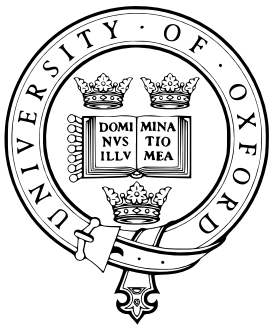

Simulating large DNA nanostructures with a coarse-grained model

Benedict E. K. Snodin

Hertford College
University of Oxford

Supervisor: Professor Jonathan Doye

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Physical & Theoretical Chemistry

Trinity 2015

Abstract

Simulating large DNA nanostructures with a coarse-grained model
Benedict E. K. Snodin
Hertford College, University of Oxford
A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Physical & Theoretical Chemistry
Trinity 2015

In this thesis we investigate the self-assembly and structural properties of DNA origamis, which are large-scale DNA nanostructures comprising thousands of DNA base pairs, using the coarse-grained model of DNA called oxDNA.

We use simulations of oxDNA, with both a "brute-force" approach and with a rare-event method, to probe how the origamis self-assemble from single strands of DNA in solution. With the brute-force approach we are able to simulate the complete assembly of a small 384-base-pair origami at a high strand concentration, and identify kinetic traps under certain conditions, as well as cooperativity between staple strands and optimal assembly windows. These findings help to rationalise certain successful design strategies. Using the rare-event method we are able to use the lower strand concentrations typical of experimental conditions and make more quantitative measurements, identifying sequential staple binding for our origami design at these conditions, and probing how a staple binds each of its domains.

We introduce oxDNA2, an enhanced version of oxDNA, which has improved predictive power for origami structure, as well as other improvements such as salt dependence. We verify that the model reproduces well the structural properties of Holliday junctions, which are a vital feature of DNA origamis. We then use the model to investigate the basic structural properties of 2D and 3D origamis, characterising their generic structural features such as, for 2D structures, the "weave pattern," where adjacent double helices push apart away from the junctions, and "corrugation," a systematic, out-of-plane bending of the double helices. We find a good agreement with experiment where data is available.

Contents

1 Introduction **5**
- 1.1 DNA Nanotechnology . . . 5
- 1.2 Systematic studies in DNA nanotechnology . . . 9
  - 1.2.1 Theoretical models of DNA . . . 9
- 1.3 The oxDNA model . . . 12
- 1.4 Thesis outline . . . 16

2 Simulation Methods **17**
- 2.1 Molecular dynamics . . . 17
- 2.2 Monte Carlo simulation . . . 19
  - 2.2.1 The virtual move Monte Carlo algorithm . . . 21
- 2.3 Umbrella sampling . . . 22
- 2.4 Forward flux sampling . . . 23

3 Direct simulation of the self-assembly of a small DNA origami **29**
- 3.1 Studies of origami self-assembly . . . 29
- 3.2 System and model . . . 31
  - 3.2.1 Simulation details and time scale estimation . . . 32
  - 3.2.2 The origami design . . . 33
- 3.3 Assembly with a large staple excess: 65 °C . . . 38
  - 3.3.1 Local features of staple binding . . . 38
  - 3.3.2 Global features of origami self-assembly . . . 42
- 3.4 Barriers to complete assembly . . . 43
  - 3.4.1 Recovering from blocking . . . 47
  - 3.4.2 Blocking in real origami self-assembly . . . 50
- 3.5 Excess-staple assembly annealed at a higher temperature: 70 °C . . . 52
  - 3.5.1 Resolution of blocking . . . 52
  - 3.5.2 Formation of new blocks . . . 53
- 3.6 Assembly with a stoichiometric strand mixture . . . 56
  - 3.6.1 General features . . . 59
  - 3.6.2 Optimal assembly window . . . 59
- 3.7 Conclusions . . . 60

4 Simulation of the self-assembly of a small DNA origami using forward flux sampling **67**
- 4.1 The model system . . . 67
- 4.2 Splitting up the assembly pathway . . . 69
- 4.3 Results . . . 74
  - 4.3.1 Average rates between assembly states . . . 75

4.3.2 Order of binding staple domains . . . 81
4.3.3 Cooperativity between staples in binding rates . . . 89
4.4 Sampling and systematic error . . . 95
4.5 Summary and conclusions . . . 98

5 Introducing an updated oxDNA model 105
5.1 A summary of the changes introduced with the new model . . . 105
5.2 Introducing different widths for major and minor DNA grooves . . . 107
5.3 Effective electrostatic interactions . . . 110
5.4 Improving structure prediction for large-scale DNA objects . . . 118
5.5 AA/TT sequence dependence . . . 121
5.6 The oxDNA2 model . . . 123
5.7 Physical properties of DNA in the oxDNA2 model . . . 124
5.8 Summary . . . 133

6 Structural properties of DNA origami 139
6.1 Previous studies of the structure of DNA objects . . . 139
6.2 Characterising Holliday junction structure in oxDNA2 . . . 142
6.2.1 Comparison of an oxDNA2 junction with experimental results . 145
6.3 Structural properties of a 2D origami . . . 150
6.3.1 The weave pattern . . . 151
6.3.2 Corrugation in 2D origamis . . . 157
6.4 Structure of the 3D "pointer" origami . . . 161
6.5 Summary . . . 165

7 Conclusion 167

References 171

Appendices

A The oxDNA2 potential 180
A.1 Introducing an electrostatic term into the model potential . . . 180
A.2 Modifying the average pitch of the model duplex . . . 181
A.3 Modifying the average coaxial stacking term in the oxDNA potential . . . 181
A.4 Modifying the hydrogen bonding and stacking interaction strengths . . . 185

B Supplementary information 188
B.1 Standard MD simulation parameters . . . 188
B.2 DNA strand sequences used for the direct simulation of origami assembly 188
B.3 Supplementary information for Chapter 4 . . . 188
B.3.1 Strand sequences used for the simulation of origami assembly using FFS . . . 188
B.3.2 Number of possible substates for an assembly state . . . 191
B.3.3 Order parameters and simulation parameters for the origami assembly FFS simulations . . . 192

B.3.4 Assembly substate breakdown for backwards transitions not shown in Chapter 4 . . . 196
B.3.5 Creating the average rate plots . . . 196
B.4 Supplementary information for Chapter 5 . . . 198
B.4.1 Measuring the global twist of an origami . . . 198
B.4.2 Measuring the torsional stiffness of a duplex in the oxDNA model 199
B.4.3 Details of the simulation protocols . . . 200
B.5 Supplementary information for Chapter 6 . . . 203
B.5.1 Holliday junction free-energy landscape . . . 203
B.5.2 Simulating 2D origami structure . . . 205
B.5.3 Calculating an average structure . . . 205
B.5.4 Comparison to experiment for the 3D "pointer" origami structure 207

Acknowledgements

The completion of this thesis would not have been possible without the help and support of many others. Firstly, I would like to thank my supervisor Jon Doye who has provided me with support and guidance for the ambitious projects he has set me during the course of my DPhil. I am also greatly indebted to Flavio Romano, who provided me with close guidance during my first couple of years as a researcher, and has remained a guiding presence for the entirety of my time in the Doye group. Others have inspired, enlightened and entertained me during long days in the office, including Aleks Reinhardt, John Schreck, Dennis Palagin, Majid Mosayebi, Ryan Harrison and Domen Prešern. I must also thank close collaborators Tom Ouldridge, Petr Šulc and Lorenzo Rovigatti, who in addition to guiding me and working with me on research projects (and being a pleasure to work with) are responsible for creating and improving the oxDNA model and simulation code on which the work of this thesis is founded. I also thank my co-supervisor Ard Louis for his guidance and encouragement during my time in the Doye group.

I am indebted to the friends and family who have supported me during my studies for my DPhil. I thank the friends I have been lucky enough to have during the time I have been in Oxford, who have patiently tolerated my desire for sports, games, and discussions on various topics from the significant to the trivial. I would also like to thank Lexie Joon for her encouragement and moral support over the last few months. And I thank my wonderful parents and sister for their continued support.

List of abbreviations

The following abbreviations are used in this thesis.

AFM	atomic force microscopy
ATP	adenosine triphosphate
CPU	central processing unit
DH	Debye-Hückel
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
FENE	finitely extensible nonlinear elastic
FFS	forward flux sampling
FRET	Förster resonance energy transfer
GPU	graphical processing unit
MC	Monte Carlo
MD	molecular dynamics
PAINT	point accumulation for imaging in nanoscale topography

RMSD	root-mean-squared displacement
SAXS	small-angle X-ray scattering
TEM	transmission electron microscopy
VMMC	virtual move Monte Carlo
WHAM	weighted histogram analysis method

1 Introduction

1.1 DNA Nanotechnology

Deoxyribonucleic acid (DNA) performs the crucial function of storing genetic information in living organisms. It is made up of repeating units called nucleotides, each of which consists of a sugar and phosphate backbone plus a base (either Adenine (A), Guanine (G), Thymine (T), or Cytosine (C)) attached to the sugar. Watson-Crick base-pairing, A with T and G with C, along with planar stacking interactions between bases and the constraints of the backbone, leads to the formation of the well-known double-helical structure of DNA.

The specificity of Watson-Crick base pairing permits the rational design of DNA objects for which the intended structure is the global free-energy minimum. This property was first exploited by Ned Seeman in 1982,[1] when he proposed linking DNA junctions together to form a lattice, signalling the birth of DNA nanotechnology. After initially steady progress, the field has grown rapidly in recent years,[2] with Paul Rothemund achieving a notable advance in 2006 when he first introduced a powerful and flexible method of DNA nanostructure design, which he christened DNA origami.[3] This approach represented the first reliable method for making large DNA nanostructures, consisting of thousands of DNA base pairs. Since then, alternative design paradigms for creating DNA nanostructures of a similar scale have been introduced, notably the "DNA brick" scheme[4] and a wire-frame approach.[5–7]

However, due to its reliability and flexibility, the DNA origami method continues

to be a staple of the DNA nanotechnology community, and the DNA origami self-assembly process and the resulting DNA nanostructures will be a particular focus of the work presented in this thesis. This method of DNA self-assembly involves folding a very long DNA strand (typically around 7.5 kilo-bases), termed the *scaffold* strand, by the action of many short strands, termed *staple* strands, each of which bind to two or more non-adjacent regions of the scaffold strand, pulling them close together (see Fig. 1.1(a)). The structures are composed of antiparallel double helices joined by *crossovers*, where staple strands cross from one region of the scaffold to another (see Fig. 1.1(b)). The method was used first to generate 2D structures,[3] and then to create 3D shapes,[8] including structures with a designed curve or twist,[9;10] as well as tensegrity structures.[11] Fig. 1.2 shows examples of these structures (as well as other non-origami DNA nanostructures), and illustrates the growth of the field measured by citation counts.

The ease and reliability of assembly, flexibility of design and addressability of DNA origami structures has opened up a variety of practical applications, which are only beginning to be explored.[13;14] A common use is as a substrate for single-molecule studies,[15] for example to probe hairpin dynamics[16] or the motion of walkers.[17] Other promising applications include use as drug delivery vehicles,[18;19] as artificial membrane channels[20;21] and nanopores,[22] and as stiff handles for single-molecule mechanical experiments with optical tweezers.[23]

Although not the focus of this thesis, the creation and study of smaller-scale DNA objects is another important aspect of DNA nanotechnology. An area of great interest is the study of DNA walkers, where a "walker" made of a few DNA strands steps along a track of DNA.[17;24;25] The walker is driven by a fuel source, which is often a species of

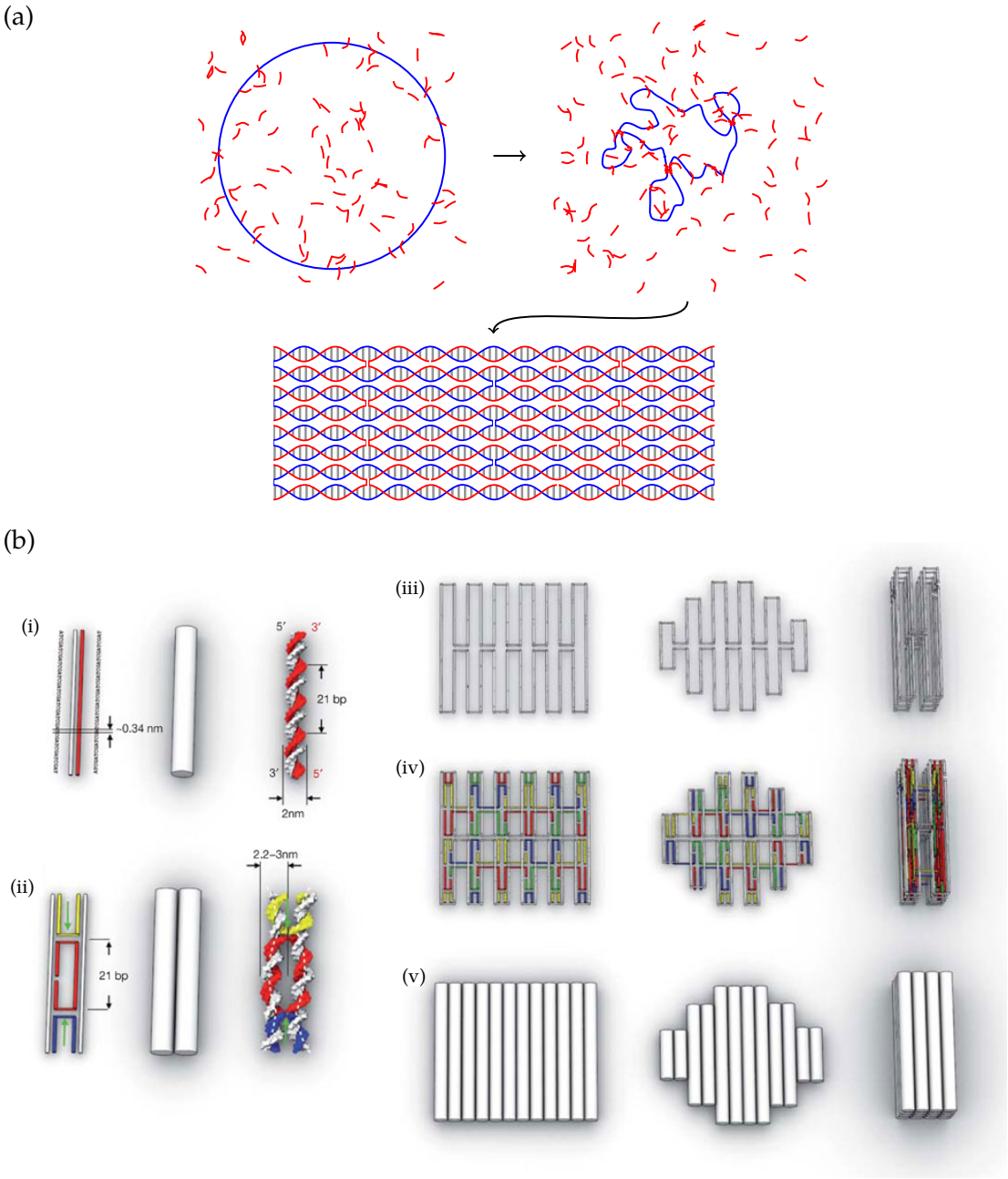

Figure 1.1: (a) A schematic showing the self-assembly process for a DNA origami structure. In the top left panel, we start with a large scaffold strand (blue) and many staple strands (red). In the top right panel, the staples are shown beginning to fold the scaffold strand. The final structure, in this case a flat rectangle, is shown in the bottom panel. (b) An illustration of DNA origami design principles, reproduced with permission from Ref. 12. (i) Shows a double helix in three different representations, and (ii) shows two double helices joined by two double crossovers, which are separated by 21 base pairs, corresponding roughly to two turns of a DNA double helix. This structure can be considered as a subunit of a DNA origami. (iii) Shows the scaffold strand routing for three different origami designs, while (iv) also shows staple strands, and (v) shows a "cylinder" representation illustrating the overall shapes.

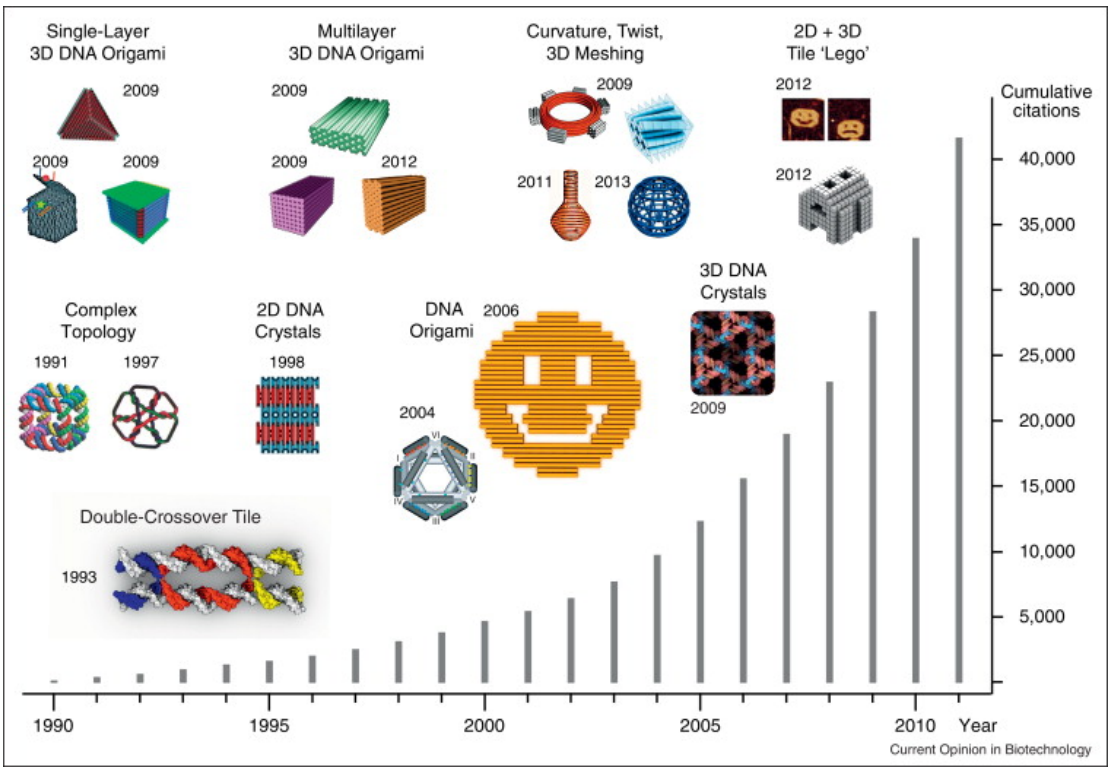

Figure 1.2: This image, taken with permission from Ref. 2, shows examples of the DNA nanostructures that have been fabricated over the years. Examples of DNA origami structures are labelled "DNA Origami," "Single-Layer 3D DNA Origami," "Multilayer 3D DNA Origami," and "Curvature, Twist, 3D Meshing." The graph shows the growth in cumulative citations in the field over time.

DNA strand added to the solution. This scheme is analogous to the biological kinesin walker, which walks along microtubules in the cell, driven by ATP. Another important subfield is DNA computing,[26;27] where DNA is used to perform logical operations[28] and store data.[29] Finally, many other small-scale nanostructures and machines have been created and studied, including DNA tweezers,[30] DNA tetrahedrons,[31;32] and novel DNA structures.[33]

1.2 Systematic studies in DNA nanotechnology

Despite, or perhaps because of, the great advances made in the field of DNA nanotechnology in recent years, there is a lot of room to develop a better understanding of the processes underlying the assembly, structural and mechanical properties, and actuation of recently-developed DNA nanodevices. Several experimental groups have conducted careful studies in an effort to better understand the DNA origami assembly process[34–38] and DNA origami structure.[9;39;40]

1.2.1 Theoretical models of DNA

Theoretical and computational approaches to modelling DNA nanotechnology have also begun to play a role in improving our understanding of DNA nanostructures and machines. At the finest level of detail, quantum chemistry calculations have been used to study the interactions between nucleotides,[41–43] although the high computational cost of such an approach limits these methods to interactions between nearest-neighbour base pairs in vacuum. Classical all-atom molecular dynamics (MD) approaches, where every atom of DNA and the surrounding solvent is modelled as a particle subject to interactions with the surrounding atoms, have been widely employed to study small DNA motifs,[44;45] and have recently been applied to larger DNA systems.[46–49]

However, simulating rare-event processes such as the breaking or formation of base pairs remains a challenge with these models, with μs time scales being the limit of what is currently accessible.[49] At the other end of the scale, theoretical approaches have been developed to understand certain large-scale properties of DNA. These include the wormlike-chain model, which treats DNA as a continuously flexible polymer.[50] While such models can provide useful insights into the physical properties of DNA, they are not detailed enough, by design, to address key processes such as duplex formation.

The middle ground between these two extremes is occupied by coarse-grained DNA models. Such models integrate out many of the degrees of freedom of the DNA nucleotide, and often do not explicitly represent the solvent molecules; these approximations inevitably imply a compromise between accuracy, generality and computational efficiency, so that care must be taken in applying these models to a given problem. However, the simplified picture presented by such an approach can be a strength, as, in addition to greatly increasing the time scale and number of nucleotides that can be studied, it can allow one to understand the generic physics governing the system more easily.

Models that represent a partially assembled origami on the domain level (a partially formed origami can conceptually be broken into domains, with each domain consisting of a continuous double helix in the completed structure) have been used to probe origami self-assembly with some success.[38;51] Other models have been used to study origami structure,[52] notably CanDo,[53;54] a specialised structural model which treats DNA origami as a set of double helices, modelled as linear elastic rods, joined by four-way junctions, modelled as somewhat flexible connections between the rods.

More detailed models have been developed with a focus on B-DNA. The model of

Savelyev and Papoian[55], which treats DNA only in its double-stranded state, represents DNA with one bead per nucleotide, each of which interacts with the 11 nearest nucleotides on the partner strand, and has been used to study nucleosome unwrapping.[56] In addition, the Nordenskiöld group have developed a model parameterised from all-atom MD simulations and representing double-stranded DNA with five beads for every two base pairs.[57] The model achieved a good agreement with experimental salt-dependent double-stranded persistence length measurements, in the physiological salt region and below. While these two models are well-suited to the biophysical systems they were developed to study, where DNA is exclusively in its double-stranded state, they are not readily applicable to DNA nanotechnology set-ups, which usually involve transitions to and from single- to double-stranded DNA, and more complex DNA motifs than simple B-DNA.

Another model with a focus on biophysics was recently developed by the Dorfman group[58] to treat single- and double-stranded DNA with particular attention to Hoogsteen bonding, using three interaction sites per nucleotide and an anisotropic potential. They were able to use it to study a thrombin aptamer, which contains G-quartets, and strand invasion during triplex formation, although these exotic structures are usually not important for DNA nanotechnology. Meanwhile, Araque *et al.* have introduced a lattice-based, one-bead-per-nucleotide model of DNA,[59] and used it to study hybridisation. This approach implies a fairly low computational cost, which in principle allows hybridization in complex systems to be studied, at the expense of realistic structural properties which are important for detailed DNA nanotechnology studies.

The Plotkin group have developed a three-sites-per-nucleotide model parameterised from all-atom simulations,[60] which they used to study the structural and mechan-

ical properties of single- and double-stranded DNA within the model as a function of temperature, salt concentration, and sequence. More recently, the Aksimentiev group introduced a model of single-stranded DNA, with an interaction site on each sugar/base group and phosphate group.[61] The model was designed to reproduce the single-stranded DNA structure seen in all-atom MD calculations, and was able to reproduce force-extension data for single-stranded DNA. Although it could potentially be extended to model B-DNA, this has not yet been done. In addition, a coarse-grained RNA model called HiRe-RNA, which uses a detailed representation with 6 or 7 interacting beads per nucleotide, was recently extended for the study of DNA systems, and was used to produce melting curves for a short DNA duplex.[62]

Finally, the 3SPN model[63] represents DNA with three independently mobile interaction sites per nucleotide, representing the base, sugar and phosphate groups, with an interaction potential that includes three-body terms. It has been used, for example, to study the wrapping of a DNA duplex around a nucleosome,[64] as well as hybridisation,[65] a process highly relevant to DNA nanotechnology. Of the models of DNA described here, 3SPN is probably the most suitable for the broadest range of investigations of DNA nanotechnology systems. However, it has not yet been applied to the systems of very large numbers of DNA base-pairs required for large-scale DNA nanotechnology, the focus of this thesis.

1.3 The oxDNA model

The results presented in this thesis were obtained using a coarse-grained model of DNA called oxDNA.[66;67] Briefly, the model represents a DNA nucleotide as a rigid body with three interaction sites, which interact with each other through a potential with effective, highly anisotropic terms. The model is designed with a heuristic, "top-

down" approach, with a focus on reproducing well-known properties of DNA (such as the helical structure of the B-DNA duplex) and experimental results (such as duplex melting temperatures), rather than, for example, building the model up by integrating out details from an all-atom representation. This treatment is sufficient to obtain good agreement with experimental data on the structural, mechanical and especially the thermodynamic properties of single- and double-stranded DNA. Reproduction of these fundamental biophysical properties of DNA allows the model to be applied in a wide variety of contexts without further fitting or adjustment of the parameters.

In particular, oxDNA has provided key insights into many different processes relevant to DNA nanotechnology[68–76] and biophysics,[77–81] and importantly, has also been shown to provide direct agreement with experimentally measured properties to which the model has not been fitted,[82] for systems such as DNA overstretching,[81] a two-footed DNA walker,[71] and toehold-mediated strand displacement.[75;83] Therefore, it seems reasonable to suggest that the model has produced greater insights into processes relevant to DNA nanotechnology and biophysics than the other available coarse-grained DNA models, which were described in Section 1.2.1. In addition, as the potential is relatively simple, simulations of very large systems of up to the order of 10,000 base pairs using a single graphical processing unit (GPU) are achievable with oxDNA.[84] This feature is essential for the studies of large-scale DNA nanotechnology systems that will be described in this thesis.

