evolution of the centers of mass (COM) distance between PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. (F) The evolution of the offset angle between PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. G) The evolution of the van der Waals energy of PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. H) The evolution of the electrostatic energy of PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. The simulations were performed at a constant temperature of 300 K.

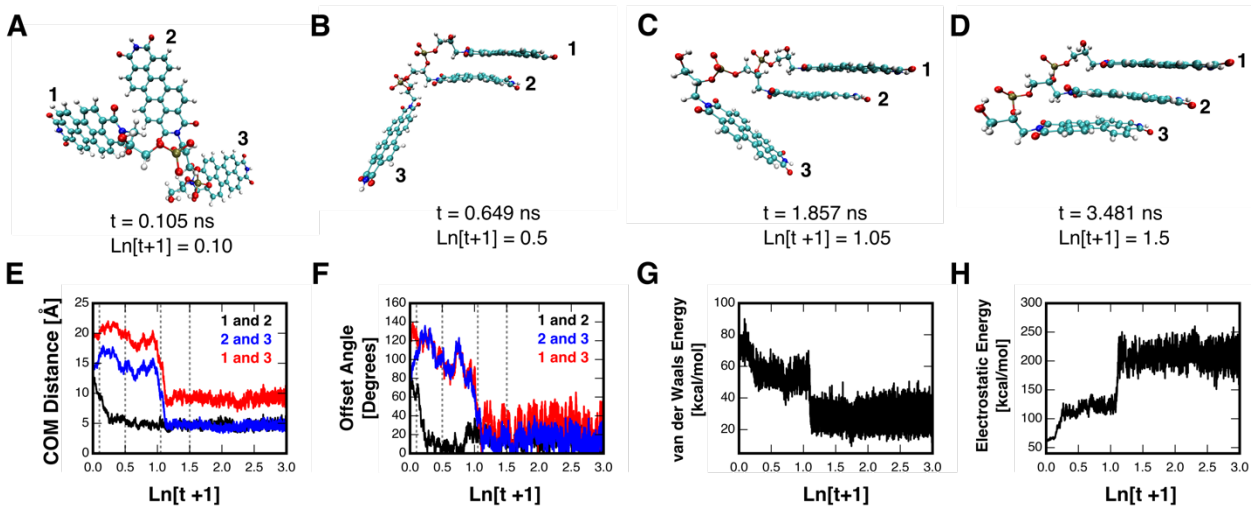

Figure S4.6: Assembly of **P3** into a stacked configuration through another possible alternative pathway, as observed during molecular dynamics simulations. Snapshots of **P3** at times of A) t = 0.105 ns, B) t = 0.649 ns, C) t = 1.857 ns, and D) t = 3.481 ns. The constituent PTCDIs of **P3** are labeled as **1**, **2**, and **3**. The sequence demonstrates the transition of **P3** from an open random to a stacked configuration. E) The evolution of the centers of mass (COM) distance between PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. (F) The evolution of the offset angle between PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. G) The evolution of the van der Waals energy of PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. H) The evolution of the electrostatic energy of PTCDIs **1** and **2** (black curve), PTCDIs **2** and **3** (blue curve), and PTCDIs **1** and **3** (red curve) as a function of time. The simulations were performed at a constant temperature of 300 K.

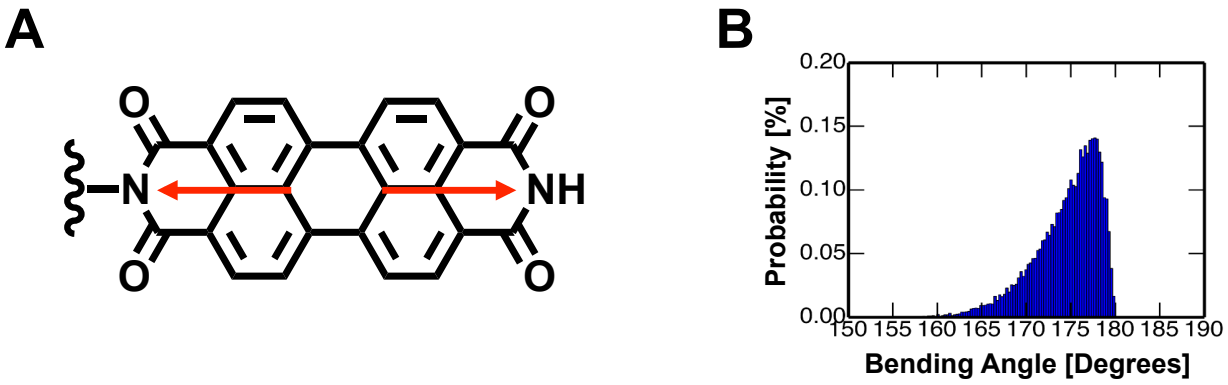

Figure S4.7: A) Structure of an isolated PTCDI moiety, where the vectors used to calculate the angle of bending are illustrated in red. B) The bending angle distribution of an individual PTCDI, the structure of which is shown in Figure S4.3A.

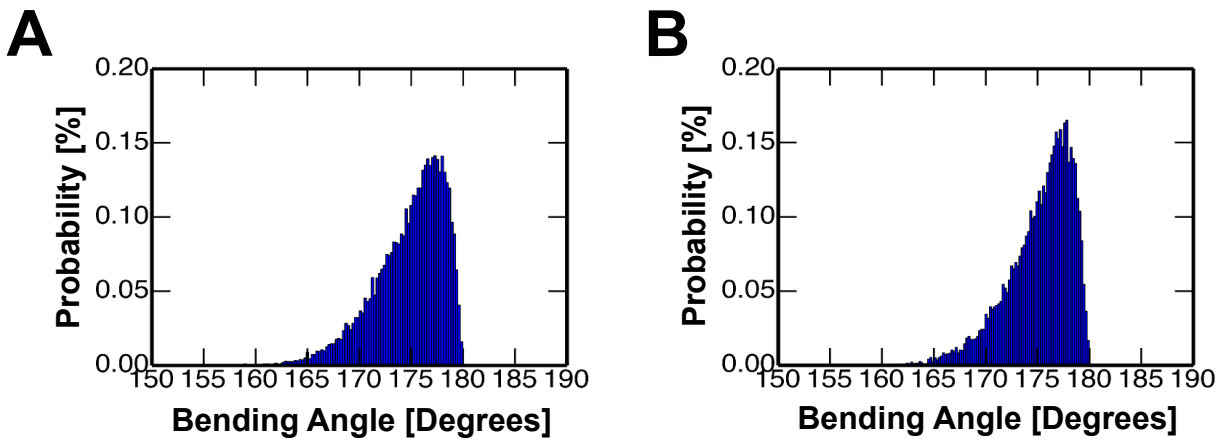

Figure S4.8: The bending angle distribution found for the A) 3' and B) 5' PTCDIs of **P2**. The data was obtained from the last 10 % of the conformation of a replica exchange simulation at 300 K.

