



Simulating folding of a DNA pyramid

Master's Thesis

Master of Science in Theoretical Chemistry

Chemistry Department
University of Oxford

written by
Raphael Peltzer
born in Cologne

working in the group of
Prof. Dr. Jonathan Doye
under the supervision of
Dr. John Schreck

Oxford, September 2014

Acknowledgments

I thank Professor Jonathan Doye for his great help and advice throughout the course of this project and for giving me the invaluable opportunity to work in his research group. I greatly thank Jonathan Doye and John Schreck for sharing their incredible insight on the project.

I am indebted to Doctor John Schreck for supporting me throughout this project and taking a lot of time on our countless discussions on the project. I would like to thank him especially for always being extremely encouraging and answering all my questions and emails, however many there were. I will also greatly miss our discussion on World politics and sports, especially American Football. I would also like to thank the entire Doye group for the help everyone has given me on my project and the group meetings and discussion.

I would also like to thank my grandparents and parents who made my year here in Oxford possible and supported me financially as well as emotionally wherever they could throughout this year. I also want to especially thank my sister for encouraging me and helping me with proofreading my thesis. I would also like to thank Balliol college for the support given me during my Master's degree program.

Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references. I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted. This paper was not previously presented to another examination board and has not been published.

first and last name

city,date and signature

Contents

List of abbreviations	v
1 Introduction	1
1.1 Structure and physics of DNA	2
1.2 DNA nanotechnology applications in modern chemistry	6
1.2.1 DNA origami	7
1.2.2 Dynamic DNA nanotechnology	7
1.2.3 DNA computing	8
1.3 DNA polyhedral nanostructures	8
1.3.1 DNA pyramid from ssDNA	10
2 Theoretical background	14
2.1 Computational methodology overview	14
2.1.1 Molecular Modelling of DNA	14
2.1.2 The oxDNA model	15
2.1.3 Representation of DNA in oxDNA	16
2.1.4 Limitations of oxDNA	18
2.2 Simulation background	18
2.2.1 Monte Carlo Simulations	19
2.2.2 Collective particle motion methods	21
2.2.3 Virtual Move Monte Carlo algorithm	21
2.2.4 Molecular Dynamics Simulations	23
2.2.5 Temperature control and kinetics	25
3 Experiments and Results	27
3.1 Experiments	28
3.1.1 Folding	28
3.1.2 Temperature induced Melting	31
3.2 Results	32
3.2.1 DNA Pyramid construction	32
3.2.2 NUPACK data	33
3.2.3 Temperature induced Melting	36
3.2.4 Folding	46
4 Summary and outlook	48
4.1 Conclusion	48
4.2 Future work	49

Contents

5 Appendix	50
5.1 Appendix A	51
5.2 Appendix B - NUPACK data	52

List of abbreviations and acronyms

ssDNA single-stranded DNA

dsDNA double-stranded DNA

OPT optimized DNA pyramid design

KH kinetically hindered DNA pyramid design

SDN Structural DNA nanotechnology

MD Molecular Dynamics

MC Monte Carlo

VMMC Virtual Move Monte Carlo

MMC Metropolis Monte Carlo

DP DNA nano-pyramid

NUPACK Nucleic Acid Package

domain DNA double helix formed from two complimentary strands

T_m Melting temperature

2D, 3D two-, three-dimensional

1 Introduction

The structure, thermodynamics and dynamics of proteins and other biological molecules play an important role in their functional behaviour. For example, polyhedral structures composed of DNA or RNA have great potential for being used for nanophotonic applications or drug delivery. The device at hand has previously been investigated by experimentalists as the structurally least complex example of a polyhedral structure and its investigation could lay ground for the introduction of polyhedral single-stranded DNA structures into biotechnological modelling.

In order to understand a molecule's function, it is necessary to study its conformation and its structural properties. Since biomolecules are often only able to fulfil their purpose in a specific conformation, it is very important to not only understand their structure, but also their thermodynamic as well as kinetic behaviour. In order to characterize structural properties of the bio-molecule, one may employ different computational and experimental approaches. Computational approaches possess the advantage that systems can be investigated at the microscopic level using models fitted to their biological application. Further, predictions can be made, that are hardly or only very costly obtainable by experiments. Also, computational experiments may shed light on the underlying thermodynamics of self-assembly, which can be compared to the kinetic processes observed experimentally. By comparing experimentally and computationally obtained data predictions can be confirmed or invalidated.

The object of this thesis is to employ a coarse-grained model of DNA, called oxDNA, in order to investigate the folding, unfolding and thermal properties of a DNA pyramid (DP) shown in Figure 1.1. This structure was proposed by Kocar Vid *et al.* and is different to similar polyhedral structures, such as those built by Turberfield^[1] and Mao^[2], in that only one long single strand of DNA is needed to build a complete structure. Polyhedral structures have been envisioned for applications in medicine and diagnostics.

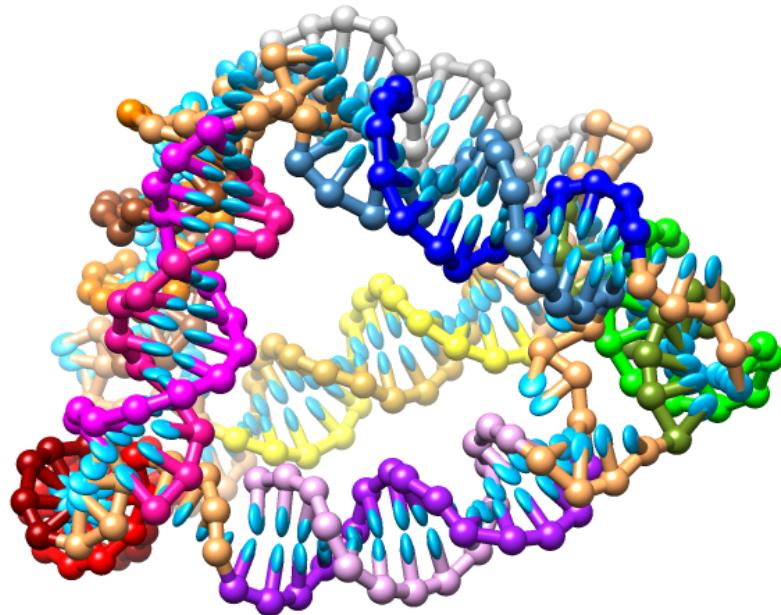


Figure 1.1 – DNA Pyramid. The colour code is given in Appendix A.

The folding pathway is thought to be controlled by tuning the sequences of the edges of the structure. In chapter 3 we will detail the experiments and our collaboration with the experimentalists, Kocar Vid and co-workers. Before we delve into the details of our results the structure and physics of deoxyribonucleic acid (DNA) will be explained in chapter 1, together with an introduction to polyhedral structures as an example of modern-day nano-structural design. In chapter 2, the computational model will be introduced. In chapter 3, the experimental procedure and the results will be explained and discussed. Finally, the conclusion shall be summarized and an outlook for future research will be presented.

1.1 Structure and physics of DNA

The biological structure of DNA has been and continues to be of huge interest, principally because of its enormous importance as the genetic information storage carrier for all living organisms. The DNA structure was first uncovered by Crick and Watson when they inferred a double-helical structure of two inter-twined DNA single strands, which they obtained from the diffraction pattern data supplied by Maurice Wilkins and Rosalind Franklin. This DNA model, for which Crick Watson and Wilkins were awarded a Noble Prize of Physiology in 1962, led to a much better understanding of the most basic elements

of biological chemistry. Their work laid ground for the investigation of very important biological processes, like DNA replication or transcription, and since then the structure has been investigated by many researchers and in a variety of disciplines. In fact our knowledge on the exact behaviour of DNA, and possible applications, is still growing.

Looking at the physical model of DNA in its helical form there are many physical interactions between the atoms. In order to mimic these interactions it is theoretically possible to simulate the most important inter-atomic forces, such as ionic, dipole-dipole or van-der-Waals interactions for all atoms. However, due to finite computing power it is practically impossible to simulate relevant time scales of all-atom-level models. The system's structure has to be investigated at a more coarse-grained level. In order to be able to directly compare the computational model to the biological structure of DNA, a coarse-grained description has to be employed.

In order to understand the DNA structure's functionality and to derive various properties from it, it is essential to first understand the functionalities of DNA's fundamental constituents and interactions. The physical interactions can be classified into four categories: i) covalent bonds between neighbouring nucleotides, on the phosphate-deoxyribose-backbone of DNA, ii) hydrogen bonding between complementary bases facing each other in the coiled structure of the double helix, iii) interactions caused by stacking the nucleobases in the centre of the double helix and iv) excluded volume interactions. These properties will be discussed in more detail next.

Each DNA strand is composed of covalently linked repeating units called nucleotides. A nucleotide consists of a pentose base, a nucleobase and a phosphate backbone. The pentose sugar, that can be seen in Figure 1.2, is hydrogenated in the 2' position, because a DNA (deoxyribose in the Figure 1.2) system is investigated. The phosphodiester bonds in the phosphate-deoxyribose-backbone are covalent bonds, which makes these bonds very stable and inductile, and thus practically unbreakable at reasonable simulation temperatures. Neighbouring mononucleotides are linked via these bonds to form polynucleotides.

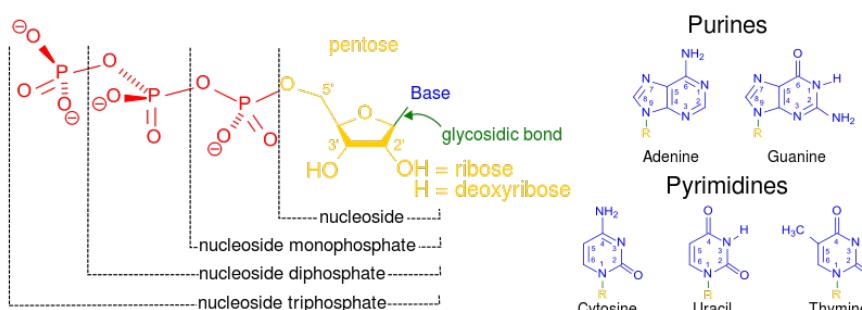


Figure 1.2 – Structure of nucleotides taken from <http://en.wikipedia.org/wiki/Nucleotide>. (Left) Illustration of the nomenclature of a nucleotide and (Right) Illustration of the different types of nucleobases.

INTRODUCTION

The main varying element of the nucleotide is the nucleobase which can be either Adenine (A), Cytosine (C), Guanine (G) or Thymine (T). ^a These purine or pyrimidine nucleobases (abbreviated "bases") form a nucleoside with the aforementioned deoxyribose sugar. Genetic information can be stored within DNA single strands by varying the order of these nucleobases. However, as can be seen from Figure 1.3, only A and T as well as C and G are complementary bases that can form hydrogen bonds between each other. This restraint is called the "Watson-Crick base pairing" rule. Due to Watson-Crick complementarity, one sequence of bases exactly determines the other sequence that will be able to form a double helix with it. These unique double-helices of complementary strand segments will be called domains. Also, Watson-Crick complementarity is essential to the biological function of DNA, especially for replication.^[3]

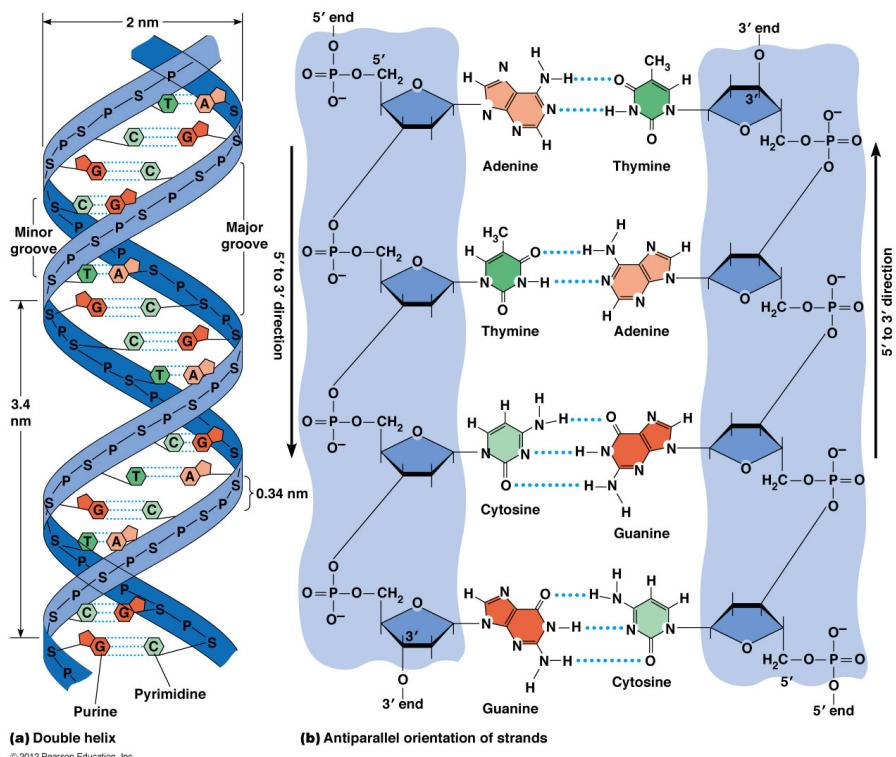


Figure 1.3 – (a) Structure of the double helix and **(b)** Planar view of 4 rungs of the DNA double helix shows how the two complementary strands have to aligned antiparallel.

Two complementary strands of DNA must be oriented anti-parallel to become hybridized. This means each hybridized strands 5' end has to be on the same side of the complementary region as the other strands 3' end, which is illustrated in Figure 1.3 (b). Further, G-C

^athe graphic Figure 1.3 also displays Uracil (U), which is the demethylated equivalent of Thymine used in RNA

interactions are more attractive than A-T interactions, because of an additional hydrogen bond that is involved in forming their planar connection.

The helical structure of double-stranded DNA stems from physical properties induced by functional groups as well as simple forces. DNA is usually found in an aqueous environment and, as for most amphiphilic compounds found in organisms, the lipophilic end - the aromatic nucleobase - is facing inwards in the plane perpendicular to the helix's axis.^b The hydrophilic ends - the negatively charged phosphate groups - face outwards, leading to an increase in solvation energy in an aqueous surrounding. The folding of the double helix naturally decreases the energy of the system, where the hydrophobic functional groups are screened from the generally hydrophilic solvent, thus solvation entropy is gained.

DNA has a very even structure, because every Purine base is paired with a Pyrimidine base and the inter-strand distance is the same for every rung. Additionally, the sugar backbone molecules are orientated at a 36° angle to one another while the helix twists about 3.4 Å with each sugar molecule. The radius of the DNA double-helix is 10 Å and its pitch is 34 Å, which is approximately equal base pairs per turn. The helicity of DNA is induced to comply with the twisting constraint restraint as well as the optimum separation of 6 Å between adjacent sugar molecules. In this form different aromatic groups are always oriented parallel to each other and the induced π -orbital interaction is known as base pair stacking. The winding leaves gaps adjacent to the nucleobases. These gaps are called the "minor groove" and the "major groove" and are shown in Figure 1.3(a) Inside the "right-handed" duplex the base pair interactions can further be summarized as interactions between non-neighbouring nucleobases and van-der-Waals forces.

Single- stranded DNA can form short segments of a helix with a higher persistence length of about 19.4 Å. Double-stranded DNA (dsDNA) is stiff and the double-helices have a persistence length of 50 nm or 150 base pairs, whereas single-stranded DNA (ssDNA) is extremely flexible.

One says the directions of both strands are oriented antiparallel. A double-helix that has two complementary strand ends connected by a flexible region is called a DNA "hairpin". If the loop region of one hair pin is equally the stem of another hairpin, the structure is called a DNA pseudoknot. Examples of a DNA hairpin and a DNA pseudoknot are shown in Figure 1.4.

The structural properties described above summarize the physical and chemical interactions present in the DNA double-helix . In the next chapter, the computational model, that will be used to study DNA, called oxDNA, will be introduced and the introduced biological interactions will be compared directly to their assigned modelling properties.

^bamphiphilic compounds are molecules with both a hydrophilic (water loving) and lipophilic (fat loving) end

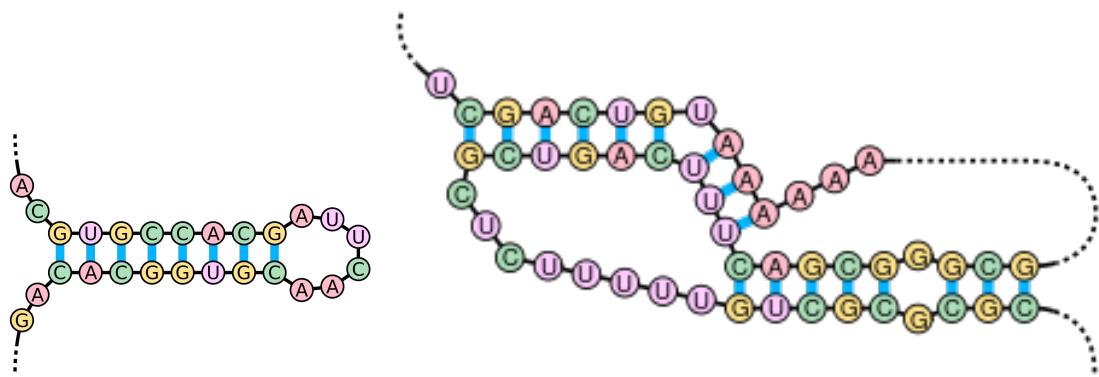


Figure 1.4 – (Left) DNA hairpin and (Right) DNA pseudoknot.