As mentioned above, each nucleotide is represented by a rigid body with three interaction sites; each strand of DNA is represented by a chain of these rigid bodies. The nucleotides interact through an effective, pairwise potential designed to capture hydrogen bonding between base pairs, stacking, backbone connectivity and excluded

volume interactions. The stacking and hydrogen-bonding terms have a complex form with angular dependence, to capture the geometric constraints of the interactions in DNA. The potential can be written as a sum of interaction terms between pairs of nucleotides

$$V = \sum_{\text{nearest neighbours}} (V_{\text{backbone}} + V_{\text{stack}} + V'_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross stack}} + V_{\text{exc}} + V_{\text{coax stack}}), \tag{1.1}$$

where the sum over nearest neighbours refers to adjacent nucleotides on a DNA strand. Further details, including the forms of the interaction terms, are given in Refs. 66 and 67.[1] The oxDNA representation of a DNA strand and the different interaction terms are illustrated in Fig. 1.3. There is no explicit solvent, with the effect of the solvent being incorporated into the interactions between the nucleotides and into the thermostat.

Due to the complexity of the model, the only practical way to obtain the properties of a particular system as predicted by the model is to use molecular simulations.[86] A computer code has been written implementing simulations of the oxDNA model with a variety of simulation algorithms. The work presented in this thesis mostly makes use of the virtual move Monte Carlo[87;88] (VMMC) and molecular dynamics (MD) algorithms, with the MD algorithm using an Andersen-like thermostat, in order to capture the diffusive motion of DNA in solution. Further details on simulation algorithms are given in Chapter 2. The computer code implementing these algorithms for the oxDNA model is publicly available from http://dna.physics.ox.ac.uk.

The computational expense of simulating the model is made more manageable by the short-ranged, pairwise (i.e. without any costly three-body or higher order terms)

[1] ref. 66 gives a description of a version of the model without the coaxial stacking term $V_{\text{coaxstack}}$, while ref. 67 is a PhD thesis which gives a complete description of the version of the model described here.

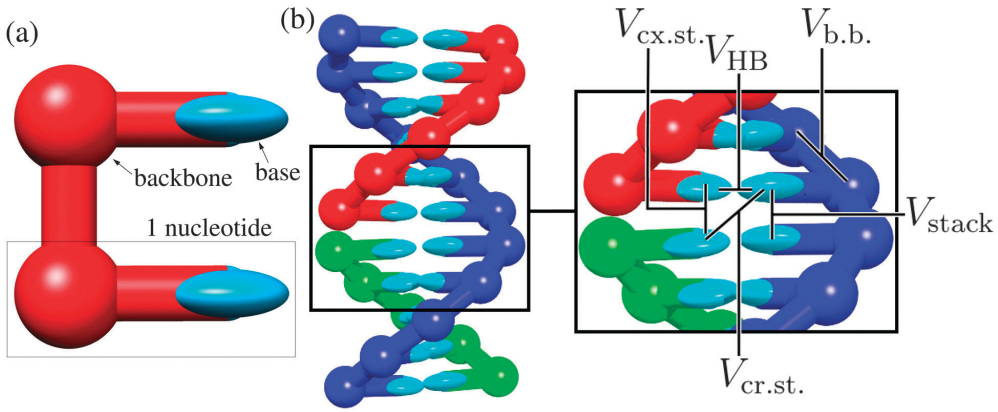

Figure 1.3: (a) A strand composed of two nucleotides and (b) a nicked duplex, as represented by the oxDNA model. Reproduced from ref. 82. Configurations in this and subsequent figures have been visualised using the USCF Chimera package.[85]

nature of the model's potential. The task of specifying the complex local behaviour of the model DNA with this kind of potential is made easier due to the fact that the multiple interactions sites representing each nucleotide are constrained to be part of the same rigid body. In addition, using a rigid body means that the extra computational effort of simulating the high frequency vibrations within the nucleotide is avoided.

We should also mention the shortcomings of the oxDNA model. A new version of the model, which is called oxDNA2[89] and which we will describe in Chapter 5, addresses some of these. The most readily apparent improvements are the introduction of differentiated major and minor grooves in B-DNA, and the introduction of an explicit electrostatic interaction, which means that DNA can be simulated at different salt concentrations. However, both the old and new versions of the model only treat canonical base pairing (i.e. A with T and G with C), and the model DNA cannot form exotic structures such as triplexes, quadruplexes, and A-DNA duplexes. In addition, the model has not yet been extended to deal with interactions with other molecules, such as RNA (although an oxDNA-inspired RNA model, oxRNA, has recently been devised[90]), metal nanoparticles, or proteins.

1.4 Thesis outline

In this thesis we are primarily concerned with studying the self-assembly and physical properties of large-scale DNA nanostructures, particularly DNA origami structures, in the context of DNA nanotechnology. In Chapter 2 we introduce the basics of the simulation methods used for this work. In Chapter 3 we study the self-assembly of a small DNA origami using direct simulation, and in Chapter 4 we study the assembly of a similar DNA origami design using a rare-events simulation method in order to gain a more detailed insight. In Chapter 5 we introduce a new, extended version of the oxDNA model, which among other improvements provides a more faithful description of the structural properties of large-scale DNA objects. Finally, in Chapter 6 we use this new model to investigate DNA origami structure.

The work described in Chapters 3 and 5 have been published.[89;91]

Although I was the largest contributor to the work reported on an extended oxDNA model in Chapter 5, the work was a collaborative effort. The contributions of other group members are outlined at the end of that chapter.

2 Simulation Methods

Although many theoretical models can be solved analytically, models are often too complicated to be solved in this way. Instead, we can turn to computers to give us numerical solutions. Specifically, given a physical model and a system (a set of particles that interact through the model's potential energy function), simulation algorithms give us a way to explore the phase space of that system numerically. Many models, including oxDNA, are designed with numerical simulation in mind.

In this chapter we discuss the simulation methods used for the work presented in this thesis. We discuss the molecular dynamics algorithm, the virtual move Monte Carlo algorithm, and two techniques for speeding up sampling, namely umbrella sampling and forward flux sampling.

2.1 Molecular dynamics

In molecular dynamics[86] (MD), the particles in the system move in discrete time steps according to Newton's equations of motion. At the beginning of each time step the particles each have well-defined positions, momenta and forces, with the forces being computed from the gradient of the potential energy function with respect to the positions, and possibly orientations, of the particles. There are a variety of algorithms commonly used to evolve the system at each time step, using the positions, momenta and forces; the simulation code used for this work uses the velocity Verlet algorithm.[92]

The most basic MD explores the phase space of the simulated system for the NVE

ensemble. To simulate the NVT ensemble, one commonly introduces a thermostat which mimics a coupling of the system to a heat bath, and the results presented in this thesis were obtained using an MD code with an Andersen-like thermostat for this purpose. This thermostat works by allowing the system to evolve in the NVE ensemble (i.e. pure MD) for a small number of steps, and then refreshing the velocities of a randomly chosen subset of the particles (each particle has the same, fixed probability of being chosen for the velocity refresh), replacing them with velocities taken from a Maxwell-Boltzmann velocity distribution at the simulation temperature. Importantly, on time scales much longer than the refresh interval the algorithm generates diffusive motion.

For the work presented in this thesis, an important advantage of the MD class of algorithms compared to Monte Carlo (MC) algorithms (which are described below) is that it is possible to run very efficient MD simulations of large systems on GPUs thanks to the parallel nature of the MD simulation step. With MD, it is possible to compute the forces between each pair of interacting particles simultaneously, which for a complex interaction potential like that of oxDNA is by far the most computationally demanding part of the algorithm.[84] Although parallel MC algorithms have been suggested,[93] they do not allow nearly as many simultaneous computations as a typical MD algorithm for a system with a large number of interacting particles.

In this thesis, we use MD simulations on GPUs for studying large systems of around 500 nucleotides or more, and MD simulations on CPUs for studying the structural and mechanical properties of smaller DNA systems. The former applies to more or less all of the systems studied in Chapters 3, 4 and 6, the simulation of which would have been unfeasible without GPUs.

2.2 Monte Carlo simulation

Metropolis Monte Carlo simulation is a method for exploring the phase space of a system with Boltzmann sampling.[94] That is, the algorithm will spend an amount of time in each state of the system (defined by the position coordinates of all the particles, $\mathbf{r}_i$) proportional to the Boltzmann weight $e^{-V/(k_B T)}$ of that state, where V is the potential energy, k_B is the Boltzmann factor, and T is the temperature.[1] This is achieved with a Markov chain Monte Carlo approach, where the system begins in a particular state and, during each simulation step, switches to a new state, or possibly stays in the same state, according to some rules. This process of possibly switching to a new state is repeated until sufficient sampling of the system is achieved. Importantly, the rules governing which state is chosen next depend only on the current state of the system. Note that MC algorithms are a powerful way of studying the equilibrium properties of a system, but, unlike when MD is used, the system does not necessarily evolve in a physically realistic way.

The scheme is:

1. Start with a state c.
2. Randomly generate a trial state t with *trial probability* $T(c \to t)$ using some rules based on the current state c.
3. Accept a move to the trial state t with *acceptance probability* $A(c \to t)$. If the move is not accepted, stay in the same state.

[1] In principle we could just as well use an algorithm to carry out (say) uniform sampling of the states, and then apply the Boltzmann weight before computing any averages. Given an infinite amount of computing power this will give an identical result to Boltzmann sampling. The point is that in order to sample the phase space efficiently, and therefore calculate things about the system with the least possible computational effort, we should concentrate on sampling the parts of phase space that the system occupies most often, rather than wasting time sampling parts of the system that are almost never occupied.

4. Repeat.

It is usual to choose the acceptance probability so that "detailed balance" is satisfied. Imagine that we have N (some very large number) samples of the system, and denote the number of times we sampled a particular state s (the population of the state s) as $P(s)$. For an equilibrium distribution of populations, that is, a Boltzmann-weighted distribution, we require that the net flow into a state must be the same as the net flow out of it; that is how we know we are in equilibrium. We can write this as

$$P(s)\Sigma_i A(s \to i)T(s \to i) = \Sigma_i P(i)A(i \to s)T(i \to s), \tag{2.1}$$

where the sums are over the neighbouring states with non-zero transition probabilities to/from the state s. This condition is called "balance." One way of satisfying this is to demand that a stricter condition is satisfied:

$$P(s)T(s \to i)A(s \to i) = P(i)T(i \to s)A(i \to s) \tag{2.2}$$

for every state s and i. This equation defines detailed balance. Although detailed balance is a stronger condition than balance, in practice it is easier to ensure that detailed balance holds. We usually have symmetric trial probabilities, so that $T(s \to i) = T(i \to s)$, and these terms cancel. We must then choose a form for $A(i \to s)$ that satisfies detailed balance. There are many possibilities, but a common choice, introduced with the original Metropolis scheme,[94] is

$$A(i \to s) = \min(1, e^{-(V_s - V_i)/(k_B T)}). \tag{2.3}$$

This means that we always accept a trial move if it will take the system to a lower energy state, and we accept a trial move to a higher energy state with a probability that decreases as the energy difference increases. Each trial move is typically a rotation or

translation of a randomly chosen particle.

2.2.1 The virtual move Monte Carlo algorithm

The Monte Carlo algorithm used for the results presented in this thesis is one called virtual move Monte Carlo. We use the variant of this algorithm described in Ref. 87 and Ref. 88. The algorithm generates trial moves involving clusters of particles. An initial particle is selected at random and then a cluster is recursively formed around it, with nearby particles each added with a certain probability based on the anticipated change in potential energy after the trial move. The trial move is either a rotation of the cluster around the initial particle or a translation of the cluster.

The VMMC algorithm is particularly appropriate for simulating small to medium-sized systems (on the order of 10 to 100 nucleotides) with oxDNA. These involve isolated clusters of strongly interacting particles, i.e. DNA complexes (usually single or double strands) consisting of chains of nucleotides bound together through FENE springs, base pairing and stacking interactions. VMMC allows the complexes to diffuse around the box very quickly, and can potentially speed up internal fluctuations, offering greatly accelerated equilibration compared to Metropolis Monte Carlo.

In this thesis, we use VMMC to simulate systems of up to a few hundred nucleotides when we do not require realistic dynamics: mostly simulations to find melting temperatures of duplexes, hairpins, and other small DNA complexes in Chapter 5. Otherwise, we use the MD algorithm as described earlier in this chapter.

2.3 Umbrella sampling

Umbrella sampling is a simulation method used to accelerate sampling of the phase space of a system, especially when there are several (meta-)stable states which are separated by one or more free energy barriers. Essentially, the method works by flattening the barrier(s) by applying a bias to the potential with the purpose of enhancing the sampling of otherwise free-energetically unfavourable states.

First, the phase space of the system is divided into discrete states using an order parameter. Then the states corresponding to each of the order parameter values are given different weights, usually with the intention of equally sampling the states[2].

To see how the method works, consider an unbiased population of states $p_u(n)$, and a set of weights $w(n)$, such that the biased population of states $p_b(n)$ is given by

$$p_b(n) = p_u(n)w(n), \tag{2.4}$$

where $n(\mathbf{r}_i)$ gives the value of the order parameter for a given state defined by $\mathbf{r}_i$, and $\mathbf{r}_i$ gives the positions and orientations of all the particles in the system. For example, this biased sampling can be achieved by simulating the system with a modified energy function

$$E_{\text{modified}}(\mathbf{r}_i, n) = E_{\text{model}}(\mathbf{r}_i) + E_{\text{bias}}(n), \tag{2.5}$$

where $E_{\text{bias}}(n) = -\log(w(n))$. After the simulation has been run, sampling the biased population distribution $p_b(n)$, the unbiased population can then be recovered simply by rearranging Eq. 2.4:

$$p_u(n) = p_b(n)/w(n). \tag{2.6}$$

[2]Equal sampling is not always the desired outcome; umbrella sampling can also be used to achieve something other than equal sampling if desired.

The unbiased population distribution can then be used to compute properly weighted average properties of the system. Of course, one does not usually know the best umbrella weights in advance (that would imply knowing the free energy landscape, usually the very thing one is trying to find!), so in practice an educated first guess of the weights is made, and subsequently updated either using an automated approach or by hand.

In this thesis we use umbrella sampling to speed up simulations of thermodynamics with VMMC in Chapter 5.

2.4 Forward flux sampling

Forward flux sampling[95] is a simulation method designed to sample rare transitions with a stable initial state (known as the A-state) and a specified final state (known as the B-state). The technique uses a series of interfaces between the A-state and the B-state to calculate a rate and transition path ensemble for the transition. We require that the A-state is stable in the sense that the time the transition takes (i.e. the time between leaving the A-state and arriving at the B-state) must be much shorter than the average time spent in the A-state. There is no requirement on the stability of the B-state.

To use FFS we need to define an order parameter λ that depends on the state of the system. In our case we define a discrete order parameter. We define the A-state by $\lambda = \lambda_{\mathrm{A}} = \lambda_0$, and the B-state by $\lambda = \lambda_{\mathrm{B}} = \lambda_n$, with the remaining intermediate order parameter values being $\lambda_1, .., \lambda_{n-1}$. We define $\lambda_{i+1} > \lambda_i$ for all i, and ensure the order parameter is defined so that the system always passes through each value in turn in going from the A-state to the B-state (although the system is allowed to go backwards and forwards on the way).

In this thesis we use the FFS variant called direct FFS. In this algorithm, transition

paths are generated simultaneously (usually on separate CPUs or GPUs) using the following procedure, which is also shown schematically in Fig. 2.1:

1. Run "flux" simulations to calculate the flux $\Phi_{\mathrm{A},k}/h_{\mathrm{A}}$ (defined below), with initial configurations drawn from an equilibrium ensemble in the A-state. Use either

   (a) Flux scheme 1: Store a configuration and count a crossing each time the simulation crosses from λ_0 to λ_1 (but not from λ_1 to λ_0).

   (b) Flux scheme 2: Store a configuration and count a crossing each time the simulation enters λ_2 having been in λ_0 more recently than λ_2.

   Scheme 1 is the most commonly used method, but can generate highly correlated configurations. Scheme 2 can reduce the correlation in the saved configurations, but requires more simulation time per configuration saved. We make use of both methods for the results in Chapter 4.

2. Randomly select one of the configurations generated at λ_k ($k = 1$ if flux scheme 1 was used or $k = 2$ if flux scheme 2 was used) and use it to initiate a trial run that stops when either the system reaches λ_{k+1} (a success) or falls back to λ_0 (a failure), with successful configurations at λ_{k+1} being stored. These simulations are called shooting simulations. The probability that a trajectory, coming from the A-state and having just reached λ_k, will reach λ_{k+1} rather than returning to the A-state, $P(\lambda_{k+1}|\lambda_k)$, can be computed from the fraction of trial runs that were successful.

3. Repeat the previous step, this time using stored configurations at λ_{k+1} to initialise shooting simulations that either stop at λ_{k+2} or λ_0.

4. Repeat until λ_n, i.e. the B-state, is reached.

Note that the transition paths generated with this method are branched, i.e. many paths may start from a single configuration at λ_1 or subsequent interfaces.

The rate for the transition, k_{AB}, is computed from this equation:

$$k_{\mathrm{AB}} = \frac{\Phi_{\mathrm{A},k}}{h_{\mathrm{A}}} P(\lambda_n|\lambda_k), \tag{2.7}$$

where $\Phi_{\mathrm{A},k}$ is the number of times trajectories leave the A-state and reach order parameter value λ_k and h_{A} is the total amount of time for which the trajectories were in the A-state more recently than the B-state. λ_k corresponds to λ_1 for flux scheme 1 and λ_2 for flux scheme 2, as above.

$P(\lambda_n|\lambda_k)$ is then computed as a product of the probabilities calculated from the shooting simulations

$$P(\lambda_n|\lambda_k) = \Pi_{i=k}^{n-1} P(\lambda_{i+1}|\lambda_i), \tag{2.8}$$

where $P(\lambda_{i+1}|\lambda_i)$ is the probability that a trajectory, coming from the A-state and having just reached λ_i, will reach λ_{i+1} rather than returning to the A-state.

FFS requires that the simulations are stochastic, as a deterministic simulation would generate identical trajectories from a given starting configuration. In addition, it is notable that FFS requires running each shooting simulation from λ_i either to λ_{i+1} or all the way back to λ_0. In some cases, such as when there is a diffusive barrier or metastable state between the A-state and B-state, the simulations can be very computationally expensive because of the long times required to return to λ_0. One way of working around this is a technique known as pruning, where some proportion of trial runs that fall back by 1 interface are terminated, and the remaining ones given a greater statistical weight to compensate. However, this approach was found not to improve computational efficiency significantly in the cases tested by Allen *et al.*,[95] and we do

not use it for this work.

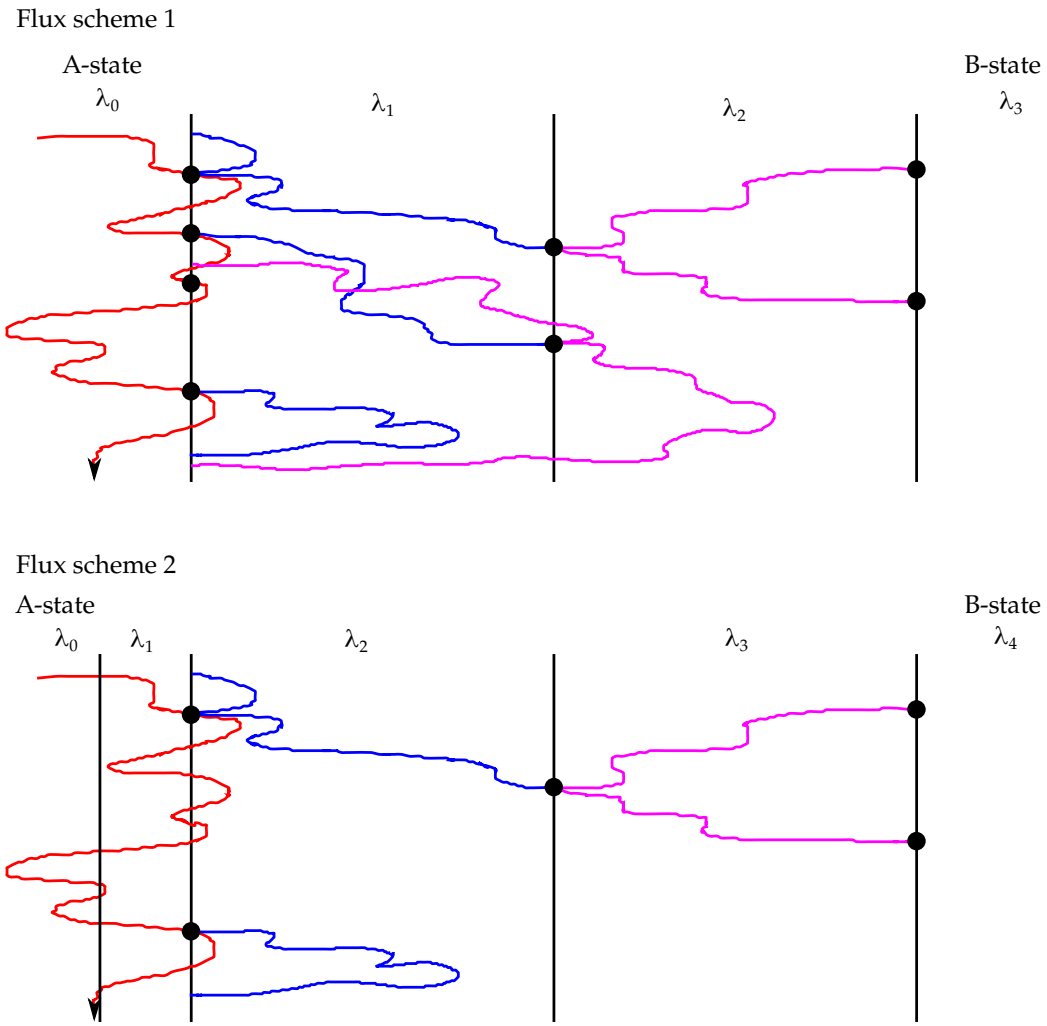

Figure 2.1: A schematic illustration of the FFS algorithm, as well as the difference between the two flux schemes described in the main text. A flux simulation started from the A-state is shown in red. In flux scheme 1, configurations are saved and crossings are counted when the system crosses from λ_0 to λ_1 (black circles at the start of the λ_1 interface). Shooting simulations from the start of λ_1 are shown in blue, and shooting simulations from the start of λ_2 are shown in pink. Black circles at the start of the λ_2 and λ_3 interfaces indicate shooting simulations completing successfully. In flux scheme 2, configurations are saved and crossings counted when the system crosses from λ_1 to λ_2, only if the system was in λ_0 more recently than λ_2.

3 Direct simulation of the self-assembly of a small DNA origami

In this chapter we use oxDNA[1] to investigate the self-assembly of DNA origami. We do this by attempting to simulate directly the assembly of a small model origami structure starting from its single-stranded components. For this origami, we are able to identify kinetic traps that hinder self-assembly as well as conditions under which we observe complete assembly. Although computational constraints limit us to considering a small origami design (384 base pairs rather than the usual ∼7000), this still represents by far the largest DNA nanostructure whose self-assembly has been directly simulated. Furthermore, we are confident that many of the insights we obtain are applicable to full-scale origamis.

3.1 Studies of origami self-assembly

The wide ranging potential applications of DNA origami have motivated efforts to systematically understand the self-assembly of these DNA origami objects. There have been attempts to better understand their self-assembly through experiment.[37] In a careful study, Sobczak et.al.[34] probed the assembly of certain 2D and 3D origami structures, and found rapid assembly in a narrow temperature window for each structure. Further work from the same group[36] attempted to quantify folding quality in origami assembly, by binding fluorescent markers to any single-stranded regions of

[1] For the study of origami self-assembly in this Chapter and in Chapter 4, we use the original oxDNA model, rather oxDNA2, the improved version introduced in Chapter 5. This is not particularly consequential as the improvements introduced with oxDNA2 are not relevant for these studies of origami self-assembly.

the scaffold strand. Similarly, Ke and coworkers[35] examined the effect of staple-break positioning on assembly and found that the yield could be very sensitive to this aspect of the design. Very recently, Dunn *et al.* designed a "dimer" origami that had multiple fully-bonded states, and thus allowed folding pathways to be inferred from the assembly products.[38] The experimental work was complemented with a domain-level kinetic model (which extended and improved upon previous such models[38;51]) whose predictions were in good agreement with the experimental results, implying that the model predicted a realistic set of folding pathways.

In Ref. 96, Cademartiri and Bishop provided an interesting classification scheme that divides self-assembly into two types, namely a "puzzle" limit in which the assembly materials are separate units, each carrying information about where they should fit in the final structure through specific interactions (e.g. "DNA bricks"[4]); and a "polymer" limit in which a polymer made of a limited number of repeating units folds up, with the topology of the polymer and weak, non-specific interactions between the polymer units determining the final structure (e.g. protein folding). DNA origami self-assembly can be regarded as a mixture of the two, with a long scaffold strand, akin to the polymer limit, which folds by the binding of many short staple strands to specific domains of the scaffold, similar to the puzzle limit. Despite the research efforts described earlier, important open questions remain regarding this mixed-type assembly. In particular, it is not clear whether the assembly is a "nucleated" process (i.e. one with a free-energy barrier), nor how and why certain changes to the origami design lead to different assembly yields.

Simulations can potentially provide insight into the self-assembly process through visualization of the assembly mechanism in a way not possible in experiment. However,

all-atom approaches are ruled out as the time and length scales relevant for origami self-assembly are far beyond those that can realistically be achieved with today's computers.[49] A potential way forward is provided by coarse-grained DNA models, as their simplified representations of DNA allow access to much longer time and length scales. In order to be useful, such models must capture the physics of DNA relevant to the assembly while still being simple enough to permit efficient computation.

The oxDNA model satisfies both of these requirements. Importantly for this study, it captures the biophysics of DNA likely to be relevant for origami self-assembly, such as hybridization,[70] displacement,[75;83] and the conformational properties of single-stranded DNA.[66] Furthermore, with the GPU version of the code implementing oxDNA it is possible to efficiently simulate systems consisting of thousands of nucleotides.