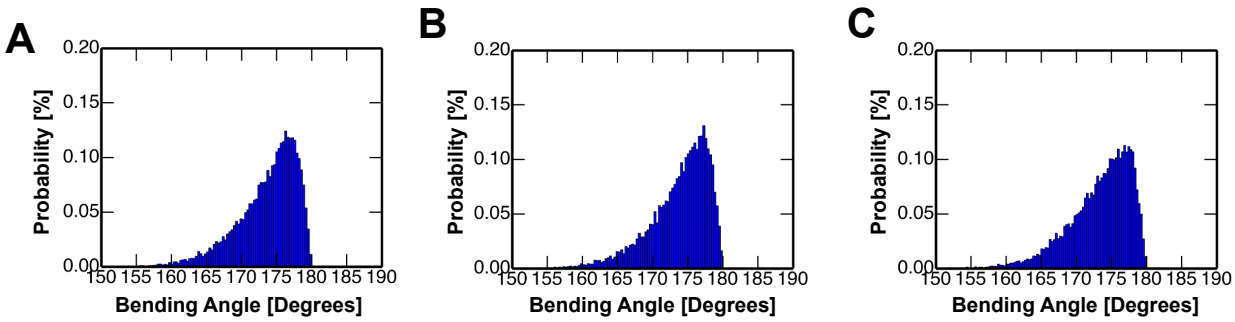

Figure S4.9: The bending angle distribution found for the A) 3' B), middle, and C) 5' PTCDIs of **P2**. The data was obtained from the last 10 % of the conformation of a replica exchange simulation at 300 K.

4.6 References

1. Herbst, W.; Hunger, K. *Industrial Organic Pigments: Production, Properties, Applications*. 2nd ed. Wiley: Weinheim, **1997**.
2. Würthner, F. Perylene Bisimide Dyes as Versatile Building Blocks for Functional Supramolecular Architectures. *Chem. Commun.* **2004**, 1564–1579.
3. Chen Li, C.; Wonneberger, H. Perylene Imides for Organic Photovoltaics: Yesterday, Today, and Tomorrow. *Adv. Mater.* **2012**, *24*, 613–636.
4. Görl, D.; Zhang, X.; Würthner, F. Molecular Assemblies of Perylene Bisimide Dyes in Water. *Angew. Chem. Int. Ed.* **2012**, *51*, 6328-6348.
5. Geng, Y.; Li, H.-B.; Wu, S.-X.; Su, Z.-M. The interplay of intermolecular interactions, packing motifs and electron transport properties in perylene diimide related materials: a theoretical perspective. *J. Mater. Chem.* **2012**, *22*, 20840-20851.
6. Kozma, E.; Catellani, M. Perylene diimides based materials for organic solar cells. *Dyes Pigments* **2013**, *98*, 160-179.
7. Jung, B. J.; Tremblay, N. J.; Yeh, M-L.; Katz, H. E. Molecular design and synthetic approaches to electron-transporting organic transistor semiconductors. *Chem. Mater*. **2011**, *23*, 568–582.
8. Huang, C.; Barlow, S.; Marder, S. R. Perylene-3,4,9,10-tetracarboxylic Acid Diimides: Synthesis, Physical Properties, and Use in Organic Electronics. *J. Org. Chem.* **2011**, *76*, 2386–2407.

9. Vura-Weis, J.; Ratner, M. A.; Wasielewski, M. R. Geometry and Electronic Coupling in Perylenediimide Stacks: Mapping Structure-Charge Transport Relationships. *J. Am. Chem. Soc.* **2010**, *132*, 1738–1739.
10. Delgado, C. R.; Kim, E.-G.; da Silva, D. A.; Bredas, J.-L. Tuning the Charge-Transport Parameters of Perylene Diimide Single Crystals via End and/or Core Functionalization: A Density Functional Theory Investigation. *J. Am. Chem. Soc.* **2010**, *132*, 3375–3387.
11. Marcon, V.; Breiby, D. W.; Pisula, W.; Dahl, J.; Kirkpatrick, J.; Patwardhan, S.; Grozema, F.; Andrienko, D. Understanding Structure-Mobility Relations for Perylene Tetracarboxydiimide Derivatives *J. Am. Chem. Soc.* **2009**, *131*, 11426–11432.
12. May, F.; Marcon, V.; Hansen, M. R.; Grozema, F.; Andrienko, D. Relationship between supramolecular assembly and charge-carrier mobility in perylenediimide derivatives: The impact of side chains. *J. Mater. Chem.* **2011**, *21*, 9538-9545.
13. Idé, J.; Méreau, R.; Ducasse, L.; Castet, F.; Olivier, Y.; Martinelli, N.; Cornil, J.; Beljonne, D. Supramolecular Organization and Charge Transport Properties of Self-Assembled π-π Stacks of Perylene Diimide Dyes. *J. Phys. Chem. B* **2011**, *115*, 5593–5603.
14. Teklebrhan, R. B.; Ge, L.; Bhattacharjee, S.; Xu, Z.; Sjöblom, J. Probing Structure−Nanoaggregation Relations of Polyaromatic Surfactants: A Molecular Dynamics Simulation and Dynamic Light Scattering Study. *J. Phys. Chem. B* **2012**, *116*, 5907−5918.
15. Wagner, C.; Wagenknecht, H.-A. Perylene-3,4:9,10-tetracarbox-ylic Acid Bisimide Dye as an Artificial DNA Base Surrogate. *Org. Lett.* **2006**, *8*, 4191−4194.
16. Baumstark, D.; Wagenknecht, H.-A. Fluorescent Hydrophobic Zippers Inside Duplex DNA: Interstrand Stacking of Perylene- 3,4:9,10-tetracarboxylic Acid Bisimides as Artificial DNA Base Dyes. *Chem. Eur. J.* **2008**, *14*, 6640−6645.
17. Baumstark, D.; Wagenknecht, H.-A. Perylene Bisimide Dimers as Fluorescent "Glue" for DNA and for Base-Mismatch Detection. *Angew. Chem., Int. Ed.* **2008**, *47*, 2612−2614.
18. Menacher, F.; Wagenknecht, H.-A. Synthesis of DNA with Green Perylene Bisimides as DNA Base Substitutions. *Eur. J. Org. Chem.* **2011**, *24*, 4564−4570.
19. Menacher, F.; Wagenknecht, H.-A. Ratiometric Molecular Beacons Based on the Perylene Bisimide as a Dimeric Internal DNA Base Substitution. *Photochem. Photobiol. Sci.* **2011**, *10*, 1275−1278.
20. Menacher, F.; Stepanenko, V.; Würthner, F.; Wagenknecht, H.- A. Assembly of DNA Triangles Mediated by Perylene Bisimide Caps. *Chem. Eur. J.* **2011**, *17*, 6683−6688.
21. Zheng, Y.; Long, H.; Schatz, G. C.; Lewis, F. D. Duplex and hairpin dimer structures for perylene diimide–oligonucleotide conjugates. *Chem. Commun*. **2005**, *38,* 4795–4797.
22. Hariharan, M.; Zheng, Y.; Long, H.; Zeidan, T. A.; Schatz, G. C.; Vura-Weis, J.; Wasielewski, M. R.; Zuo, X.; Tiede, D. M.; Lewis, F. D. Hydrophobic Dimerization and Thermal Dissociation of Perylenediimide-Linked DNA Hairpins. *J. Am. Chem. Soc.* **2009**, I, 5920−5929.
23. Zeidan, T. A.; Carmieli, R.; Kelley, R. F.; Mi, Q.; Lewis, F. D.; Wasielewski, M. R. Charge-Transfer and Spin Dynamics in DNA Hairpin Conjugates with Perylenediimide as a Base-Pair Surrogate. *J. Am. Chem. Soc.* **2008**, *130*, 13945−13955.
24. Carmieli, R.; Zeidan, T. A.; Kelley, R. F.; Mi, Q.; Lewis, F. D.; Wasielewski, M. R. Excited State, Charge Transfer, and Spin Dynamics in DNA Hairpin Conjugates with Perylenediimide Hairpin Linkers. *J. Phys. Chem. A* **2009**, *113*, 4691−4700.
25. Wilson, T. M.; Zeidan, T. A.; Hariharan, M.; Lewis, F. D.; Wasielewski, M. R. Electron Hopping Among Cofacially Stacked Perylenediimides Assembled by Using DNA Hairpins. *Angew. Chem., Int. Ed.* **2010**, *49*, 2385−2388.
26. Wohlgamuth, C.; McWilliams, M.; Mazaheripour, A.; Burke, A.; Lin, K.; Doan, L.; Slinker, J.; Gorodetsky, A. Electrochemistry of DNA Monolayers Modified with a Perylenediimide Base Surrogate. *J. Phys. Chem.* **2014**, *118,* 29084−49090.
27. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multifunctional and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357−1377.