1.2 DNA nanotechnology applications in modern chemistry

In this section, a brief overview of the field of DNA technology will be given. The design of DNA nanostructures was first proposed as the combination of immobile junctions as well as interstrand (sticky end base pair) connections by Nadrian Seeman in 1982.^[4] The concept was revolutionary and the natural folding of DNA into double-helices significantly simplified the human effort needed to craft such nanostructures. Many different structures like cubes, triangles or hexagons, have been synthesised in an incredibly cost efficient and relatively simple way. There are now a variety of more complex structures available, for example tweezers^[5] and walkers.^[6] Many structures rely on self assembly and many active devices require ssDNA as fuel for strand displacement reactions. The DNA pyramid at hand is very small and only requires proper folding of the structure.

The focus shall be on structures that can be simulated on reasonable computing time scales using models such as oxDNA. This is exactly the scale on which DNA nanotechnology has its greatest advantages over traditional synthetic chemistry as well as micro structuring applications.^[7] Over the past three decades a huge variety of DNA nanostructures and nanodevices have been developed and the development of more applications, such as drug delivery diagnostics and therapeutics for human health, have been hypothesized.^[7] For example, polyhedral structures have been supposed as containers for drug delivery. Possible applications and further background to DNA polyhedra will be discussed in the next chapter. The field of DNA nanotechnology is large and is growing very fast. I separated the overview into broader categories, namely DNA origami in structural DNA nanotechnology (SDN), dynamic DNA nanotechnology and DNA computing.

1.2.1 DNA origami

A modern approach to assembling nanostructures , called DNA-origami, was developed by Rothemund.^[8] His approach was to direct the folding of a long "scaffold" strand using a number of short "staple" strands and to hold the structure in place by the rigidity of the junctions in between scaffold and staple strands. This method allows for the synthesis of much more complicated structures and severely limits the possibility of yielding misbond-forming structures.

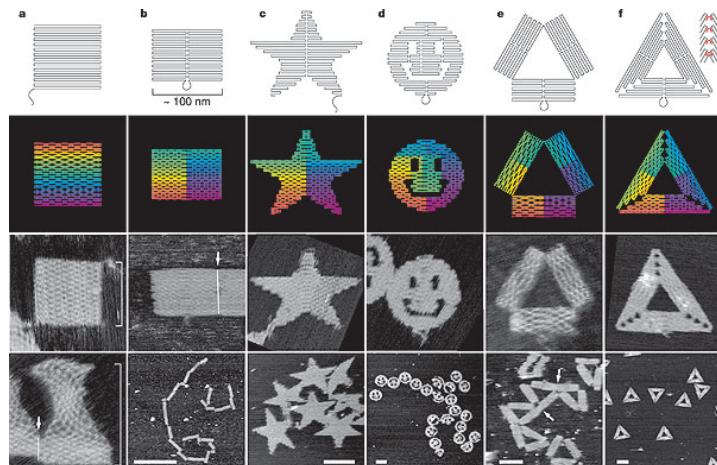


Figure 1.5 – Structures made from DNA using the DNA-origami method taken from (Rothemund, 2006).

DNA origami has many applications. For example complex shapes can be produced, such as boxes, cubes and containers, that can used for single-molecule detection^[9] and material organization. Furthermore, a two-dimensional origami has been applied as a template on which proteins can be efficiently assembled.^[10] Another very interesting example is using DNA origami as a template for the folding of carbon nanotubes into 2D cross-junctions. These devices could be used as a "programmable nanobreadboard", which would allow the development of further nanostructures.^[11]

1.2.2 Dynamic DNA nanotechnology

The field of dynamic DNA nanotechnology is based on synthesizing nanostructures with dynamic functionalities. Nanoscale dynamics can be used to create so called "nanorobotics". Upon a stimulus, such as a pH-change or hybridization with another complementary strand, a nanodevice may change its conformation, making it possible to carry out actions on a mechanically inaccessible length scale. The combination of a number of different highly specified motions and molecular recognition properties may enable these DNA devices to carry out important tasks in vitro. DNA nanodevices include

INTRODUCTION

"tweezers"^[5], "walkers"^[6], "gears"^[12], as well as stretching and rotation devices^[2], that are all running on different sorts of chemical fuel (protons for triggers^[13], ssDNA for strand displacement^[14]), which can be regained later on. It has also been shown that these devices can run autonomously^[15] and can be used as switches^[16]. A cascade of switches and logically stacked events can then be used to create autonomously running computing devices.^[17] Most recently a DNA nanorobot has gained public interest, that was reported to be able to cloak itself against immune defence reactions. This device could possibly "use logic to diagnose cancer earlier and more accurately than doctors can today".^[18]

1.2.3 DNA computing

DNA computing machines are nanoscale, autonomous, programmable computers for which the input and output, as well as software and hardware, are completely consisting of biomolecules. DNA computing has become popular because of its unique ability to store data at an incredibly low volume. For specific problems there are already DNA computing solutions that do a better job than classical computers in size and speed. The most famous of these solutions is surely the use of DNA computing to solve the Hamiltonian-path problem in DNA strands by Adleman *et al.* 20 years ago.^[19] The major advantages of this method are: 1) The obtained information can be directly transmitted in its biological form and does not have to be translated electronically at every intermediate stage, allowing it to run at a much higher maximum speed; 2) a high parallelisation of operations can be achieved; 3) biologically stochastic events can be mimicked perfectly using the stochastic nature of biomolecular events.

The technology does underlie a number of limitations: Firstly, the number of errors in the self-folding process is currently too high for high-precision processes. Secondly, creating the DNA input for these computing devices is costly. And thirdly, very large problems cannot be tackled using DNA computing, because the amount of DNA required to store the solution of an exponentially growing problem is not feasible. Despite of the number of disadvantages of this method, there are a number of simple nanoscale applications, such as an universal programmable computer^[20] and an autonomous DNA nanorobot^[21], that have been developed recently.

1.3 DNA polyhedral nanostructures

In 2005 Goodman *et al.* developed a self-assembling DNA tetrahedron, that could be joined with other tetrahedra to form larger nanostructures. A number of three-dimensional nanostructures have been designed in the past 10 years (and in fact even before that), further expanding the database of 3D-DNA nanostructures.^[22]^[23]^[24] Many polyhedral design have been folded using ssDNA as well as tile-based methods^[2]. Tile based assemblies require more than one strand to build a complete structure.

Traditional DNA-based self-assembly is based on the design of many complementary strands that are then mixed in equimolar stoichiometry and processed biochemically. It is crucial that the chosen stoichiometry is perfect, which is underlying experimental error, that can lead to a low yield. To circumvent this problem it is possible to build more simple structure out of single-stranded DNA, guaranteeing that the stoichiometry is always perfect. This approach has however not been successfully applied to larger structures owing to the complexity of DNA folding. Also, if it is possible to assemble from a single strand, proximity in solution of the double helix forming domains of DNA to another is guaranteed even on a nanoscopic level and good yields can be achieved even at low concentrations.

Further, single-stranded DNA objects have been folded successfully, as shown for a tetrahedron developed by Li and co-workers.^[25] They have already pointed out that single-stranded DNA assembly provides a number of advantages over traditional DNA folding. In addition to the simplicity of their stoichiometry, ssDNA can be circularized much more efficiently to grant exonuclease resistance, prolonging the life time of in vivo applications, such as drug delivery designs. Furthermore, DNA nanostructures can be multiplied and thus extended efficiently using biological methods, making ssDNA designs very attractive.^[26]

It has recently been shown that DNA nano-pyramids can be used for protein detection and to increase device sensitivity. Just this year, Setyawati *et al.* modified a DNA pyramid and loaded the structure with Actinomycin D, a very common antibiotic.^[27] The modifications made to the DNA nanostructure were red-emitting gold nanoclusters that were used for tracking and data analysis. This allowed the researchers to determine when the DNA pyramid entered the bacterial cells. In their studies it was shown that the DNA nanopyramid was able to increase the drug's efficiency in killing their targets (*E. coli* and *S. aureus*) and improved its transportation properties. These designs have a great advantage over inorganic-based delivery structures, which have been shown to underlie toxicity issues.^[28] If the cloaking of the drug can be combined with further delivery efficiency this might offer a direct approach to taking on the great challenge of battling drug-resistant bacteria.

1.3.1 DNA pyramid from ssDNA

The work of this thesis stands in cooperation with a reported development of a ssDNA pyramid by Kocar Vid from the National Institute of Chemistry in Ljubljana, Slovenia. The ideally folded DP has the form of Figure 1.6.

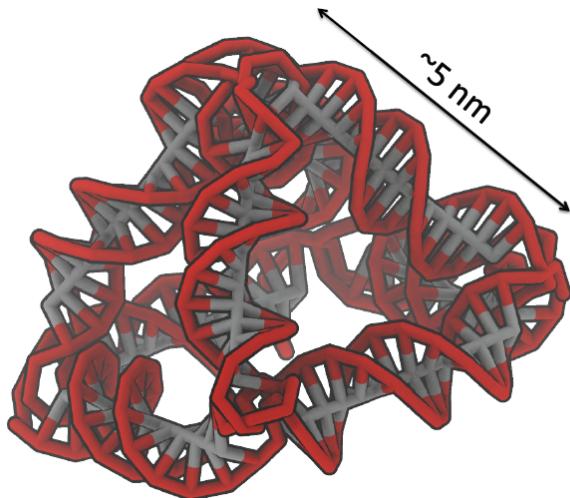


Figure 1.6 – Ideal fold of a symmetrical ssDNA Pyramid. Both isostructures, kinetically hindered and optimized, would ideally yield the same folded geometry. The figure was taken with permission from Kocar Vid.

Polyhedral structures of two folding pathways for different isoforms have been developed and are shown in Figure 1.6. The four-sided pyramid, henceforth referred to as "pyramid", is folded according to Figure 1.7. In order for the structure to be stable it has to possess at least one anti-parallel double-trace topology. This means that the single strand has to be wound in a way that every strand is aligned antiparallelly to its complementary region. As described in section 1.1, one strand has to be aligned in $3' \rightarrow 5'$ direction while the other one has to be aligned in $5' \rightarrow 3'$ direction. For the DNA pyramid there are only two circular paths, of which one is chosen as the initiation point and is shown in Figure 1.7.

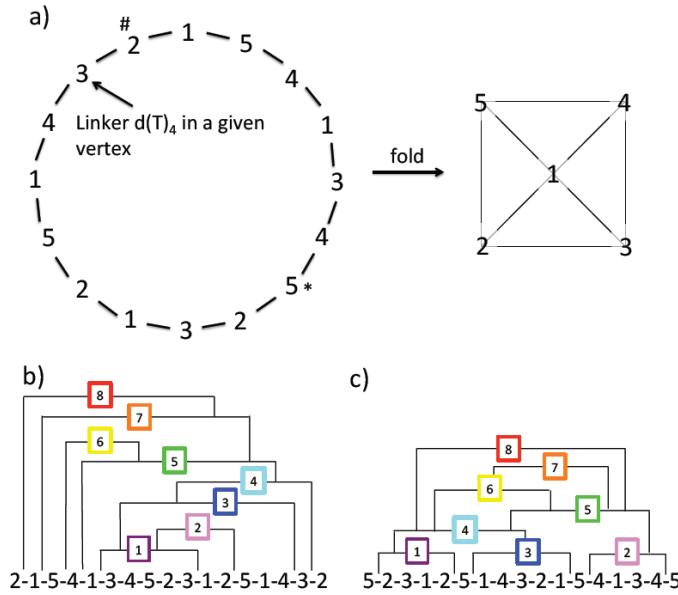


Figure 1.7 – Topological properties of the optimized and kinetically hindered DNA pyramid folding pattern. (a) Circular pattern of the DNA sequence and the corresponding fold (b) and (c) show the geometrical ordering of all domains; two complementary strand regions are connected by a bridge. The number next to each dash represent the corners in between which the domain is located (for the folded geometry). The numbers on each bridge connecting two dashes represent the order of double helix formation hypothesized by the experimentalists.

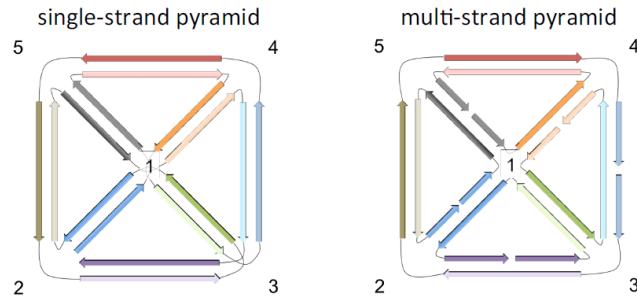


Figure 1.8 – (Left) Illustration of the ssDNA path through the structure. The strands must be antiparallel at every edge. (Right) Multi stranded assembly geometry.

The numbers in Figure 1.7 and Figure 1.8 represent the edges of the pyramid as well as the $d(T)_4$ linkers, a region of "TTTT" that is not involved in any bonding, but allows the rigid double helices to be connected to a pyramidal structure. The black solid lines in Figure 1.7 and coloured arrows in Figure 1.8 represent the folding domains. Two single

INTRODUCTION

stranded variants can be chosen to select and maximise certain topological effects. The "optimized" version, labelled OPT, is obtained by cutting the circular pattern by leaving out the linker 2#, which is shown in Figure 1.7(b). The cyclic permutation in Figure 1.7(c) is created by omitting the linker 5* and was named "kinetically hindered", labelled KH. Both designs shall be addressed in the following as OPT and KH respectively.

The domains posses different content of G-C bonds and are therefore expected to form at different temperatures. Gradually lowering the temperature from some temperature T higher than the domains' melting points, the domain labelled 1 is suspected to form first, while domain 8 is expected to form last. Throughout the thesis specific domains shall be addressed by their prospected folding order.

The experimentalists designed the two DNA pyramids to behave differently during folding and melting of the structures. The OPT version was designed to unfold from the strand ends during a melting process, while the KH version was designed to show a concerted melting behaviour. In the DP each domain can be seen as a hairpin, whose loops consist of other domains and $d(T)_4$ linkers. This makes the entire DP essentially one giant pseudoknot. Folding a pseudoknot of this size is highly complicated. In Figure 1.7 one can see how for the OPT version the smaller hairpins in the middle of the sequence are projected to form first and the hairpins are expected to form afterwards from the middle of the DNA sequence to the ends of the sequence. When gradually and slowly lowering the temperature, all domains are expected to fold just below their melting points.

For the KH design not only was another folding pattern used, but also the DNA sequence was changed. Further, the domains of highest melting point are not located in the middle of the sequence, but rather at the strand ends. During a thermal annealing process the domains at the strand ends with the highest G-C content are expected to fold first, making it physically very difficult for domains of lower G-C content to form successfully. The domains may form when the temperature becomes low enough, but stage-wise folding during the annealing process would make it difficult for the structure to form.

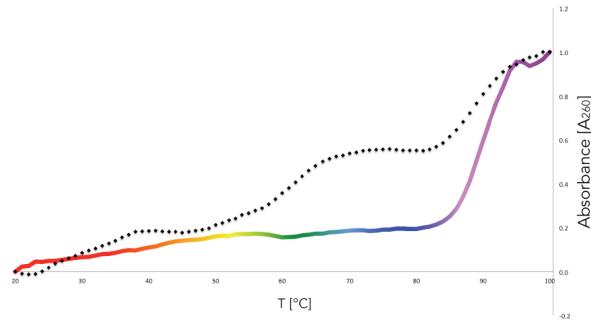


Figure 1.9 – Thermal annealing profile of KH (black dotted) and OPT (coloured) design obtained by measuring the hyperchromic effect of both designs. A folded geometry will show less absorbance than an unfolded geometry. Taken with permission from Kocar Vid.

In Figure 1.9 the hyperchromic effect of the optimized (coloured) and kinetically hindered (black dotted) DNA pyramid during thermal annealing is displayed. The graphic was provided by our collaborators and a high absorbance value corresponds to an unfolded geometry, whereas a small absorbance value corresponds to a folded geometry. However, the data is preliminary and does not give conclusive evidence on the geometry of the folded structure. The Figure 1.9 illustrates how on the one hand the peeling version is likely to form without greater complications. Only one small temperature range is visible for the transition from unfolded to folded structure. On the other hand the thermal annealing profile of the KH version shows multiple, broad transition regions. The experimentalists suggested that the KH design would have to undergo multiple structural rearrangements to reach the final, folded geometry.

2 Theoretical background

2.1 Computational methodology overview

In this Chapter an overview of molecular modelling and the DNA model employed shall be given. The DNA model, called oxDNA, was used to study the tetrahedral structure of Vid *et al.*. The oxDNA model was developed by Ouldridge *et. al.*. Its advantages and limitations will be highlighted in the following sections. The oxDNA can model will be compared with experimental properties of DNA. Lasst, an introduction to the sampling techniques and dynamics employed in the study will be given.

2.1.1 Molecular Modelling of DNA

Molecular Modelling techniques aim to describe the properties and dynamics of biological, chemical and physical systems on a microscopic length scale. Also, molecular modelling permits the investigation of systems on time scales up to milliseconds. Understanding the microscopic processes from macroscopic experimental observations is not always possible and drawing conclusions can be difficult. Often experimental predictions are in need of assessment by computational models for a microscopic explanation of the observed measurements. Furthermore, molecular modelling allows for the manipulation of an experiment in a very artificial way, in order to find out which forces are responsible for the observed micro-dynamics. The ease with which the modelling conditions can be changed is a lot greater than within experimental capabilities. As experimental capabilities rise, the accessible amount of data on DNA-based systems and the experimental accessibility of nanoscale systems is rapidly increasing. Therefore, the need for computational modelling solutions to study the self-assembly of DNA nanostructures is greater than ever.