3.2 System and model

The parameterisation of the oxDNA model we use here is an "average-base" one in which the strength of the stacking and Watson-Crick base-pairing interactions are independent of the identity of the bases. We use this parameterisation, rather than a sequence-dependent one,[97] because we are interested in the generic properties of the self-assembly rather than the specifics of how it depends on sequence. We also note that the model does not include the possibility of any non-Watson-Crick base pairs. However, this should not affect our results as these interactions are unlikely to be relevant at the high temperatures where assembly takes place. The oxDNA model is parameterised to a salt concentration of 0.5M, which is representative of the high salt conditions typically used for origami assembly where the electrostatic interactions are largely screened.

3.2.1 Simulation details and time scale estimation

To simulate origami self-assembly, we use MD coupled to an Andersen-like thermostat, which we described in Chapter 2. Even at our level of coarse graining, the simulations presented here are computationally very demanding: each of the runs presented required several months of computer time on a GPU. The simulation parameters used are given in Table B.1. The initial configurations for the assembly runs were generated by simulating the appropriate single strands in a periodic box (with artificial polyA sequences to prevent any base pairing) for 10^7 simulation steps. The assembly simulations were run for up to 6×10^{10} simulation steps. In order to make simulating thousands of nucleotides with oxDNA feasible, the algorithm was run on GPUs, devices suitable for parallel computation of pairwise forces, giving up to a 25× speed-up over a typical CPU.

Inferring the physical time elapsed during an assembly run simulated with an Andersen-like thermostat is not straightforward.[82] The energy and length scales within the model suggest a conversion from simulation time units to physical units of 3.03×10^{-12} s per simulation time unit. However, this is likely to lead to an underestimate of the times relevant to self-assembly because it does not take into account the reduction in time scale separations often inherent in coarse graining, e.g. speeding up diffusion relative to microscopic time scales.[98] In fact, the appropriate time unit conversion for the model will likely be different for different processes.

Here, we use a conversion factor obtained by considering the rate of hybridization of a 16-base-pair domain, as this is likely to be the important time scale for the assembly. For the experimental figure, we use an approximate rate constant suggested by experimental measurements[99–105] for a 16-base-pair duplex of $10^6\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$. We then

compare this to a similar transition computed using oxDNA, namely the binding of the first scaffold domain (16 base pairs long) in the self-assembly of the origami design that we consider in Chapter 4, which is similar to the current design. Clearly, the simulated system (one stage of an origami assembly) is somewhat different to the experimental one (formation of a simple 16-base-pair duplex), but we believe that this difference is not too important for this very rough time scale estimation.

The simulated on-rate is 1.0×10^{-10} per simulation time unit, measured at a concentration of 1.0×10^{-8} M. The experimentally determined on-rate at this concentration is $10^{6}\,\mathrm{M}^{-1}\,\mathrm{s}^{-1} \times 10^{-8}\,\mathrm{M} = 10^{-2}\,\mathrm{s}^{-1}$. Thus, the unit conversion is given by

$$\frac{1.0 \times 10^{-10} \text{ per simulation time unit}}{10^{-2}\,\mathrm{s}^{-1}} \approx 10^{-8}\,\mathrm{s} \text{ per simulation time unit.} \qquad (3.1)$$

In this chapter we give simulation times in seconds, using this conversion.

3.2.2 The origami design

In this chapter, we study the assembly of a small, 384-base-pair origami from single strands in solution. The design of the origami is shown in Fig. 3.1. Note that the helices in this structure are not perfectly parallel. This is because it is favourable for the helices to bend away from each other at the junctions, which leads to the weave pattern that is also seen in micrographs of DNA origami structures – we will consider the origins of this weave pattern in detail in Chapter 6. In our example the relatively small size of the origami and the low number of constraining junctions leads to a further splaying out of the helices, an effect which has been observed experimentally for regions of an origami that are sparsely populated with junctions.[39]

The scaffold strand can be divided up into 24 16-base domains (labelled 1–24 in Fig. 3.1 starting from the 3′ end) with each staple strand binding to two of these domains. There are twelve different staples, which can be divided into four types (Fig. 3.1). There are two "end" ("E") staples that bind to two consecutive domains at the two ends of the scaffold in a single double helix. The rest have two double-helical domains, and one staple crossing, where we use staple crossing to refer to the point at which the staple crosses from one region of the scaffold to another, folding the scaffold into the target shape. The non-E staples are classified as "L", "M" and "R", corresponding to whether they are on the left, middle or right of the origami when oriented as in Fig. 3.1. In this design, there are two types of four-way junction: one type involving the M and L staples, which has two staple crossings at the junction, and the other type having just a single staple crossing from one of the R staples.

In order to make origami self-assembly more feasible on our computational time scales, we designed the DNA strands' sequences to minimize secondary-structure formation in the single strands, as it is known that hairpins can significantly slow down the rate of duplex formation.[74;99] One consequence of this choice is that misbonding between scaffold and staple strands is somewhat more likely than for random sequences. All the sequences are given in Table B.2 in the appendix.

In this chapter we consider assembly under two different concentration ratios between the staples and scaffold: firstly, we consider a system with a large excess of staples (namely 17-fold) and secondly, a system with a stoichiometric mixture (i.e. only one copy of every staple in our simulation).

In our simulations, we consider the assembly of a single origami, i.e. we have only one scaffold strand in our simulation box. We use a scaffold strand concentration of

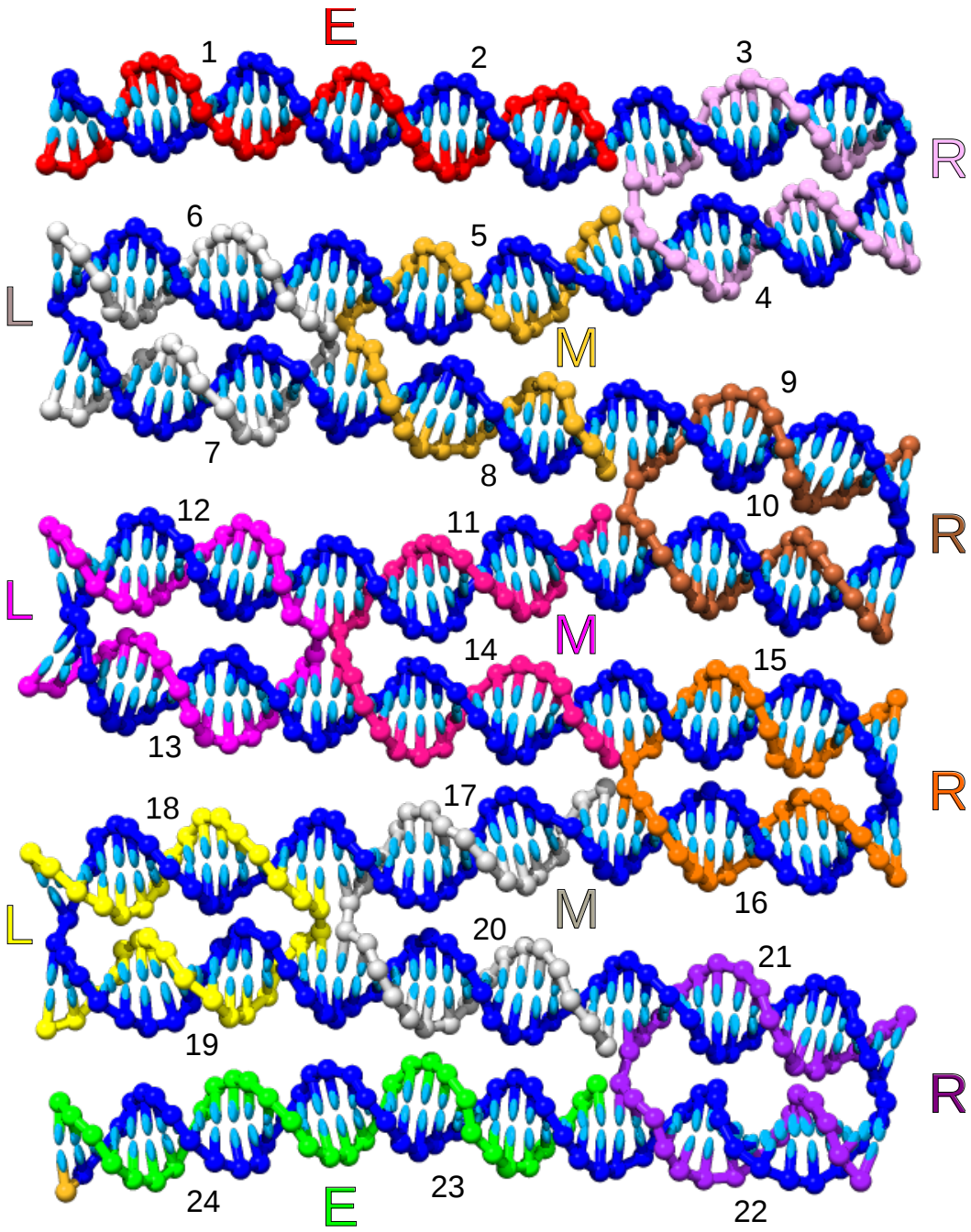

Figure 3.1: The complete origami, adsorbed onto a surface to help visualization. The 384-base scaffold strand is blue (with the 5′ nucleotide gold), while each of the twelve 32-base staple strands are depicted in a unique colour, in a colour scheme which is maintained throughout this chapter. The scaffold is divided up into 24 binding domains, and the staples are labelled according to their type, "E", "R", "M" and "L", based on their general position in the complete origami structure.

7.7 μM, and staple strand concentrations of 7.7 μM and 130.9 μM for the stoichiometric and excess-staple system, respectively. These conditions correspond to an approximately 1000$\times$ higher concentration than typical experimental concentrations (for example, Dietz and coworkers[9] used a scaffold strand concentration of 20 nM to assemble 3D origamis). Without such a high strand concentration it would not be feasible to observe self-assembly in our simulations, as almost all the simulation time would be spent waiting for each of the staple strands to find their scaffold domains by diffusing around an enormous simulation box. Fig. 3.2 shows typical configurations at the beginning of an excess-staple assembly simulation and a stoichiometric assembly simulation.

It is likely that there is a relatively narrow temperature window for optimal assembly, as evidenced in recent experiments.[34] If the temperature is too low secondary structure in the staples and scaffold can occur, as well as misbonding (that is, Watson-Crick base pairing between bases that are not paired in the target structure) between staples and scaffold. Both these effects may result in kinetic traps which will slow down or even prevent correct assembly. The window will be bounded from above by the melting temperature of the origami, but even below this temperature assembly may be hampered by large nucleation free-energy barriers, a slower attachment rate and an increased rate of melting for partially or fully bound staples.

These latter effects are particularly problematic for the direct simulation of origami assembly. In practice, it may only be possible to see assembly in our simulations on reasonable time scales when the correct binding of a staple to a domain is relatively irreversible, placing our simulations of assembly at the lower end of (or perhaps below, where kinetic trapping is relevant) this optimal temperature window. An important reference temperature in this regard is the oxDNA melting temperature for a single

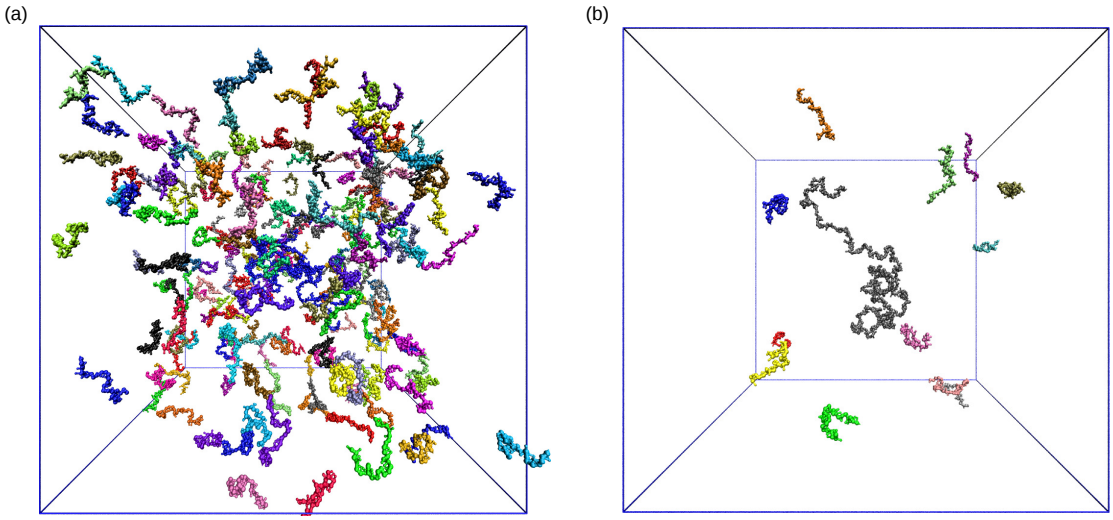

Figure 3.2: Snapshots from the beginning of an assembly simulation for (a) the excess-staple assembly and (b) the stoichiometric assembly. The box shows the location of the periodic boundaries. The high concentration of DNA strands can clearly be seen, especially for the excess-staple assembly.

16-base-pair duplex ($T_{16\text{mer}}$ = 66 °C and 72 °C for the stoichiometric and excess-staple assembly DNA strand concentrations, respectively) and so we generally chose to run our simulations close to these temperatures.

To summarize the results of our self-assembly simulations we use kymographs, such as the one illustrated in Fig. 3.3, which show the state of each of the 24 scaffold domains as a function of time. We designate each base pair between each domain and any staples that have fully or partially bound to it as belonging to one of three categories: (i) *native (without blocking)*, indicating a native base pair, i.e. one which is present in the target structure, in the absence of any blocking (see below) (ii) *misbonded*, indicating a non-native base pair, i.e. one which is not present in the target structure, and (iii) *native (with blocking)*, indicating a native base pair, but one where rather than one staple strand binding to both its scaffold domains, two copies of the same staple strand each bind to only one of the scaffold domains, preventing either of them from reaching their correct fully-bound state until one of them melts away or is displaced. This blocking effect is described in further detail below. Note that it is possible for a

scaffold domain to simultaneously experience misbonding and partial native bonding (either blocked or not blocked).

3.3 Assembly with a large staple excess: 65 °C

A kymograph for an assembly with an excess of staples and at a temperature of 65 °C is shown in Fig. 3.3. Of the four assembly runs simulated under these conditions, this one had the most correctly bound domains by the end of the simulation, and we will focus our discussion on this run. Kymographs from the other three simulation runs are shown in Fig. 3.4. From Fig. 3.3 it is clear that most of the origami is able to assemble correctly, with 20 of the 24 scaffold domains correctly bound at the end of the simulation (see Fig. 3.5 for a simulation snapshot). There are a number of important features about the nature of the origami self-assembly that can be gleaned from Fig. 3.3.

3.3.1 Local features of staple binding

Although correct binding of a staple to its corresponding domain on the scaffold is a rare event — the two have first to diffuse close enough to interact — when it does occur it does so relatively rapidly, as seen by the step-like increases in the number of native bonds in a domain from zero to approximately 16. Thus, even at the relatively high concentrations we consider, there is a clear separation between the intermolecular time scales associated with strand diffusion and the intramolecular time scales associated with going from the formation of an initial contact to complete binding of a domain. This behaviour is what we expect based on prior studies of hybridization in oxDNA.[71] In addition, correct binding of a 16-base-pair region is effectively irreversible at this temperature on our time scales.

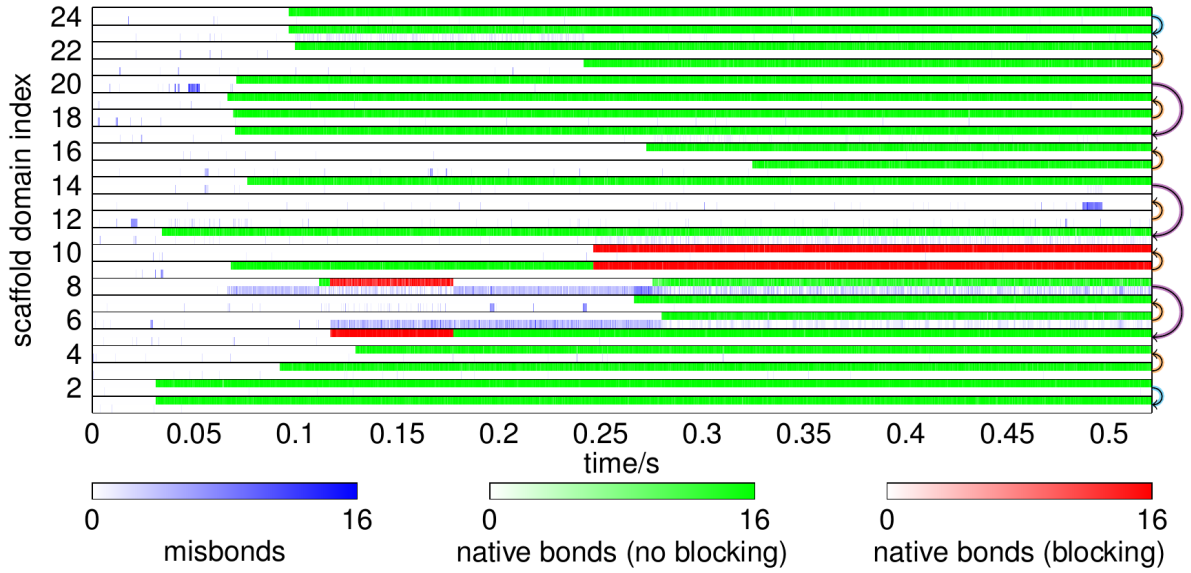

Figure 3.3: A kymograph at 65 °C for a single assembly simulation with a 17× excess of staples, starting in an unbonded configuration. Each row shows the state of a domain of the scaffold strand as a function of time. The top half of each row is either white, red or green. Red and green both indicate native base pairing (i.e. the presence of base pairs which are also found in the target structure) in that domain. Red indicates that the staple bound to that domain is blocked by another staple (see main text), while green indicates that it is not. White indicates no native base pairing. The bottom half of each row is either white or blue, with blue indicating misbonding (i.e. the presence of base pairs which are not found in the target structure) and white indicating no misbonding. In all cases, the intensity of the colour corresponds to the fraction of base pairs within the domain which are in that state. On the right, the curly arrows show the binding pattern of the staple strands, with the arrow head showing where the 5′ end of the staple binds.

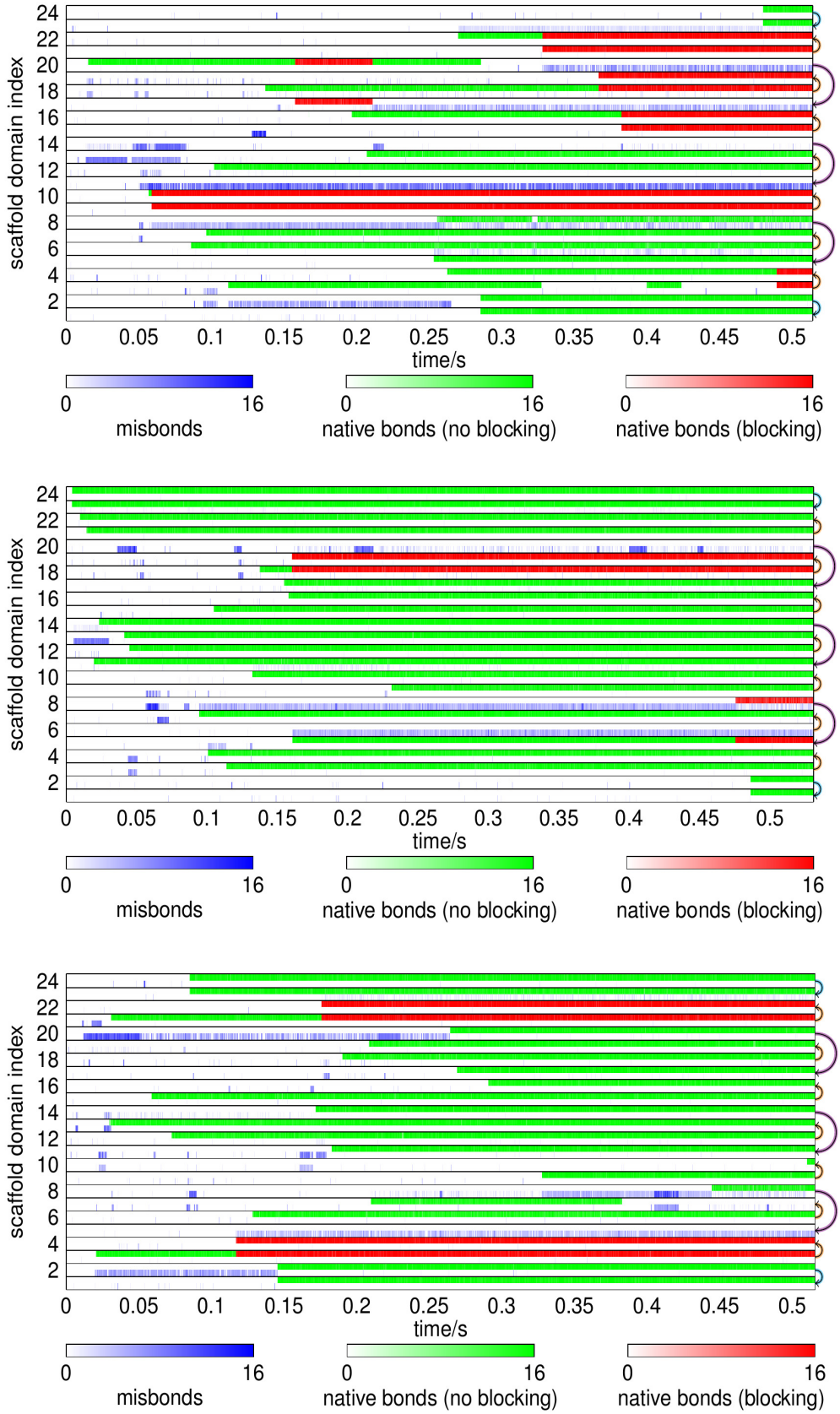

Figure 3.4: Kymographs for the three further simulations at 65 °C with a 17× excess of staples.

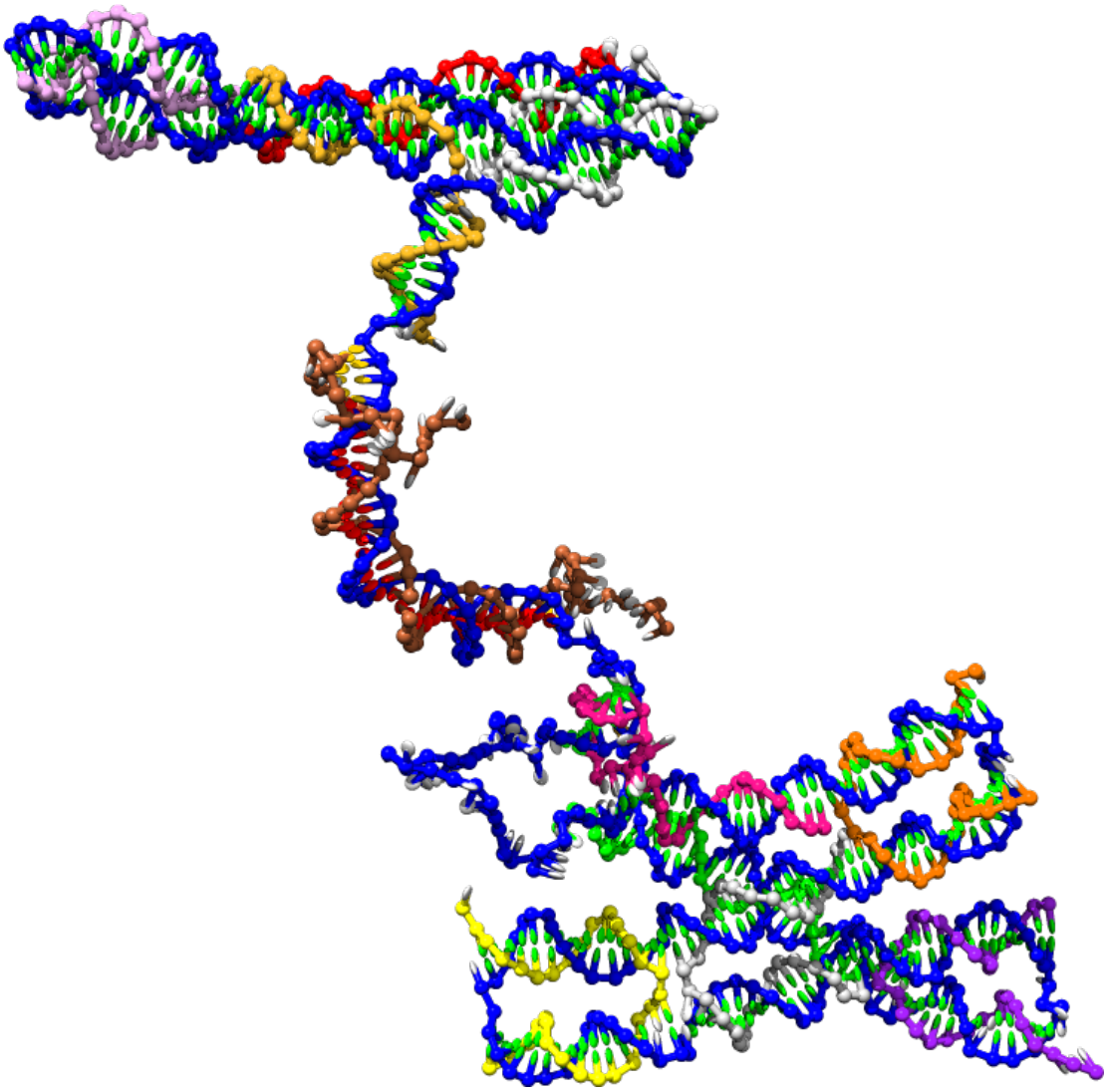

Figure 3.5: A snapshot from the end of the excess-staple 65 °C assembly whose kymograph is shown in Fig. 3.3. The base sites are coloured according to the classification of their base pairing (while the rest of the nucleotide is coloured according to its strand). Unpaired base sites are white, natively-bonded, unblocked base sites are green, misbonded base sites are yellow, and blocked base sites are red.