28. Gaussian 09, Revision C.01. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.;Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.
29. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
30. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
31. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
32. Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Calfisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecule Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
33. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26,* 1781–1802.
34. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
35. Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105–124118.
36. Clark, A. E.; Qin, C.; Li, A. D. Q. Beyond Exciton Theory: A Time-Dependent DFT and Franck-Condon Study of Perylene Diimide and Its Chromophoric Dimer *J. Am. Chem. Soc.* **2007**, *129*, 7586–7595.

Chapter 5 The Role of PEGylated Tail on Stacking of Perylenediimide DNA Base Surrogates

5.1 Abstract

Perylene-3,4,9,10-tetracarboxylic diimides (PTCDIs) are a well known class of materials. Recently, these molecules have been incorporated within DNA as base surrogates, finding ready applications as probes of DNA structure and function. Previously, we examined the assembly dynamics and kinetics of these PTCDI base surrogates. Herein, we conjugated each PTCDI base surrogate with a short polyethylene glycol (PEG) chain to elucidate the effect of the PEG chains on the arrangement of the perylene bases and compare with both experimental results and the previously obtained simulations of perylene without PEG. We employ constant temperature molecular dynamics simulations to gain an improved understanding of the assembly of PEGylated PTCDI dimers, trimers and tetramers. We find that PEG chains helps PTCDIs to stack together without interfering with their assembly kinetics due to the weak attraction between PEG chains and between each PEG and its PTCDI compared to the strong attraction between individual PTCDIs. Our studies provide insight into the equilibrium configurations of multimeric PTCDIs and hold implications for the construction of DNA-inspired systems from PTCDI-based organic semiconductor building blocks.

5.2 Introduction

Perylene-3,4,9,10-tetracarboxylic diimide (PTCDI) derivatives are a class of organic molecules that have studied using both experimentally and computationally [1-5]. Due to PTCDI's having desirable electronic properties they have been applied to making organic

electronic devices such as transistors and solar cells [6-9]. Recently, PTCDIs have been incorporated within DNA as base surrogates, finding ready applications as probes of DNA structure and function. However, the assembly dynamics and kinetics of these PTCDI base surrogates have received little attention to date. Previously, we employed constant temperature molecular dynamics simulations to gain an improved understanding of the assembly of PTCDI dimers and trimers. We also used replica-exchange molecular dynamics simulations to elucidate the energetic landscape dictating the formation of stacked PTCDI structures. Our studies provided insight into the equilibrium configurations of multimeric PTCDIs and hold implications for the construction of DNA-inspired systems from PTCDI-based organic semiconductor building blocks.

Experiments conducted by the Gorodetsky group currently use short polyethylene glycol (PEG) chains attached the perylene base in order to increase the solubility of the nanowires in water (**Figure 5.1**). X-Ray diffraction data obtained from synchrotron indicates that the perylenes are very ordered in their structure. Therefore, this work will investigate the effect of the PEG chains on the arrangement of the perylene bases and compare with both experimental results and the previously obtained simulations of perylene without PEG. The results from simulations will give a clearer picture of what is occurring at the molecular level in developing models for how charge transport occurs in these molecular nanowires.

Herein, we present a molecular dynamics study of multimeric PTCDI DNA base surrogates that are conjugated with short polyethylene glycol (PEG) chains. We first synthesize and characterize oligonucleotides featuring one, two, three and four covalently attached PEGylated PTCDI moieties. We then parametrize the oligonucleotides' PEGylated

PTCDI subunits and formulate an atomistic model of these compounds. We subsequently employ constant-temperature molecular dynamics simulations to develop an improved understanding of the assembly kinetics of PEGylated PTCDI dimers, trimers and tetramers. Our findings may hold implications for the design of DNA-inspired systems and materials from not only PTCDIs but also other organic semiconductor building blocks.

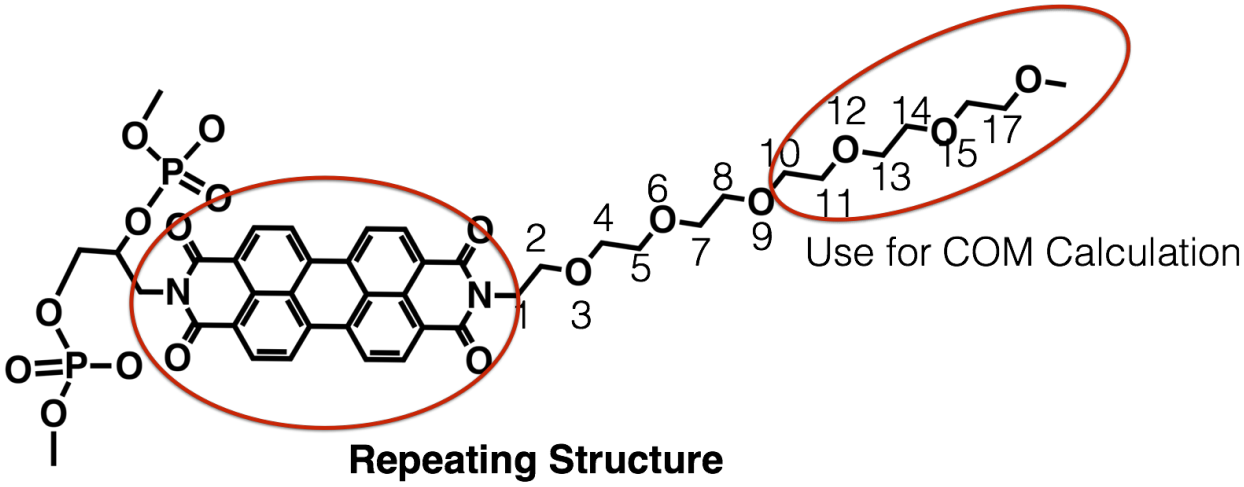

Figure 5.1: The chemical structures of each PEGylated PTCDI subunit used for molecular dynamics simulations. The center of mass (COM) of the second half of each PEG tail is used to determine its dynamics and interactions with other chemical moieties.

5.3 Methods

5.3.1 Parameterization of the Perylenediimide Base Surrogates

For the simulations, three separate residues (Figure S5.1B-D) were designed according to established literature protocols [10]. Subsequently, the geometry of one structure similar to all three (Figure S5.1A) was optimized by using the Hartree-Fock method with the 6-31G(d) basis set in Gaussian 09 [11]. After convergence of the structure, the electrostatic potential were determined with the Merz-Kollman scheme in Gaussian 09 [11]. In turn, the point charges necessary for molecular dynamics simulations were

obtained via the two-step Restricted Electrostatic Potential (RESP) method [27]. The residue libraries necessary for construction of our oligonucleotides' PTCDI subunits (Figure 5.3) were generated in LEaP by combining the point charges with the Generalized AMBER Force Field (GAFF) and deleting the necessary atoms to make the connections [12-14].