In the fast growing field of DNA nanotechnology there is an increasing need for simulation models to fully understand the underlying physics. In the past, a variety of different models have been introduced to the field,^{[29][30]} ranging from all-atom models, such as AMBER^a[31] and CHARMM^b[32], and intermediate coarse-grained models, such as the one by Santa Lucia,^[33] and NUPACK,^c[34] which is based on the Santa Lucia model, to worm-like chain models.^[35] The main drawback of fully-atomistic models is that they are

^aAMBER: Assisted Model Building with Energy Refinement

^bChemistry at HARvard Macromolecular Modelling

^cNUPACK: NUclei acid PACKage

currently unable to simulate large systems over relevant time scales, while simple models often cannot provide the necessary insight.

NUPACK is another simple model for analysing the melting of DNA double helices. The NUPACK software was used to estimate the melting range of individual duplexes relative to a reference state, and can be found under reference 25. Here, the NUPACK software provided an easy-to-use sampling method for estimating the melting temperatures of duplexes and hairpins, rather than using Umbrella Sampling, which can be computationally expensive.^d However, NUPACK is unable to predict the pyramidal structure, but oxDNA can investigate pseudoknots and does well with predicting the melting temperatures of hairpins and duplexes. Also, oxDNA can characterize the structure of various nanostructures on much more relevant time scales up to milliseconds.^[36] A month-long simulation of a system of 7000 nucleotides on a single GPU yields a 55 microseconds trajectory, using a time step of 9 femtoseconds.

2.1.2 The oxDNA model

For our system a model is needed that can model single-stranded (ssDNA) and double-stranded DNA (dsDNA) well. Also the model needs to be able to simulate the folding process of ssDNA and it needs to properly characterise the thermodynamics of melting. Further, bending at junctions and at unbound base pairs must be possible in order to allow the folding of the tetrahedron without creating very large stress inside the system. Of course it is important that the model provides a good three-dimensional structure of our DNA nanostructure. The model needs to be able to simulate relevant time scales for large systems, that usually require a lot of simulation time. There are a great number of coarse-grained models, that could theoretically be used to study DNA polyhedra^{[37] [38] [39] [40] [41] [42] [43] [44] [45]} and the oxDNA model by Ouldridge *et al.* is chosen, because it seems the most correct for our application.

The oxDNA model is a coarse-grained representation at the nucleotide level that uses a "top-down" approach to describe DNA. Top-down approach means that the model is parametrised to fit macroscopic properties, rather than microscopic properties, which is the case for bottom-up approaches. The melting behaviour has been fitted to the results of melting short duplexes in the Santa Lucia model.^[33] The model was parametrized to fit the stacking transitions of ssDNA^[46] and the system mechanics were fitted to the persistence lengths of ssDNA and dsDNA, while and the elastic properties of dsDNA. The oxDNA model was first published in 2010 and has been reworked and extended since then. Up to this date oxDNA supports a variety of features, which will be the topic of the next section. On the oxDNA homepage^e the model is well explained and also freely available for download. The oxDNA model has recently been applied to a

^dUmbrella sampling can be applied where sampling of the entire energy landscape is hindered by large energy barriers

^ehttps://dna.physics.ox.ac.uk/index.php/Main_Page

THEORETICAL BACKGROUND

variety of different nano-technological systems,^[36] such as DNA nanotweezers^[47] and DNA nanowalkers.^{[48][49]} Furthermore, there have been studies on the hybridization kinetics^[50] and the sequence dependence of the model's thermodynamics.^[51]

2.1.3 Representation of DNA in oxDNA

The physical DNA potential is a summation of various interactions, which are illustrated in Figure 2.1. Each nucleotide in oxDNA is taken as a single rigid body where the sugar-phosphate backbone is treated as one interaction site, and another site describes the nucleobase. The ellipsoidal form represents the planar orientation of nucleotides in the ladder-like double helix structure. In the oxDNA model the orientational dependence of interactions allows for the use of only one simulation site for the backbone and one for the nucleotide. In respect to other models,^[44] which usually use more than one site representation for the backbone, this significantly decreases the computational effort needed. Because the simulated DNA pyramid structure does not have spatial areas in which backbone proximity interactions are of influence, the representation in the oxDNA model is extremely suitable for this project.

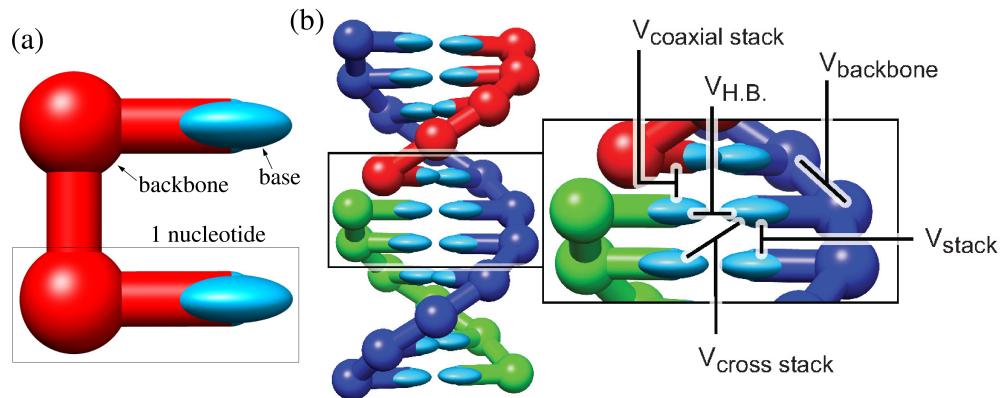


Figure 2.1 – The coarse-grained oxDNA model of DNA (a) A visualisation of the rigid nucleotides that make up the basic unit of oxDNA. (b) An 11 base-pair DNA duplex is visualised using the Chimera software. The backbone sites on three strands are represented as *red*, *green* and *blue* spheres respectively and are connected with tubes to neighbouring backbone sites of the same strand, and the bases are represented by disks. The forces are represented by abbreviations and their interacting sites indicated by connecting lines. Annotated are, $V_{b.b.}$, the interactions between neighbouring backbone sites, V_{stack} , the interactions between neighbouring bases and $V_{c.stack}$ and $V_{H.B.}$, interactions between bases of complementary strands. Not explicitly shown, but included in each particle are isotropic excluded volume interaction V_{exc} . Taken with permission from reference 62.

The oxDNA potential is,

$$V_{oxDNA} = \sum_{\langle ij \rangle} (V_{bb} + V_{stack} + V'_{exc}) + \sum_{i,j \neq \langle ij \rangle} (V_{HB} + V_{cr.st.} + V_{exc} + V_{cx.st.}), \quad (2.1.1)$$

where the first sum is over all sites that are nearest neighbours along the backbone and the second sum is over all non-nearest neighbour sites.

Physical terms 1. A backbone interaction V_{bb} is a covalent bonds, and modelled using FENE^f springs.

2. The stacking interaction between parallel oriented nucleotides adjacent to each other V_{stack} are represented by a SCOMP^g. The well minimum is set to be at 3.4 Å
3. Coaxial stacking interactions $V_{cx.st.}$ behave the same as stacking interactions, however are only applied to nucleotides that are not directly connected via the backbone. These interactions will not need to be applied for the single-strand simulation at hand.
4. Cross-stacking Terms $V_{cr.st.}$ between diagonally adjacent nucleotide sites are accounted for using smoothly-cut off quadratic wells.
5. complementary base hydrogen bond interactions V_{HB} in each rung of the double helix are given by SCOMPS.
6. Excluded volume interactions V_{exc} are simulated by a repulsive Lennard-Jones potential.

(1.) In oxDNA covalent bonds link the backbone sugar molecules with each other through a phosphate group. (2.) The stacking interactions can exist between neighbouring bases on the same sequence, or (3.) any two bases that are not nearest neighbours. The later are called coaxial stacking interactions. Stacking and coaxial stacking interactions incorporate a number of effects including π -orbital interactions between the aromatic sites of stacked nucleotides. (4.) The cross-stacking interactions are approximate for all non-nearest neighbour interactions in the double helix, especially those of charges and dipoles. (5.) The hydrogen bonding terms in an average-base model represent the hydrogen bonding interactions in biological DNA. (6.) The excluded volume terms practically account for the repulsions due to overlap and charge-charge repulsion. An in-depth description of those can be found in the PhD Thesis of Thomas Ouldridge^[52] or in Appendix A of reference 45.

Because this thesis mainly employs a temperature dependent method it is important to note how temperature dependence is handled within the system. In the oxDNA model temperature dependence is incorporated in the stacking interactions in order to

^fFENE: Finitely-Extensible Nonlinear elastic

^gSCOMP:Smoothly Cut-Off Morse Potential

simulate the geometric sources of entropy. These interactions also account for temperature dependent DNA-solvent interactions.

2.1.4 Limitations of oxDNA

When applying the oxDNA model it is also important to note its limitations to be able to evaluate which predictions are physically justified. oxDNA is currently only parametrized to a specific salt concentration of 500 mM. At this high salt concentration positive repulsion of DNA is suppressed, which allows for incorporation of electrostatic interactions into excluded volume interactions. The sequence dependence of oxDNA solely changes the relative strengths of G-C and A-T bonds and there is no difference between the sizes or relative positions of the nucleotides.

In this thesis both a sequence-dependent and an average-base-pair model will be used. The average-base-pair model assumes that every base-pairing is of the same strength, while equally only allowing Watson-Crick base pairing (C-G, A-T). This model will be employed to study the differences between the optimized and kinetically hindered designs of the DNA nanopyramid. Having the average-base model is not redundant, because it allows for the distinction between effects induced by structural differences and effects due to changes in the domain strength (changing the G-C content of each domain). Later on, application of the sequence-dependent model allows for a direct comparison with experimental results and gives further insight into the exact melting process and the folding design of both structures.

2.2 Simulation background

In the course of this project Metropolis Monte Carlo as well as Brownian Molecular dynamics are used, as these constitute the basics of molecular simulations. Simulations are carried out on a microscopic scale whereas experimental observations are based on macroscopic measurements. Macroscopic observables are statistically linked to their microscopic observables by the *Ergodic hypothesis*.^[53] The ergodic hypothesis, originally described by Boltzmann in his lectures on "gas theory", states that all thermodynamic systems behave randomly and thus for sufficiently long time periods, for almost all variables, the average value over time $\langle A \rangle_{N,V,T}$ is the same as the average value over a number of ensembles $A(\bar{t})$. So one has

$$A(\bar{t}) \stackrel{!}{=} \langle A \rangle_{N,V,T}, \quad (2.2.1)$$

where the ensemble and time average are defined as

$$A(\bar{t}) = \lim_{t' \rightarrow \infty} \int_0^{t'} A[r^N(t), p^N(t)] dt \quad (2.2.2)$$

and

$$\langle A \rangle_{N,V,T} = \frac{\int \int_{N,V,T} A[r^N, p^N] dr^N dp^N}{\int \int_{N,V,T} dr^N dp^N} \quad (2.2.3)$$

respectively, where $\{r^N\}$ are the Cartesian coordinates of the N particles in the system and $\{p^N\}$ are the particle's momenta. The subset N,V,T stands for a number of particles, volume and temperature consistent ensemble, also known as the canonical ensemble. In the canonical ensemble both Monte Carlo and Molecular Dynamics simulations can be used to obtain observables.^[54] Both simulation types can be used within the oxDNA software and are applied according to their specific advantages and disadvantages. Both methods will now be addressed in turn.

2.2.1 Monte Carlo Simulations

Monte Carlo is generally a random number generator based method, that uses the principle of ergodicity to calculate equilibria. The Monte Carlo algorithm may sample a large number of microstates and derives observables by integration over the system replica. The Monte Carlo algorithm is a special stochastically based integration method suitable for solving integrals of high dimensions. This property is essential to solving averages determined by classical statistical mechanics. An observable A has an average value given by

$$\langle A \rangle_{NVT} = \frac{\int A(\mathbf{r}^N, \mathbf{p}^N) \exp(-\beta E) d\mathbf{r}^N d\mathbf{p}^N}{Q(N, V, T)}, \quad (2.2.4)$$

where E is the energy of a microstate and the denominator is the canonical partition function. However to calculate averages using the resulting distribution requires prior knowledge of the partition function. This predicament can be overcome using a method introduced by Metropolis, which uses a random walk through the configuration space - also called a Markov chain of configurations-without memory of the previous history of moves. In the Metropolis algorithm it is not necessary to know the absolute probability of visiting a microstate, but the ratio of probabilities of final and initial states. The basic Metropolis Monte Carlo (MMC) approach is:

1. start with initial configuration o ;
2. Create a new configuration n , by applying a trial move to the system;
3. Accept or reject the move with some transition probability $\pi(o \rightarrow n)$;
4. Calculate properties of interest for the current configuration and update the average;
5. Repeat steps 2-4 until a sufficiently large trajectory is acquired.

Before the application criterion of the Metropolis scheme is discussed a few simplifications can be made to the above equations. Firstly, the integral over the momenta can be carried out, as these are generally analytical and the integral evaluation does not explicitly depend on these coordinates. The partition function can be rewritten as 2.2.5.

THEORETICAL BACKGROUND

$$Q(N, V, T) = \frac{1}{N! \Lambda^{3N}} Z(N, V, Z), \quad (2.2.5)$$

where $\Lambda = h/\sqrt{2\pi k_B T}$ is the De Broglie thermal wavelength and Z the configurational partition function, given as

$$Z(N, V, T) = \int \exp[-\beta V(r^N)] dr^N. \quad (2.2.6)$$

Secondly, in order to be able to sample states from the Boltzmann distribution it is necessary that this distribution remains stationary. Therefore, the net probability for a transition has to be the same in both directions of a state change. The product of probability of being in the initial state $P(o)$ with the transition probability has to be the same as the product of final state probability and reverse transition probability, a principle called "detailed balance". The transition probability can subsequently be broken down into two contributions, $\alpha(o \rightarrow n)$, the probability of initiating a trial move from o to n , and $acc(o \rightarrow n)$, the probability that a trial move is accepted. As already used by Metropolis, the move probability matrix α is taken to be symmetrical, therefore for the acceptance ratio equation 2.2.7 must be obeyed:

$$\frac{\pi(o \rightarrow n)}{\pi(n \rightarrow o)} = \frac{acc(o \rightarrow n)}{acc(n \rightarrow o)} = \exp[-\beta(E_n - E_o)]. \quad (2.2.7)$$

The acceptance criterion originally used by Metropolis (only one of many solutions) is thus

$$acc(o \rightarrow n) = \begin{cases} \exp[-\beta(E_n - E_o)] & \text{if } E_n > E_o \\ 1 & \text{if } E_n \leq E_o. \end{cases} \quad (2.2.8)$$

All moves that decrease the system's energy are accepted and all moves that increase the system's energy are accepted with a probability decreasing with the energy difference between initial and final state. If the trial move is rejected the original state o is counted in the ensemble average calculation again and a new trial move is initiated. MC calculations are easily implemented and most suitable for simple single particle movement, such as translation or rotation, but do not do a good job of sampling large free energy differences between different minima.

Finally choosing the maximum allowed displacement is important. On the one hand, if the maximum displacement is too small then most moves are accepted, but the phase space is sampled too slowly (consecutive states are highly correlated, the time span and number of configurations necessary to generate a correct average becomes very large). On the other hand if the maximum displacement is too large then most of the trial moves are rejected and important configurations in the phase space might be completely omitted. Thus a heuristic acceptance ratio of about 50% is usually chosen.

2.2.2 Collective particle motion methods

As highlighted in the previous section, the simple Monte Carlo algorithm uses single-particle moves. However, for folding and melting processes a lot of diffuse motion is simulated. Due to the random nature of single-particle moves a lot of moves that involve strongly interacting nucleotides as well as backbone sites would be rejected without adding any information to the calculation. Therefore, simple MC is not computationally efficient for the simulation of large systems. Because single moves on a large system would take a long time to sample, a collective particle Monte Carlo method has to be employed. In oxDNA two types of algorithms for moving clusters of nucleotides have been implemented, the Cluster move Monte Carlo (CMMC) and Virtual-Move Monte Carlo (VMMC). The CMMC algorithm is very fast and simple,^[55] whereas the VMMC algorithm is more precise. The VMMC algorithm, introduced by Whitelam and Geissler,^[56] tries to identify particles that shall be moved in concert and selects clusters according to given criteria, based on potential energy gradients. The newest version^[57] of the VMMC code by Whitelam and Geissler is implemented in oxDNA and has successfully been applied in the oxDNA framework.^[36] Also, in oxDNA the proposed trajectories of clusters is stored in addition to their current trajectory. In the next section the VMMC algorithm used by oxDNA is discussed.

2.2.3 Virtual Move Monte Carlo algorithm

For every cluster move there has to be a "seed" particle from which neighbouring particles are added probabilistically. A detailed description of the selection algorithm can be found in most recent version of reference 56. In oxDNA the seed move used is:

1. Rotation of a nucleotide about its backbone site, with rotational vector axes uniformly taken from the unit sphere and the angle taken from simple Gaussian distribution with σ of 0.22 radians and μ of zero.
2. Translation of a nucleotide, where the displacement is taken out of another Gaussian distribution with σ of 0.15 simulation units of length (0.1277 nm) and μ of zero for each Cartesian axis.
3. If the "pseudocluster" after the seed initiation is larger than 7 particles, the corresponding move is automatically rejected, to speed up the calculation.