3.3.2 Global features of origami self-assembly

Firstly, assembly occurs in a relatively parallel manner. In particular, the assembly that is happening in distant sections of the scaffold is relatively independent, as can be seen from the absence of obvious long-range cooperativity in the kymographs for the different assembly runs at 65 °C (Fig. 3.3 and 3.4). This is to be expected given the small number of crossovers in our relatively simple origami design, meaning that the formation of each domain (along with any associated crossovers) only affects other domains that are close to it. Furthermore, the kymographs also suggest that there is no set order in which the staples bind, although this may partly reflect the relatively low temperatures we have chosen. In particular, the initial binding of each staple appears unaffected by both the nearby and distant state of the partially formed origami.

Secondly, there are some signs of cooperativity in the binding of a staple to the second of its scaffold domains, i.e. the binding of the second domain is not necessarily independent of which adjacent domains are occupied by other staples. Generally, binding the second staple domain can be relatively slow because the relevant parts of the staple-scaffold complex have to diffuse close enough to allow binding. For example, in Fig. 3.3 a staple correctly binds to scaffold domain 22 after around 0.1 s, but only manages to bind to scaffold domain 21 after a further 0.15 s. Note that these time scales would be better separated for the lower concentrations typical of experimental origami assembly; at these concentrations the binding of the first domain would be much slower while the rate of binding for the second domain would be unchanged.[106] We will see this with simulations of oxDNA at experimental strand concentrations in Chapter 4.

For the four scaffold domains adjacent to each of the double-crossover junctions (which are bound by an L staple and an M staple), we see a much faster binding of the

fourth domain than is typical for an equivalent binding elsewhere in the assembly. This is likely to be because the scaffold strand is constrained by the first three domains to be in a geometry that makes binding of the fourth domain relatively easy. This can be seen separately for scaffold domains 5–8 and 17–20 in the kymograph in Fig. 3.3, and simulation snapshots illustrate the process for scaffold domains 17–20 in Fig. 3.6.

3.4 Barriers to complete assembly

(i) *Misbonding.* The misbonding observed for the assembly simulations can be divided into two types, the first of which is not much of a hindrance to assembly and the second of which is more disruptive. Firstly, it is possible for a staple strand to bind to the scaffold purely by misbonding. This behaviour can be seen in the kymograph in Fig. 3.3 as the short stretches of blue scattered throughout the graph. Such events are common in the initial stages of the simulation when most of the scaffold is still available for binding. However, these misbonded configurations only persist for a relatively short amount of time before melting away and, during the time they are present, the number of base pairs fluctuates a lot as the duplexes fray and reform. For these reasons, they offer little resistance to correct assembly of the origami at this temperature.

The second type of misbonding can occur when a staple is correctly bound to one of its scaffold domains. It is then sometimes favourable to extend the double helix beyond the end of this domain as it may be able to form some base pairs with the next domain (for example, see Fig. 3.7(a)). This type of misbonding is more persistent (see scaffold domains 6 and 8 in Fig. 3.3), because the strands involved are unlikely to detach from the scaffold due to their correctly bound domain, and is thus potentially more detrimental to the assembly.

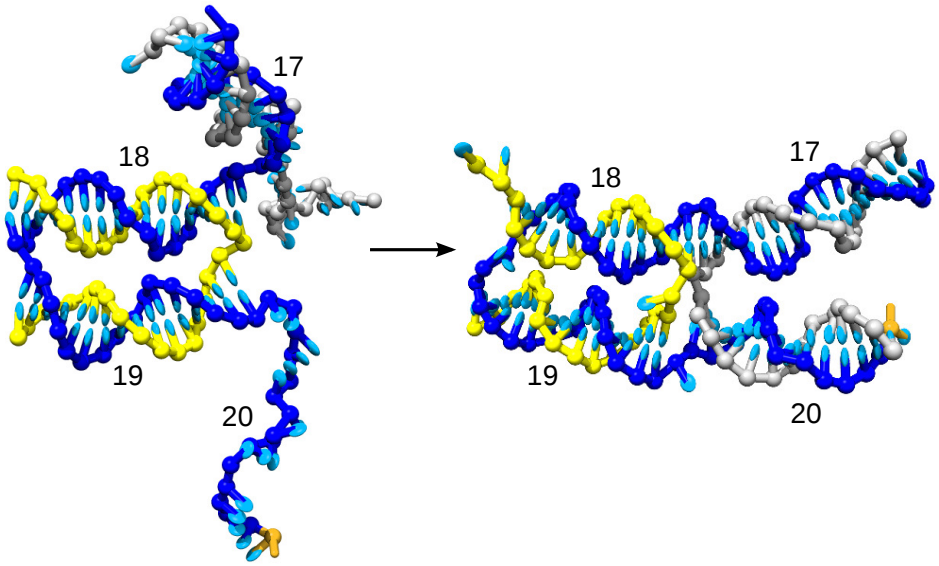

Figure 3.6: Cooperative binding of a staple to form a double crossover. Only domains 17-20 of the scaffold are depicted. The partially bound grey staple is in a favourable configuration to bind to scaffold domain 20 because of the correctly bound yellow staple.

This persistent misbonding can interfere with origami assembly in two ways. Firstly, it will be harder for the misbonding scaffold domain to be bound by the correct staple, because the domain is partially blocked by the misbonding staple. Note that this is not too much of a hindrance, because the correct staple still has the possibility of beginning to bind to some of the free scaffold bases, after which it will potentially be able to displace the misbonded staple. For example, in Fig. 3.3, the correct staple is able to bind to domain 8 at around 0.11 s despite the misbonding in domain 8 due to the staple bound to domain 9, as illustrated in Fig. 3.7(a) and Fig. 3.7(b) (although in this case it happens to melt later in the simulation, probably partly due to misbonding from the domain 9 staple).

Secondly, persistent misbonding makes it harder for the strand that is misbonded to find its own correct second scaffold domain. Again, this only slows down rather than stops assembly, because there is still some of the staple unbound and available (this amount can fluctuate considerably) to bind to the correct domain, although clearly the

misbonding will often impose severe geometrical constraints on the staple bases that remain free. For example, in Fig. 3.3, the staple correctly bound to domain 5 misbonds with domain 6, but is still able to bind later to domain 8. Relevant simulation snapshots are shown in Fig. 3.7(e) and Fig. 3.7(f).

We should note both that, as mentioned earlier, the sequence we have used makes misbonding more likely than for a random sequence, and that oxDNA tends to overestimate the thermodynamic stability of misbonded states somewhat.[66] Therefore, we expect that misbonding is less likely to be a problem for real DNA at an equivalent temperature. Nevertheless, in both simulation and experiment, as the temperature is decreased the misbonding that occurs is expected to become more persistent, until at sufficiently low temperature it can be effectively irreversible, and hence prevent complete assembly. This will be explored more when considering stoichiometric assembly, where we have performed simulations at lower temperatures.

(ii) *Blocking.* The major obstacle to complete assembly of the origami in our simulations is what we have termed "blocking". This denotes the case where two copies of the same staple each bind to one of the two scaffold domains for that staple, thus preventing either copy from reaching its fully-bound state. An example configuration for such a state is shown in Fig. 3.7(e). These blocking states are long-lived on our simulation time scales and at 65 °C we have never observed a blocked configuration involving two R staples being resolved.

These blocking events will be most prevalent when the time scale for the intramolecular process of a staple binding to its second scaffold domain is of the same order as that for the intermolecular process of a staple initially binding to the scaffold. In the examples of blocking that we have observed, misbinding often plays a role in hindering

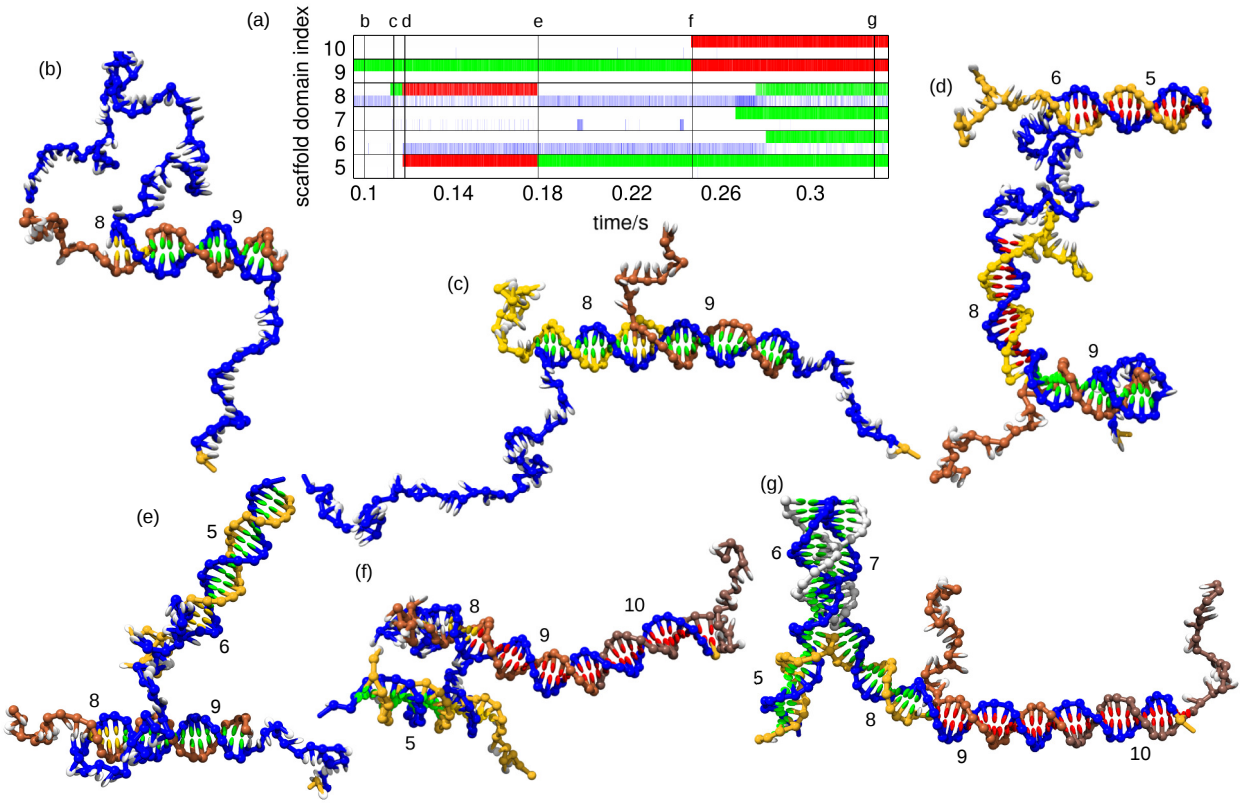

Figure 3.7: A series of snapshots from the trajectory corresponding to the kymograph in Fig. 3.3 illustrating misbonding and blocking. Only the scaffold domains 5–10, and any staples with at least one base pair with those scaffold domains, are depicted. Note that some staple domains are bound to a scaffold domain that is not shown. The base sites are coloured white if unpaired, green if natively-bonded and unblocked, yellow if misbonded and red if blocked. (a) An extract from the kymograph in Fig. 3.3 is shown, with letters along the top showing where the snapshots appear in the kymograph. (b) A staple is natively bound to domain 9 and partially misbonded to domain 8. (c) A correct staple comes in to bind to domain 8, although there is still some misbonding in domain 8 due to the domain 9 staple. (d) Before the domain 8 staple can complete its binding by binding to domain 5, a second copy of the same staple binds to domain 5, as well as partially misbonding to domain 6. (e) The original staple, which bound to domain 8, melts off, partly due to the domain 9 staple misbonding with domain 8. The domain 5 staple is now no longer blocked. (f) A copy of the domain 9 staple comes in and binds to domain 10, blocking the domain 9 staple. (g) A new staple comes in and binds to domains 6 and 7, while the domain 5 staple binds to domain 8, finishing its binding. Although domains 5–8 are correctly bound, the blocking in domains 9 and 10 remains.

the binding of a staple to its second scaffold domain. This could either be because the staple's second domain is itself involved in misbonding (for example, see the kymograph in Fig. 3.3 and the snapshots shown in Fig. 3.7, where a staple is correctly bound to scaffold domain 9 and misbonds to scaffold domain 8, eventually allowing a second staple to come into scaffold domain 10 and block the original staple) or because the second scaffold domain required for complete binding is partially blocked by misbonding by a different staple.

We performed four simulation runs in total at this temperature (65 °C) – kymographs for the other runs are shown in Fig. 3.4, and show similar patterns of behaviour. In all four runs blocking prevents complete assembly, with the run described above getting closest to completion. At the end of the simulations, the other three runs had 10, 18 and 18 correctly bound domains, with 10, 4 and 4 domains blocked, respectively, compared to 20 correctly bound domains and 2 blocked domains for the simulation run described above. Longer simulations confirm that the blocking involving R strands is effectively irreversible on our time scales at this temperature.

3.4.1 Recovering from blocking

Conceptually, the most straightforward way for the partial assembly to recover from blocking is by one of the blocking staples melting away from the scaffold. However, this is rare at 65 °C, because we have chosen to work in the "downhill" temperature regime, i.e. where the binding of every staple domain is downhill in free energy (as mentioned earlier, observing assembly in simulations at higher temperatures is less tractable because of the reduced binding rates and increased detachment rates). At higher temperatures, this mechanism is likely to be more important, for example where a staple is only stable when its second domain has bound. Blocking is likely to be less of

a problem at these temperatures, and in particular at the temperature where a complete origami first becomes stable on annealing.

A potentially easier way to recover from blocking at 65 °C is through one blocking staple displacing the other. The displacing staple cannot form a continuous helix when initiating this displacement. In this sense, it is a "remote-toehold" displacement.[107] The likelihood of this form of displacement will in part depend on any geometrical constraints stemming from the rest of the origami structure. For example, for blocks involving the M (or L) staples, if the adjacent L (or M) staple is correctly bound then the crossover will bring the unbound domains of the blocking staple strands in a favourable position to initiate displacement (e.g. Fig. 3.6 shows a fully bound L staple creating a more favourable geometry for the neighbouring M staple).

By contrast, for blocks involving R staples, the rest of the origami provides no such constraints, as in our small origami the R staples form single crossovers (Fig. 3.1). Instead the most likely configuration is one where there is coaxial stacking between the bound ends of the staples, so that the unbound domains are as far away as possible from each other. In this case, coaxial stacking refers to stacking between the adjacent ends of two staples. An example of this is seen for domains 9 and 10 in Fig. 3.7(f). Finally, misbonding of the blocking staples to other scaffold domains will again make displacement less likely.

The melting of one of the blocking strands can be assisted by misbonding from an another staple with a free domain. A possible example of this can be seen in Fig. 3.3, where the staple natively bonded to scaffold domain 9 competes for base pairs in scaffold domain 8 with a blocking staple from about 0.1 s to about 0.17 s, after which the blocking staple falls off. This is illustrated with simulation snapshots in Fig. 3.7(c) and

Fig. 3.7(d). By forming some misbonds with the scaffold domain, the misbonding strand can partially compensate for the lost base pairs on melting the blocking strand, in a process somewhat analogous to strand displacement. Although this kind of blocking resolution involves a reduction in base pairing, it is geometrically easier than displacement by the other blocking strand because the misbonding strand is adjacent (along the scaffold strand) to the domain to which it misbonds.

The majority of the blocking that we observe involves the R staples, and at 65 °C seems to be effectively irreversible. By contrast, blocking involving the M or L staples is more transient. Note that, while L and R staples can enhance M staple melting by misbonding, and M staples can do the same for L staples, R staples cannot be subject to any such effect. This is because the free ends belonging to the staple strands adjacent to the R staples are too far from the R staples' domains, a consequence of the single crossover that the R staple forms.

In fact, we have never seen blocking involving M (L) staples when the adjacent L (M) staple is fully bound. There are a number of factors that potentially contribute to this behaviour: (i) As discussed above, if a fully bound L (M) staple is already in place, an adjacent M (L) staple that subsequently binds its first domain will rapidly bind its second, leaving little opportunity for a blocking strand to bind. (ii) If a blocking pair of L (M) staples are in place, binding of the second domain of an adjacent M(L) staple may be hindered by misbonding from the free end of one of the blocking strands. (iii) In the above situation, the cost of forming the junction when binding the second domain will also be increased somewhat due to the presence of the single-stranded tails of the blocking strands. The tails may have to break stacking at the junction and will have reduced conformational entropy due to steric constraints caused by the presence

of other nearby strands (a somewhat similar free-energy penalty has been found for initiating displacement because the displacement generates a second single-stranded tail[75]). Interestingly, in separate simulations of an origami with a blocking pair of L (M) staples and an adjacent fully bound M (L) staple not only did we find that sometimes one of the blocking strands would displace the other, but that sometimes the fully bound staple would melt off.

3.4.2 Blocking in real origami self-assembly

Clearly, the relatively high concentrations that we are using play a role in the frequency of these blocking events, as they increase the rate of intermolecular binding without affecting intramolecular binding. However, although experimental origamis are typically assembled at much lower concentrations, there are strong reasons to suggest that blocking might still be a hindrance for assembly. Firstly, a full-size origami has many more staples, and a small number of key staples failing to fully bind due to blocking could be highly detrimental. Secondly, the longer scaffold strand and potentially multiple scaffold domains per staple strand will increase the time scale for intramolecular binding compared to our simple design, boosting the chance of staple blocks occurring. The longer scaffold strand leads to designs which often involve staples with domains that are much further apart along the scaffold strand than for our design, so that completing the binding of each staple can take much longer. Meanwhile, when staples that bind to three or more scaffold domains are used it is possible for the domains at the two ends of the staples to bind first, in which case the binding of the middle domain may be hindered by the absence of a free end to facilitate wrapping of the two strands around each other.[108] Thus it may be necessary for one of the end domains to dissociate before binding can be completed, a potentially slow process. We will think

about such processes in more detail in Chapter 4.

Effects analogous to staple blocking have been seen in experiments with DNA tweezers, with dimerisation between tweezers being observed.[30;109] In addition, in the DNA walker experiments of Tomov *et al.*,[25] increasing the fuel strand concentration was found to decrease the stepping yield for one of the designs, even though it increased the initial rate of stepping. Simultaneous binding of two copies of the fuel strand was identified as the cause of this reduction in yield. For that system, the effect was observed for fuel concentrations as low as 50 nM, while the distance between the two domains to which a single fuel strand was intended to bind was 5-10 nm. This observation suggests that staple blocking might be relevant at staple strand concentrations of $\sim$50 nM and perhaps even lower, especially when remote domains of the scaffold must be bound by a single staple.

We note that, while the amount of time required might be prohibitively long in practice, displacement or melting of the blocking staples will eventually resolve the blocking we observe. However, one or more of the blocking staples may bind to another scaffold in the meantime, which would cause unwanted aggregates of multiple partially-formed origamis to form. The experiments of Bae *et al.*,[110] where origami assembly is initiated from a scaffold strand that is both under tension and fully loaded with staples that are only bound to a single domain, provide interesting insights into these processes. When the tension is reduced, assembly mediated by the displacement of blocking strands is observed, but the heterogeneity in the assembly products (as evidenced by the wide range of extensions observed when tension is applied to the product) indicates the possibility of persistent staple blocking. Note that the immobilization of the scaffold strand on a surface makes aggregation unlikely in this experiment.

3.5 Excess-staple assembly annealed at a higher temperature: 70 °C

We have seen that the assembly run with a large staple excess was hindered by staple blocking, preventing complete assembly on the time scales simulated. One way to resolve the blocks might be to increase the temperature somewhat, so that both partially-bound staples are more likely to melt off the scaffold during the simulation, and displacement is more likely due to enhanced fraying. We tested this idea by performing some runs beginning from the end points of previous lower temperature runs (which had been run for approximately 0.8 s), and raising the temperature to 70 °C.

A kymograph of one of these runs is shown in Fig. 3.8. From this graph it is clear that, although the annealing is allowing blocks to be resolved, new blocks are also being formed, and so overall there is no net progress towards forming a complete origami.

3.5.1 Resolution of blocking

Examples of the three possible ways of resolving blocking mentioned above are evident in the kymograph of Fig. 3.8, with snapshots illustrating these processes shown in Fig. 3.9.

Firstly, we see an example of blocking resolution by displacement, in this case for an R staple, in scaffold domains 21 and 22 at around 1.3 s. The difficulty of this kind of displacement is highlighted in the simulation snapshots in Fig. 3.9(a), where the scaffold strand has both to break coaxial stacking and to bend quite sharply, a rare configuration, in order for the displacement to begin.

Secondly, we see an example of blocking resolution by melting for the R staple whose target is scaffold domains 15 and 16. At around 1.3 s the blocking staple bound

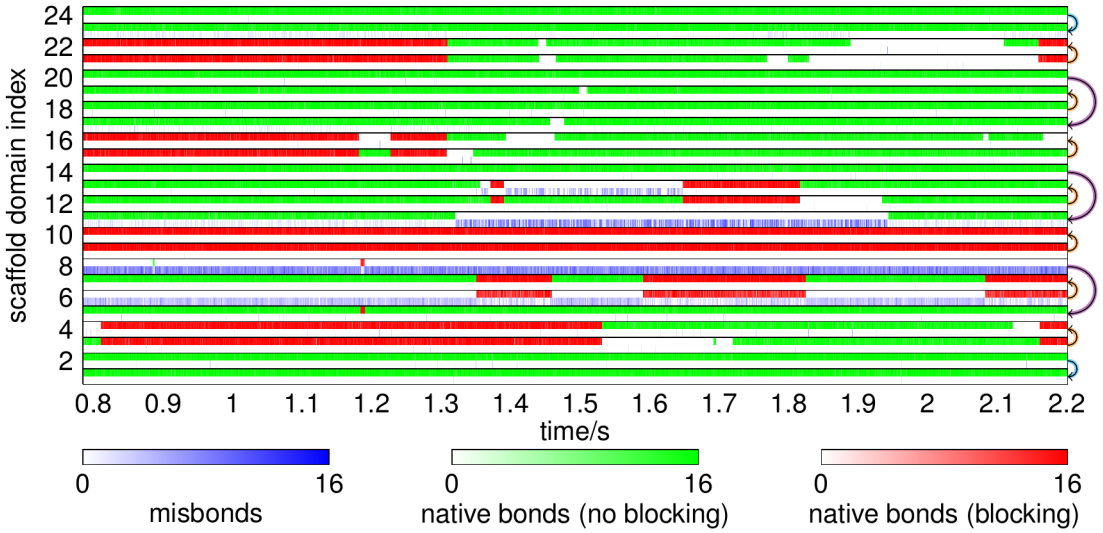

Figure 3.8: A kymograph for the annealing at 70 °C of an assembly produced by a 0.8s simulation at 65 °C with a large staple excess.

to domain 15 melts off, and then approximately 0.05 s later the other staple binds to domain 15, completing its binding to the scaffold (Fig. 3.9(b)).

Thirdly, an example of misbonding aiding blocking resolution occurs for scaffold domain 8, after slightly less than 1.2 s. Here we see the loss of a blocking staple in scaffold domain 8 at the same time as the staple bound to scaffold domain 7 misbonds with domain 8 (Fig. 3.9(c)). It is likely that this misbonding plays a role in accelerating the melting of the blocking staple.

3.5.2 Formation of new blocks

However, at this higher temperature we also see instances of fully bound staples becoming partially unbound and then subsequently blocked. This behaviour is perhaps not surprising, as the higher temperature will slightly destabilise correctly bound staples as well as blocking ones.

An example of melting of a fully bound staple followed by blocking is given by the R staple bound to domains 3 and 4. At around 2.12 s the part of the staple bound to

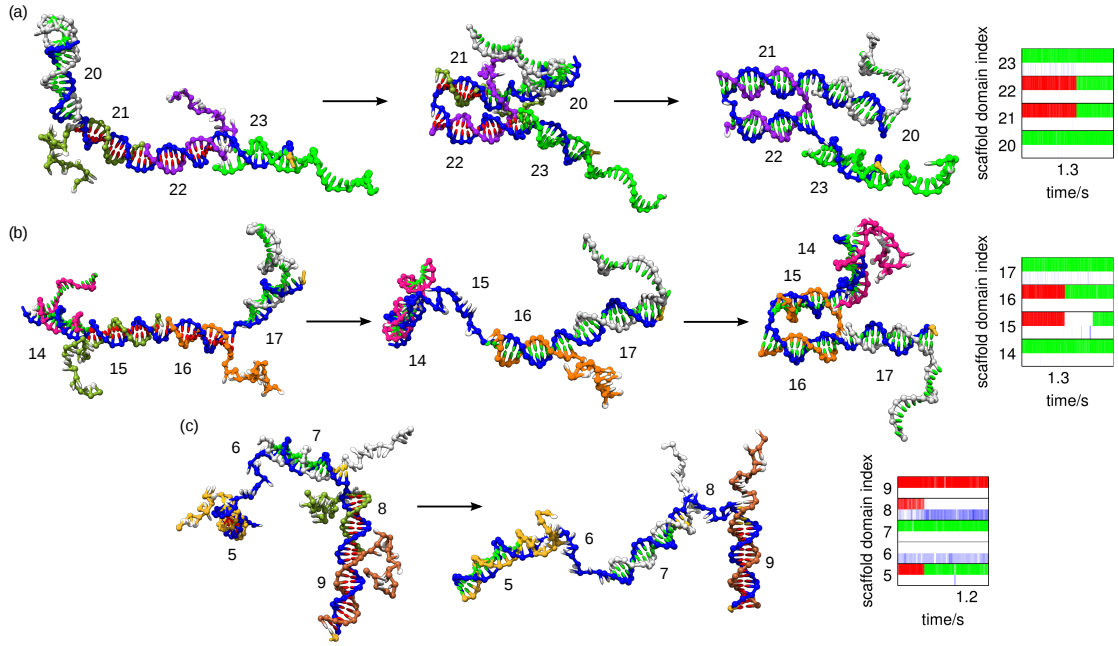

Figure 3.9: Simulation snapshots and corresponding kymographs (right) showing resolution of blocking by various mechanisms during annealing at 70 °C with a large staple excess. For clarity, only the relevant section of the scaffold is shown in each case. Note that some staple domains are bound to a scaffold domain that is not shown. The nucleotides are coloured as in Fig. 3.5 and Fig. 3.7. (a) Blocking resolution by displacement. The purple strand bound to domain 22 displaces its competing blocking strand bound to domain 21, thus achieving complete binding. The middle panel shows an intermediate in the strand-displacement reaction. (b) Blocking resolution by melting. The olive blocking strand bound to domain 15 melts off, making this domain available for binding. About 0.05 s later, the orange staple binds to domain 15, completing its binding to the scaffold. (c) Blocking resolution facilitated by misbonding. In the left panel, the olive strand (bound to domain 8) and golden strand (bound to domain 5) are blocking each other. The gray strand (natively bound to domain 7) is misbonding with scaffold domain 8, preventing the olive strand from fully binding to that domain. In the right panel, the olive strand has melted off, having been unable to fully bind to domain 8. The golden strand is now no longer blocked and is free to bind to domain 8. Note that the brown strand (bound to domain 9) is also misbonding with the now vacated scaffold domain 8.

domain 4 melts off, and shortly after that a copy of the staple comes in and binds to domain 4, forming a block.