5.3.2 Molecular Dynamics Simulations of Perylenediimide Base Surrogate Stacking Kinetics

Molecular dynamics simulations of PTCDI stacking kinetics (30 total) were performed with the Generalized AMBER Force Field (GAFF) in NAMD 2.10 [12, 15]. The simulations employed the Generalized Born Implicit Solvent model (GBIS) and a monovalent salt concentration of 0.115 M [15-16]. For each simulation, the starting configuration was obtained by turning off the attractive van der Waals interactions in the force field and setting the temperature to 500 K, thereby ensuring that all PTCDI moieties were completely separated from one another in an unstacked random open configuration. To initiate the simulation, the attractive van der Waals interactions were turned on and the initial temperature was set to 300 K. All of the simulations were performed at a constant temperature of 300 K for 40 ns, ensuring that steady state was reached. The simulations were analyzed by monitoring the relative centers of mass (COM) distances and offset angles for every pair of PTCDIs. The COM distances were calculated from the atomic coordinates and atomic mass of the individual PTCDIs. The offset angles were calculated by constructing a vector from the nitrogen closest to the backbone to the nitrogen farthest from the backbone for the individual PTCDIs. The dot product of these vectors for every

pair of PTCDIs yielded their offset angles. The COM distances and offset angels indicated the relative separation and alignment of the PTCDIs, respectively. As an example, if the COM distance of two PTCDIs is ~ 3.4 Å and their offset angle is 0°, the two molecules are stacked and perfectly aligned on top of one another. The simulations also yielded the van der Waals and electrostatic interactions for each pair of PTCDIs. These interactions were monitored as a function of time to gain insight into the factors driving the self-assembly of the PTCDIs.

5.4 Results and Discussion

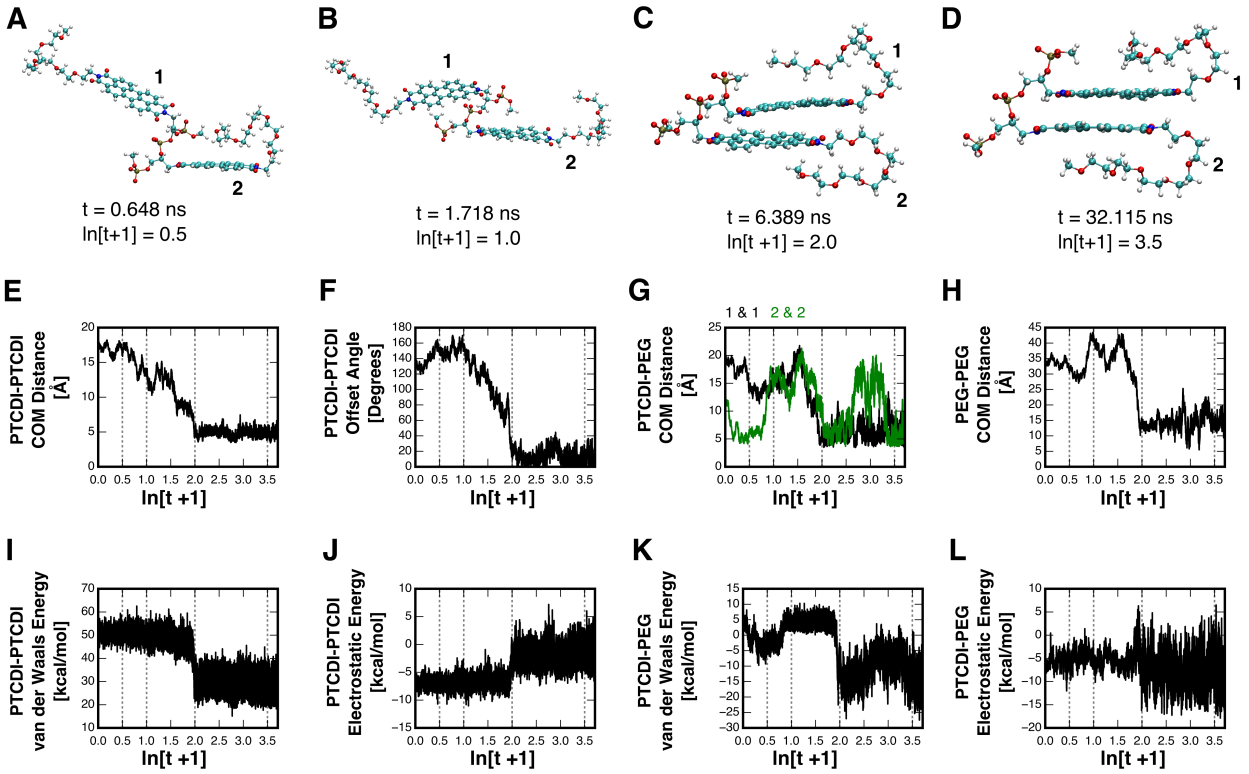

Figure 5.2: Snapshots of **P2** during a molecular dynamics simulation at times of A) t = 0.648 ns, B) t = 1.718 ns, C) t = 6.389 ns, and D) t = 32.115 ns. The sequence demonstrates the transition of **P2** from an open random to a stacked configuration. As a function of time, the evolution of: E) the centers of mass (COM) distance between the two PTCDIs, (F) the offset angle between the two PTCDIs, (G) COM distance between each PTCDI and its PEG tail**,** (H) COM distance between two PEG tails, (I) the van der Waals energy of PTCDIs, (J) the electrostatic energy of PTCDIs, (K) the van der Waals energy between each PTCDI and its PEG tail, and (L) the electrostatic energy between each PTCDI and its PEG tail of **P2**. The four vertical dashed lines in (E)-(L) correspond to the times used for the snapshots in A), B), C), and D). The simulations were performed at a constant temperature of 300 K.

The first set of MD simulations discerns the stacking kinetics of two PTCDI moieties of **P2**, as illustrated for a typical simulation in **Figure 5.2**. At t = 0.648 ns, the two PTCDIs maintained a random open configuration (Figure 5.2A), with a COM distance of 17.5 Å (Figure 5.2E) and an offset angle of 160.0° (Figure 5.2F). The PEG tail of the top PTCDI was relatively flexible moving around and staying away from its PTCDI whereas the PEG tail of

the bottom PTCDI flipped over and stayed closer to its PTCDI (Figure 5.2G), hindering the top PTCDI from moving closer to the bottom PTCDI. After 1.718 ns, the top PTCDI started moving closer to the bottom PTCDI (Figure 5.2B), leading to a slight decrease in the COM distance to 12.0 Å (Figure 5.2E); however, the two PTCDIs remained in opposite direction with an offset angle of 170° (Figure 5.2F). At this point, the PEG tail of the bottom PTCDI moved away from its PTCDI (Figure 5.2G). After 6.389 ns, each PEG tail made a bend along its length (Figure 5.2C) and kept a distance away from the other PTCDI (Figure 5.2G) and the other PEG tail (Figure 5.2H), allowing the top PTCDI to orient itself towards the bottom PTCDI, leading to a further decrease in the COM distance to 6.0 Å (Figure 5.2E); by this time, it had started to flip around leading to a reduction in the offset angle to 20.0° (Figure 5.2F). Then the two PTCDIs continued to orient themselves to further reduce the COM distance and offset angle for the remainder of the ~ 34 ns simulation. Over the last 5 ns when the system had already reached equilibrium as shown at 32.115 ns (Figure 5.2D), the two PTCDIs exhibited an average COM distance of 4.929 ± 0.418 Å (Figure 5.2E) and an average offset angle of 10.371 ± 6.449° (Figure 5.2F). These values of the average COM distance and offset angle falls within the region of the lowest free energy on the energy landscape that we had previously calculated for the stacking of two un-PEGylated PTCDIs.