In oxDNA detailed balance is obeyed and the acceptance criterion is given by:

$$acc(o \rightarrow n) = D_R(o \rightarrow n) \min\left\{1, \prod_{i,j}^{k,l} \exp(-\beta(\varepsilon_{i,j}^n - \varepsilon_{i,j}^o))\right\}, \quad (2.2.9)$$

where $D_R(o \rightarrow n)$ is a factor which is zero if there are any partial links between particles inside and outside of the cluster and unity otherwise. The product $\prod_{i,j}^{k,l}$ is taken over all

THEORETICAL BACKGROUND

pairs of particles which are non-interacting in state o and have positive energy in state n and vice versa. $\varepsilon_{i,j}^n$ and $\varepsilon_{i,j}^o$ are pairwise energies between particles i and j in the new and initial configuration respectively.

A visual representation of equation 2.2.9 is given in Figure 2.2.

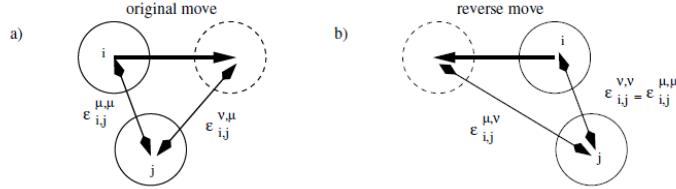


Figure 2.2 – Cluster building in oxDNA using the VMMC algorithm. Particle i is randomly selected and a random move is initiated. (a) Particle j interact with particle i and the pairwise energies for i and j are calculated before and after the move. A particle link is formed between i and j with a probability given in equation 2.2.10. The same procedure is carried out for the reverse operation (b). If a link between i and j is formed in (a) and (b), the particle is added to the cluster. This procedure is iterated for all particles that interact inside the cluster. Taken with permission from reference 54.

In oxDNA links with all particles, with which the seed particle interacts, are initiated with the probability:

$$P_{link}(i, j) = \max(0, 1 - \exp(-\beta(\varepsilon_{i,j}^n - \varepsilon_{i,j}^o))). \quad (2.2.10)$$

Using these cluster move techniques possesses a few drawbacks. Firstly, reproducing physical diffusion rates is very difficult, especially because moves that would normally be accepted are rejected in cluster-move algorithms. Secondly, with VMMC it is difficult to compare simulation times of differently sized systems. Both these points are only of small matter to this work, because only one ssDNA molecules is studied and there is no desire to capture physical diffusion rates. For the folding process large VMMC moves might also cause trouble, as these could introduce topological frustration to the structure. The simulation is also slowed down a lot, because the 284-nucleotide system at hand is too large to be sampled efficiently. Therefore, it is advisable to also consider the application of Molecular Dynamics simulations, which will be done in the next section. MD simulations are generally able to be parallelized, which makes MD usually a lot more suitable to describe DNA dynamics of large systems.

2.2.4 Molecular Dynamics Simulations

Molecular dynamics is a deterministic method that employs computational techniques to study the time dependent behaviour of a many-body molecular system. Other than MC simulations Molecular Dynamics (MD) simulations also contain information about the transport properties of the system. In MD Newton's equation of motion are solved iteratively starting from an initial configuration Γ .^[58] The MD simulation algorithm follows the scheme:

1. Start with the initial configuration Γ_0 by setting the initial positions $\{r^N(t = 0)\}$ and initial velocities $\{v^N(t = 0)\}$;
2. Calculate all forces for each particle from the potential energy function $V(r^N)$

$$\mathbf{F}_i(t) = -\nabla_{\mathbf{r}_i^N}[V(\mathbf{r})], \quad (2.2.11)$$

$$V(\mathbf{r}) = \sum_i^N \sum_{j>i}^N V(\mathbf{r}_i, \mathbf{r}_j), \quad (2.2.12)$$

where ∇_{r_i} is the gradient of particle i ;

3. Integrate equations of motion, the most basic choice being Newton's equation of motion below, from t to $t = t_0 + \Delta t$ in order to obtain new positions and velocities:

$$\mathbf{F}_i(t) = m_i \times \frac{d\mathbf{v}_i(t)}{dt}, \quad (2.2.13)$$

$$\mathbf{v}_i(t) = \frac{d\mathbf{r}_i(t)}{dt}, \quad (2.2.14)$$

4. Reiterate steps 2-3 until a trajectory of sufficient size is generated

This iteration can be stopped as soon as enough data for the measurement of the wanted macroscopic entity is completed. Typically a certain number of iterations is pre-set. After this equilibration period, the time average \bar{A} for the microscopic function $A(\Gamma)$ can be obtained from M discrete points of the MD phase trajectory, rather than from calculating a complicated integral time average, simply by solving

$$\bar{A} = \frac{1}{M} \sum_{m=1}^M A(\Gamma(t_m)), \quad (2.2.15)$$

where $\Gamma(t_m)$ is the ensemble representation at a given point m and Δt is the simulation time step.

To apply this scheme to any MD calculation, the correct time step for integration has to be chosen, as it is a key element to every MD simulation. When setting up an algorithm it is important to weigh up simulation accuracy and computation time. On the one hand, a

THEORETICAL BACKGROUND

time step that is too long will yield a less accurate representation of physical time through computation time, because the errors of integration scale as $\mathcal{O}(\Delta^n)$. Here n is the lower limit number of integrations that need to be retained to give physical meaning to any MD results. On the other hand time steps that are too small may cause the calculation to be too slow and computationally expensive.

For the first initialization step, the simulation box size is fixed and the particles are initialized in such a way, that the associated Boltzmann weight is non-zero and the total momentum of all particles is zero. For every specific temperature the particle velocities are rescaled according to their kinetic energy representation, that is

$$E_{kin} = \sum_{i=1}^N \frac{m\mathbf{v}_i^2}{2} = 3Nk_B T. \quad (2.2.16)$$

For the oxDNA model another constraint is imposed, being that the nearest neighbour backbone inter-site distance on the same single-strand may not exceed a certain defined value. If this distance is to large, the calculation immediately fails.

For the second step of the algorithm, the set of equations of motion (EOMs) has to be chosen appropriately. In the oxDNA model the solvent molecules are coarse-grained away and therefore have to be considered. Solvent influence is simulated using a solvent noise and a damping term. In this work a Brownian algorithm is used, for which solely a random direct force $\mathbf{R}_i(t)$ has to be added to the inter-particle force $\mathbf{F}_{ip,i}(t)$ on each particle.

$$\mathbf{F}_i(t) = \mathbf{F}_{ip,i}(t) + \mathbf{R}_i(t) \quad (2.2.17)$$

In the third step of the algorithm the Velocity Verlet algorithm is used. The Verlet integration scheme is carried out using Taylor expansions of the particle positions around $t + \Delta t$ and $t - \Delta t$:^[59]

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{\Delta t^2}{2m_i}\mathbf{f}_i(t) + \frac{\ddot{\mathbf{r}}_i(t)\Delta t^3}{6} + O(\Delta t^4) \quad (2.2.18)$$

$$\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \mathbf{v}_i(t)\Delta t + \frac{[\delta t]^2}{2m_i}\mathbf{f}_i(t) - \frac{\ddot{\mathbf{r}}_i(t)\Delta t^3}{6} + O(\Delta t^4) \quad (2.2.19)$$

where \mathbf{f}_i is the force on particle i and the velocities are $\mathbf{v}_i = \dot{\mathbf{r}}$. Adding these two equations together gives

$$\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) + \frac{[\delta t]^2}{2m_i}\mathbf{f}_i(t) + O(\Delta t^4). \quad (2.2.20)$$

Equation 2.2.21 can be rewritten as

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{r}_i(t - \Delta t) + \frac{[\delta t]^2}{2m_i}\mathbf{f}_i(t) + O(\Delta t^4), \quad (2.2.21)$$

which is known as the Verlet integration scheme. The velocity 2.2.22 and the updated velocity 2.2.23 can be derived from the verlet scheme.^[60]

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t)}{2\Delta t} + O(\Delta t^2) \quad (2.2.22)$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{[\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)]}{2m_i} \times \Delta t \quad (2.2.23)$$

The Velocity Verlet scheme is used in the Brownian MD simulations from which trajectories are generated in oxDNA. In comparison to the VMMC algorithm introduced earlier, MD simulates DNA dynamics in a very simple fashion.

2.2.5 Temperature control and kinetics

In oxDNA the solvent is not taken into account explicitly, rather an implicit method called an Anderson-like thermostat that tries to mimic solvent effects is used.^[61] With an Anderson-like thermostat the velocities of random particles are drawn from the Boltzmann distribution corresponding to the simulation temperature. This behaviour mimics random collisions with a heat bath. The equations of motion are integrated by Verlet integration with a discrete time step δt . Using this simple thermostat, the positions, velocities, angular velocities and orientations of the nucleotides would be recalculated at each time step. However, in reality DNA undergoes Brownian motion because of DNA-solvent interactions. To model Brownian dynamics the velocity is recalculated with a probability of $p_v = 0.02$ and the angular velocities with a probability of $p_v = 0.0067$ every 103 time steps. The integration time step was chosen as $\delta t = 1.52 \times 10^{-14}s$ for all MD simulations. This gives a diffusive constant $D_{Sim} = 1.19 \times 10^{-8} m^2 s^{-1}$ ^[62] that is about 100 times higher than experimental measurements.^[63] Higher diffusion constants can be used to speed up diffusion. Accelerated diffusion can smooth out energy profiles on a macroscopic scale, which leads to acceleration of certain processes. Faster processes can possibly allow for the application of oxDNA to more complicated systems. The time step of $\delta t = 1.52 \times 10^{-14}s$ cannot be used to compare experimental rates and computational rates directly. It is important to note that the motion is artificially accelerated and that only relative rates can be compared to the experimental ones.

Summary

In this chapter the fundamental techniques of many-body computation as well as more in-depth methodologies currently applied in modern simulations have been introduced. These methodologies are the underlying essentials from which results can be obtained. In this thesis VMMC and MD calculations are extensively applied.

Also, for later repeated use in the Result section, a few terms have to be introduced. Firstly, the term "fully folded" structure shall refer to the global energy minimum of the pyramidal structure. Secondly, the term "topology problems" shall refer to a structure that is topologically frustrated and is practically unable to achieve the global energy minimum in a reasonable amount of simulation time, because escaping the misbonded state it is currently in would require more energy than is available to the system. Usually this barrier is presented by the formation of previous domains, which cannot be broken at the simulation temperature, because the energy of favourably rearranging the structure is much smaller than the energy necessary to break apart an already formed double-helix. The system is "stuck" in a local minimum and this configuration can only be overcome via a large temperature increase or very large trial move. Even if all domains are formed, the corner junctions of the pyramid may be topologically frustrated, which leads to distortion of the DNA pyramid's geometry. These distorted geometries will also be addressed as "topology problems" and will later be pointed out and explained in the Experiments and Results section.

3 Experiments and Results

In this section the structure and the folding, melting and annealing of a single-stranded DNA pyramid fold shall be investigated using Molecular Dynamics as well as Monte Carlo simulations. The aim of this section is to give an insight into the annealing and melting process of two possible isoforms for the tetrahedron and to find similarities and differences between the folds of these two structures. The obtained results shall be compared to experimental results to develop a better understanding of the measurements.

The results can be divided into three major stages upon which the simulations are compared to experiments. First, a simulation model of the DNA Pyramid (DP) has to be created that is a fully functional structure, i.e. there are no topology problems. This structure ideally corresponds to the global energy minimum for the DP at hand, because it has the maximum number of base pairs. Next, the whole structure has to be melted to gain information about the melting properties of the different domains. Furthermore, looking at the melting process of a complete pyramid structure can indicate how a folding cascade might look. Analysis of the melting process can also yield insight into the structural issues, if any, concerned with the structural design (e.g. by revealing topology problems). There are many types of topology problems and one of them is depicted in Figure 3.1. Last one has to look at the folding process of the DP. However, as folding involves many rare events, it requires a lot of computation time and there is no guarantee domain folding can be observed, let alone good statistics on folding can be obtained. Since the computational time granted within the framework of this thesis is limited, the full folding of the two given structures lies outside the possibilities of this work.

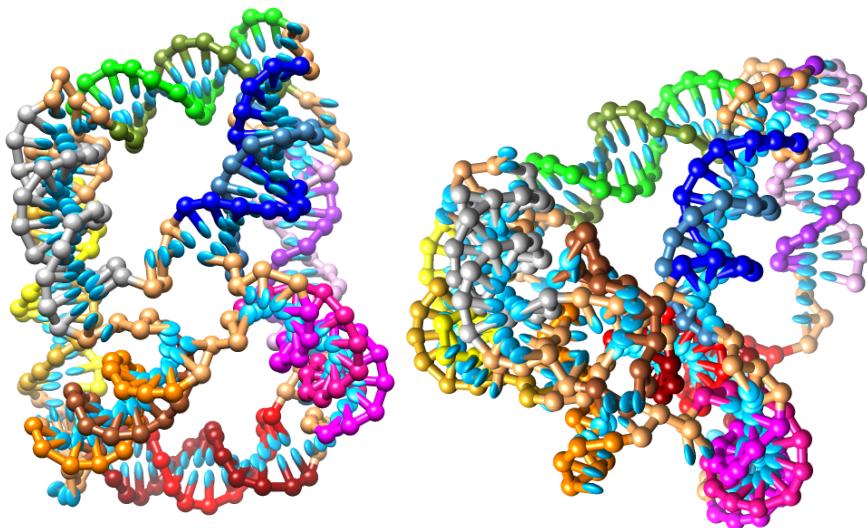


Figure 3.1 – Top view on the folded DNA Pyramid without (left) and with (Right) topology problems. In the pyramid on the right the orange domain is unfolded and encloses the yellow and grey domain.

3.1 Experiments

3.1.1 Folding

At the start of each melting simulation it is necessary to use a fully folded structure to base further analysis on. Within the computational possibilities, it was most suitable to artificially build the structure in some way. To achieve folding of the structure, calculations utilized so called “traps”. These traps consist of two spring forces between two particles. These particles can be chosen arbitrarily in the oxDNA model, but for the folding process they are inserted between nucleotides at the ends of two complementary regions. In order to prevent any misbonds and potential topological problems, a unique topology was generated that only allowed bonds in between complementary strands (i.e. only correct domains can be formed). Both restrictions presented above are unphysical however, the sole purpose of using traps is to obtain a fully folded DP, thus traps provide a more economical solution to the folding process. Of course, one has to be totally clear that the determined folding pathway is unphysical and cannot be compared to experimental data. Also, one has to be aware that the use of traps as such can also cause topological problems, but it was found that the economic advantages granted by the use of “traps” clearly outweigh the disadvantages.

The folding of the optimized sequence was initiated at a simulation temperature of 42°C, which is significantly below the melting temperature of each domain, to grant the system enough kinetic energy to allow for efficient folding, while equally maintaining a high

energetic advantage of a folded structure over a free, unbound geometry. First there was a number of replicas launched using VMMC. The folding attempts did not yield the desired structure, if no bonding restrictions other than forbidding misbonds were imposed. The system was observed to get stuck in energy minima, which did not allow for the folding of all domains. A different approach, which was later repeated with MD simulations, involves only allowing traps and bonds between a certain domain and all previously formed domains, also forbidding other native bonds. The domains were then folded in the order 1-8 as given in Figure 1.7.

To obtain the correct structure a topology was created that only allows bonds to be formed between nucleotides of domain 1.^a The corresponding traps were initiated, while all other bonds were prevented. After domain 1 was formed, a new topology that only allows bonds for domains 1 and 2 was created, and traps on domains 1-2 were introduced. Multiple replicas were run and after each iteration one configuration was chosen, that did not possess any topology problems. This configuration was then chosen as the initiation point for further folding simulations. This process was iterated until all domains were fully formed. Using this method a few different pyramidal structures were obtained using VMMC as well as MD simulations. Out of all formed DP, the one with the least number of topology problems was chosen. It was made sure that all domains were correctly formed and there were only topological problems at the DP corners. An example of a topologically frustrated corner is shown in Figure 3.2.

One can see from Figure 3.2 that the pyramid is distorted due to the topology problems at corner, where the orange and yellow domain are linked together. The domains are linked together with a flexible region of ssDNA with the sequence "TTTT", which will be referred to as $d(T_4)$ linkers. In order to generate the intended conformation of the DP, the topologically frustrated corners were identified. At these edges the structure was cut between the $d(T_4)$ linkers and one selected domain.

^aThe experimentalists suggested this domain would form first

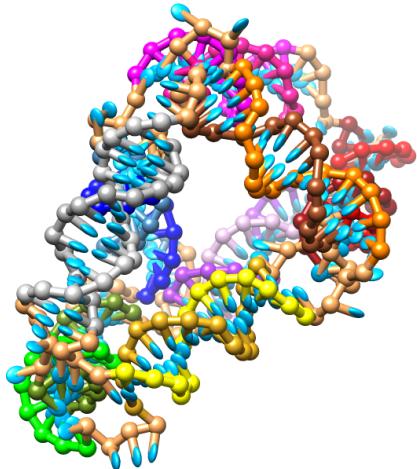


Figure 3.2 – Illustration of a topologically frustrated corner (orange and yellow domain cross section). The orange domain should connect to the corner from the inside of the DP, rather than from the side facing away from the center of the DP.