At the start of the simulation an L staple is correctly bound to scaffold domains 12 and 13. However at around 1.3 s domain 13 melts, probably facilitated by misbonding from the staple bound to domain 14. This allows another copy of the original staple to bind to domain 13, thus generating a blocked configuration.

Overall, our annealing simulations at 70 °C with an excess of staples showed no progress, that is, on average there are no fewer blocking staples at the end of the simulation than there were at the beginning. In other words, the enhanced rate of resolving blocking appears to be roughly equal to the increased rate of forming new blocking pairs. If this really is the case, it would imply that the thermodynamically favourable state for the system is a partially assembled origami with some blocking strands. This may be partly due to the small simulation box and large staple excess (equivalently, very high strand concentration). The scaffold strand gains configurational entropy when the origami has blocks rather than being correctly bound, as an origami with blocks has fewer junctions constraining the scaffold strand. However, the system also loses translational entropy (equivalent to the translational entropy one of the staples would have had if it could diffuse freely) for each blocking pair. By raising the strand concentration, we reduce this loss of translational entropy (without significantly affecting the configurational entropy of the scaffold strand), shifting the balance in favour of staple blocking.

3.6 Assembly with a stoichiometric strand mixture

Having established that, for our high-concentration, large-staple-excess simulations, staple strand blocking prevents completed assembly, at least on the time scales accessible to us through simulation, the obvious question is how we can we get around this problem. We have seen above that simply annealing such assemblies at higher temperature is not enough.

One approach would be to lower the staple concentration to something comparable to the concentrations used in experiment. This would have two positive effects: (i) it would increase the translational entropy penalty for blocking, and (ii) it would lead to a greater separation between the time scales for a staple binding its second domain and the initial binding of a staple to the scaffold, thus making it more likely that each staple would fully bind before it could be blocked by a second staple. It is not feasible to observe assembly with our direct simulations under these concentrations, as the vast majority of the computational effort would be spent simulating strands diffusing around the box. However, it is possible to probe assembly at these concentrations using rare events simulation methods, an option we explore in Chapter 4.

We take an alternative approach here: while keeping the same simulation box size, we simulate a stoichiometric assembly, with only one copy of each staple in the simulation box. The design of the origami remains unchanged. Fig. 3.10 shows example kymographs for simulations at 50 °C, 55 °C and 60 °C, while Fig. 3.11 shows examples at 65 °C and 70 °C.

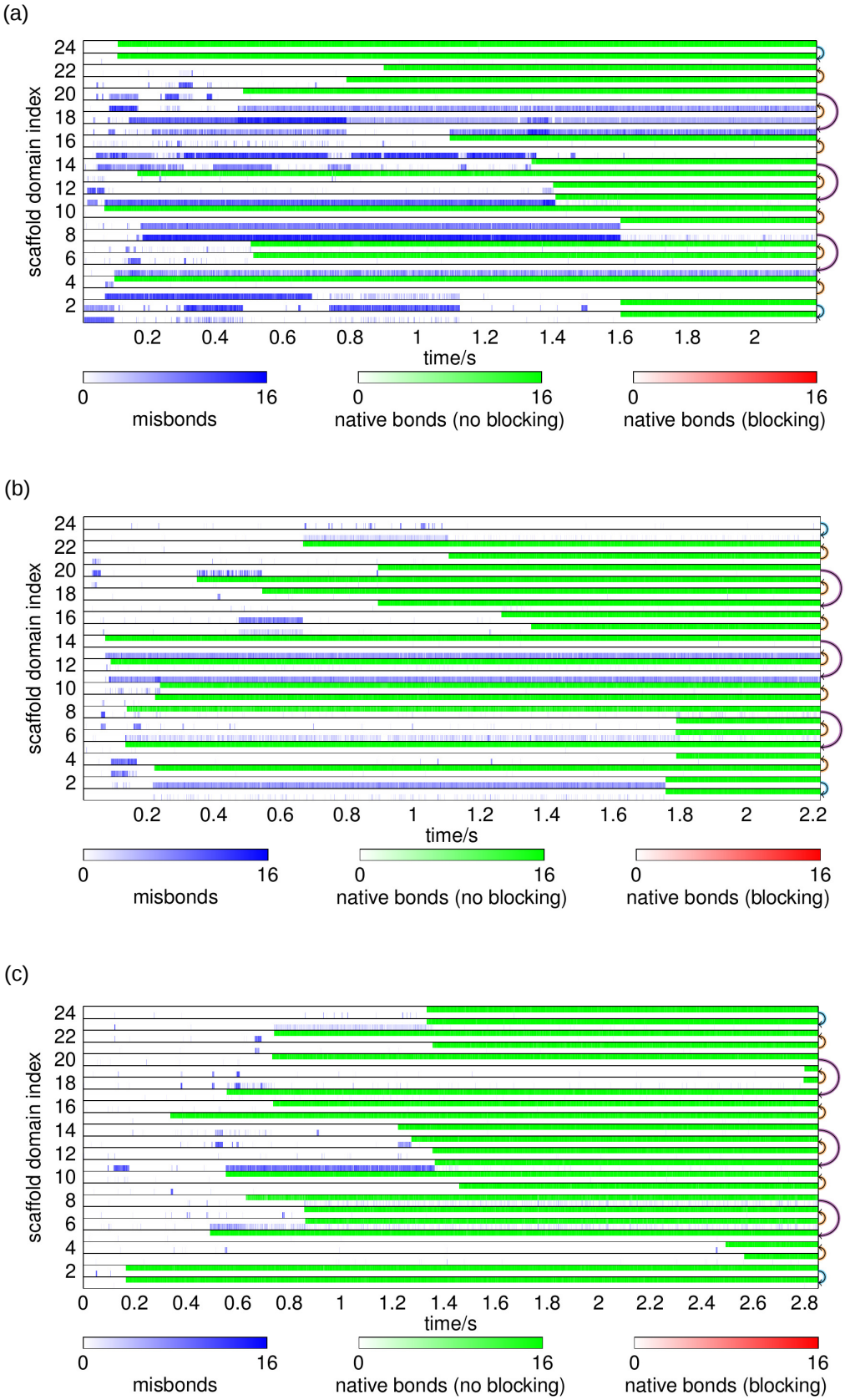

Figure 3.10: Kymographs for stoichiometric assembly simulations at (a) 50 °C, (b) 55 °C and (c) 60 °C.

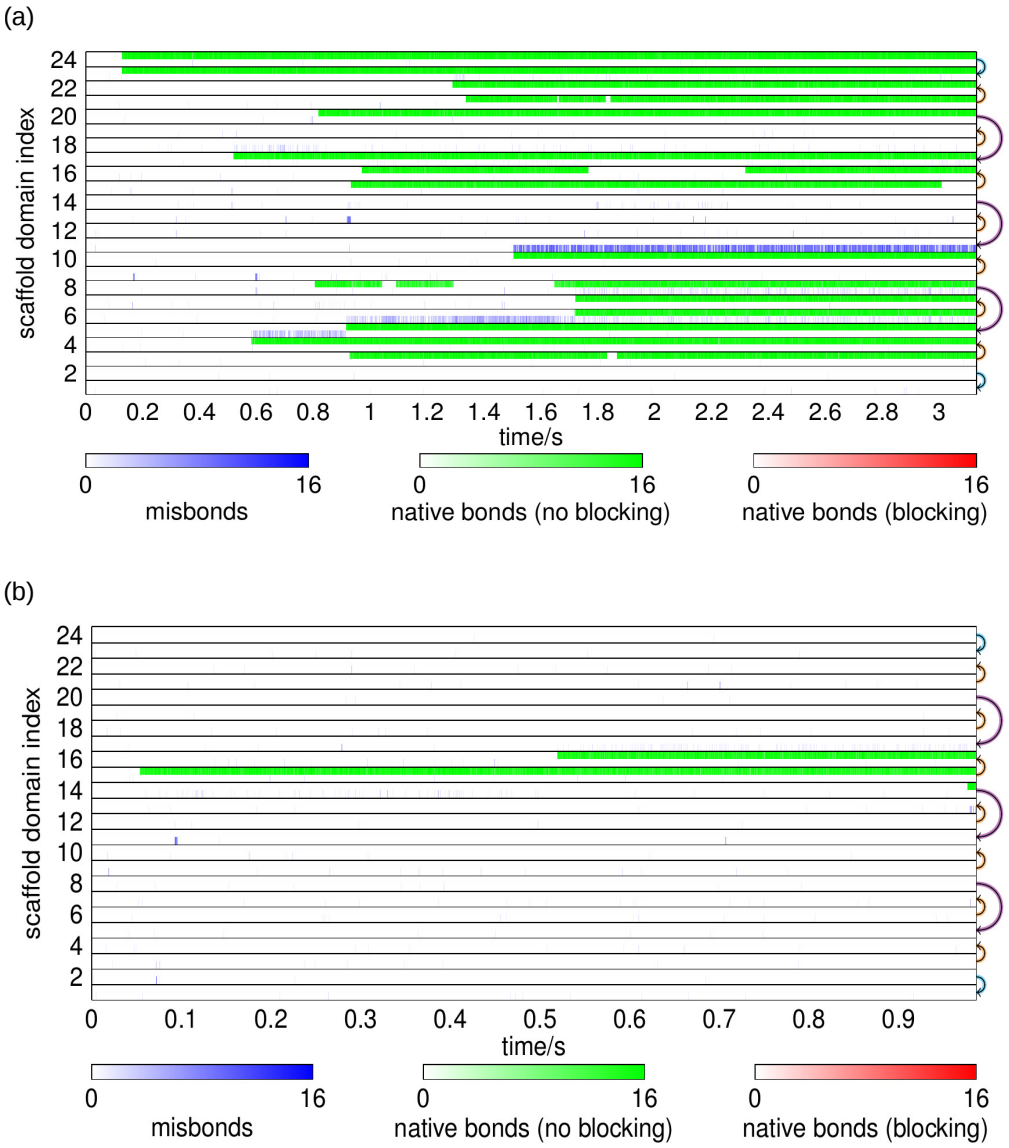

Figure 3.11: Kymographs for stoichiometric assembly simulations at (a) 65 °C and (b) 70 °C.

3.6.1 General features

The features of the stoichiometric assembly are in some ways very similar to those of the excess-staple assembly. The staples can bind more or less in parallel, misbonding is still evident, and the time scales for intra- and inter-molecular binding are still comparable, even though binding the first staple domain is now significantly slower due to the lower staple concentration (fortunately the lower number of nucleotides in the simulation box means that each simulation step is cheaper, so that simulating the assembly is still feasible). As with the excess-staple assembly, binding the second staple domain can be slow, especially at 50 °C where there is a lot of misbonding, and in the absence of staples bound to nearby scaffold domains which might help to provide a favourable scaffold geometry (cooperativity in forming the fourth domain around a four-way junction with two crossovers is still very clear).

However, by design the most detrimental effect for our assembly simulations, staple blocking, is no longer possible. As a result, a complete assembly was achieved, at 60 °C. The kymograph for this simulation is shown in Fig. 3.10(c) and snapshots illustrating this pathway are shown in Fig. 3.12. Note that this trajectory was the only one to reach completion of the four simulations run at 60 °C, all of which ran for roughly the same length of time (2.8 s). However, it seems very likely that the others would have also completed their assembly relatively easily if given more time as there are no obvious kinetic traps.

3.6.2 Optimal assembly window

As expected, we find that there is an optimal temperature window for the stoichiometric assembly on our simulation time scales. Misbonding is observed more frequently at

lower temperatures, as can clearly be seen from the regions of blue in Fig. 3.10(a). The misbonding hinders the assembly, because it makes it harder for both the correct staple to bind and for the staple involved in misbonding to bind to its correct domains. On the other hand, at higher temperatures (see e.g. Fig. 3.11(a) and Fig. 3.11(b)) the staple binding rate is reduced and the rate of melting correctly bound staples increases. These competing effects lead to assembly being fastest at intermediate temperatures. For the stoichiometric mixture, we ran simulations at 50 °C, 55 °C, 60 °C, 65 °C and 70 °C. We found that not only was the only complete assembly observed at 60 °C, but on average simulations at this temperature showed the largest number of correctly bound domains for any given simulation time.

3.7 Conclusions

The work presented in this chapter represents a significant milestone in the simulation of DNA nanotechnology: we have simulated the complete assembly of by far the largest DNA nanostructure ever studied using a nucleotide-level, continuous-space model. In the course of our investigations we have identified a number of interesting features of origami self-assembly.

We have identified staple blocking as the biggest barrier to assembly for our system. It is important to note that our assembly is atypical in two ways: the strand concentration is roughly 1000× larger than for a typical assembly, and the origami structure is roughly 20× smaller than a typical origami. The very high concentration will make blocking more likely, as it increases the binding rate of each staple's initial domain without affecting the binding rate for its second domain. Additionally, in order to see assembly on reasonable computational time scales, we had to work at a temperature where the binding of any staple to any scaffold domain is free-energetically downhill,

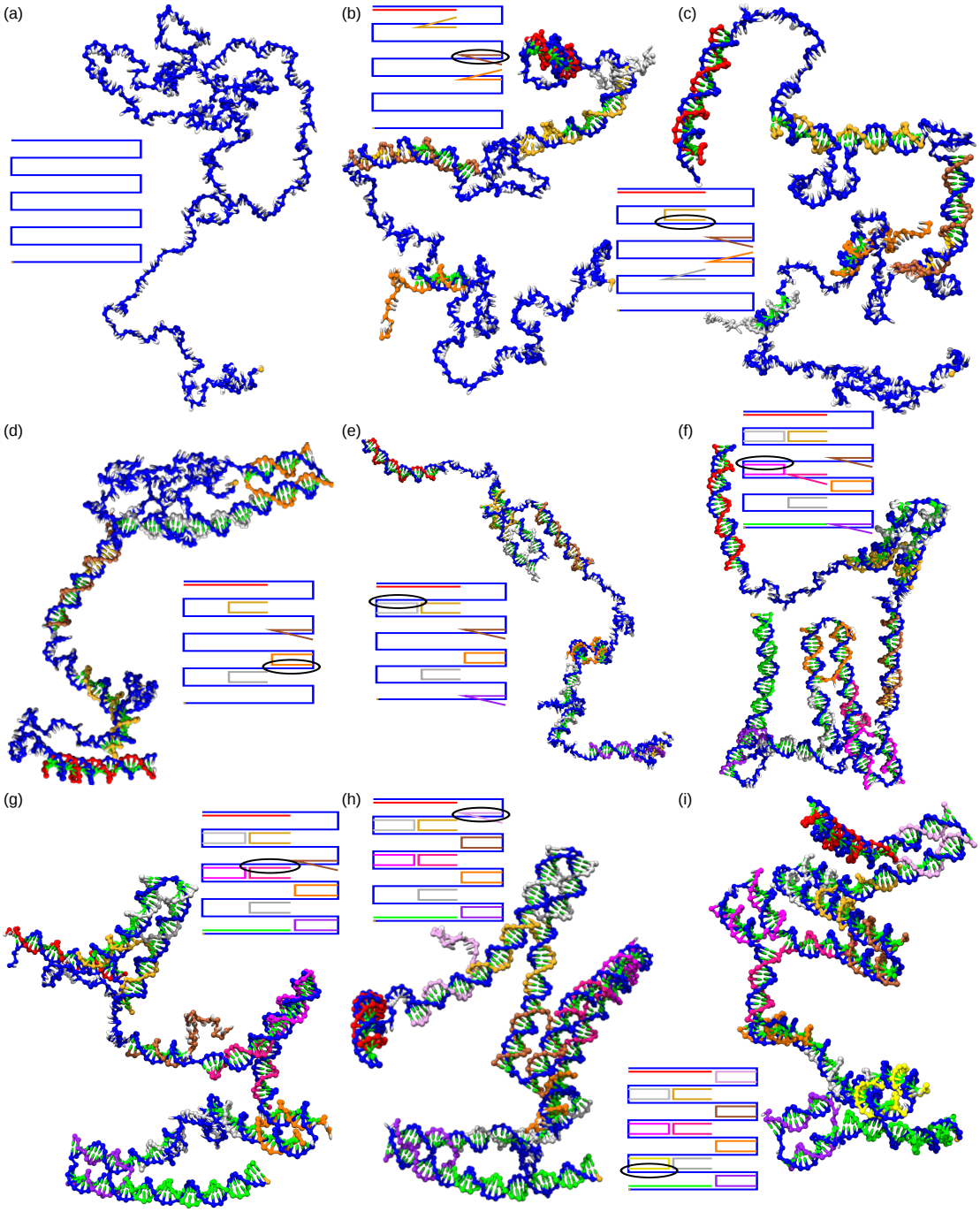

Figure 3.12: Snapshots along the pathway to complete assembly for the trajectory of Fig. 3.10(c). The completely assembled origami is shown in (i) – note that due to the small number of crossovers the structure is not very well constrained and so looks quite disordered compared to the snapshot shown in Fig. 3.1, where the origami is shown adsorbed onto a surface. The scaffold strand is shown in blue and staple strands are shown in other colours. Unpaired base sites are white, natively-bonded base sites are green, and misbonded base sites are yellow. Only staple strands with at least one base pair formed with the scaffold are shown. *Cadnano*[111]-style schematics are shown, illustrating the assembly progress, with ovals highlighting the most recently bound scaffold domain.

thus helping to stabilize the blocked state.

Nevertheless, there are several factors which suggest that staple blocking might also be important for experimental origami self-assembly. They are all related to the fact that experimental origamis are typically much larger and more complex than the simple design we have simulated. Firstly, in our design, a partially bound staple must only bind to one more scaffold domain, and that domain will not be too far away from the first, so that it is not too difficult for a staple to complete its binding. On the other hand a typical origami design, which is much larger and more complicated, will imply a much slower binding completion rate for each staple, leaving a much larger window of opportunity for a competing staple to bind to the scaffold in the mean time. In addition, a typical origami will need to incorporate many more staple strands, so that even a low probability of a given staple becoming blocked could lead to the blocking of a significant number of staples. Finally, staples which bind particularly distant stretches of the scaffold together are the most likely to be hindered by staple blocking, and are also likely to be the most vital for the proper folding of the origami.

Therefore, it seems reasonable to conclude that staple blocking may be an important obstacle to assembly for some designs, and certainly for unusually high strand concentrations. As noted earlier, experimental results for a DNA walker[25] show an effect analogous to blocking occurring at strand concentrations as low as 50 nM.

We emphasise that, whereas in our simulations a partially formed origami with a few staple blocks appeared to be the thermodynamic equilibrium near to the melting point, this will almost certainly not be the case at the low strand concentrations typical of experiment ($\sim$10 nM). In a low-concentration assembly, any blocking staples that bind will be slowly replaced by correct binding, as long as the assembly is cooled sufficiently

slowly to allow this to happen. We can see this from the very large free-energy penalty suffered from a blocking pair of staples (compared to a single fully bound staple), due to the loss of translational entropy of one of the staples. Thus it should be possible to overcome staple blocking with sufficiently slow annealing. However, if the system is cooled too quickly, or deliberately quenched to a lower temperature as in Ref. 34, blocking is one of the likely mechanisms preventing robust assembly of the target structure, and may lead to aggregation.

Ke and coworkers[35] found that positioning staple breaks so that each staple had a long (14-base-pair) middle domain and many shorter domains permitted high yields, while positioning the longest domains at the ends of the strands or having equal lengths for all domains led to little or no successful assembly. It is interesting to note that the success of the high-yield strategy is consistent with our findings regarding staple blocking. For an annealed assembly with staples with long middle domains (the high-yield strategy), we would expect each staple to experience the following (and therefore avoid being blocked): (i) initially, in a certain temperature window one copy of the staple is able to bind its long domain, leaving only shorter domains available for competing copies, so that any competing copies which happen to bind will quickly melt off (ii) either at that temperature or as the temperature is lowered further, this partially bound staple will be able to bind to the other, shorter scaffold domains before any copies can come in and block it.

In general the assembly we have studied is relatively parallel in nature – that is, assembly in distant parts of the structure occurs relatively independently, with no set order in which the staples bind. This is again probably in part due to the necessity of simulating the assembly at a temperature where binding of a staple to any scaffold

domain is stable, in order to make assembly computationally tractable. Having said this, there is also evidence for cooperativity between staples in certain cases, which occurs when staples bring together distant parts of the scaffold strand, enhancing the geometry for other staples to bind their second domains. This inter-staple cooperativity, which has been incorporated in theoretical studies of origami assembly,[106] is particularly evident in the case of the four-way junction formed by the L and M staples.

For a typical experimental design, which will be larger and potentially more complex than our simple origami, there is a greater potential for cooperativity, as (i) a partially bound staple can have more neighbouring staples than for our design, which can lead to an even more constrained, and so more favourable, local scaffold geometry, (ii) scaffold domains which are to be connected by a single staple may be separated by a much greater distance along the scaffold strand than for our small origami design, so that larger gains are made by constraining the scaffold geometry, and (iii) on annealing in experiment assembly is likely to first occur when there is a nucleation barrier to assembly, and where it is more favourable to first bind staples that involve less loss in scaffold entropy. This enhanced cooperativity suggests the possibility of there being nucleation sites in partially assembled origamis, where closing one staple accelerates the closing of the neighbouring staple, which on closing aids its neighbour, and so on.

In addition, we expect a clear intra-staple cooperativity effect at experimental strand concentrations, which we confirm by simulating origami assembly at these concentrations in Chapter 4. Although in the simulations presented in this chapter the time scales for binding the first staple domain are similar to those of binding the second staple domain, this is not the case at experimental concentrations, which greatly reduce the rate of binding the first staple domain while leaving the rate of binding the second

domain unchanged. Thus, once a staple has bound its first domain to the scaffold, the second domain will bind relatively very fast for the origami we have considered.

Optimal temperature windows for origami assembly have been identified in experiment.[34] We see similar features in our simulations. At lower temperatures, misbonding is more prevalent, which hinders correct binding. For the more complex designs typical of experiment, it is likely that this effect will be even more pronounced, as there is more opportunity for misbonding to cause kinetic traps if the assembly is cooled too rapidly. Meanwhile, at high temperatures staples form native base pairs with the scaffold more slowly, and staples that have partially or fully bound to the scaffold are more likely to melt off.

However, in our simulations, in order to see assembly on computationally feasible time scales, the temperature had to be sufficiently low that the native binding of a staple to a scaffold domain was always free-energetically downhill. This condition, though, leads to, for example, blocking being a potential kinetic trap hindering assembly. In experiment, it is likely that assembly occurs with highest fidelity at somewhat higher temperatures where there is a nucleation free-energy barrier to assembly that helps prevent the formation of incorrect structures. Performing assembly simulations in this regime is a challenge that we address in the next chapter.

4 Simulation of the self-assembly of a small DNA origami using forward flux sampling

In Chapter 3 we examined the assembly of a DNA origami by brute force simulations using the oxDNA model. Such simulations are limited both to high concentrations and to temperatures where domain binding is almost irreversible, if significant progress with the assembly is to be observed. In this chapter we study the assembly of a similar origami with oxDNA using the FFS rare-event simulation method[95] described in Section 2.4. The method, which uses checkpointing to allow large free-energy barriers to be crossed more easily, enables efficient estimation of assembly rates and efficient generation of thermodynamically representative assembly trajectories. Thus, this approach has the potential both to consider assembly over a much wider range of conditions and to provide more rigorous statistical information about the assembly process, rather than a few trajectories that we hope to be typical.

4.1 The model system

The origami design used in this chapter is shown in Fig. 4.1 and is similar to that used in Chapter 3. As with the earlier design, it consists of a 384-base scaffold with 24 domains of length 16 bp each. However, in this design there are 8 staples of 3 domains each, rather than 12 staples of 2 domains each, and the scaffold strand is circular. We use a design with 3 domains per staple in order to permit a larger and more varied set of possible transitions and intermediate assembly states, and we use a circular scaffold strand to reflect the widespread usage of circular scaffold strands in

experimental origami designs. With more crossovers, the complete origami also has a more well-defined structure.

In Fig. 4.1, each domain is given a unique label X*n*y, where X can be "A", "B", "C" or "D", *n* ranges from 1 to 3, and y can be "a" or "b". Each staple is uniquely defined by its values for X and y, with *n* differentiating the three domains of the staple. Due to the rotational symmetry of the origami, an X*n*a domain is identical to an X*n*b domain (i.e. in terms of the local geometry around the domain), except for sequence. We use random DNA sequences for this study, which are given in the appendix in Table B.3.

Without using a rare-event method, it would be very difficult to simulate origami assembly (or even simple duplex formation) at strand concentrations typical of experimental assembly protocols ($\sim$10 nM), because most of the simulation time would be taken up waiting for the strands to find each other before they have the chance to bind. This is why we used relatively high strand concentrations when studying assembly with direct simulation in Chapter 3. However, using FFS allows us to simulate strand association at experimental concentrations without the computational cost being prohibitively high. Thus, we can use experimental strand concentrations for this study, allowing us to relate any concentration-dependent effects we see in our simulations to experiment more easily. We use a scaffold strand concentration of 10 nM and a staple strand concentration of 50 nM for each staple species, which are the strand concentrations used for the experimental origami assembly protocols in Ref. 9.