Additional insight into the kinetics of stacking for the two PTCDI moieties of **P2** can be gained by analyzing the energetics driving their self-assembly. When the two PTCDIs collapsed to a stacked configuration between ~ 6.0 ns and ~ 7.0 ns (Figures 5.2B-C), there was a sharp decrease in the van der Waals energy of the system, due to pi-pi stacking between the molecules' aromatic cores, (Figure 5.2I) and an unfavorable sharp increase in the electrostatic energy of the system, due to repulsion between the molecules' carbonyl

groups (Figure 5.2J). The competition between these two sets of interactions appeared to dictate the final arrangement of the system, where there was substantial overlap between the perylene cores but a slight misalignment of the carbonyl groups (Figure 5.2D).

These relatively small fluctuations of the two PTCDIs in the stacked conformation at equilibrium are in contrast to the significantly large mobility of each PEG tail that kept moving around its PTCDI (Figure 5.2G) and the other PEG tail (Figure 5.2H). This is due to the relatively weak attraction between each PEG tail and its PTCDI (Figure 5.2K) and between the two PEG tails (Figure 5.2L) compared to the strong attraction between the two PTCDIs. Therefore, each PEG tail never inserted itself between the two PTCDIs once they already stacked.

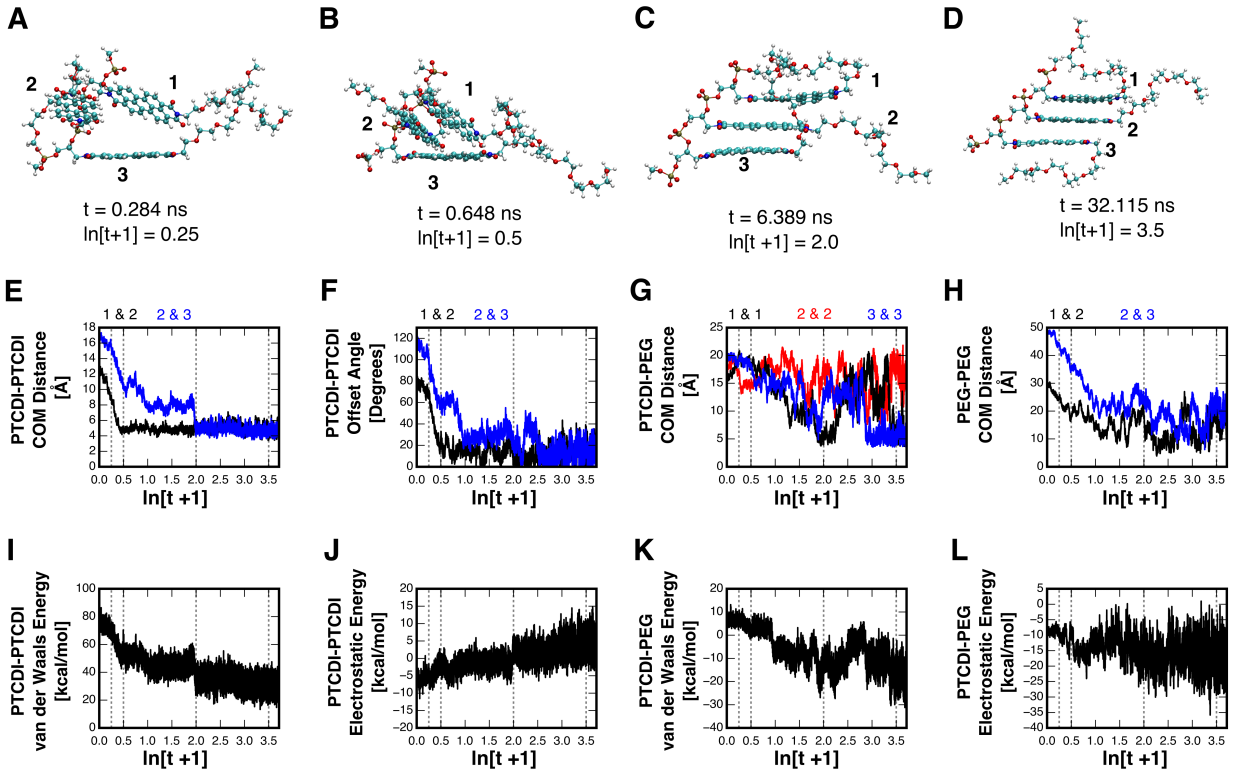

Figure 5.3: Snapshots of **P3** during a molecular dynamics simulation at times of A) t = 0.284 ns, B) t = 0.648 ns, C) t = 6.389 ns, and D) t = 32.115 ns. The sequence demonstrates the transition of **P3** from an open random to a stacked configuration. As a function of time, the evolution of: E) the centers of mass (COM) distance between the two PTCDIs, (F) the offset angle between the two PTCDIs, (G) COM distance between each PTCDI and its PEG tail**,** (H) COM distance between two PEG tails, (I) the van der Waals energy of PTCDIs, (J) the electrostatic energy of PTCDIs, (K) the van der Waals energy between each PTCDI and its PEG tail, and (L) the electrostatic energy between each PTCDI and its PEG tail of **P3**. The four vertical dashed lines in (E)-(L) correspond to the times used for the snapshots in A), B), C), and D). The simulations were performed at a constant temperature of 300 K.

We next investigated the stacking kinetics of the three PTCDI moieties of **P3**. In our experiments, we observed that self-assembly of **P3** into the final stacked arrangement necessitated a two-step mechanism. First, two of the PTCDIs formed a dimer analogous to the one found for the final configuration of **P2**. Second, the remaining third PTCDI stacked with this dimer to form a trimer. Although our simulations revealed multiple possibilities for the PTCDIs' stacking order, this general mechanism was highly reproducible, with

similar results obtained for ten independent simulations.

As an example, **Figure 5.3** illustrates a typical **P3** stacking kinetics simulation, which demonstrates the aforementioned multi-step mechanism and the most likely assembly pathway. At t = 0.284 ns, the three constituent PTCDIs (denoted as **1**, **2**, and **3**) were unstacked with a random open configuration (Figure 5.3A). Subsequently, during the first step of the stacking mechanism at t = 0.648 ns (Figure 5.3B), **1** and **2** formed a dimer, with a corresponding sharp drop from 13.0 Å to 5.0 Å for their COM distance (Figure 5.3E) and from 80.0° to 4.0° for their offset angle (Figure 5.3F). In turn, during the second step of the stacking mechanism at t = 6.389 ns, **3** stacked on the bottom of the dimer to form a trimer (Figure 5.3C). For this step, there was a sharp drop from 9.0 Å to 5.0 Å in the COM distance (Figure 5.3E) and from 63.26° to 13.76° in the offset angle (Figure 5.3F) for **1** and **2**. There also was an accompanying drop from 10.92 Å to 9.29 Å in the COM distance (Figure 5.3E) and from 50.0° to 20.0° in the offset angle (Figure 5.3F) between **2** and **3**. Although the three PTCDIs remained in a stacked arrangement for the remainder of the ~ 34 ns simulation as shown at 32.115 ns (Figure 5.3D), we again observed variability in their relative COM distances and offset angles (Figure 5.3E,F) of the three PTCDIs, indicating some dynamic character for the final configuration.