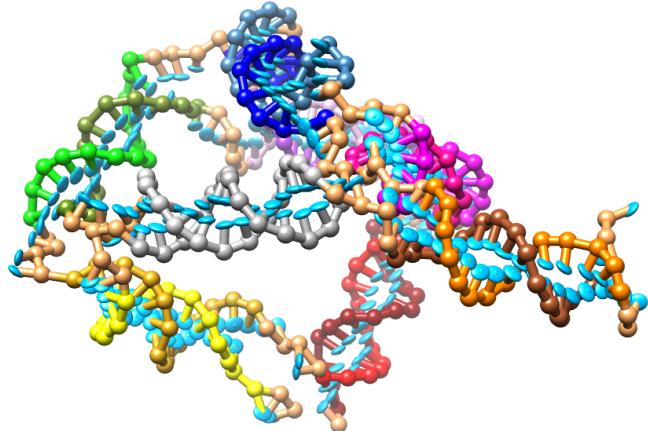


Figure 3.3 – Illustration of the DNA pyramid where the structure has been cut between the orange and yellow, and orange and red domains.

The structure was run until the corner had relaxed and then the cut regions were reconnected using traps. The final, topology-problem-free structure was then relaxed for 10^8 steps at 50°C without any external forces applied and the structure of lowest total energy was chosen as the starting geometry for commencing melting processes starting from 50°C .

In order to obtain the “kinetically hindered” version the “energetically favoured” DP was circularized, i.e. a new linker “TTTT” strand, was introduced and both ends of the ssDNA were connected using this linker. From this circularized DP, both the KH as well as the OPT version can be obtained by omitting a certain linker in graphic Figure 1.7. Linker 5* was omitted, to obtain the geometry of the KH version, as described in the DP section “DNA pyramid from ssDNA”. To fully realise the KH version, the new topology had to replace the old nucleotide information stored with the OPT version. This step was important to change for later analysis with sequence dependent version of the oxDNA model, in which the influence of the rearrangement of domains with different melting points was investigated.

3.1.2 Temperature induced Melting

From here on, all melting procedures were carried out using MD simulations because the simulations need to run on much longer time scales than the folding process and the use of GPU nodes was necessary. In the oxDNA model this is most efficiently realised using MD simulations. For the melting process a variety of different replicas were launched that all followed the same simple algorithm. The temperature was set at an initial temperature and gradually increased after a certain simulation time, until a specified final temperature was reached. At the start, all calculations were set running at 50°C , because the fully formed DP was relaxed at this temperature. Runs with different temperature steps and different steps per temperature were initiated. On the one hand, lower steps per temperature allow for faster screening of the melting temperature range, but can also cause superheating of the structure, a process in which the simulation temperature has significantly surpassed the optimum melting temperature ^b before strands had time to sample relevant states at that temperature. On the other hand, a large number of steps per temperature makes the calculation infeasible.

To investigate the melting process a variety of different set-ups were chosen. First, the melting range was screened with large temperature steps and using the average base model. The melting runs of equal time steps and equal design can further be divided into subcategories according to, if those were created using an average base/sequence dependent model and, if those were using a topology that did/did not allow misbonds. For the latter category only small changes in the melting behaviour were suspected and I had to determine if those are negligible. The average base model was utilized to obtain information on the influence of the order of the domains in the sequence on the DP’s

^bthe temperature at which the DP domains would melt with infinite equilibration time

melting behaviour. Also, the average-base model was used as a reference to determine which contributions to the melting behaviour arise from the DP's geometry and which arise from the different G-C content of the domains. The sequence dependent model is expected to provide different data, because the G-C and A-T base pairs are of different strengths. The sequence-dependent model was then also used to compare simulation data to experimental data. Using both the sequence-dependent and the average-base model allows one to distinguish between effects originating from the geometric design and the pyramid's sequencing.

3.2 Results

3.2.1 DNA Pyramid construction

As reported earlier, a fully formed kinetically hindered pyramid could only be obtained from the already formed OPT pyramidal structure. Trying to obtain the KH fully formed DNA pyramid always resulted in more or less strongly topologically frustrated structures. The first few domains fold with ease. However, later on the system becomes twisted and knotted. Therefore, it is unable to form further domains. In summary, the forced-folding simulation replicas only yielded partially formed pyramids. Those runs, that formed structures that involved bonds in all domains, did not show a pyramidal form.

Comparing MD simulation runs for the KH and OPT showed that the OPT version forms much more rapidly due to the different folding order. Also the folding process “gets stuck” (is unable to escape a local minimum) much more frequently for the KH version than for the OPT version. The domains that are suspected to form first ^c are located at the strand ends for the KH version, whereas these are located in the middle of the strand for the OPT version. This topological difference causes inflexibility in the KH structure. To form the KH version without structural rearrangements the higher order domains would have to have folded in an optimized geometry even before the most outer domains form. Because the domains of lower order are expected to form first, the structure would need to undergo rearrangements that require more energy than freely available to the system at a simulation temperature of 42°C. The starting geometries for the melting process are shown in Figure 3.4 and Figure 3.5.

^cderived from NUPACK data in section NUPACK data later

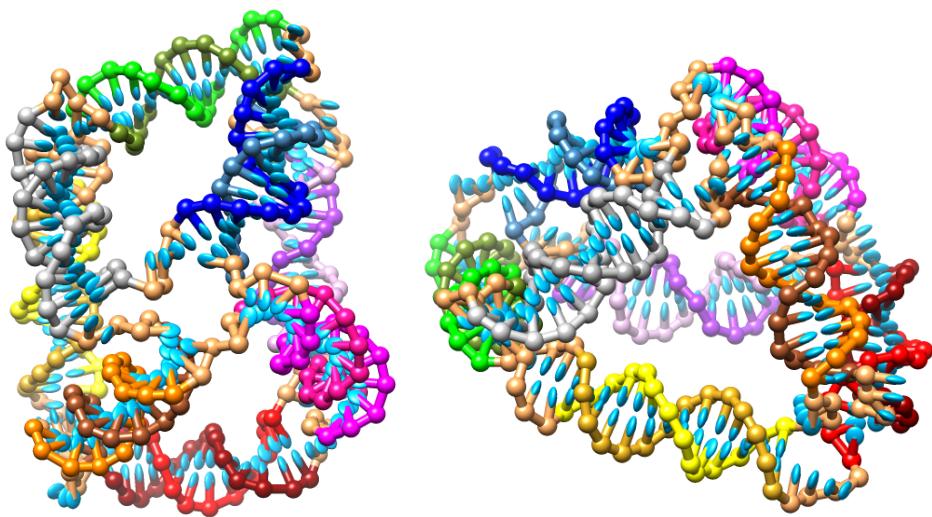


Figure 3.4 – Initialized DP geometry for the OPT design.(Left) View from the top and (Right) view from the side of the pyramid.

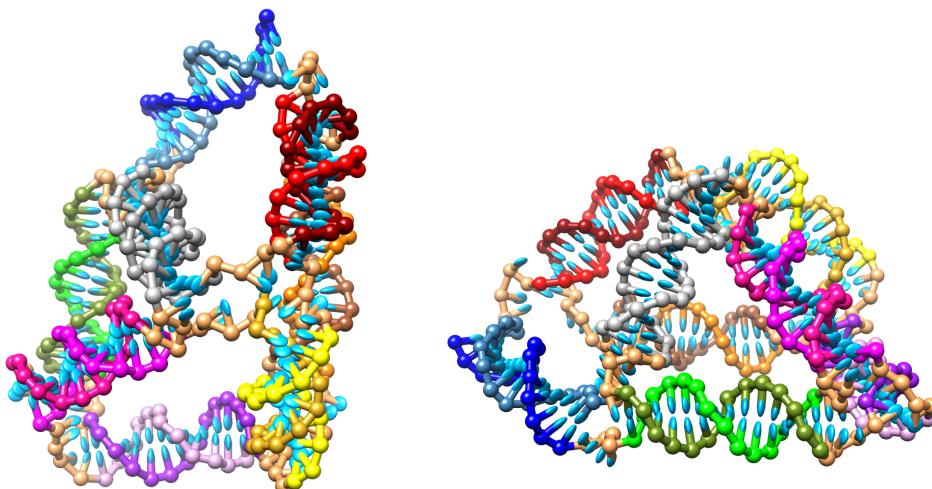


Figure 3.5 – Initialized DP geometry for the KH design.(Left) View from the top and (Right) view from the side of the pyramid.

3.2.2 NUPACK data

In order to get estimates of the domain melting temperatures and the order in which those domains should melt, the melting behaviour of the corresponding DNA hairpins was analysed for different loop lengths. To ensure the data gained is comparable only the

EXPERIMENTS AND RESULTS

two complementary strand regions for every domain are taken and connected by multi-T linker. The left graphic in Figure 3.6 below displays an example of an $d(T)_{18}$ -linker hairpin formed for domain Nr. 8. The right graphic displays the melting curve of the corresponding hairpin as predicted by NUPACK.

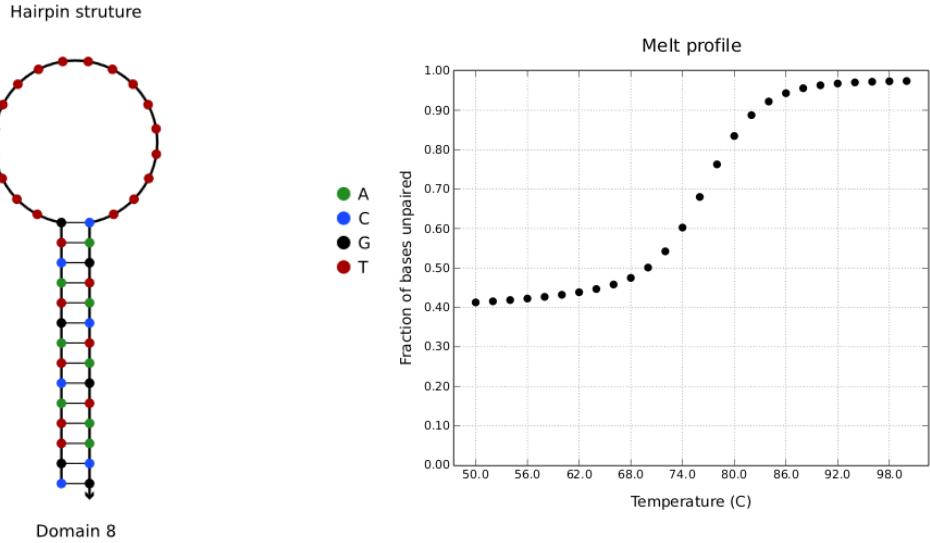


Figure 3.6 – (Left) NUPACK hairpin and (Right) visualisation of the temperature profile for Domain 8 of the optimized version.

Loop lengths of 18 and 36 were chosen to get a good upper and lower estimate of the expected domain melting temperature. The system loses tension and rigidity during the melting process and therefore a broader range of loop lengths has to be analysed. During the melting process the hairpin loop length is constantly changing, no two domains can be described by exactly the same hairpin model. Also, preformed domains in the pseudo-loop connecting two complementary strand regions affect the melting behaviour in an entirely different way than the free bases in the loop would. During the melting process the tension decreases and other previously formed duplexes will melt. Therefore, it is expected that with an increasing number of domains melted, an increasing loop size will be able to approximate the system best. For the OPT structure, this is generally expected to lead to a relatively increased melting temperature for the outer domains that melt first, and for a relatively lower melting temperature for those domains that melt last. The explicit effect on each domain could be predicted from annealing data. Also, the tension exerted on every domain develops in a complex fashion with the melting of previous domains.

In table Table 3.1 the average melting temperatures are given for variable multi-T loop lengths. The average melting temperature is defined here as the temperature at which exactly half of the maximum number of bonds in a domain are broken. The temperature is calculated by linearly extrapolating the melting curve data shown in Figure 3.6. The

raw data is given in Appendix B.

Table 3.1 – Melting Temperatures[°C] of DNA hairpins for OPT design as predicted by NUPACK; all A-T and all G-C data for reference

Domain	$d(T_{18})$	$d(T_{36})$	Avg.-base T_m	Seq.-dep. T_m
1	95.1	88.9	80.2	91.0
2	97.7	92.1	80.0	88.3
3	88.5	82.8	77.9	78.3
4	86.4	81.16	76.1	74.8
5	86.0	79.1	82.1	79.8
6	78.2	71.7	82.3	79.4
7	80.0	74.1	79.4	76.7
8	76.3	70.8	74.8	72.1
all A-T	58.0	50.9		
all G-C	109.0	103.3		

Overall the NUPACK data gives a good idea of the melting temperatures for each domain. Furthermore, this data is very useful for the prediction of the DP's domain melting order, because it constitutes a summary of the influence the G-C content has on a domain's melting behaviour. For the kinetically hindered design, the same loop lengths were investigated and the underlying argument is the same. However, the KH version is expected to melt near the melting temperatures of each domain. The domains are expected to melt in concert rather than stage-wise, like in the peeling version, because the parts of the structure that need to melt before other regions have the highest melting temperatures, by design. One can see that the range of melting temperatures predicted by NUPACK is very broad, whereas the observed melting range is very small.

Table 3.2 – Melting Temperature[°C] of DNA hairpins for KH design, as predicted by NUPACK

Domain	$d(T_{18})$	$d(T_{36})$	Seq.-dep. T_m
1	95.4	91.9	86.3
2	97.7	92.1	83.6
3	88.4	82.9	86.3
4	86.3	81.1	86.1
5	82.1	76.4	83.6
6	77.6	71.6	86.3
7	80.5	74.7	85.6
8	77.4	71.6	85.0

In the following, the observed differences between NUPACK data and simulation data will be discussed.

EXPERIMENTS AND RESULTS

3.2.3 Temperature induced Melting

As one can see the NUPACK melting temperatures from section "NUPACK data" do not match exactly the suspected reverse folding order of the domains given in Figure 1.7. Over the course of this section the experimentally suspected formation order will be compared to simulation results for the OPT version.

Two simulation factors have to be determined first. On the one hand choosing the correct time step is essential. On the other hand, obtaining an approximate melting range is necessary to judge the validity of the obtained melting data. A summary of eight average base replicas that all took 10^8 steps per temperature iteration and in which the temperature was increased by 5°C from 50°C to 100°C , is given on the left of Figure 3.7. A more detailed graph for a representative run is given in Appendix B. To get a closer estimate, another four replicas of the same calculation with an increased temperature sampling density around the previously discovered range were launched.^d The results can be seen on the lower graphic below. One simulation time step is equivalent to 1.02×10^{-14} seconds. The error bars show the standard deviation.

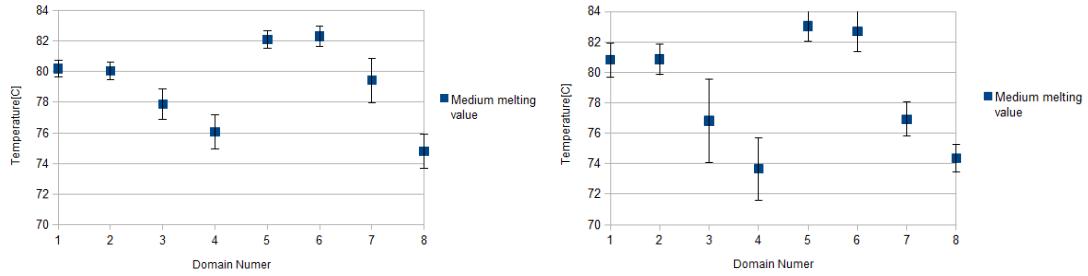


Figure 3.7 – Average melting temperatures for an avg.-base model of the OPT version. (Left) Analysis with 5°C intervals and (Right) with 2°C intervals over the temperature range where melting is observed.

^dThe temperature profile was altered, so 2°C temperature steps were initiated after the system temperature had reached 70°C rather than 5°C intervals.

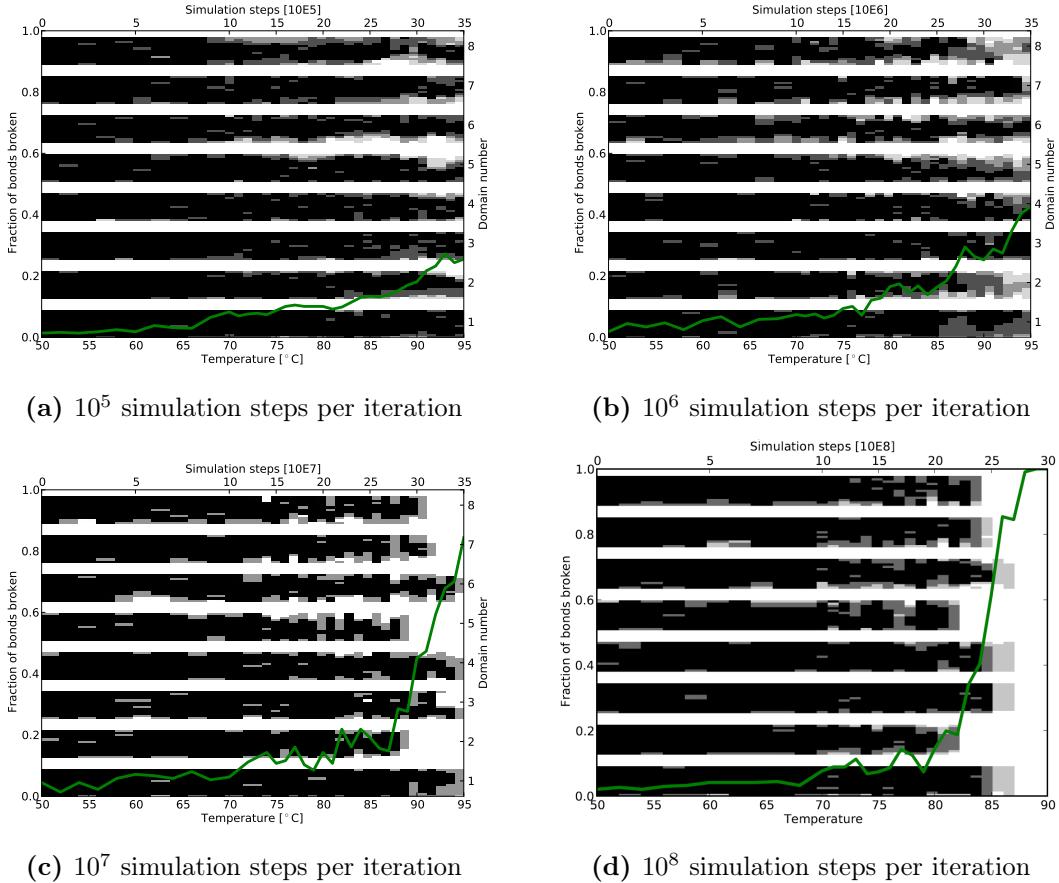


Figure 3.8 – Melting test runs for kinetically hindered version. The green line represents the percentage of bonds broken. The grey-scaled bars show, if a certain bond is still intact in the final configuration of each temperature step. Black corresponds to a formed bond in all replicas, whereas a white area shows zero bonds in all replicas. A grey-scaled result indicates there were replicas in which the bond was broken and some in which bonds were still intact. The domains are in order 1-8 from bottom to top and the white spaces in between represent the non-bonding $d(T)_4$ linkers.