We choose a temperature of 70 °C for the simulations, which is roughly equal to the estimated melting temperature of the binding-unbinding transition for a single staple strand on the scaffold. In experimental assembly protocols the temperature is typically lowered from around 90 °C to room temperature over many hours. We chose to work

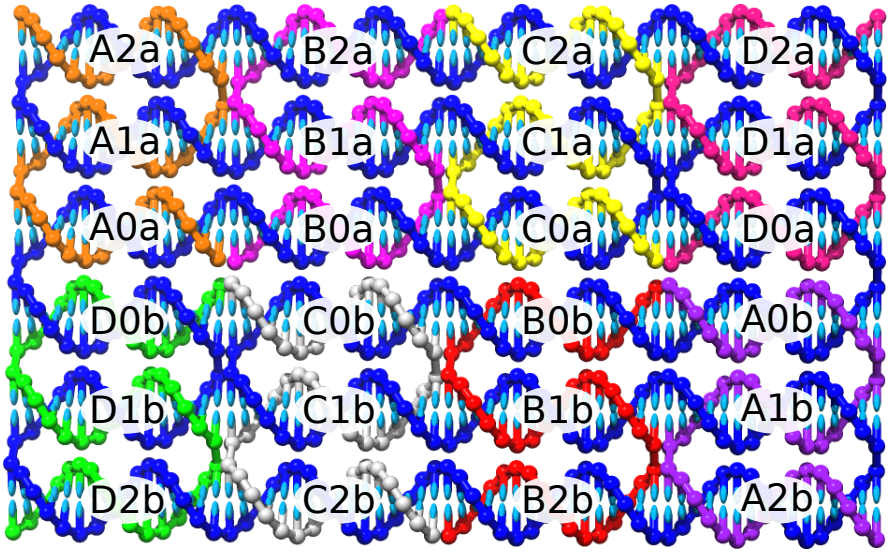

Figure 4.1: The origami design chosen for the origami assembly study using FFS. For clarity an "ideal structure" is shown, rather than a representative equilibrium structure. Each domain is given a unique label X*n*y.

at 70 °C under the assumption that the assembly takes place at a single temperature (or a very narrow temperature range) where staples bind to and unbind from the scaffold relatively often, which should correspond approximately to the melting temperature of a staple. This idea is consistent with the constant temperature assembly protocols successfully used in Ref. 34. The melting temperature for a staple was estimated as being roughly 10 K lower than the melting temperature of an average-sequence 48-bp duplex at a concentration of 10 nM (as given by the SantaLucia model[112]), as each staple forms 48 base pairs with the scaffold strand, but will be less stable than a 48-bp duplex because it must close two scaffold loops in order to hybridize all three of its domains.

4.2 Splitting up the assembly pathway

Using the FFS algorithm requires us to define an initial and final state for the process we want to simulate. These are commonly known as the A-state and the B-state, respectively. Note that in general, the A-state must be a well-defined stable state. For

our study of the origami self-assembly, the simplest choice conceptually would be to define the A-state as the state with all staple and scaffold strands unbound, and the B-state as the completed origami. However, this definition would be very problematic for simulating origami assembly with oxDNA, because of the many metastable states along the assembly pathway. These metastable states correspond to some domains on the scaffold strand being fully bound by staple domains, and some not being bound (as opposed to intermediate states where one or more staples are partially bound to a scaffold domain). These are a problem because, in FFS, the accelerated sampling only happens in the forwards direction: simulations run from intermediate points in the assembly must be able to return to the A-state without the help of checkpointing or any biasing, and any metastable states between the current intermediate point and the A-state will greatly slow down the simulation of these processes.

To avoid this problem, we split the assembly into separate FFS calculations from one metastable state to the next, as shown in Fig. 4.2. From now on we will refer to these metastable states, plus the initial and final state of the assembly, as "assembly states." Although this splitting up of the assembly necessitates many FFS calculations in order to fully sample the assembly, each of these calculations is tractable.

We noted above that in FFS the A-state must be a stable state – but because we have split up the assembly as described, we must now use metastable states for our A-states. The important condition here is that the time spent in the A-state must be large compared to the amount of time a transition takes (i.e. the amount of time between leaving A and arriving at B for trajectories that make it from A to B), and fortunately it turns out that this condition is satisfied for our A-states and transitions.

We label the (meta)stable assembly states as Sn according to the number of fully

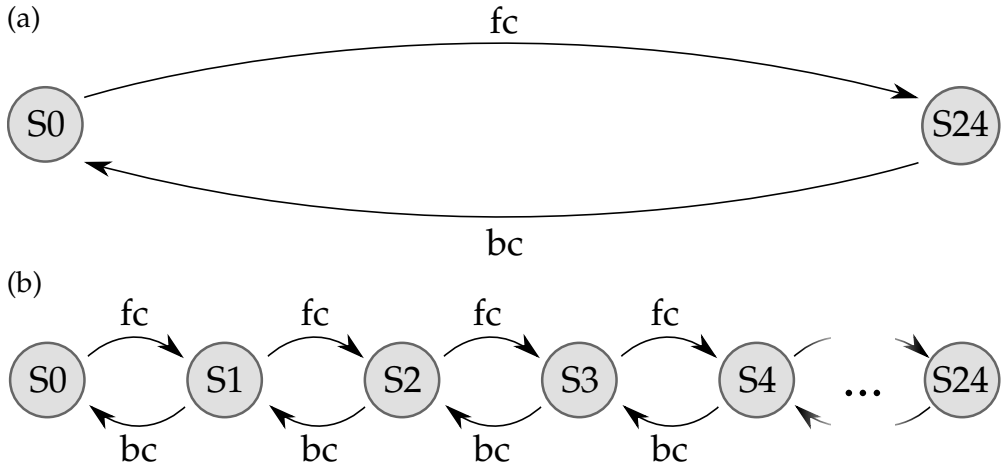

Figure 4.2: Two possible schemes for studying origami assembly with oxDNA using FFS. (a) shows a scheme where just one forwards and one backwards calculation is performed. (b) shows the scheme used for this study, where the assembly is broken into 25 "assembly states," requiring 50 FFS calculations in total. The assembly states are labelled Sn, and each one corresponds to a state with exactly n domains bound, with no partially bound domains. "fc" denotes a forwards FFS calculation, and "bc" denotes a backwards FFS calculation.

bound staple domains n in that state, so the beginning of the assembly process, where the scaffold is completely free of staples, is called S0, after one staple domain has bound the partially formed origami is in assembly state S1, and so on.

In most cases, each assembly state Sn has a very large number of possible combinations of bound scaffold domains. For example, for S8 there are there are 735 471 possible combinations of bound domains. We might try to reduce this number by assuming that each staple fully binds before the next one binds to the scaffold, so that each assembly state has between 0 and 8 fully bound staples and up to 1 partially bound staple. It turns out that this almost always holds in our simulations. In addition, we can use the symmetry of the origami under rotations of 180 $^\circ$ in its plane, which leave it unchanged except for sequence, to further reduce the number of combinations by one half. However, even with these simplifications there are still 252 possible combinations for S8, for

example.[1]

As we cannot hope to properly sample transitions from all possible combinations of bound domains to all others, instead we must use a subsample. The method we used was as follows. We started by generating equilibrated configurations in the S0 assembly state, the assembly state with all scaffold domains unbound. We then ran a FFS calculation from the S0 assembly state to the S1 assembly state. For the calculation of the S1 to S2 transition, we chose a random subsample (with each configuration having an equal probability of being chosen, and multiple picks of the same configuration being permitted) of the configurations at S1 generated by the S0 to S1 FFS calculation. We then used a subsample of the configurations generated at S2 for the calculation of the S2 to S3 transition, and so on.

One might hope that the ensemble of configurations generated in this way at each assembly state Sn would be representative of the ensemble of Sn substates generated by the true dynamics of the system. However, this may not be the case both for a fundamental and a practical reason.

To understand the fundamental reason, consider a case where S9 is the highest free-energy metastable intermediate. In this case, for $n \geq 9$ the transitions to higher n will be free-energetically downhill, and so the ensembles for each of the assembly states should be predominantly determined by the forwards rates, and our approach should be a good approximation. However, one would expect assembly states Sn with $n < 9$ to be in quasi-equilibrium, so that the appropriate ensemble of configurations for each of these states should reflect both the forwards and backwards rates into and out of a state, whereas in our scheme the Sn population is just governed by the forwards

[1] Details of how the number of combinations are calculated are given in Appendix B.3.2.

rates. If the backwards rates were all equal, this would not be a problem. Indeed, for origamis, one possible reason for the different stability of substates within a particular assembly state S*n* is differences in scaffold entropy, with bound domains that bring more distant parts of the scaffold together involving a greater loss of entropy. These differences would be expected to be mainly evident in the forwards rates rather than the backwards rates, because the transition state already has the first few base pairs formed. However, we know from previous studies that differences in stability can sometimes be more apparent in the backwards rates, e.g. for the dynamics of hybridization between strands that can form hairpins.[74;99]

This picture of a free-energy barrier to assembly is supported by experimental studies[34;113] that have shown significant hysteresis between origami assembly and melting, despite the relatively slow heating and cooling ramps used. These results imply that there are significant free-energy barriers to origami assembly at the equilibrium origami melting temperature, and only once the system is supercooled (in the case of assembly) or superheated (in the case of melting) is the barrier sufficiently small for that process to occur.

The practical reason that the S*n* ensembles we generate might not be representative of the real dynamics is simply that there are so many possible states that, without extremely extensive sampling, it is likely that the dominant substates will be determined by which of the transitions happen to have been more sampled, rather than those that are the part of the fastest pathways for the state space. This is probably the greater limitation for this study.

Thus, any computation of the overall rate based on the scheme in Fig. 4.2(b) will have a hard-to-quantify error associated with the approximate character of our Sn en-

sembles. Furthermore, the dominant pathways through our ensembles of states should be interpreted with caution. However, this is not disastrous, as we can still compute rates and reactive trajectories from the individual substates of Sn to the assembly state Sm. For example, for our S1 to S2 calculation, we can compute the rate and trajectories for the transition from the substate with A1a domain bound to the assembly state with two domains bound, and so on. This in itself can provide us with a wealth of information.

We ran the FFS flux and shooting runs by simulating the oxDNA model with a MD algorithm. We use the same time unit conversion factor that we used in Chapter 3, 10^{-8} s per simulation time unit. For the forwards FFS calculations, we defined our discrete order parameter using conditions that, early on in the process, required the relevant staple and scaffold domains to be within a certain distance of each other, and subsequently required a certain number of correct base pairs (i.e. base pairs present in the target structure) to be formed. For the backwards FFS calculations we used conditions first requiring fewer correct base pairs to be formed, and then requiring the relevant domains to be further apart. A complete description of the order parameters and the other simulation parameters that we used is given in Appendix B.3.3.

4.3 Results

We refer to an FFS calculation where the B-state has one more domain bound than the A-state as a *forwards calculation*, and a calculation where the B-state has one fewer domain bound than the A-state as a *backwards calculation*. Using a considerable amount of computational resources, we were able to perform forwards FFS calculations from S0 to S15, and backwards FFS calculations from S3 to S0. Fig. 4.3 shows a snapshot from the initial state of the assembly, with no staples bound to the scaffold (contrast this snapshot

with the initial assembly state snapshots from the direct assembly simulations in Fig. 3.2, which show a far more densely packed system), while Fig. 4.4 shows snapshots of various other assembly states.

We emphasise that the error bars shown in the figures in this section should be taken as very rough estimates of the uncertainty in the measurements. They show the standard error on the mean (i.e. the value of the data point) for up to four independent estimates from independent FFS rate calculations. We note that this estimate of the error is necessarily rough when only up to four samples are available. For example, if two samples are available and happen to lie close together, a very small error bar will be drawn, greatly overstating the precision of the measurement. Separate to this consideration is the possibility of systematic errors, which we return to at the end of the chapter. With these points in mind, we are careful about drawing strong conclusions from the data without evidence from another source such as separate simulations.

Details of how the data for each plot was calculated are given in Appendix B.3.5.

4.3.1 Average rates between assembly states

The average forwards rates are shown in Fig. 4.5, with each point representing an average over all sampled transitions between the substates of the initial and final assembly states. Almost all of the transitions had A-states with at most one partially bound staple, shown as blue in the plot. For these points, there is a clear separation in time scales between transitions which involve binding a new staple (S0 to S1, S3 to S4, etc.), which occur at a rate of approximately $10^{-3}\,\mathrm{s}^{-1}$, and those which involve binding further domains of a staple that is already partially bound (S1 to S2, S2 to S3, etc.), which occur at a rate of approximately 1 to $10\,\mathrm{s}^{-1}$. However, the S3 to S4 rate

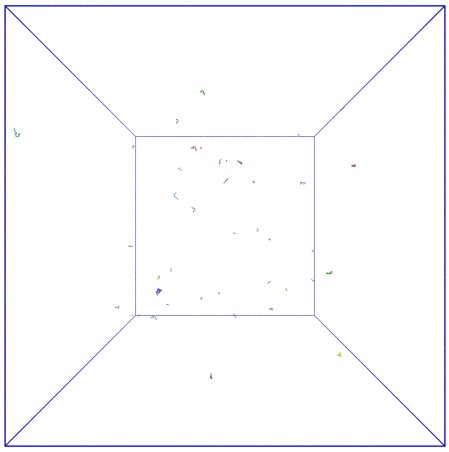

Figure 4.3: A snapshot showing the entire simulation box at the beginning of the assembly (assembly state S0). Contrast this very sparsely populated system with the densely populated ones shown in Fig. 3.2 for the direct assembly simulations of Chapter 3.

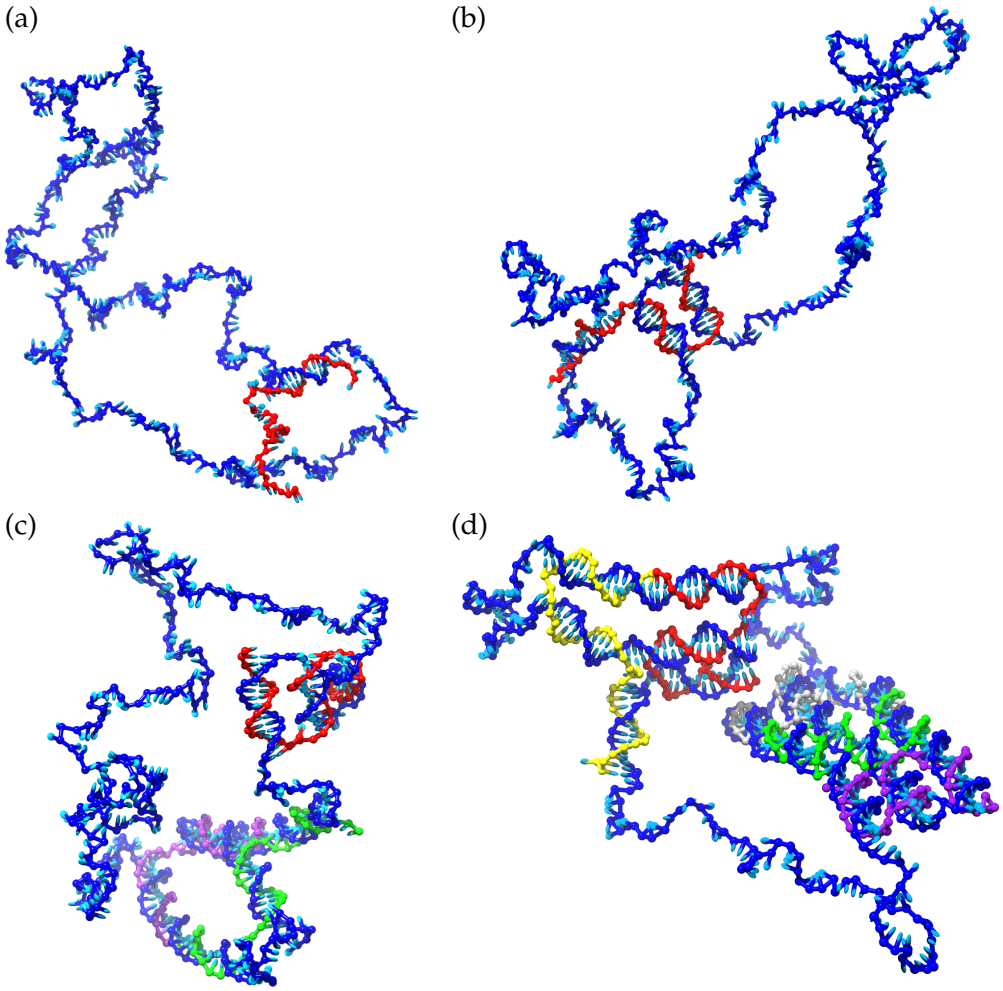

Figure 4.4: Snapshots showing the scaffold strand (blue) and any staple strands bound to it (other colours) for assembly states (a) S1, (b) S3, (c) S9, and (d) S15.

calculation happened to include transitions from an S3 assembly state involving two partially bound staples. These transitions could be completed without a new staple coming in from the solution to bind to the scaffold; hence the rate of binding for these transitions (green point in the plot) is much higher than for the other transitions between substates of S3 to substates of S4.

The separation of time scales between initial staple binding and the formation of subsequent domains originates from the qualitative difference in the processes underlying these transitions. For initial binding, the staple and scaffold strands must happen to come into close proximity with each other. By contrast, a staple that is already partially bound to the scaffold has its remaining single-stranded domain(s) localised close to the scaffold strand. Clearly, the rate of initial binding will be concentration dependent, with a higher concentration increasing the rate, while the rate of binding the staple's subsequent domains will be concentration independent. Indeed, for the relatively high strand concentrations used in Chapter 3 ($\sim 10\,\mu$M) we found that the time scales of initial and subsequent staple domain binding were comparable.

The time scale separation seen in this study indicates that the assembly mostly proceeds with each staple fully binding in turn, rather than several partially bound staples occupying the scaffold strand at once. This is also reflected in the nature of the B-state configurations generated by the FFS calculations, which followed this pattern in the great majority of cases, with the exception of a few configurations generated for the S2 to S3 transition as noted above. Comparing time scales computed with a coarse-grained model is not straightforward;[82] but as both processes require a diffusive search for the complementary domain, albeit one process being intermolecular and the other intramolecular, we expect this time scale separation to be robust. Therefore,

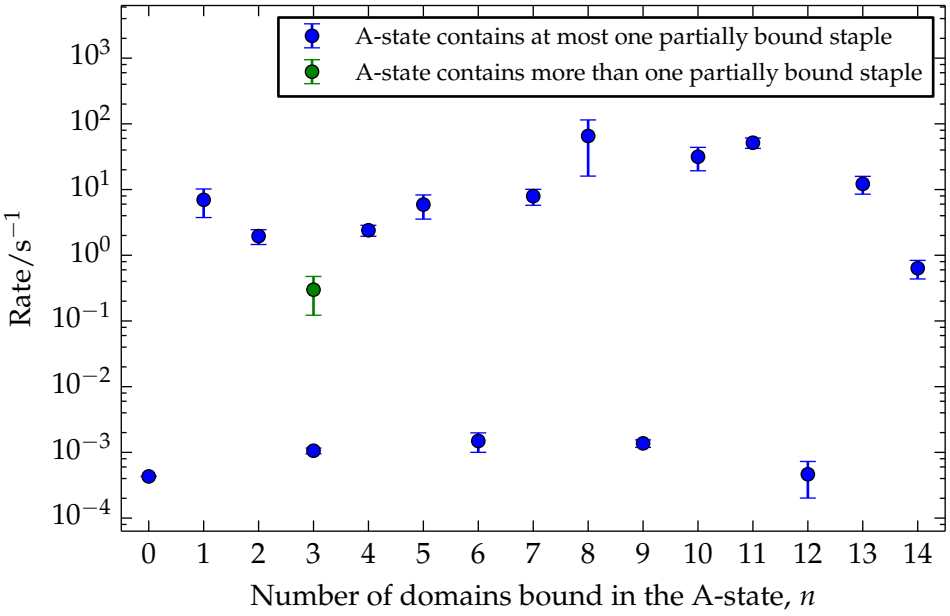

Figure 4.5: The average forwards rates for the origami assembly computed from FFS calculations. Error bars show the standard error on the mean calculated from four independent FFS calculations (not applicable to the $n = 0$ rate). There was at most one partially bound staple in the A-state (blue points), except for some of the S3 to S4 transitions, for which the A-state consisted of two partially bound staples (green point), boosting the rate.

the "blocking" events we observed in the direct assembly simulations in the previous chapter are unlikely to be relevant for such small origamis at concentrations typical of experiment.

For the assembly of full-size origamis of ∼7000 base pairs, however, this time scale separation may not be so clear cut. The staple strands present in a full-size origami may take much longer to complete their binding, for example due to difficulties with threading[114] or with closing very large scaffold loops, while the rate of initial binding should be unchanged. Thus, the gap in time scales may be sufficiently compressed that there is no longer a clear separation, and so staple blocking is much more likely to be relevant for full-size origamis.

It is tempting to read too much into the overall trend of the second and third staple

domain binding rates, for example the fall in rate for the S13→S14 and S14→S15 transitions. We have already emphasised that the error bars presented should not be taken as rigorous estimates of the real error for each measurement, and we have noted that not all possible transitions between all of the substates of the relevant assembly states can be sampled using our FFS scheme. Further, the number of substates represented at each assembly state decreased as the forwards calculations went on. For example, the S5 to S6 transition was calculated for 11 different substates from assembly state S5, while the S14 to S15 transition was calculated for only 3 different substates from S14, even though many more substates exist for S14 than for S5. Taken together, this means that the overall error may be larger than that suggested by the error bars. For these reasons, we do not feel confident in drawing any conclusions from an apparent trend in second and third staple domain binding rates.

The average backwards rates are shown in Fig. 4.6. The rates for all three backwards transitions are on the order of $1\,\mathrm{s}^{-1}$, comparable to the second- and third-staple-domain binding rates, and much faster than the first-domain binding rates (Fig. 4.5). This indicates that the binding of the first staple is uphill in free energy for the concentrations and temperature considered here, with the average forwards and backwards rates giving an estimate of the effective ΔG for binding the first staple of $\sim +9k_{\mathrm{B}}T$. Note that the S3→S2 transition was sampled with 10 times less simulation time than the other two assembly state transitions (and the other transitions presented in this chapter), and hence there is a larger uncertainty associated with the rate for the transition.

As with the forwards rates, not too much should be read into the small apparent rise in the melting rate with the number of domains bound. A very similar process to these transitions is duplex melting, for which secondary structure formation [74;99] and

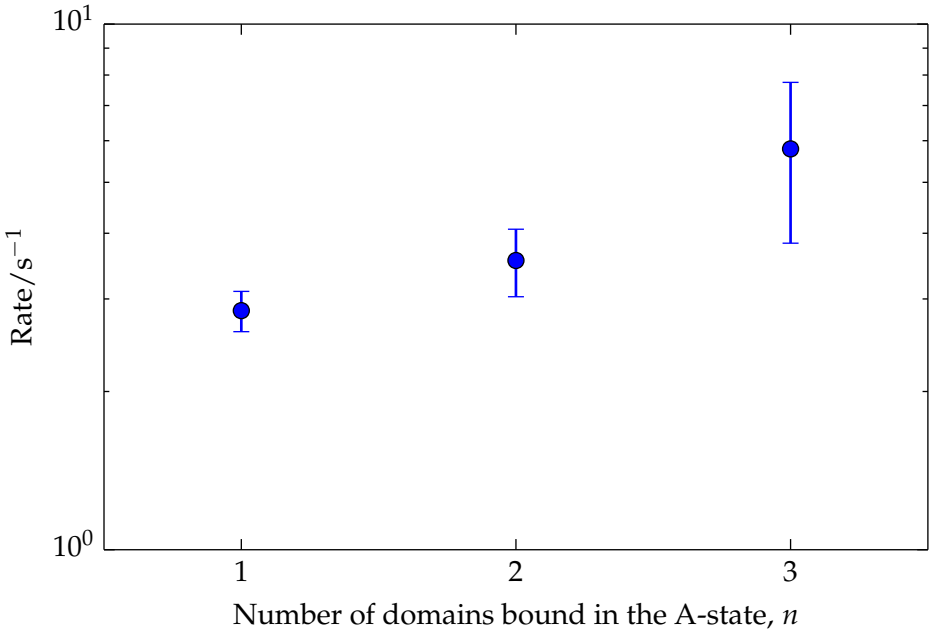

Figure 4.6: The average backwards rates for the origami assembly computed from FFS calculations, showing an approximately uniform rate for the three transitions.

sequence[70;115] can play a significant role. For the temperature considered (70 °C) and the random DNA sequences we have used, secondary structure is unlikely to play a role, while sequence effects should be mostly eliminated for this study because we used the average-interaction-strength version of oxDNA.

This leaves (anti-)cooperative effects between domains as factors likely to lead to differentiated backwards rates. These might include coaxial stacking between double-stranded domains, which is likely to somewhat stabilize the bound state and thus slow melting, and geometric effects such as needing to thread a long tail through a small loop (discussed in the next section). Another consideration may be how constrained the relevant scaffold strand region is, which may affect the scaffold strand entropy gained on breaking each base pair of the domain. This will depend on which other scaffold domains are bound.

As we only considered melting of the first staple, we did not sample any transitions

where inter-strand cooperativity might play a role. However, we do see one example of intra-strand anti-cooperativity, which we will discuss in the next subsection. With the exception of this example, we observe approximately uniform rates for the backwards transitions between substates that we have sampled, as shown in Fig. B.1, Fig. 4.11 and Fig. B.2.

4.3.2 Order of binding staple domains

We expect the rates we calculate for transitions between different substates for a given initial and final assembly state to depend somewhat on the geometry involved. To take an extreme example, it may be that some transitions are topologically forbidden, without first melting a domain or breaking covalent bonds in the DNA backbone. In this section we examine some of these geometric effects.