To further understand the stacking kinetics of the three PTCDI moieties of **P3**, we analyzed the energetics driving their self-assembly. In the first step of the mechanism, the formation of the dimer was driven by a decrease in the van der Waals energy and an increase in the electrostatic energy (Figure 5.3I,J), as also observed for **P2** (Figure 5.2I,J). In the second step of the mechanism, a similar interplay of favorable and unfavorable interactions appeared to drive formation of the trimer, as indicated by similar changes in

the van der Waals and electrostatic energies (Figure 5.3I,J). In their equilibrium structure, **P3**'s three constituent PTCDIs adopted a slightly twisted arrangement, where they were stacked with a small offset with respect to one another (Figure 5.3D). In fact, the COM distance between **1** and **2** is 4.587 ± 0.502 Å and between **2** and **3** is 4.528 ± 0.357 Å while the offset angle between **1** and **2** is 16.712± 7.004° and between **2** and **3** is 16.974 ± 6.586°. As also observed for **P2**, these values of the average COM distance and offset angle falls within the region of the lowest free energy on the energy landscape that we had previously calculated for the stacking of three un-PEGylated PTCDIs.

These relatively small fluctuations of the two PTCDIs in the stacked conformation at equilibrium are in contrast to the significantly large mobility of each PEG tail that kept moving around its PTCDI (Figure 5.3G) and the other PEG tail (Figure 5.3H). This is due to the relatively weak attraction between each PEG tail and its PTCDI (Figure 5.3K) and between the two PEG tails (Figure 5.3L) compared to the strong attraction between the two PTCDIs. Therefore, each PEG tail never inserted itself between the two PTCDIs once they already stacked.

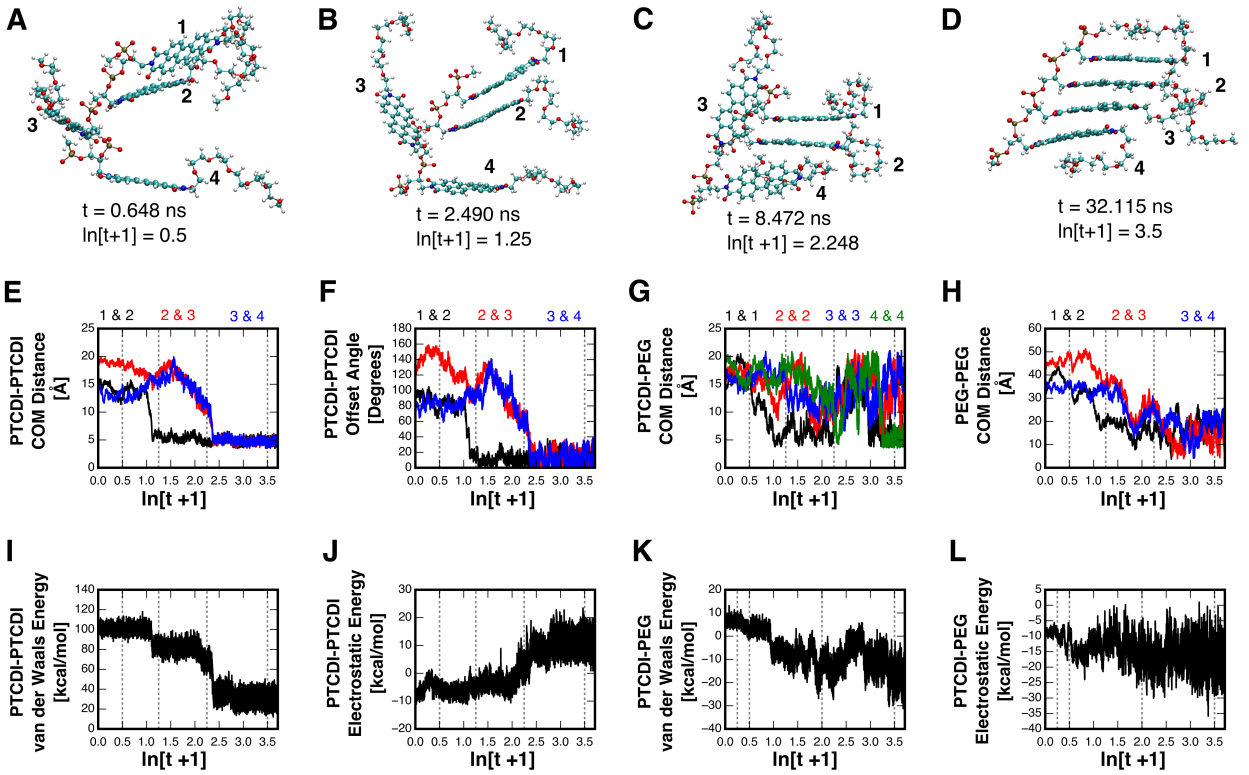

Figure 5.4: Snapshots of **P4** during a molecular dynamics simulation at times of A) t = 0.648 ns, B) t = 2.490 ns, C) t = 8.472 ns, and D) t = 32.115 ns. The sequence demonstrates the transition of **P4** from an open random to a stacked configuration. As a function of time, the evolution of: E) the centers of mass (COM) distance between the two PTCDIs, (F) the offset angle between the two PTCDIs, (G) COM distance between each PTCDI and its PEG tail**,** (H) COM distance between two PEG tails, (I) the van der Waals energy of PTCDIs, (J) the electrostatic energy of PTCDIs, (K) the van der Waals energy between each PTCDI and its PEG tail, and (L) the electrostatic energy between each PTCDI and its PEG tail of **P4**. The four vertical dashed lines in (E)-(L) correspond to the times used for the snapshots in A), B), C), and D). The simulations were performed at a constant temperature of 300 K.

Finally, we investigated the stacking kinetics of the four PTCDI moieties of **P4**. In our experiments, we observed that self-assembly of **P4** into the final stacked arrangement necessitated a three-step mechanism involving multiple pathways. One popular pathway involves the formation of a dimer with any two consecutive PTCDIs before the third PTCDI stacked on top or bottom of the dimer to form a trimer, which is followed by the final PTCDI stacked on top or bottom of the trimer to form a tetramer. Another pathway involves the formation of two separate dimers by the top two PTCDIs and the bottom two

PTCDIs; these two dimers then stacked together to form a tetramer.

As an example, **Figure 5.4** illustrates the latter pathway to form a stacked **P4** structure. At t = 0.648 ns, the four constituent PTCDIs (denoted as **1**, **2**, **3**, and **4**) were unstacked with a random open configuration (Figure 5.4A). Subsequently, during the first step of the stacking mechanism at t = 2.490 ns (Figure 5.4B), **1** and **2** formed a dimer, with a corresponding sharp drop from 15.0 Å to 5.0 Å for their COM distance (Figure 5.4E) and from 100.0° to 8.0° for their offset angle (Figure 5.4F). In turn, during the second step of the stacking mechanism at t = 8.472 ns, **3** and **4** stacked together to form a dimer (Figure 5.4C). For this step, there was a sharp drop from 10.0 Å to 5.0 Å in the COM distance (Figure 5.4E) and from 70.0° to 10.00° in the offset angle (Figure 5.4F) for **3** and **4**. There also was an accompanying drop from 10.92 Å to 9.29 Å in the COM distance (Figure 5.4E) and from 50.0° to 20.0° in the offset angle (Figure 5.4F) between **2** and **3**. Although the three PTCDIs remained in a stacked arrangement for the remainder of the ~ 34 ns simulation as shown at 32.115 ns (Figure 5.4D), we again observed variability in their relative COM distances and offset angles (Figure 5.4E,F) of the four PTCDIs, indicating some dynamic character for the final configuration as also observed in **P2** and **P3**.

The three-step mechanism of stacking four PTCDI moieties of **P4** is also observed as three separate transitions in the van der Waals energy (Figure 5.4I) and the electrostatic energy (Figure 5.4J). In every step of the mechanism, the formation of the dimer, trimer or tetramer was driven by a decrease in the van der Waals energy and an increase in the electrostatic energy (Figure 5.4I,J). This interplay of favorable and unfavorable interactions correspond to due to the formation of attractive pi-pi interactions between aromatic cores and repulsion between the carbonyl groups (Figure 5.4I,J).