From Figure 3.7 we estimate that further analysis should be focused around 65°C - 95°C . Furthermore the time step analysis indicated superheating, i.e. domains clearly not melting beyond their estimated melting point, for number of simulation steps of 10^5 , 10^6 , 10^7 and 10^8 for every 1°C temperature interval. 3.8a, 3.8b and 3.8d are based on four replicas each, 3.8c is based on two replica. For these runs no misbonds were allowed and a sequence dependent model was used.

All graphs show the same tendency for the melting order, however the melting temperature is decreasing from top left to bottom right. This cannot be true, because the average

EXPERIMENTS AND RESULTS

melting temperature is an independent property. Therefore the structure had not reached equilibrium before the temperature was increased again, which is called "superheating". As a conclusion it was necessary to run all further calculations with 10^8 simulation steps for every 1°C of temperature increase. It was proposed that a higher number of replicas and a comparison with runs of 2×10^8 simulation steps per iteration would be necessary to determine if a longer equilibration time would be needed. Calculations of this extent would go far beyond the scope of this work. Looking at the small differences in melting between different replicas shown in figure 3.8d, it was concluded that 10^8 simulation steps suffice as the equilibration time. Furthermore, a more detailed melting behaviour analysis as well as a comparison between both OPT and KH design are the main focus of this work. An improvement of the data documentation can be achieved by analysing the entire trajectory rather than just the final configuration of every temperature iteration.

Melting the optimized DNA Pyramid - Average base vs sequence dependent model

Average base pair-based model

For the melting behaviour of the OPT structure it is interesting not only to investigate the process, but also to analyse the origin of certain influences on the melting properties. The data shown in Figure 3.7 was obtained using an average base model, where G-C and A-T bonds are equal. This clearly shows that the most outer domains in Figure 1.7 (b) melt first, as they might have been expected.

Melting a structure from the strand ends leads to the greatest entropy gain, because the free strand ends gain a lot of freedom. Therefore, melting from strand ends is greatly favoured for an increase in temperature. In the average base model, as soon as domains 4 and 8 have melted off, the neighbouring domains (in perspective from the free strand ends) can melt next. It also seems that over a large number of replicas domains 4 and 8 are expected to melt at around the same temperature. Melting of those domains will then yield the highest gain in entropy. If a domain is melted from the unbound strand ends it only has one local restriction after the melting process. If their neighbouring domains have already melted than there is no local restriction. After a domain, whose neighbouring domains are still formed, has melted the complementary strands are still locally constrained and the gain of entropy is smaller. For the average base model melting from the strand ends can be observed almost symmetrical.

It is necessary to simulate the structure using a sequence dependent model, to be able to compare the simulation data with experimental data, because in nature G-C bonds are stronger than A-T bonds.

Sequence dependent based model

After running a sequence dependent model of the DP, the domain order melting is altered, because G-C bonds are now stronger than A-T bonds and the G-C content of different domains also varies.

In Figure 3.9 the results of simulations of the optimized design using a topology that did (Left) and did not allow (Right) misbonds are shown. The graphs are very similar and the melting order observed was the same. Misbonds do not seem to influence the melting process and it was concluded that only intended bonds were necessary for further analysis.

The melting order of Figure 3.9 neither matches the reverse folding order based on the experimentalist's predictions,^e nor does it match the order provided by NUPACK. The folding order assumed by the experimentalists is equivalent to the domain number and

^eHere it is assumed that the melting order will be the reverse of the folding order. This is not mandatory, but will be assumed as a reference geometry.

EXPERIMENTS AND RESULTS

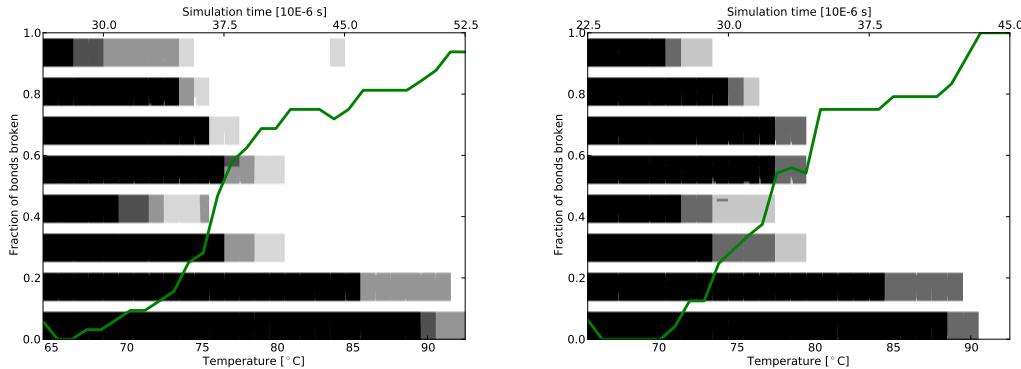


Figure 3.9 – Melting data for the OPT design with (Left) and without (Right) misbonds allowed. The warm-up (50°C - 65°C) is not shown. The domains are shown from 1 (bottom) to 8 (top) for both graphs.

is ordered from 8 assumed to melt first to 1 assumed to melt last. By comparison of the average-base model results, Figure 3.7, and the sequence-dependent model results, Figure 3.9, the melting points for the sequence dependent model have changed. Generally these values have increased for low number domains, which is a natural effect of the sequence dependent model for domains containing a high content of G-C hydrogen bonds and decreased for higher number domains, which is expected, because higher number domains have fewer G-C base pairs. The melting temperature of lower numbered domains has increased more than the ones of higher number domains, which is due to the increasing number of G-C base pairs towards lower numbers.

Finally, looking at selected MD movies of the trajectories did not show any explicitly recognisable processes that were unable to finish within one temperature iteration, such as major breaking of bonds. This reassures the assumption that the chosen time interval per temperature increase of one centigrade per 10^8 simulation steps is sufficient. However, one still needs to check if using longer time per temperature iteration would lead to different measurements.

Now considering the annealing process, the structure will start to form with the domains of highest average melting point first, domains 2 and 1. The structure is able to peel off the domains in an ordered fashion during the melting process and it suggests itself that this behaviour could be observed in reverse during the folding process. From the melting data there does not seem to be any obvious topological restriction on the structures reformation, other than the structure always having to melt from free strand ends. Since the order of formation expected from NUPACK data is nearly the reverse order of melting the folding process should not be kinetically hindered.

For this and all other investigated melting it is visible that the domains melt from the strand ends. In other cases partial melting from the domain end facing other still-intact neighbouring domains could be observed. Interestingly, in most of these cases the process

did not complete and the double-helix returned to its initial state before finally melting from its outward-facing end. In Figure 3.9 this process is captured by small lighter shaded areas inside the darker shaded block. The greater the lighter shaded area is the longer the region stays partially melted. This observation undermines the idea that melting from the free end of each domain is highly favoured. The free energy gained from the newly accessible degrees of freedom is smaller than the energy gained by formation of the duplexes.

The domains were observed to melt in the order 8-4-7-6-5-3-2-1. The melting order of domain 4 and 7 as well as 5 and 3 was seen to be reversed, when looking at the detailed trajectory files of each run. However, melting from the strand ends was observed in all analysed trajectories and the domains whose melting order was reversed are located at different strand ends. This suggests that melting of one domain does only trigger melting of adjacent domains for the OPT design.

Comparing the data to the NUPACK predictions domain 6 and 7 are expected to melt in the reverse order. However, the melting points of the hairpins corresponding to both domains are so similar that both melting curves overlap. Taking into account that melting from strand ends must be expected due to entropic effects, the observed melting order is favoured. Further, domains 3 and 4 would be expected to show melting after domains 5 and 6 have melted. Because domain 4 is located at one of the free strand ends, as can be seen in the topology Figure 1.7 (b), the entropic gain of breaking these bonds is immediately quite large, the partially melted structure gains even more energy.

There are two small differences between the observed melting order and the proposed folding order. Firstly, domains 3 and 4 are seen to melt out of order, but after initiation of domain 1 and 2 the free strand ends underlie spatial limitations and the DP must fold in from these two ends. Here it practically does not matter if the strand end containing domains 3,4 or the one containing domains 5,6,7,8 is folded in first. Secondly, the NUPACK order of 6,7 is reversed to the prospected melting order. At the stage, where all domains other than 6,7,8 should be formed, there is only one more free strand end. This free strand end is already spatially confined by previously formed domains to the proximity of domain 6. This would heighten the annealing temperature of domain 6. Of course, in experiment there is always the possibility that the structure gets stuck in a topologically frustrated region of phase space. Therefore, one has to take care not to anneal to quickly.

Another possible effect is that removing the double helix from the connection of corners 2-3 also grants a large increase in freedom of deforming the rest of the pyramidal structure and thus gaining further entropy. Removing the first double helix destabilises the structure and breaking further domain was observed to totally break the structures symmetry. This structural change is illustrated below. The three different stills were taken at 70°C, 78°C and 86°C respectively.

EXPERIMENTS AND RESULTS

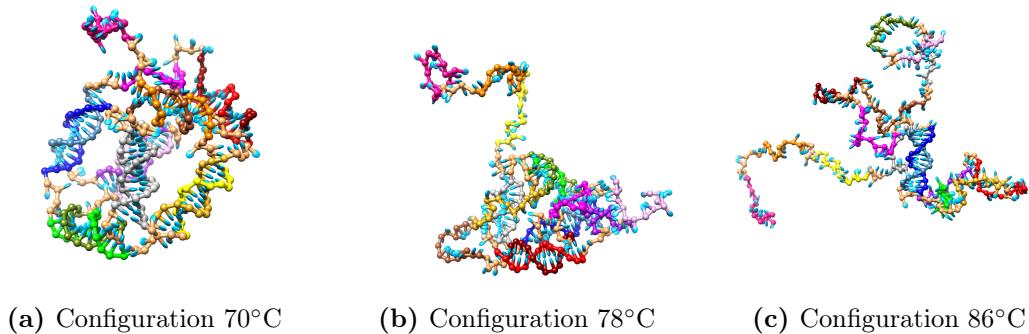


Figure 3.10 – Stages of the DNA pyramid structure during the melting process.

Lastly, the calculated melting data is generally below or equal to the average melting temperature suggested by NUPACK. A reason for this is probably, that melting a domain in the pyramid also has beneficial effects on the degrees of freedom of the remaining structure. In NUPACK this effect cannot be incorporated, because only single hairpins are simulated.

From our results we can conclude that the melting occurs in stages, which seems to agree with the experimentalists suggestions.

Below a series of stills shows excerpts of the unwinding process, that is taking place, when a double helix melts, to provide an insight into the kinetics of staged melting.

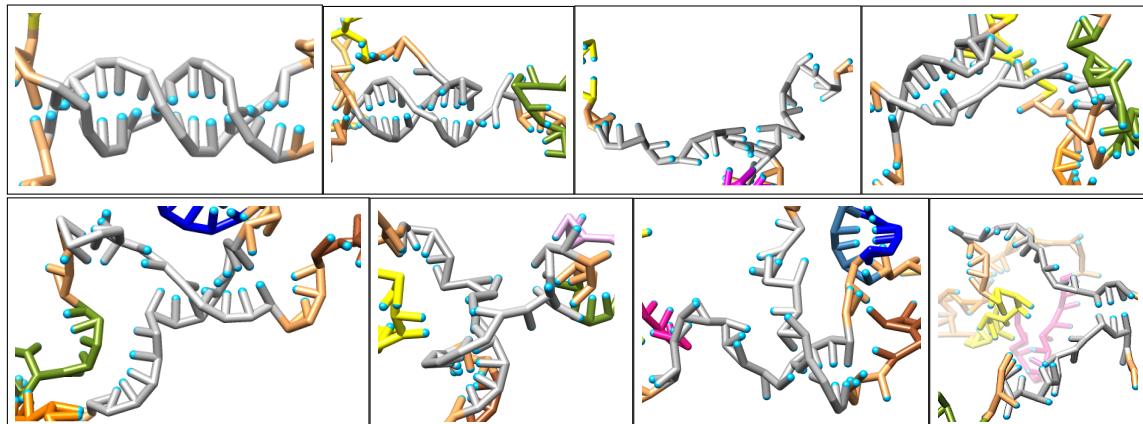


Figure 3.11 – Melting process of a single domain (grey). The visualisation of oxDNA is simplified to give a better overview. From top left to bottom right the kinetics of a melting of a domain can be observed. The domain partially melts and reforms during the melting process.

Finally, we want to give a visual impression of the melting of the optimized version, which is shown in Figure 3.12, to give the reader an impression of the trajectory.

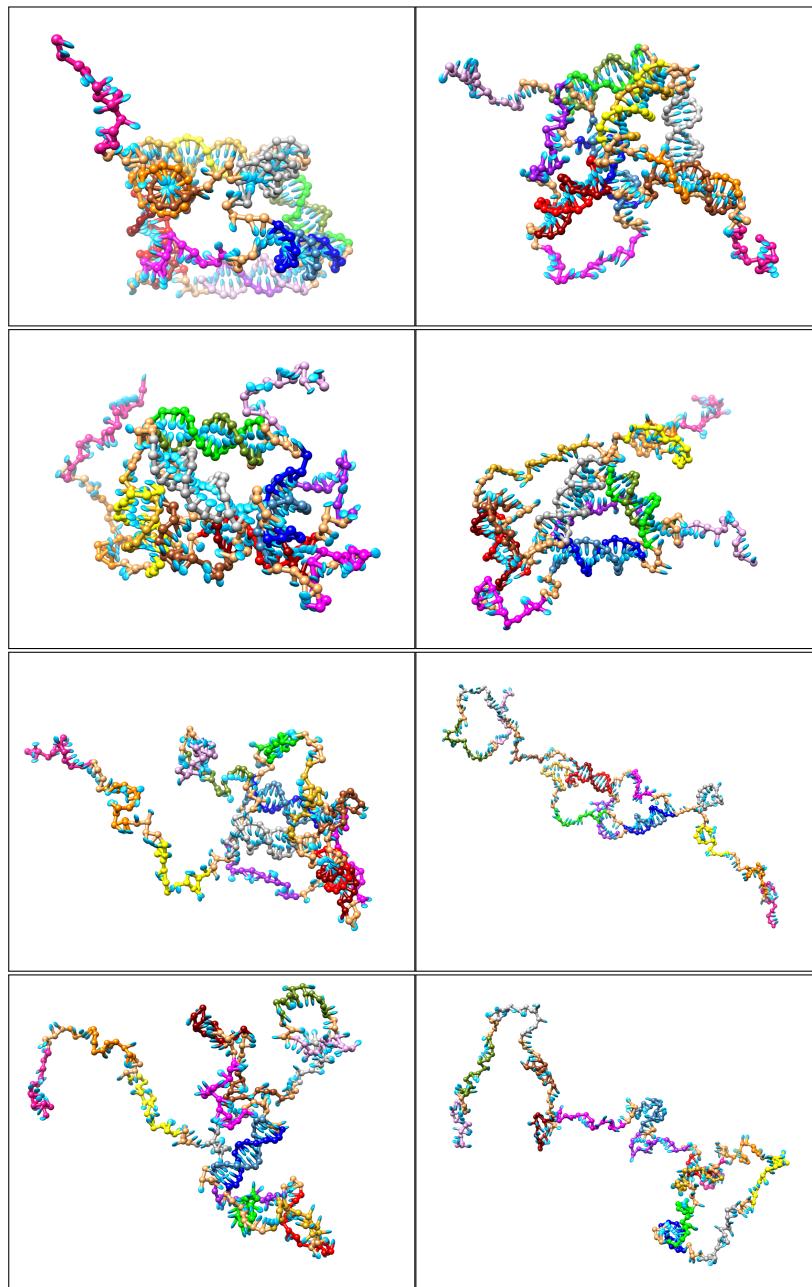


Figure 3.12 – Stage-wise melting process of the DNA nanopyramid. From top left to bottom right the melting process is shown domain by domain. The melting order for this trajectory is 8-4-7-6-3-5-2-1.