For the following analysis, we consider only the transitions for which the initial and final assembly states contained at most one partially bound staple. This means that staple binding is always sequential, so that each staple completes its binding before another one binds to the scaffold.

Each staple strand has two end domains and one middle domain. One could imagine that it would be easier for the initial binding to the scaffold strand to occur at one of a staple's end domains than at the middle domain, because completely binding the domain then requires less of the staple strand to be wrapped around the scaffold strand. Also, for duplex formation in oxDNA it has been shown that initial contact formation is more likely to occur near to a chain's end.[70] Here, we investigate potential differences in behaviour between end and middle domains by splitting the forward transitions from initial assembly states with no partially bound staples into two groups,

one where the final configurations had a newly bound staple with an end domain bound to the scaffold, and one where the final configurations had a newly bound staple with a middle domain bound to the scaffold. We then calculate the average rates in each group, giving each transition equal weight. The resulting data is shown in Fig. 4.7 and suggests that binding a staple end domain may be slightly faster than binding the middle of the staple, although any effect is relatively small (with an average speed-up by a factor of 2.1).

A similar analysis can be made for the transitions involving binding of the second staple domain. In particular, we investigate whether the rate of binding the second staple domain is faster when the first domain is in the middle of the staple or when it is at one of the staple ends. The average rates for the two groups are compared in Fig. 4.8, and indicate that binding the second staple domain may be faster when the initial domain is in the middle of the staple. Plausible reasons for this speed-up are (i) both of the staple ends have less DNA to wrap/thread than if the staple was initially bound by its end domain, and (ii) because both of the other domains are adjacent to the bound middle domain, thus potentially making binding easier because the complementary scaffold domain is likely to be closer to the staple.

To explore case (i) further, we examine particular instances where a long tail is likely to be detrimental because it must be threaded through a small loop. Such instances will always occur (regardless of other staple species being bound to the scaffold) when a staple binds its first domain to A2y and its second to A1y; or first C0y and then C1y; or first D0y and then D1y; or first D2y and then D1y, where y can be either a or b for a particular transition. The results in Fig. 4.9 show that transitions involving threading of a long tail through a small loop are indeed noticeably slower than the other

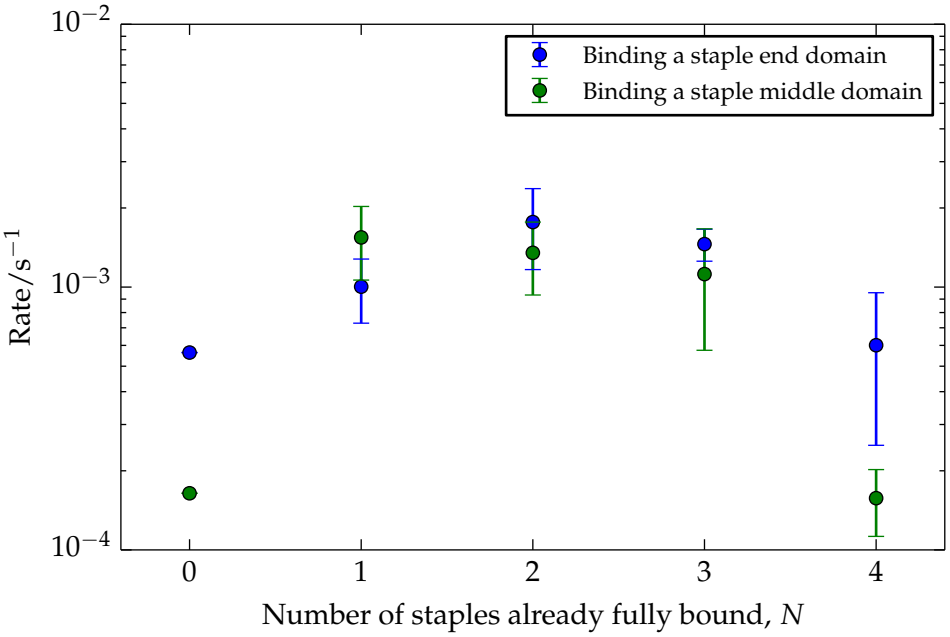

Figure 4.7: The average rate of binding the first domain of a staple, with the transitions split into two groups: one for which the domain being bound is at the middle of the staple (green), and one for which it is at one of the staple ends (blue). Error bars show the standard error on the mean calculated from four independent FFS calculations (not applicable to the $n = 0$ rate). The data suggests that binding an end domain may be slightly faster than binding a middle domain.

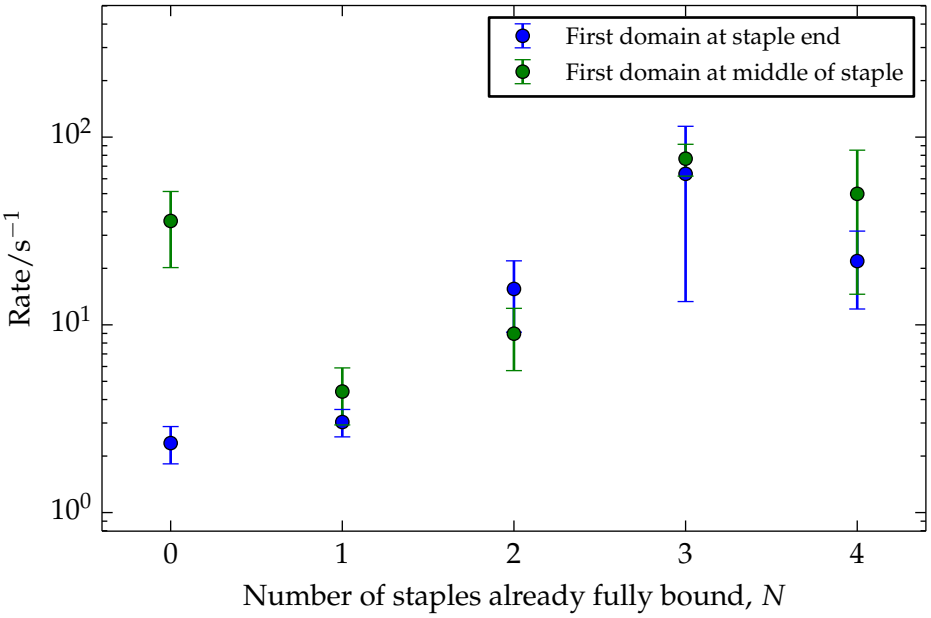

Figure 4.8: The average rate of binding the second domain of a staple, with the transitions split into two groups: one for which the first domain which bound was at the middle of the staple (green), and one for which it was at one of the staple ends (blue). There is a suggestion that binding the second domain may be faster when the middle domain was the first to bind.

transitions, with an average transition requiring threading being 33× slower than the others. Note that these results are taken from the S1 to S2 transition only – the signal is washed out for the later transitions, possibly because the extra staples bound to the scaffold introduce extra constraining effects on the geometry of the scaffold strand that make comparing like with like difficult. Simulation snapshots illustrating the threading process are shown in Fig. 4.10.

We see a suggestion of the same effect for the backwards transition S2→S1 (Fig. 4.11). There is not much data available for the relevant transitions (one of the four independent runs sampled one of the transitions, and the rest sampled none), but this is suggestive in itself, because slower transitions will on average be sampled less often under the scheme we have used. This is also consistent with a picture where threading issues are just a kinetic effect, and so should not affect the thermodynamic stability of the fully

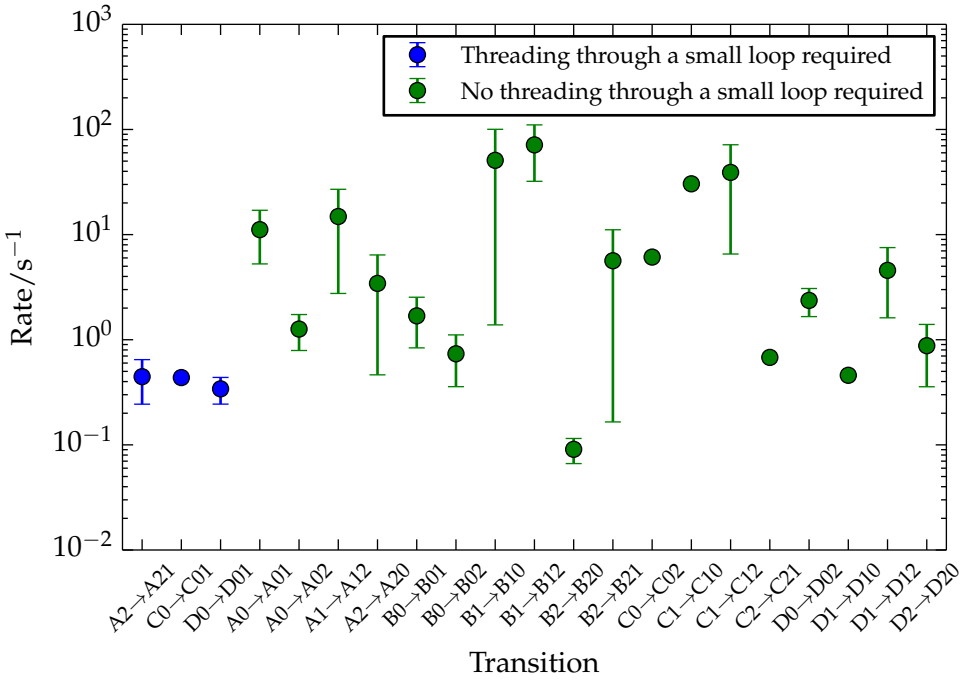

Figure 4.9: The rates for the transitions from each of the substates of S1 to the substates of S2. The x-axis labels indicate the substates that are involved in the transition. A compressed notation is used, where, for example, A0→A01 means the A-state has A0y bound and the B-state has A0y and A1y bound. Points in blue correspond to cases where threading through a small loop formed by the scaffold and partially bound staple is required (see Fig. 4.10 for simulation snapshots).

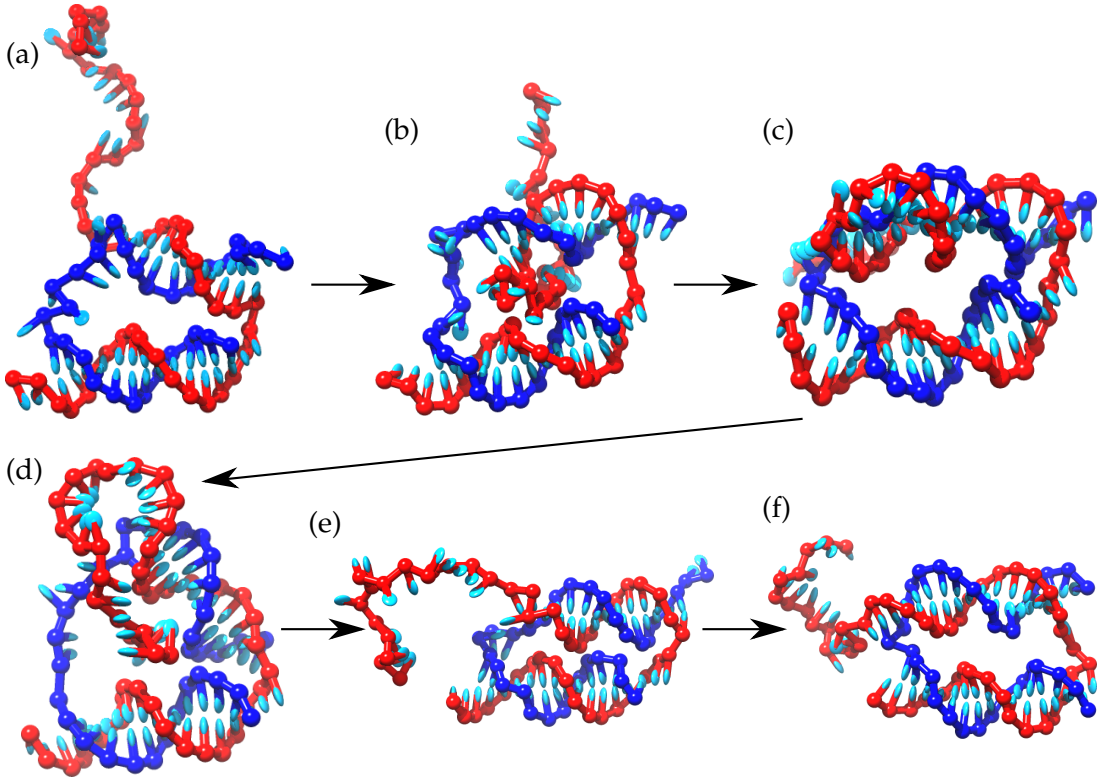

Figure 4.10: Simulation snapshots showing a staple strand threading a one-domain tail through a small scaffold-staple loop to complete the binding of its second domain. The staple strand is red and the scaffold strand is blue, with only the two relevant domains of the scaffold strand shown. (a) The second domain has formed most of its base pairs but is unable to complete its binding. (b) – (d) the staple strand pushes its near end through the loop formed by the staple and scaffold. In (e) the staple strand is completely through, and in (f) it takes advantage of the now correct strand wrapping to complete most of its base pairs. Note that a significant amount of fraying of the base pairs at either end of the scaffold-staple loop is evident, especially for (a) – (e), suggesting that this fraying may be necessary for the process to happen.

bound state. As expected from this picture, the size of the slow downs in the backward and forward rates look comparable.

We examine case (ii) by grouping the transitions involving the binding of a second staple domain into two groups: one for which the second domain to bind is adjacent along the staple strand to the first, and one for which the two domains are at either end of the staple. The average rates for the two groups are plotted in Fig. 4.12. We see that adjacent staple domains do bind significantly faster than non-adjacent domains, which, everything else being equal, implies that the staples should on average bind their second domain more quickly when the first domain is in the middle of the staple.

Fig. 4.14 shows the average rate of binding for the final domain of a staple. Again we split the transitions into two groups, one where the final domain to be bound is in the middle of the staple, and one for which it is at one of the staple ends. The plot suggests that binding a middle domain may be faster than binding an end domain. This result may seem surprising, because it might at first sight appear that wrapping the middle domain around the corresponding region of the scaffold is impossible without melting one of the other staple domains to free one of the staple ends. However, it is possible to do this in some cases, as shown in Fig. 4.13. In addition, even if the topology of the system means that is not possible to correctly form the middle domain without introducing some incorrect wrapping somewhere in the structure, it may be that there is still enough geometric freedom to allow this middle domain to form sufficient base pairs to be deemed to have reached the B-state. Although one might envisage that the mechanism for middle domain binding is kinetically slower than wrapping of a single-stranded tail, this putative effect may be more than compensated for by the fact that, for the origami we are studying, the middle staple domain can be expected to be

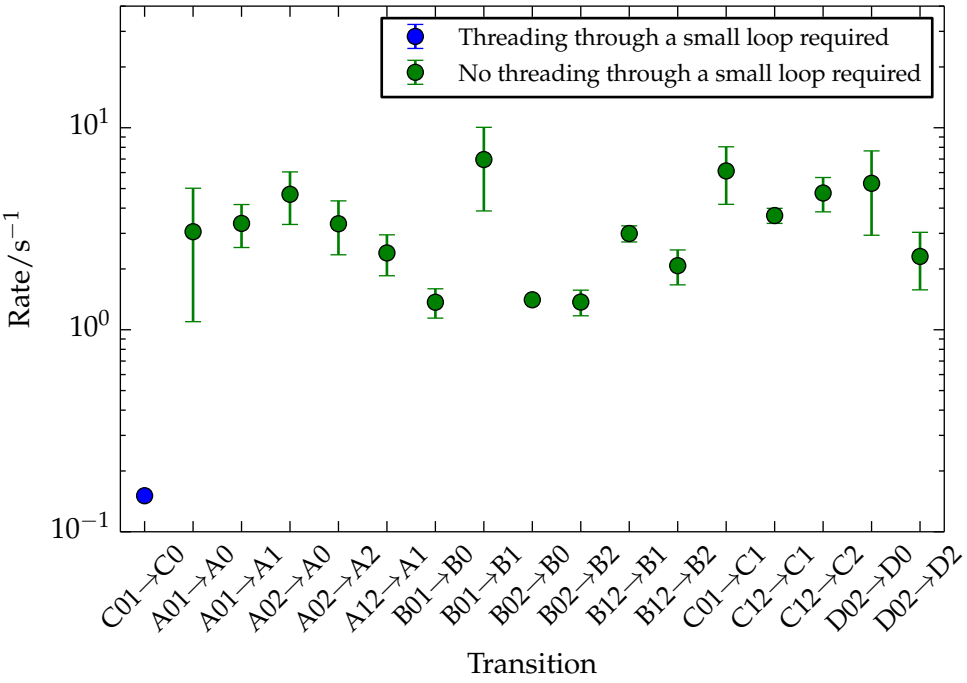

Figure 4.11: The rates for the transitions from each of the substates of S2 to the substates of S1. A compressed notation is used, where, for example, A0→A01 means the A-state has A0y bound and the B-state has A0y and A1y bound. Although the transition requiring the unthreading of a long staple tail through a small loop (blue) is poorly sampled, it suggests that this unthreading reduces the rate.

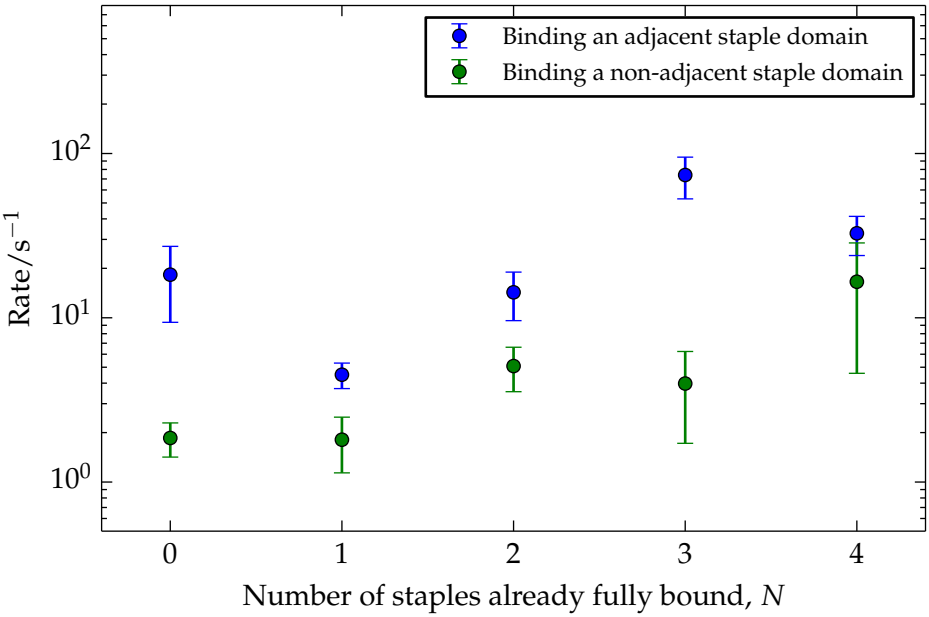

Figure 4.12: The average rate of binding the second domain of a staple, with the transitions split into two groups: one for which the second domain which is bound is adjacent on the staple to the first domain (blue), and one for which it is non-adjacent (green). The data indicates that adjacent domains bind faster than non-adjacent ones.

close to its complementary scaffold domain most of the time when the staple ends are bound.

4.3.3 Cooperativity between staples in binding rates

We now consider cooperativity between staples, by examining the rate at which a staple binds its second and third domains in the presence and absence of a neighbouring staple. As we are interested in neighbouring staples, we only consider the presence or absence of staples in the same "row" (either "a" or "b") in Fig. 4.1.

Note that, for the purposes of this section, we refer to a staple with domains A0a, A1a and A2a, or with domains A0b, A1b, A2b, as an A staple, with B, C and D staples defined in an analogous way. We also used a compressed notation for specifying transitions between particular substates, where, for example, A0→A01 means the A-

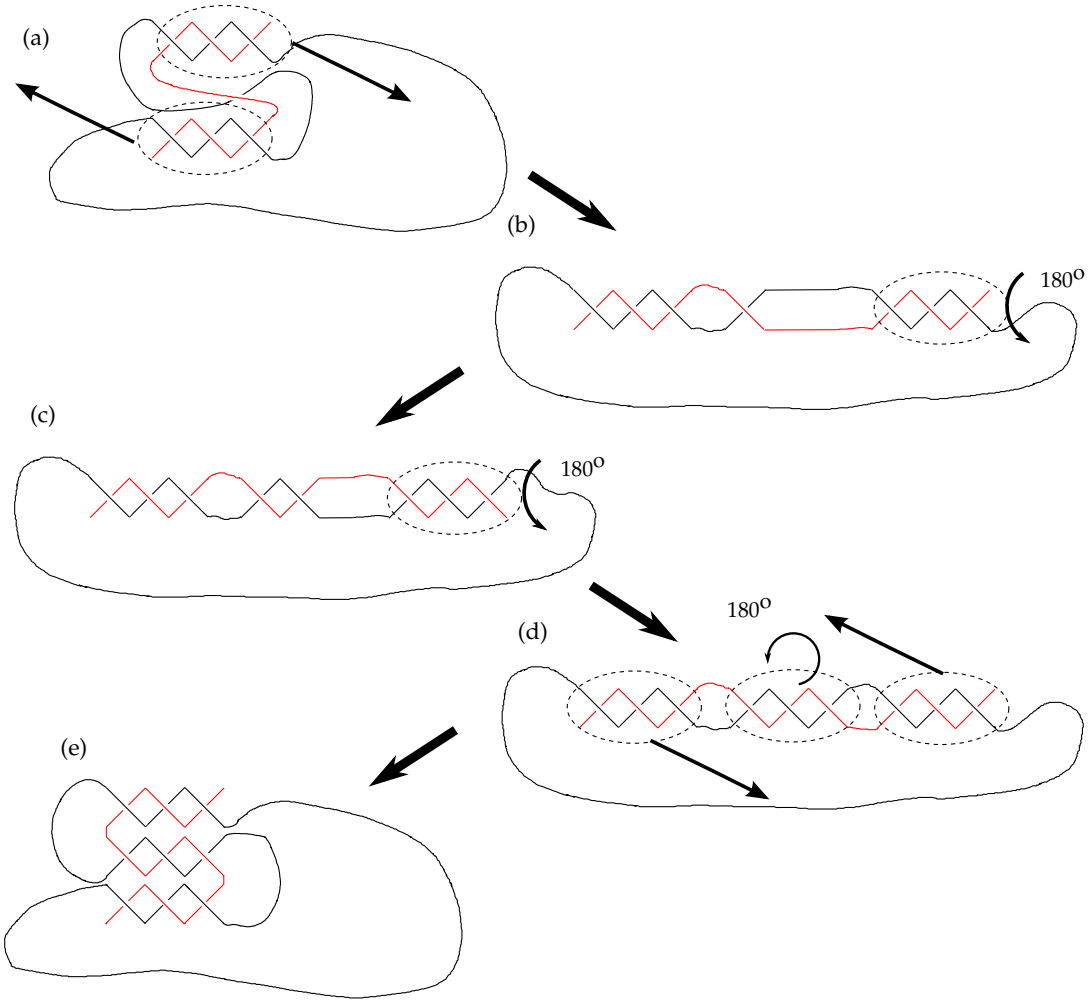

Figure 4.13: An illustration of a procedure for binding the middle domain of a staple after two end domains have bound, without breaking base pairs. The topology shown here corresponds to that of a B staple in our origami design. The black line represents the scaffold strand and the red line the staple strand. Duplexes are represented by zig-zagging lines, which thread around each other with the same topology as the real system. Duplexes with dashed lines around them are to be moved in the next step, with straight arrows indicating a translation and curly arrows indicating a rotation by a specified angle. (a) A partially formed origami with one staple partially bound. We translate the duplexes to make the procedure easier to follow. In (b) and (c) one of the end domains is rotated around its helical axis in steps of 180°, creating a half-turn in the central duplex each time. In (d) the middle domain of the staple strand is now sufficiently wrapped around the scaffold strand to achieve full base pairing. We then rotate the middle domain by 180° in the plane of the page and translate the end domains to return the structure to something recognisable from Fig. 4.1

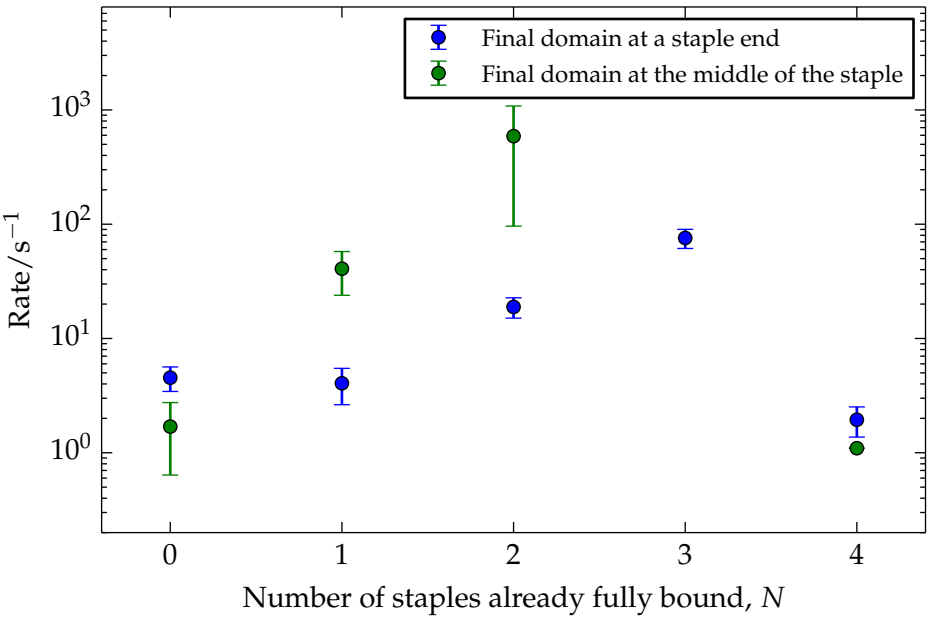

Figure 4.14: The average rate of binding the final domain of a staple, with the transition split into two groups: one for which the final domain that is bound is in the middle of the staple (green), and one for which it is at one of the staple ends (blue). Note that there was little or no sampling for the case where the final domain to be bound is in the middle of the staple for the $N = 3$ and $N = 4$ cases. The data suggests that binding a middle domain may be faster.

state has A0*y* bound and the B-state has A0*y* and A1*y* bound, as well as some number (possibly zero) of other fully bound staples.