In their equilibrium structure, **P4**'s three constituent PTCDIs adopted a slightly twisted arrangement, where they were stacked with a small offset with respect to one another (Figure 5.4D). In fact, the COM distance between **1** and **2** is 4.722 ± 0.432 Å, between **2** and **3** is 4.694 ± 0.378 Å, and between **3** and **4** is 4.746 ± 0.345 Å. The offset angle between **1** and **2** is 12.556 ± 7.482°, between **2** and **3** is 14.886 ± 8.310°, and between **3** and **4** is 11.742 ± 7.944°. These values of the average COM distance and offset angle falls within the region of the lowest free energy on the energy landscape that we had previously calculated for the stacking of two or three un-PEGylated PTCDIs.

These relatively small fluctuations of the two PTCDIs in the stacked conformation at equilibrium are in contrast to the significantly large mobility of each PEG tail that kept moving around its PTCDI (Figure 5.4G) and the other PEG tail (Figure 5.4H). This is due to the relatively weak attraction between each PEG tail and its PTCDI (Figure 5.4K) and between the two PEG tails (Figure 5.4L) compared to the strong attraction between the two PTCDIs. Therefore, each PEG tail never inserted itself between the two PTCDIs once they already stacked.

5.5 Conclusions

In this work we investigate PEGylated PTCDI stacking using molecular dynamics simulations. Our simulations show similar results to un-PEGylated PTCDI, however with an added complexity of the PEG chain during stacking. We find the PEG chain to only interfere in PTCDI stacking and once the surrogates are stacked the PEG chain does not adversely affect the stacking. PEG chains are exposed to solution and not inserted between the PTCDI's. For example, P3 with the top PEG chain is twisted on top of the PTCDI's and the

bottom PEG chain on the bottom of the PTCDI's. The middle PEG chain is free in solution and not interfering with the stacking between the PTCDI's. We have also demonstrated the presence of multiple stacking pathways of PTCDI's in solution: sequential addition or separated addition as shown by P4 with two separate PTCDI's stack and then come together to an equilibrium structure.

Additionally, simulations will elucidate the effect of adding the substituents the core of the perylene with a phosphate backbone can similar intermolecular spacing and offsets to that of the unsubstituted. Using free energy analysis tools, and quantum mechanics calculations the molecular orbitals can be calculated based on the lowest free energy structures and determine if there are significant differences to the non-PEG and unsubstituted perylene core.

5.6 Supplementary Information

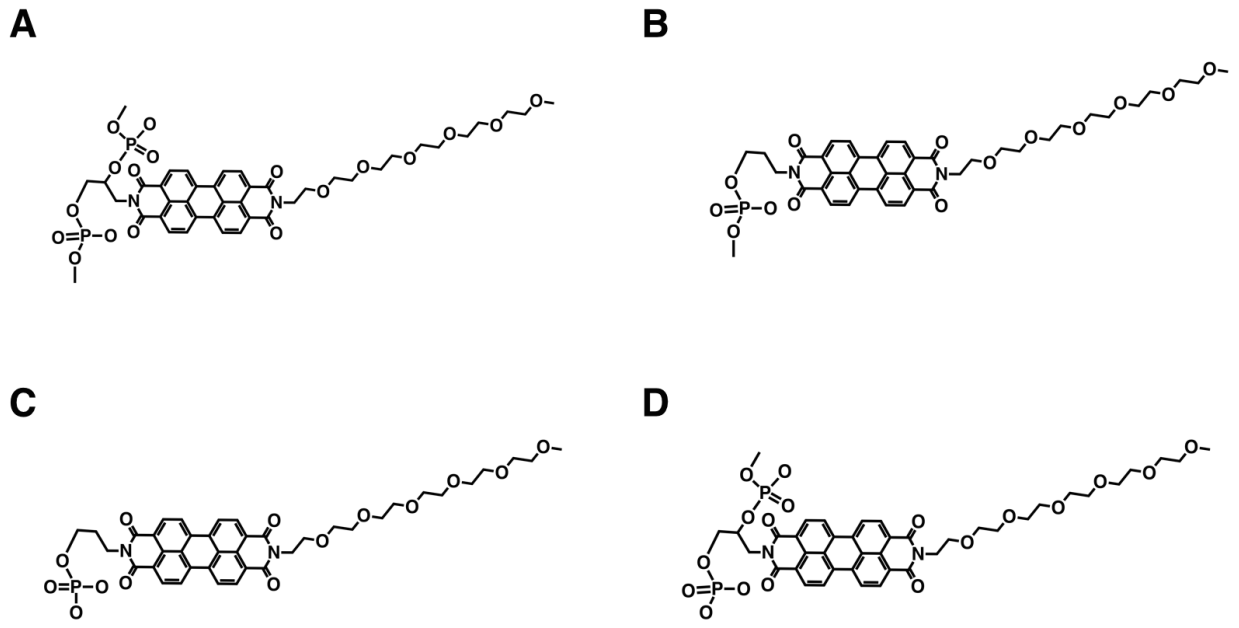

Figure S5.1: PTCDI-PEGylated Residues. (A) Used for charge optimization. (B) (C) (D) are modified versions of (A) by removing excess atoms. B-D are used to be able to make PTCDI nanowires of any length.

5.7 References

1. Herbst, W.; Hunger, K. *Industrial Organic Pigments: Production, Properties, Applications*. 2nd ed. Wiley: Weinheim, **1997**.
2. Würthner, F. Perylene Bisimide Dyes as Versatile Building Blocks for Functional Supramolecular Architectures. *Chem. Commun.* **2004**, 1564–1579.
3. Görl, D.; Zhang, X.; Würthner, F. Molecular Assemblies of Perylene Bisimide Dyes in Water. *Angew. Chem. Int. Ed.* **2012**, *51,* 6328-6348.
4. P. M. Kazmaier and R. Hoffmann. A Theoretical Study of Crystallochromy. Quantum Interference Effects in the Spectra of Perylene Pigments. *J. Am. Chem. Soc*. 116, 9684-9691 (1994).
5. Geng, Y.; Li, H.-B.; Wu, S.-X.; Su, Z.-M. The interplay of intermolecular interactions, packing motifs and electron transport properties in perylene diimide related materials: a theoretical perspective. *J. Mater. Chem.* **2012**, *22*, 20840-20851.
6. Chen Li, C.; Wonneberger, H. Perylene Imides for Organic Photovoltaics: Yesterday, Today, and Tomorrow. *Adv. Mater.* **2012**, *24*, 613–636.
7. Kozma, E.; Catellani, M. Perylene diimides based materials for organic solar cells. *Dyes Pigments* **2013**, *98*, 160-179.
8. Jung, B. J.; Tremblay, N. J.; Yeh, M-L.; Katz, H. E. Molecular design and synthetic approaches to electron-transporting organic transistor semiconductors. *Chem. Mater*. **2011**, *23*, 568–582.
9. Huang, C.; Barlow, S.; Marder, S. R. Perylene-3,4,9,10-tetracarboxylic Acid Diimides: Synthesis, Physical Properties, and Use in Organic Electronics. *J. Org. Chem.* **2011**, *76*, 2386–2407.

10. Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multifunctional and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
11. Gaussian 09, Revision C.01. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.;Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2009.
12. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
13. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
14. Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
15. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26,* 1781–1802.
16. Brooks, B. R.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Calfisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecule Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

CHAPTER 6 Summary and Future Direction

6.1 Summary

In this dissertation, a novel coarse-gained model was developed to study DNA hybridization processes at high concentrations. The model is capable of capturing the mechanical and thermal properties of DNA compared to experiments. It has been implemented in our newly developed BioModi simulation package for simulating large systems of peptides, nucleic acids and polymers for long time scales. In Chapter 2, our results indicate that initial contacts between strands of DNA splits the kinetic pathways to either a zipping-like mechanism or a slithering-like mechanism. This kinetic partitioning, which is strongly dependent on the sequence, greatly affects the kinetic pathway and thus determines the yield of desired double-helices. Our large-scale simulations at a high DNA strand concentration demonstrate that DNA self-assembly is a robust and enthalpically driven process in which the formation of double helices is deciphered to occur via multiple self-assembly pathways including the strand displacement mechanism. It is demonstrated that sequence plays a complex role in self-assembly and that the model can be used for studying DNA nanotechnology systems. In Chapter 3, the coarse-grain model is used to elucidate the effects of sequence, temperature and concentration on DNA self-assembly. Simulations of many ssDNA demonstrate the presence of many kinetic pathways in self-assembly that depend on their environmental conditions. A method for evaluating aggregate properties during the course of the simulation is used to demonstrate trends in aggregate lifetimes, creation times, and base pairing. Kinetic modeling of the self-assembly process was performed and an experimentally accurate activation energy was found for the

formation of dimers. This work demonstrates a framework in which can be used to study more complex systems such as those found in DNA nanotechnology systems.

In addition to developing a coarse-grained model, a DNA base surrogate nanowire was parameterized and simulated using an all-atom approach. In Chapter 4, perylene-3,4,9,10-tetracarboxylic diimides (PTCDIs) are parameterized and simulated using constant temperature molecular dynamics obtain an improved understanding of the self-assembly of PTCDI dimers and trimers. Using an advanced sampling technique, it is shown that the equilibrium configurations are similar to those expected from experiments. In Chapter 5, PTCDI structures are PEGylated and demonstrate that equilibrium structures are similar to non-PEGylated structures. PEGylated PTCDI's are shown to have a more complex kinetic pathway due to the inclusion of the PEG chain. The length of the PTCDI nanowires was also determined to affect the kinetic mechanisms with more emerging with longer length.

6.2 Future Direction

6.2.1 Biomodi DNA Model and DMD Base Code

The foundation of using BioModi for the study of DNA hybridization processes at large scale has been provided by this work. This leaves a number of possible applications of the model and improvements. The field of dynamic DNA devices is rapidly growing and the need for understanding the processes at the molecular level has not been fully elucidated [1]. BioModi could provide insight into how to improve logic gate circuits, such as those proposed by Qian & Winfree and others [2-6], by identifying molecules that cross-talk and

investigating how to reduce them. Preliminary work done on these logic gates using BioModi has been promising. Additionally, applications to drug delivery using DNA nanostructures could be investigated such as the work by Douglas *et al*. using a DNA nanobox with a molecular payload [7]. An example investigation would be to study how the box is unlocked, how the box opens and how the molecular pay load is released into solution. BioModi could also be applied to study dynamic DNA nanostructures that grow and shrink based on the amount of fuel or anti-fuel is in the system [8].

In building and testing of the current BioModi DNA model, a number of improvements were realized to further the applications. One improvement would be to fine tune the persistence length and thermal melting properties to even more closely capture the DNA structure at real temperatures. This would include scaling of the base pairing strength and other interactions to scale from reduced temperatures to real temperatures. Simulating at real temperatures will enable closer comparisons to experiments and enable further improving the model to capture the dynamics and structures of complex DNA structures. The model could be further extended to support the transition between A-form and Z-form DNA as the current model was primarily parameterized for only B-form DNA. Additionally, a number of improvements were discovered for the DMD implementation in BioModi. To enable larger scale simulations and achieve longer time scales the DMD code will need to be parallelized. Khan *et al*. have parallelized their DMD code [9], thus showing a possible method in which to parallelize the BioModi code. To make the DMD code more maintainable and expandable the code should be refactored into an object-oriented design, similar to DynamO [10]. This would enable additions of other important biomolecules such as poly-saccharides, easier to implement.

6.2.2 DNA Base Surrogates

Substituting the core of PTCDIs with molecules such as chlorine has been show to disrupt the typical planar nature due to steric hindrance. Furthermore, it has been shown by Delgado *et al*. that adding chlorine groups to the core of the perylene structure will greatly affect the LUMO energies (~0.4 eV) [11]. Therefore, it is important to look at how the structure will change when bonded to the backbone of DNA. An example of the 3' is shown below in **Figure 6.1**. Additionally, it is not known how many PTCDI nanowires in solution would interact with one another. One possible study would be to take randomized unstacked nanowires and put them in a simulation box to study if mis-aggregates form. It is also important to understand the interactions between individual nanowires while chemisorbed to a surface. One could simulate a set of PTCDI nanowires connected to a large surface and study the effects of packing.

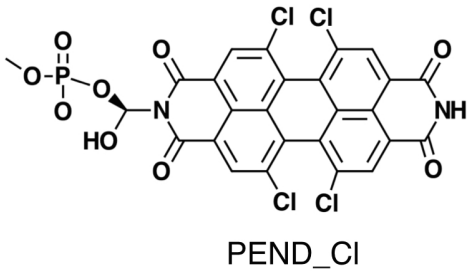

Figure 6.1: Example PEND (3' end) tetrachlorinated perylene base residue to be parameterized for molecular dynamics studies.

6.3 References

1. D. Y. Zhang and G. Seelig. Dynamic DNA Nanotechnology using strand displacement reactions. *Nature Chem*., **2011**, 3: 103-113.
2. Qian, L.; Winfree, E. Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades. *Science* **2011**, *332*, 1196–1201.
3. L. Qian and E. Winfree. A simple DNA gate motif for synthesizing large-scale circuits. *J. R. Interface*, **2011**, 8, 1281-1297.
4. H. Lederman, J. Macdonald, D. Stefanovic, and M. N. Stojanovic. Deoxyribozyme-Based Three-Input Logic Gates and Construction of a Molecular Full Adder. *Biochemistry*, **2006**, 45: 1194-1199.
5. Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro. An autonomous molecular computer for logical control of gene expression. *Nature*, **2004,** 27: 423-429.

6. C. C. Santini, J. Bath, A. J. Turberfield, A. M. Tyrrell. A DNA Network as an Information Processing System. *Int. J. Mol. Sci.*, **2012**, 13, 5125-5137.
7. S. M. Douglas, I. Bachelet, and G. M. Church. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science*, **2012**, 335, 831-834.
8. R. P. Goodman, M. Heilemann, S. Doose, C. M. Erben, A. N. Kapanidis, and A. J. Turberfield. Reconfigurable, braced, three-dimensional DNA nanostructures. *Nat. Nanotech.,* **2008**, 3: 93-96.
9. Md. Ashfaquzzaman Khan, and M. C. Herbordt. Parallel discrete molecular dynamics simulation with speculation and in-order commitment. *J. Comput. Phys.,* **2011**, 230: 6563-6582.
10. M. N. Bannerman, R. Sargant, and L. Lue. DynamO: A Free O(N) General Event-Driven Molecular Dynamics Simulator. *J. Comput. Chem.,* **2011**, 32: 3329-3338.
11. M. C. R. Delgado, E.-G. Kim, D. A. da Silva Filho, and J-L Bredas. *J. Am. Chem. Soc.,* **2010**, 132, 3375-3387.