EXPERIMENTS AND RESULTS

Melting of the kinetically hindered version

For the kinetically hindered version, the main interest lies within understanding the melting behaviour and comparing it to the optimized version. The experimentalists deliberately put those domains of highest G-C content at the strand ends. Thus, only a sequence dependent model is of interest. In 3.13a and 3.13b the probability that a base pair in a domain is formed as a function of temperature is visualised . In 3.13a misbonds were prevented while in 3.13b the possibility of misbonds was permitted. The data in (a) and (b) was generated by 6 and 3 parallel simulations respectively. The statistics are very small, but there does not seem to be any notable trend that would allow for differentiation in domain melting temperature between simulations with or without misbonds allowed. It is however visible that a lot of misbonds form, when they are permitted. This stands in contrast with the results for the optimized version, where the graphs with and without allowed misbonds were more similar. A possible conclusion is that because the KH design unfolds rapidly, the unbound nucleotides are still in proximity to other unbound nucleotides and the formation of misbonds is favoured with respect to the optimized design.

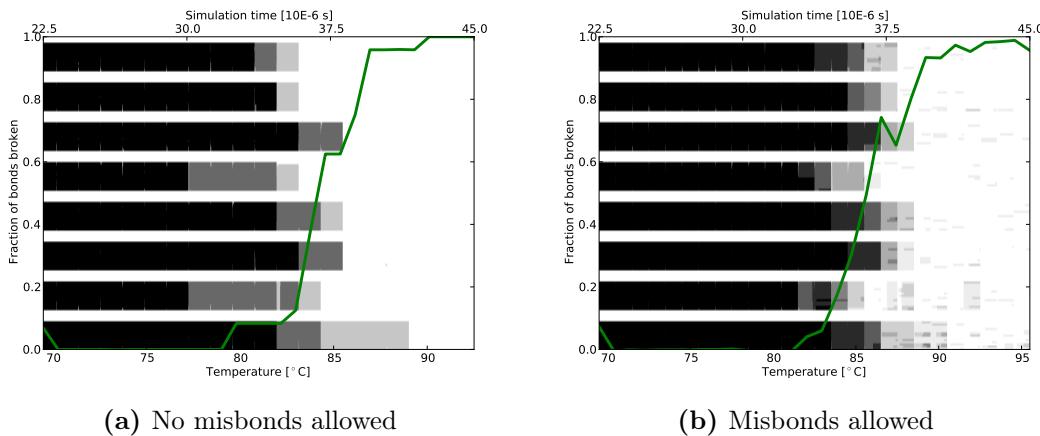


Figure 3.13 – Melting procedure for the KH design. The warmup (50°C - 70°C) is not shown.

Comparing the melting data of the kinetically hindered design summarized in Table 3.2 to the melting data for the peeling version given in Table 3.1, there is obviously a large difference. The melting process does not seem to progress in stages, but over a smaller temperature range when comparing to the peeling version. From the trajectories one can see that domains 2 and 5 melt together. After domain 2 melts the adjacent domain number 5 melts immediately. From Figure 1.7 (c) one can see that one of the complementary regions that form domain 2 is located at one of the strand ends. By the same argument as in the previous section, melting from the strand ends is entropically favoured.

The first melting initiation takes place at a much higher temperature than what could

be observed for the peeling version. Also, the fact that domain 5 melts so rapidly after domains 2 suggests that melting of the latter imposes a rather strict constraint on melting of the prior. Again, even though domain 2 has a quite high expected melting temperature, melting commences from the strand ends and at much lower temperatures than predicted by NUPACK simulations.

Like in the OPT version, melting of domains inside the structure can be observed in 3.13a and 3.13b, but these domains quickly reform if the neighbouring domains are still fully formed. It is also notable that the overall pyramidal structure is maintained quite well, even after the first two domains are completely broken. The difference between the OPT and the KH version is that the top view (the four domains constituting the top of the pyramid) is intact even after two domains breaking. This seems to freeze the DP's configuration. Melting becomes rapid after domain 1 and domain 8, which is the other strand end, are melted. It was observed that both domains seem to act as triggers. Melting of domain 8 immediately allows for melting of domain 7. After domain 1 is melted the melting of domains 4,6 and 3 in this order can commence.

However, it is also possible for the structure to completely melt from one strand end and it has been observed that in one of the replicas domain 1 melted last, even though it is also located at one of the sequence's ends. As soon as a domain is no longer constrained by other duplexes on both sides, it can melt from the end of the domain easily. NUPACK data also shows that the melting point of simple hairpins of domains 4,6 and 3 are much lower than the melting temperature of domain 1, and thus should have broken apart if it had not been geometrically constrained.

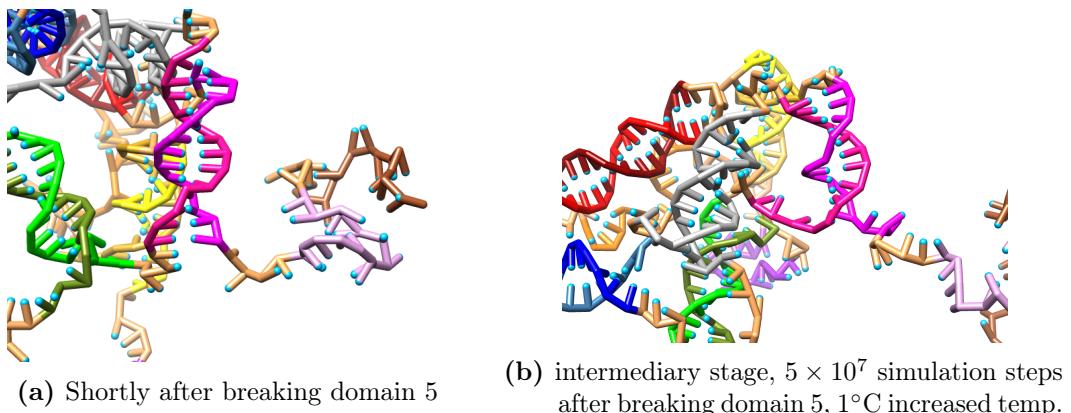


Figure 3.14 – A side view onto the free end of domain 8 (magenta) during melting of the KH design.

It is very surprising that domain 8 does not melt immediately after one of its ends has become free, because its predicted melting point is much lower than the melting temperature of domain 2 and 5. The melting process of domain 8 is geometrically complicated and before the process is finally completed the domain passes intermediary

EXPERIMENTS AND RESULTS

stages in which it partially melts and reforms. The simulation time spent by domain 8 in and out of these intermediary stages exceeds the simulation time of 10^8 steps per one temperature iteration and therefore it seems likely that the system would need more time to relax and that melting of the DNA pyramid is kinetically hindered. Due to the computational limitations of this project, this hypothesis could not be tested and remains as a suggestion for further analysis.

The above limitations to the melting of the structure explain the overall increased melting temperature of the structure compared to the OPT design. It is important to notice that for the KH version a direct connection between folding data and melting data is unclear, because the melting process is hindered quite strongly by topological restrictions. Certainly during an annealing process the domains with the highest melting transition would form first. Because these domains are located on the outside of the strand, the fold would have to undergo multiple structural rearrangements before being able to achieve the "fully formed" DNA Pyramid structure. This is suggested to be the reason why a single properly formed DNA pyramid could not be obtained in the simulations when thermal annealing was tried, and also when folding was accelerated using fictitious forces.

3.2.4 Folding

Finally, the folding pattern of the DP shall be investigated. Domain folding is a rare event and its observation is essentially random, meaning there is no guarantee domain formation will be observable. The simulation set-up was kept very simple:

1. A random initial configuration was chosen, that has no hydrogen bonding present, and the simulation temperature is set to 95°C.
2. The simulation was run for a certain number of simulation steps.
3. The simulation temperature was decreased by 1°C.
4. Stage 2 and 3 were repeated until the simulation temperature reached 50°C.

In order to get an estimate of the probability of observing such a rare event 250 replicas with 10^5 simulation steps and another 50 replicas with 10^6 simulation steps for stage 2 were run for the optimized DP design, but only little folding could be observed. Most of these 300 runs did not yield any results, no domains were folded. The 3 most successful runs resulted in the formation of two partially formed domains or one fully formed domain. Further 2 runs each for KH and OPT design were run with 10^8 simulation steps for stage 2 and a final temperature of 70°C. These runs also resulted in the folding of one domain at most.

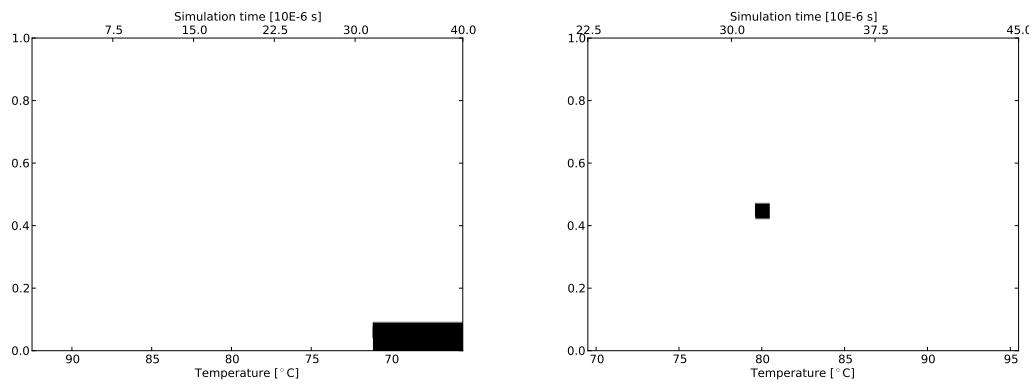


Figure 3.15 – Example of an annealing simulation for the kinetically hindered version (Left) and the optimized version (Right).

To simplify the simulation no misbonds were permitted. In Figure 3.15 two examples for the observed annealing profile are shown. The results shown in Figure 3.15 are not chosen to represent the KH and OPT version respectively, but to exemplify the results obtained for the annealing simulations. To successfully analyse the thermal annealing behaviour of the DP, significantly longer simulation times would have been necessary and must be left as a suggestion. Another suggestion is to also allow for misbonding in the structure, because misbonded conformations can lead to correctly bound conformations.

4 Summary and outlook

4.1 Conclusion

In this study I have used the coarse-grained model of DNA, called oxDNA, developed by T. Ouldridge *et al.* to investigate the folding and melting processes of a single-stranded DNA Pyramid. Specifically the melting and folding processes of two different designs of a ssDNA pyramid have been studied using Molecular Dynamics simulations.

The simulations were designed to analyse the difference between the melting behaviour of an "optimized" and a "kinetically hindered" design. The oxDNA model reproduces the temperature dependence of the OPT design and agrees with melting temperatures from NUPACK . The observations on the melting process seem to be in good agreement with experimental data.

When analysing the temperature influence on the kinetically hindered and optimized version of the DNA pyramid, quite different behaviour was observed for both melting processes. On the one hand, the OPT design showed a stage-wise melting process, melting from both strand ends and one domain after the other. The domains melted approximately as predicted by NUPACK. On the other hand, the kinetically hindered design melted in a completely different order than expected from hairpin stability data from NUPACK. The results show that peeling and KH structures are good examples of stage-wise versus concerted melting of a polyhedral DNA structure.

These findings stand in good agreement with the experimentalist's predictions of the melting behaviour for the KH version and for the OPT version. The results of this work show the unique capabilities of the oxDNA model to study the properties and temperature behaviour of single-stranded DNA and double-stranded DNA and more complex structures composed of these basic motifs. The model systems provided by Kocar Vid *et. al* are interesting examples of DNA nanostructures. It is quite remarkable how seemingly simple manipulation of the DNA sequence can lead to such different melting temperatures for both fully folded structures. The application span of these designs may be large, because a controlled temperature behaviour is a very desirable property. Structural rearrangements can possibly be steered using temperature control for the optimized design. Such structural rearrangements can be useful in drug delivery and similar applications where active and inactive states are utilized. Folding a geometry from a single strand also allows the system to change its conformation very rapidly even at low concentrations because spatial proximity is always granted. Single-stranded designs with

a reliable folding pattern could be especially useful when passing a narrow. The unfolded structure can be passed through a gate and then folded given a temperature gradient between the systems.

The kinetically hindered design also has a very special field of application. The single-stranded nature of the fold allows domains to stay intact far above their predicted melting points. The melting process is concerted and can be realised within a very small temperature range. If only a small temperature gradient is accessible within the system such a design could be applied. Here the design would have to prove itself against multi-stranded self-assembly design.

4.2 Future work

In this work the main focus was set on the analysis of the melting behaviour of the ssDNA pyramid. The temperature induced melting was simulated by increasing the temperature step-wise within the computational capabilities. For future work testing a more shallow temperature profile would need to be tested, i.e. increase the number of simulation steps per temperature unit to see if the melting temperatures would be shifted.

Furthermore, the annealing process could not be studied properly because the simulation time accessible was too small to simulate the rare events of thermal annealing. No statistics on the annealing behaviour could be obtained. We suggest that the thermal annealing process could be investigated using longer simulation times. Also, the use of rare-event methods such as Forward Flux Sampling or Umbrella Sampling may be advantageous.

5 Appendix

List of Figures

1.1	Ideal fold of a DNA Pyramid	2
1.2	Structure of DNA nucleotides	3
1.3	DNA double helix visualisation	4
1.4	(Left) DNA hairpin and (Right) DNA pseudoknot.	6
1.5	Structures made from DNA using the DNA-origami method taken from (Rothemund, 2006).	7
1.6	Ideal fold of a symmetrical ssDNA Pyramid	10
1.7	Topological properties of the optimized and kinetically hindered DNA pyramid folding pattern	11
1.8	(Left) Illustration of the ssDNA path through the structure. The strands must be antiparallel at every edge. (Right) Multi stranded assembly geometry.	11
1.9	Thermal annealing profile of KH and OPT design	12
2.1	Visualisation of the the coarse-grained oxDNA model of DNA	16
2.2	Cluster building in oxDNA using the VMMC algorithm	22
3.1	Top view on the folded DNA Pyramid without and with topology problems.	28
3.2	Illustration of a topologically frustrated corner.	30
3.3	Illustration of a "sliced" DNA pyramid.	30
3.4	Initialized DP geometry for the OPT design.(Left) View from the top and (Right) view from the side of the pyramid.	33
3.5	Initialized DP geometry for the KH design.(Left) View from the top and (Right) view from the side of the pyramid.	33
3.6	(Left) NUPACK hairpin and (Right) visualisation of the temperature profile for Domain 8 of the optimized version.	34
3.7	Statistics on the average domain T_m of the OPT design using an average-base model.	36
3.8	Melting runs for the kinetically hindered version using different time steps per one °C temperature increase.	37

3.9	Melting data for the OPT design with and without misbonds allowed.	40
3.10	Stages of the DNA pyramid structure during the melting process.	42
3.11	Melting process of a single domain.	42
3.12	Stage-wise melting process of the DNA nanopyramid.	43
3.13	Melting procedure for the KH design.	44
3.14	A side view onto the free end of domain 8 (magenta) during melting of the KH design.	45
3.15	Example of an annealing simulation for the kinetically hindered version (Left) and the optimzed version (Right).	47

List of Tables

3.1	Melting Temperatures[°C] of DNA hairpins for OPT design as predicted by NUPACK; all A-T and all G-C data for reference	35
3.2	Melting Temperature[°C] of DNA hairpins for KH design, as predicted by NUPACK	35
5.1	Colour Code for Chimera illustrations	51
5.2	Hairpin NUPACK data for a $d(T)_{18}$ loop for OPT design	52
5.3	Hairpin NUPACK data for a $d(T)_{36}$ loop for OPT design	53
5.4	Hairpin NUPACK data for a $d(T)_{18}$ loop for KH design	54
5.5	Hairpin NUPACK data for a $d(T)_{36}$ loop for KH design	55

5.1 Appendix A

Table 5.1 – Colour Code for Chimera illustrations

	Optimized design	Kinetically hindered design
Domain1	blue	blue
Domain2	red	orange
Domain3	green	red
Domain4	purple	green
Domain5	grey	purple
Domain6	yellow	grey
Domain7	orange	yellow
Domain8	pink	pink

Table 5.2 – Hairpin NUPACK data for a $d(T)_{18}$ loop for OPT design

5.2 Appendix B - NUPACK data

T[°C]	Domain 8	Domain 7	Domain 6	Domain 5	Domain 4	Domain 3	Domain 2	Domain 1
50	0.402	0.412	0.410	0.416	0.439	0.444	0.430	0.413
52	0.403	0.413	0.411	0.418	0.443	0.450	0.433	0.415
54	0.404	0.414	0.413	0.420	0.447	0.456	0.437	0.419
56	0.405	0.416	0.415	0.422	0.451	0.463	0.440	0.422
58	0.406	0.417	0.417	0.425	0.456	0.471	0.444	0.427
60	0.407	0.419	0.419	0.427	0.461	0.480	0.449	0.432
62	0.408	0.421	0.422	0.431	0.466	0.490	0.454	0.438
64	0.410	0.424	0.424	0.434	0.473	0.502	0.459	0.447
66	0.411	0.426	0.427	0.438	0.480	0.515	0.466	0.458
68	0.413	0.429	0.431	0.442	0.488	0.531	0.475	0.475
70	0.415	0.432	0.435	0.447	0.497	0.551	0.487	0.501
72	0.418	0.435	0.439	0.453	0.509	0.575	0.504	0.542
74	0.421	0.439	0.445	0.461	0.523	0.607	0.531	0.603
76	0.424	0.443	0.452	0.472	0.540	0.645	0.571	0.680
78	0.428	0.447	0.463	0.490	0.561	0.691	0.627	0.763
80	0.433	0.453	0.479	0.516	0.587	0.739	0.695	0.835
82	0.439	0.459	0.504	0.556	0.618	0.786	0.765	0.888
84	0.448	0.466	0.544	0.612	0.655	0.826	0.826	0.922
86	0.462	0.476	0.602	0.681	0.695	0.857	0.871	0.943
88	0.484	0.489	0.676	0.752	0.734	0.881	0.901	0.956
90	0.519	0.508	0.754	0.815	0.770	0.897	0.920	0.963
92	0.574	0.537	0.823	0.863	0.801	0.909	0.932	0.968
94	0.648	0.579	0.875	0.896	0.824	0.917	0.939	0.971
96	0.731	0.637	0.910	0.917	0.840	0.922	0.943	0.972
98	0.807	0.705	0.932	0.931	0.850	0.926	0.945	0.973
100	0.865	0.774	0.946	0.939	0.856	0.929	0.947	0.974

Table 5.3 – Hairpin NUPACK data for a $d(T)_{36}$ loop for OPT design

Temp. [°C]	Domain 8	Domain 7	Domain 6	Domain 5	Domain 4	Domain 3	Domain 2	Domain 1
50	0.570	0.578	0.576	0.582	0.597	0.602	0.592	0.579
52	0.571	0.579	0.577	0.584	0.600	0.606	0.594	0.581
54	0.572	0.580	0.579	0.585	0.603	0.611	0.597	0.584
56	0.572	0.581	0.580	0.587	0.606	0.617	0.600	0.587
58	0.573	0.583	0.582	0.589	0.609	0.624	0.604	0.591
60	0.574	0.584	0.584	0.592	0.613	0.632	0.608	0.597
62	0.575	0.586	0.586	0.594	0.617	0.643	0.613	0.606
64	0.576	0.587	0.588	0.597	0.623	0.658	0.621	0.622
66	0.577	0.589	0.591	0.601	0.629	0.679	0.632	0.648
68	0.579	0.592	0.594	0.606	0.638	0.708	0.651	0.692
70	0.580	0.594	0.598	0.612	0.649	0.745	0.680	0.753
72	0.582	0.597	0.605	0.622	0.666	0.787	0.724	0.821
74	0.585	0.600	0.614	0.638	0.690	0.829	0.778	0.880
76	0.587	0.603	0.631	0.664	0.721	0.865	0.833	0.922
78	0.592	0.608	0.658	0.702	0.759	0.893	0.880	0.948
80	0.598	0.613	0.700	0.752	0.799	0.912	0.912	0.963
82	0.609	0.621	0.757	0.807	0.836	0.925	0.933	0.971
84	0.629	0.634	0.819	0.858	0.866	0.934	0.946	0.975
86	0.663	0.653	0.874	0.897	0.886	0.939	0.953	0.978
88	0.716	0.683	0.913	0.924	0.899	0.943	0.957	0.979
90	0.785	0.725	0.939	0.941	0.906	0.946	0.960	0.980
92	0.851	0.779	0.954	0.951	0.909	0.948	0.961	0.980
94	0.902	0.834	0.963	0.957	0.910	0.949	0.962	0.981
96	0.934	0.881	0.969	0.961	0.909	0.950	0.962	0.981
98	0.952	0.914	0.972	0.963	0.906	0.950	0.962	0.981
100	0.962	0.936	0.974	0.965	0.903	0.951	0.962	0.981

Table 5.4 – Hairpin NUPACK data for a $d(T)_{18}$ loop for KH design

Temp. [°C]	Domain 8	Domain 7	Domain 6	Domain 5	Domain 4	Domain 3	Domain 2	Domain 1
50	0.427	0.432	0.435	0.420	0.416	0.410	0.412	0.402
52	0.430	0.435	0.439	0.422	0.418	0.411	0.413	0.403
54	0.433	0.439	0.443	0.425	0.420	0.413	0.414	0.404
56	0.437	0.442	0.448	0.428	0.422	0.415	0.416	0.405
58	0.441	0.447	0.454	0.431	0.425	0.417	0.417	0.406
60	0.445	0.451	0.461	0.435	0.427	0.419	0.419	0.407
62	0.450	0.456	0.469	0.439	0.431	0.422	0.421	0.408
64	0.457	0.461	0.478	0.444	0.434	0.424	0.424	0.410
66	0.466	0.467	0.490	0.450	0.438	0.427	0.426	0.411
68	0.478	0.475	0.505	0.457	0.442	0.431	0.429	0.413
70	0.497	0.485	0.527	0.466	0.447	0.435	0.432	0.415
72	0.527	0.500	0.557	0.479	0.453	0.439	0.435	0.418
74	0.573	0.522	0.597	0.497	0.461	0.445	0.439	0.421
76	0.638	0.556	0.649	0.526	0.472	0.452	0.443	0.424
78	0.715	0.606	0.707	0.567	0.490	0.463	0.447	0.428
80	0.790	0.672	0.764	0.623	0.516	0.479	0.453	0.433
82	0.852	0.745	0.813	0.689	0.556	0.504	0.459	0.439
84	0.894	0.814	0.852	0.756	0.612	0.544	0.466	0.448
86	0.921	0.868	0.879	0.814	0.681	0.602	0.476	0.462
88	0.938	0.906	0.898	0.858	0.752	0.676	0.489	0.484
90	0.947	0.930	0.911	0.889	0.815	0.754	0.508	0.519
92	0.953	0.945	0.920	0.910	0.863	0.823	0.537	0.574
94	0.956	0.954	0.925	0.923	0.896	0.875	0.579	0.648
96	0.958	0.960	0.929	0.931	0.917	0.910	0.637	0.731
98	0.959	0.963	0.931	0.936	0.931	0.932	0.705	0.807
100	0.960	0.965	0.933	0.940	0.939	0.946	0.774	0.865

Table 5.5 – Hairpin NUPACK data for a $d(T)_{36}$ loop for KH design

Temp. [°C]	Domain 8	Domain 7	Domain 6	Domain 5	Domain 4	Domain 3	Domain 2	Domain 1
50	0.588	0.592	0.596	0.585	0.582	0.576	0.578	0.570
52	0.591	0.595	0.599	0.587	0.584	0.577	0.579	0.571
54	0.593	0.598	0.603	0.590	0.585	0.579	0.580	0.572
56	0.596	0.601	0.608	0.592	0.587	0.580	0.581	0.572
58	0.600	0.604	0.614	0.595	0.589	0.582	0.583	0.573
60	0.605	0.608	0.621	0.599	0.592	0.584	0.584	0.574
62	0.613	0.613	0.631	0.604	0.594	0.586	0.586	0.575
64	0.625	0.619	0.646	0.610	0.597	0.588	0.587	0.576
66	0.645	0.628	0.668	0.618	0.601	0.591	0.589	0.577
68	0.679	0.644	0.700	0.630	0.606	0.594	0.592	0.579
70	0.730	0.668	0.742	0.650	0.612	0.598	0.594	0.580
72	0.792	0.707	0.791	0.679	0.622	0.605	0.597	0.582
74	0.853	0.760	0.838	0.720	0.638	0.614	0.600	0.585
76	0.900	0.819	0.876	0.771	0.664	0.631	0.603	0.587
78	0.931	0.873	0.903	0.823	0.702	0.658	0.608	0.592
80	0.949	0.913	0.921	0.868	0.752	0.700	0.613	0.598
82	0.959	0.940	0.933	0.902	0.807	0.757	0.621	0.609
84	0.965	0.956	0.940	0.924	0.858	0.819	0.634	0.629
86	0.968	0.965	0.945	0.938	0.897	0.874	0.653	0.663
88	0.969	0.970	0.948	0.947	0.924	0.913	0.683	0.716
90	0.970	0.973	0.950	0.952	0.941	0.939	0.725	0.785
92	0.971	0.974	0.951	0.956	0.951	0.954	0.779	0.851
94	0.971	0.975	0.952	0.958	0.957	0.963	0.834	0.902
96	0.971	0.976	0.952	0.959	0.961	0.969	0.881	0.934
98	0.971	0.976	0.953	0.960	0.963	0.972	0.914	0.952
100	0.971	0.976	0.953	0.961	0.965	0.974	0.936	0.962

Bibliography

- [1] Erben, C. M.; Goodman, R. P.; Turberfield, A. J. *Journal of the American Chemical Society* **2007**, *129*, 6992–3.
- [2] Mao, C.; Sun, W.; Shen, Z.; Seeman, N. C. *Nature* **1999**, *397*, 144–6.
- [3] Watson, J. D.; Crick, F. H. C. *Cold Spring Harbor Symposia on Quantitative Biology* **1953**, *18*, 123–131.
- [4] Seeman, N. C. *Journal of Theoretical Biology* **1982**, *99*, 237–47.
- [5] Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L. *Nature* **2000**, *406*, 605–8.
- [6] Shin, J.-S.; Pierce, N. a. *Journal of the American Chemical Society* **2004**, *126*, 10834–5.
- [7] Pinheiro, A. V.; Han, D.; Shih, W. M.; Yan, H. *Nature Nanotechnology* **2011**, *6*, 763–72.
- [8] Rothemund, P. W. K. *Nature* **2006**, *440*, 297–302.
- [9] N.V. Voigt, T. Tørring, A. Rotaru, M.F. Jacobsen, J.B. Ravnsbæk, R. Subramani, W. Mamdouh, J. Kjems, A. Mokhir, F. Besenbacher, K. G. *Nat Nanotechnol* **2010**, *5*, 200–203.
- [10] Kuzyk, A.; Laitinen, K. T.; Törmä, P. *Nanotechnology* **2009**, *20*, 235305.
- [11] H.T. Maune, S.P. Han, R.D. Barish, M. Bockrath, W.A. Goddard, P.W.K. Rothemund, E. W. *Nat Nanotechnol* **2010**, *5*, 61–66.
- [12] Tian, Y.; Mao, C. *J. Am. Chem. Soc.* **2004**, *126*, 11410–11411.
- [13] Deng, Z.; Chen, Y.; Tian, Y.; Mao, C. In *Nanotechnology: Science and Computation*; Chen, J.; Jonoska, N.; Rozenberg, G., Eds.; Natural Computing Series; Springer Berlin Heidelberg, 2006; pp 23–34.
- [14] Zhang, D. Y.; Seelig, G. *Nature chemistry* **2011**, *3*, 103–13.
- [15] Chen, W. . M. *Angew. Chem. Int. Ed.* **2004**, *43*, 3554–6.
- [16] Simmel, F.; Yurke, B. *Physical Review E* **2001**, *63*, 041913.
- [17] Benenson, Y.; Gil, B.; Ben-Dor, U.; Adar, R.; Shapiro, E. *Nature* **2004**, *429*, 423–9.

- [18] *Cloaked DNA nanodevices survive pilot mission.* <http://wyss.harvard.edu/viewpressrelease/150/cloaked-dna-nanodevices-survive-pilot-mission>.
- [19] Adleman, L. *Science* **1994**, *266*, 1021–1024.
- [20] Shapiro, E. *Interface Focus* **2012**, *2*, 497–503.
- [21] Douglas, S. M.; Bachelet, I.; Church, G. M. *Science (New York, N.Y.)* **2012**, *335*, 831–4.
- [22] Bhatia, D.; Mehtab, S.; Krishnan, R.; Indi, S. S.; Basu, A.; Krishnan, Y. *Angewandte Chemie* **2009**, *121*, 4198–4201.
- [23] Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L. P.; Pedersen, J. S.; Birkedal, V.; Besenbacher, F.; Gothelf, K. V.; Kjems, J. r. *Nature*, *459*, 73–6.
- [24] Mastroianni, A. J.; Claridge, S. A.; Alivisatos, A. P. *Journal of the American Chemical Society* **2009**, *131*, 8455–9.
- [25] Li, Z.; Wei, B.; Nangreave, J.; Lin, C.; Liu, Y.; Mi, Y.; Yan, H. *Journal of the American Chemical Society* **2009**, *131*, 13093–8.
- [26] Lin, C.; Rinker, S.; Wang, X.; Liu, Y.; Seeman, N. C.; Yan, H. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 17626–31.
- [27] Setyawati, M. I.; Kutty, R. V.; Tay, C. Y.; Yuan, X.; Xie, J.; Leong, D. T. *ACS applied materials & interfaces* **2014**.
- [28] Setyawati, M. I.; Khoo, P. K. S.; Eng, B. H.; Xiong, S.; Zhao, X.; Das, G. K.; Tan, T. T.-Y.; Loo, J. S. C.; Leong, D. T.; Ng, K. W. *Journal of Biomedical Materials Research. Part A* **2013**, *101*, 633–40.
- [29] Pérez, A.; Luque, F. J.; Orozco, M. *Accounts of Chemical Research* **2012**, *45*, 196–205.
- [30] *List of software for molecular mechanics modeling.* http://en.wikipedia.org/wiki/List_of_software_for_molecular_mechanics_modeling.
- [31] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.
- [32] Brooks, B. R.; et al. *Journal of Computational Chemistry* **2009**, *30*, 1545–614.
- [33] SantaLucia, J. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 1460–5.
- [34] Caltech; *NUPACK*; 2014. <http://nupack.org/>.
- [35] Bustamante, C.; Marko, J.; Siggia, E.; Smith, S. *Science* **1994**, *265*, 1599–1600.

Bibliography

- [36] Doye, J. P. K.; Ouldridge, T. E.; Louis, A. a.; Romano, F.; Šulc, P.; Matek, C.; Snodin, B. E. K.; Rovigatti, L.; Schreck, J. S.; Harrison, R. M.; Smith, W. P. *J. Physical Chemistry Chemical Physics : PCCP* **2013**, *15*, 20395–414.
- [37] Starr, F. W.; Sciortino, F. *Journal of physics. Condensed matter : an Institute of Physics journal* **2006**, *18*, L347–53.
- [38] Ouldridge, T. E.; Johnston, I. G.; Louis, A. a.; Doye, J. P. K. *The Journal of Chemical Physics* **2009**, *130*, 065101.
- [39] Araque, J. C.; Panagiotopoulos, A. Z.; Robert, M. A. *The Journal of Chemical Physics* **2011**, *134*, 165103.
- [40] Linak, M. C.; Tourdot, R.; Dorfman, K. D. *The Journal of Chemical Physics* **2011**, *135*, 205102.
- [41] Svaneborg, C. *Computer Physics Communications* **2012**, *183*, 1793–1802.
- [42] Drukker, K.; Wu, G.; Schatz, G. C. *The Journal of Chemical Physics* **2001**, *114*, 579.
- [43] Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. *The Journal of Chemical Physics* **2010**, *132*, 035105.
- [44] Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. *Journal of Chemical Theory and Computation* **2010**, *6*, 1711–1725.
- [45] You, M.; Chen, Y.; Zhang, X.; Liu, H.; Wang, R.; Wang, K.; Williams, K. R.; Tan, W. *Angewandte Chemie (International ed. in English)* **2012**, *51*, 2457–60.
- [46] Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; Record, M. T. *Biochemistry* **1999**, *38*, 8409–22.
- [47] Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. *Physical Review Letters* **2010**, *104*, 178101.
- [48] Ouldridge, T. E.; Hoare, R. L.; Louis, A. A.; Doye, J. P. K.; Bath, J.; Turberfield, A. J. *ACS Nano* **2013**, *7*, 2479–90.
- [49] Šulc, P.; Ouldridge, T. E.; Romano, F.; Doye, J. P. K.; Louis, A. A. *Natural Computing* **2013**.
- [50] Ouldridge, T. E.; Sulc, P.; Romano, F.; Doye, J. P. K.; Louis, A. A. *Nucleic acids research* **2013**, *41*, 8886–95.
- [51] Šulc, P.; Romano, F.; Ouldridge, T. E.; Rovigatti, L.; Doye, J. P. K.; Louis, A. A. *The Journal of Chemical Physics* **2012**, *137*, 135101.
- [52] Ouldridge, T. E.; Ph.D. thesis; University of Oxford; 2011.
- [53] von Meyenn, K. *Berichte zur Wissenschaftsgeschichte* **1982**, *5*, 263–264.

- [54] Doye, J. P. K.; *Lecture: Applications of Statistical Mechanics: Molecular Simulation Methods in Chemistry*; 2014.
- [55] Satoh, A. *Journal of Colloid and Interface Science* **1992**, *150*, 461–472.
- [56] Whitelam, S.; Geissler, P. L. *The Journal of chemical physics* **2007**, *127*, 154101.
- [57] Whitelam, S.; Feng, E. H.; Hagan, M. F.; Geissler, P. L. *Soft Matter* **2009**, *5*, 1251.
- [58] Vlugt, T. J. H. *Introduction to Molecular Simulation and Statistical Thermodynamics*, 2009.
- [59] Verlet, L. *Physical Review* **1967**, *159*, 98–103.
- [60] Swope, W. C. *The Journal of Chemical Physics* **1982**, *76*, 637.
- [61] Russo, J.; Tartaglia, P.; Sciortino, F. *The Journal of Chemical Physics* **2009**, *131*, 014504.
- [62] Schreck, J. S.; Ouldridge, T. E.; Romano, F.; Sulc, P.; Shaw, L.; Louis, A. A.; Doye, J. P. K. **2014**, 1–19.
- [63] Lapham, Jon and Rife, JasonP. and Moore, PeterB. and Crothers, D. *Journal of Biomolecular NMR* **1997**, *10*, 255–262.