Cooperativity between A and B staples in binding a staple's second domain is examined in Fig. 4.15, which looks both at the binding of an A staple in the presence or absence of a B staple, and the binding of a B staple in the presence or absence of an A staple. Note that this is a collection of all the rate data for transitions where either A or B was the only partially bound staple. We see a clear cooperative effect for the transitions A1→A12, A2→A12, B1→B12, and B2→B12 (red box), which in the presence of a fully bound neighbouring staple corresponds to binding the fourth domain at a double crossover (see Fig. 4.16 for simulation snapshots). We noted this effect in our study of the assembly of a similar origami using direct simulation in Chapter 3. We also note that, with the exception of a poorly sampled point for the B1→B01 transition, binding the fourth domain at a double crossover is the fastest kind of transition. This is also in agreement with our conclusions from the direct assembly simulations, albeit now put on a much more quantitative, rather than somewhat anecdotal, footing.

In Fig. 4.17 we look at cooperativity between the B and C staples for binding the second staple domain. Here there are again hints of cooperativity around the junction (red box), but the data is less clean than for the A and B staple case.

Cooperativity in binding a staple's third domain is examined for the A and B staples in Fig. 4.18 and for the B and C staples in Fig. 4.19, with the transitions that form a junction shown in red and purple boxes. There is a fairly clear indication of cooperativity for the data in the red boxes. The transitions shown in the purple boxes involve binding a middle domain when the two staple ends are already bound, complicating the picture due to the potential geometric or topological problems the presence of other staples

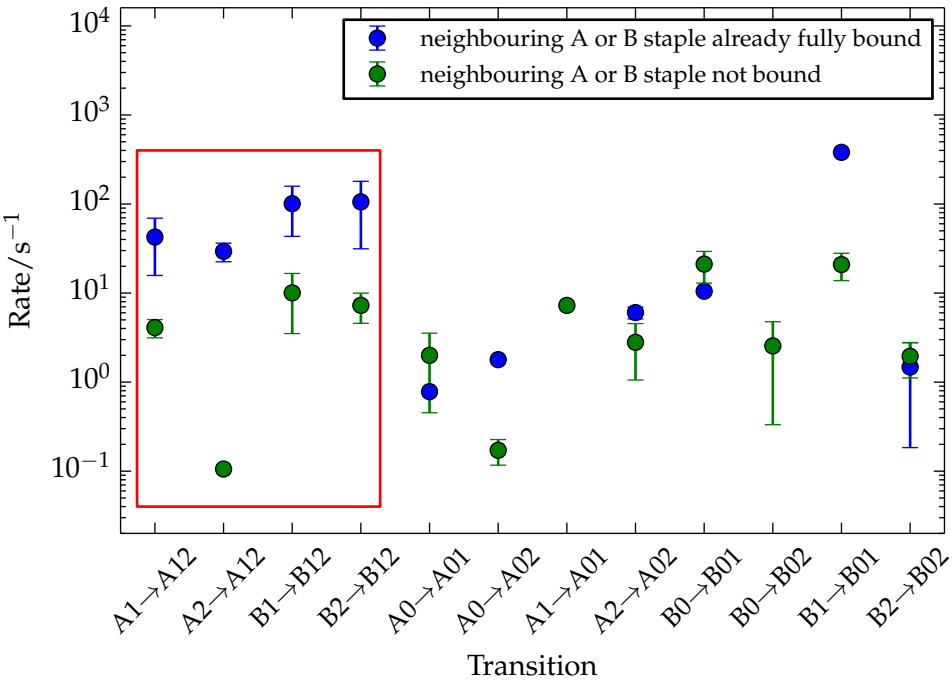

Figure 4.15: Cooperativity between A and B staples when binding the second A or B staple domain in the presence of a neighbouring B or A staple, respectively. All data for the relevant transitions where either an A or B staple was the only partially bound staple are used for this figure. A clear cooperative effect is seen for the transitions which form the fourth domain of a double crossover, which are shown in a red box.

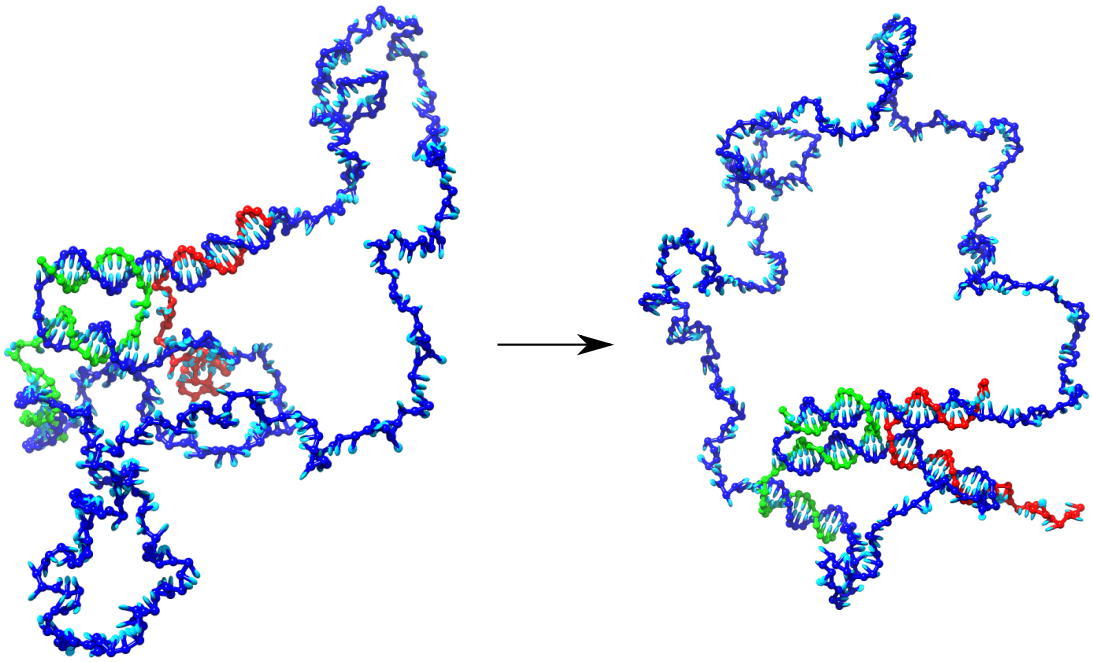

Figure 4.16: A cooperative junction-forming transition, in this case B1→B12 in the presence of an A staple, with the scaffold strand in blue, the A staple in green and the B staple in red. See also Fig. 3.6.

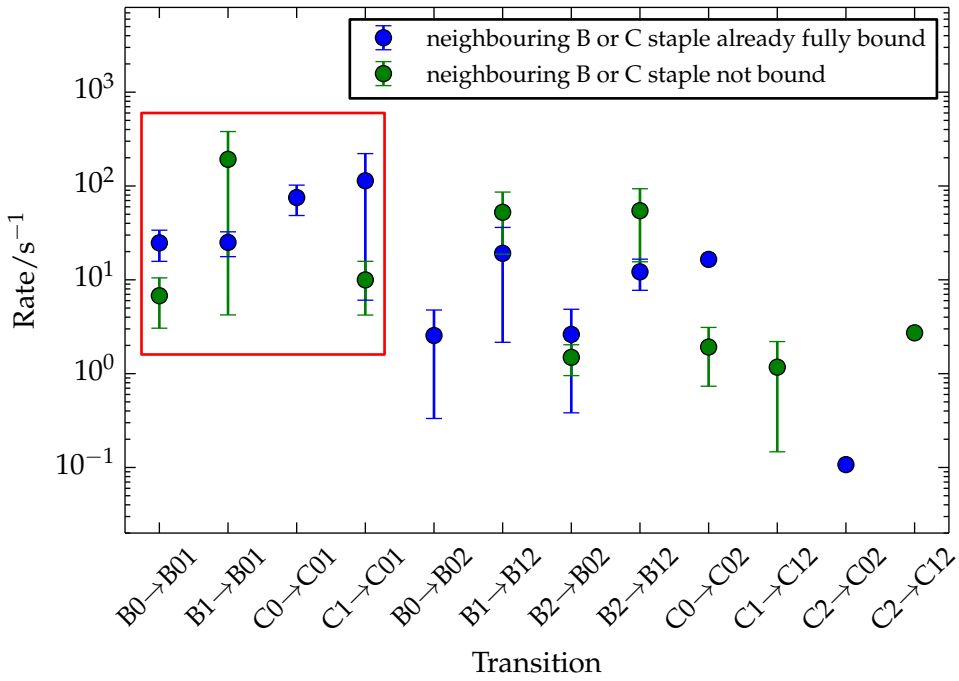

Figure 4.17: Cooperativity between B and C staples when binding the second B or C staple domain in the presence of a neighbouring C or B staple, respectively. All data for the relevant transitions where either an B or C staple was the only partially bound staple are used for this figure. There is some indication of a cooperative effect for the transitions which form the fourth domain of a double crossover, which are shown in a red box.

might introduce. Unfortunately the noisy data does not provide a consistent story for these transitions.

4.4 Sampling and systematic error

A theme of our analysis of the results presented in this chapter is that the transitions are not well sampled. The sampling is also not uniform: each successive forwards calculation was begun from the configurations generated by the previous one, and this meant that "winners" began to dominate the ensemble of configurations at each assembly state. The initial hope was that these dominant substates would reflect the substates associated with the fastest forward rates for the transitions used to generate them. However, it seems just as likely that they came to dominate by chance.

One culprit here is the discrete order parameter, which defines the conditions that the system must reach within each simulation in order to proceed. With a good order parameter, it should be possible to have relatively high success probabilities ($\sim$0.3 to 0.5) associated with each set of shooting simulations from one order parameter value to the next. However, this is not the case for our order parameter, and in particular, for every forwards rate calculation, simulations launched from configurations that had just reached order parameter value $\lambda = \alpha$ had a success probability of only $\sim$0.05. Here $\lambda = \alpha$ corresponds to one base pair in the relevant staple and scaffold domains being "nearly base-paired" (see the definition of n_{near} in Appendix B.3.3), with the shooting simulations run from that interface needing to gain one base pair between the domains in order to reach $\lambda = \alpha + 1$. It turns out that a very small number of the configurations that reach $\lambda = \alpha$ tend to have a very high success probability, and constitute the bulk of the successes launched from that order parameter value, because they are particularly suited to forming the first base pair. So even if thousands of

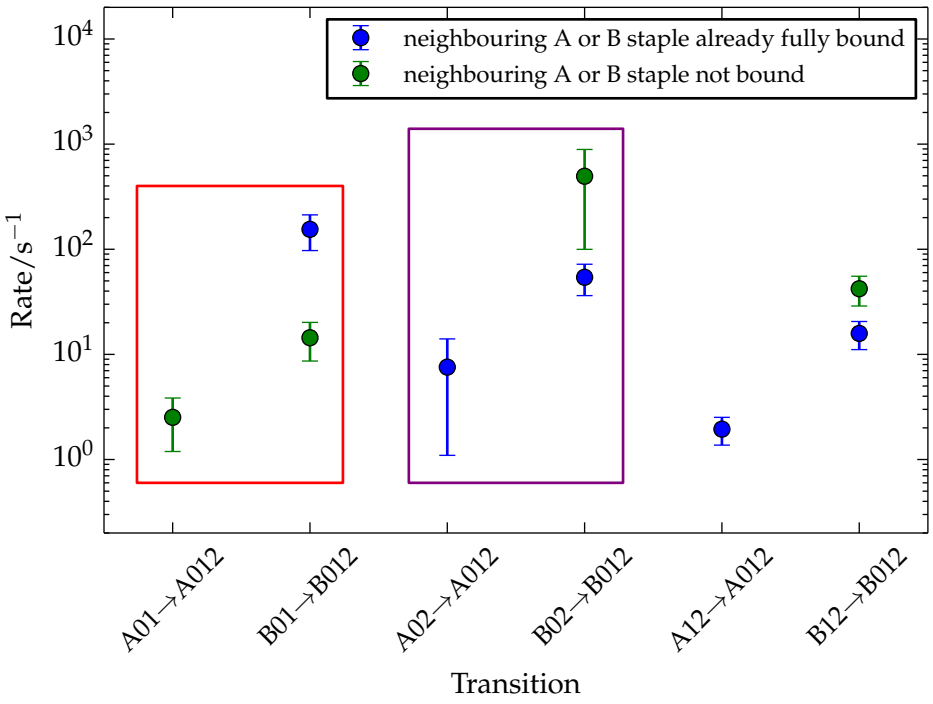

Figure 4.18: Cooperativity between A and B staples when binding the third A or B staple domain in the presence of a neighbouring B or A staple, respectively. All data for the relevant transitions where either an A or B staple was the only partially bound staple are used for this figure. The transitions which form the fourth domain of a double crossover are shown in red and purple boxes, with the former, corresponding to transitions that do not involve the potential topological problems associated with binding a middle staple domain, showing a cooperative effect.

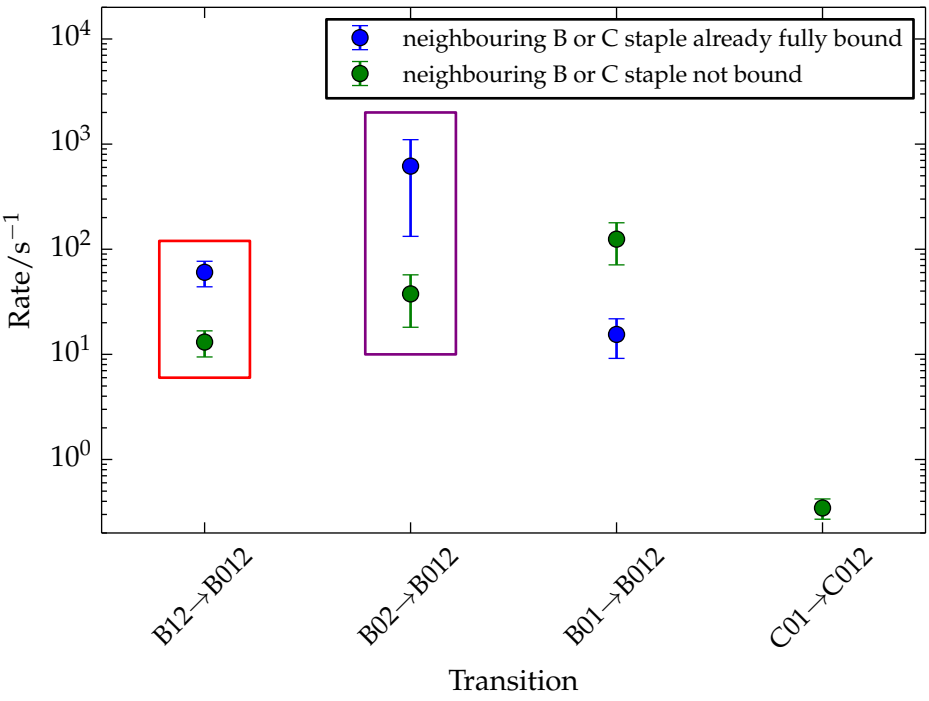

Figure 4.19: Cooperativity between B and C staples when binding the third B or C staple domain in the presence of a neighbouring C or B staple, respectively. All data for the relevant transitions where either an B or C staple was the only partially bound staple are used for this figure. The transitions which form the fourth domain of a double crossover are shown in red and purple boxes, the former showing a transition that does not involve binding a middle staple domain and the latter showing a transition that does.

configurations are generated in the B-state, the majority have usually originated from a few configurations at $\lambda = \alpha$. Naturally, this greatly reduces the diversity of substates generated.[2] An order parameter defined with a more suitable set of conditions that facilitate this key stage of the binding transition would therefore be of great help in improving the sampling.

There is another aspect of the order parameter that we have used that may also introduce a bias into the rates. For the forwards calculations, the B-state is defined by a new domain having 14 base pairs formed (of a possible 16). This condition usually ensures that we are in a metastable state as desired – however, it is sometimes possible for complexes with geometric or topological problems to transiently form 14 base pairs, thus adopting a configuration that satisfies the B-state definition but which will quickly fall back to the A-state if simulated for a short amount of time, and so is not truly a metastable state. The configurations were equilibrated before being used for subsequent FFS calculations, reducing the chance of starting calculations from configurations that were not really in the relevant assembly state, but they still introduced a bias to the rate for the transition that generated them. Again, this problem could potentially be avoided with an order parameter that takes into account the geometry of the partially formed origami – but formulating such an order parameter is not straightforward.

4.5 Summary and conclusions

The simulations presented in this chapter have highlighted a number of aspects of origami self-assembly that build on and complement the results in Chapter 3. For the experimental strand concentrations we have used, we see a clear time scale separation

[2] We did make one attempt to mitigate against this effect: After every full staple binding (i.e. at the S3, S6, S9 and S12 assembly states), we ensured that each available substate was equally sampled for the flux simulations for the subsequent forwards FFS calculation.

between the rate of binding a staple's first domain and the rate of binding its subsequent domains. This indicates that, for the origami we are studying here, the assembly will occur in series, with each staple fully binding to the scaffold in turn. However, we emphasise that the assembly of full-scale origami designs, which are likely to have greater distances along the scaffold strand between a staple's binding domains and more complicated requirements for staple strand threading, may involve staples that have significantly slower rates of binding their second and subsequent domains than we observe here for our small model origami. Thus, this time scale separation may be significantly reduced in full-size origamis, making the staple blocking effects seen in Chapter 3 again relevant.

For the backwards transitions from S3 to S0 that we have studied, we have found that the majority of transitions have rates similar to the internal (i.e. second- and third-domain) staple binding rates, under the conditions we have used. While most of the transitions have a roughly uniform rate, there is a suggestion that transitions that require a long staple tail to unthread through a small staple-scaffold loop have a reduced rate.

We have examined how the rate of binding a staple's domains is sensitive to the order of binding for our 3-domain-per-staple origami design. We have seen evidence that, for the first staple domain, the domains at either end of the staple bind faster than the middle domain. On the other hand, we have seen that a staple can bind its second domain more quickly if its middle domain was the first domain to be bound, rather than one of the end domains. This is likely to be primarily due to the reduced rate of binding non-adjacent staple domains, and partly due to the reduced rate of binding domains where the transitions requires threading a long staple tail through a small staple-scaffold loop. Interestingly, we have seen hints that, for final domain binding,

binding the middle staple domain may be faster than binding an end domain, despite the putative potential for topological problems due to the absence of a single-stranded tail to achieve wrapping. We have verified that this binding is possible with correct strand threading for some cases where no other staples are bound to the scaffold, while the likely proximity of the middle staple domain to its target domain on the scaffold might explain the observed increase in rate.

Finally, we have examined the forwards rate data for evidence of cooperativity between neighbouring staples. Inter-strand cooperativity was identified for transitions which bind the fourth domain around a Holliday junction, but not when the junction was next to a seam (as is the case for the Holliday junction formed by the C and D staples). The same effect was suggested by the direct simulations of origami assembly in Chapter 3.

We attempted to choose a temperature corresponding to that at which assembly might occur in a typical experimental protocol. One possibility that we envisaged was that, at this temperature, there would be a free-energy barrier to assembly, so that the first few staples might be thermodynamically unstable, but that later staples might become thermodynamically favourable, with the final stages of the assembly proceeding downhill in free energy. Although we have not measured rates for the whole assembly, the data we do have suggests that, for the origami design and assembly conditions we have used, the assembly proceeds uphill in free energy up to and including the binding of the fifth staple (of eight). This is because the on-rates are either much slower than (for initial staple domain binding) or comparable to (for second- and third-domain binding) the off-rates. Meanwhile, the cooperativity between staples that we do see does not appear strong enough for the on-rates to dominate the off-rates for the binding of the

final three staples. The strongest cooperative effect we have seen is the binding of the last of the four staple domains around a Holliday junction – but this is a rather local effect and is seen as early as the binding of the second staple, and there is no reason to think that it will have a significantly greater impact further along the assembly pathway. Furthermore, within the noise limits of our data there is no obvious acceleration in the forward rates as the assembly progresses. Unfortunately we do not have data for cooperativity between staples for staple domain melting, but the probable sources of this cooperativity, coaxial stacking between domains and reduced scaffold strand entropy, seem unlikely to be strong enough to significantly change the story.

Our results thus suggest that we may actually have chosen a temperature for our simulations that was above the melting temperature of the origami. One potential way to have probed the suitability of the chosen assembly temperature beforehand would have been to simulate the fully formed origami: if it began to melt, this would have indicated that the entire assembly pathway was uphill in free energy at this temperature. However, an absence of any melting would have not necessarily implied the converse, namely that the last assembly steps were downhill in free energy, because it is not clear that the accessible simulation time scales would be sufficiently long if the origami was not much above its melting temperature. The basic difficulty in choosing the assembly temperature is that tractable approaches to accurately compute the melting temperatures of large DNA assemblies are not currently available.

Our results also suggest that the free-energy landscape for the assembly of our small model origami might have an approximately constant slope as a function of the number of staples bound (being roughly flat at the origami melting temperature). If this were the case, the lack of an overall barrier to assembly (of course, there would still be individual

barriers associated with the binding of each staple domain) would perhaps imply that the assembly of an origami of this size should occur relatively reversibly and without significant hysteresis, in contrast to what has been seen for experiments on full-scale origamis.[34;113] Thus, the source of the cooperativity that leads to the "nucleation-like" assembly dynamics in large origamis remains an open question.

The main source of cooperativity that we have identified is associated with the Holliday junctions. Thus, we would expect to see a stronger cooperative effect for a design with a higher density of cooperatively formed junctions. Although the junction density for the design we have used ($\sim$ 0.021 junctions per base pair) is comparable to that of full scale origamis such as the 2D tile and 3D pointer discussed in Chapter 6 ($\sim$ 0.027 junctions per base pair), four of the junctions in this design (those at the interface between the C and D staples) are atypical due to the presence of nicks or other adjacent junctions, which will greatly reduce the cooperative effect. So we might expect a stronger cooperative effect for typical, full scale origamis. Another factor to bear in mind is that for a full-scale origami, because of the larger number of staples involved, the curvature in the free energy versus number of staples curve does not need to be as large to lead to a significant free-energy barrier to assembly as for our small model origami.

The findings in this chapter suggest two contrasting directions for future study of origami assembly using FFS. One potentially fruitful avenue would involve using FFS to examine transitions of particular interest between certain substates. This approach should permit a good level of sampling, which would allow more subtle effects to be teased out, and could be used for larger, perhaps even full-scale, origami designs, allowing a more direct comparison to experiment. The second possibility, which is

far beyond current computational capabilities, would be to systematically perform FFS rate calculations for the transitions between all possible substates (or perhaps a representative subsample) and use these rates to inform a kinetic model of the assembly. Our results suggest ways that such simulations might be sped up. For example, we deliberately did not make assumptions about the assembly pathway when running our simulations, but we clearly found that for the small origami considered here, assembly virtually always occurs one staple at a time. Thus, after a staple makes initial contact with the scaffold, there is no real need to still simulate all of the unbound "spectator" staples that are simply diffusing around the box.

5 Introducing an updated oxDNA model

The oxDNA model has provided key insights into many different processes relevant to DNA nanotechnology and biophysics. However, there are some areas where the model can be improved. In this chapter we present an improved version of the oxDNA model, which we call oxDNA2. The new model introduces a salt-dependent term to the interaction potential, improved predictive power for the structural properties of large DNA nanostructures, and a different interaction strength for consecutive adenine bases compared to consecutive thymine bases.

5.1 A summary of the changes introduced with the new model

The original oxDNA model was parameterised to $[Na^+] = 0.5\,M$, a high salt concentration similar to that used for many applications in DNA nanotechnology. However, the ability to study DNA behaviour as a function of salt would allow quantitative comparison with a greatly expanded set of experiments, and, in particular, if we wish to apply oxDNA to biological systems we would like to work at physiological salt ($[Na^+] \approx 0.15\,M$). At the same time, the wealth of experimental data for the thermodynamic and mechanical properties of DNA as a function of salt concentration[112;116–118] makes fitting the salt-dependent properties of an extended model possible.

While the detailed effect of salt on DNA electrostatics can be highly complex,[119;120] for oxDNA2 we use a simple Debye-Hückel interaction term, as first implemented for oxDNA by Wang and Pettitt in Ref. 78, to model the way salt screens repulsive interac-

tions. This coarse-grained description is commensurate with the level of approximation used in general in the oxDNA framework. After adding the Debye-Hückel interaction term, we carefully parameterise the model to an extended set of experimental data for melting temperatures and persistence lengths, including the behaviour of single strands.

A second area that merits attention is the performance of the model in simulating the structure of large (kilobase-pair) DNA objects. The model value of the B-DNA pitch did not come under too much scrutiny in the original parameterisation as there is some disagreement in the literature about the precise value of the average pitch.[118;121–125] However, improved experimental techniques in fabrication and imaging of large DNA objects[9;39;126] have presented an opportunity to finely tune this value, because small adjustments to the duplex pitch can result in significant changes to the global twist of a large-scale DNA nanostructure. In addition, simulating these large-scale structures has illustrated the potential importance of nicks and junctions for the effective duplex pitch, so that improving the model's description of these effects has become a priority.

In addition, the original model duplex had grooves with equal widths, whereas B-DNA is known to have a larger major groove and a smaller minor groove. This implies that the positions of the backbone sites in the model, which are directly related to the groove widths, could be more realistic. This detail could be relevant, for example, in origami structure, where the precise backbone positions are known to be important for junction placement.[9]

The oxDNA model was previously given sequence-dependent hydrogen-bonding and stacking strengths[97] by fitting to the duplex thermodynamics of the SantaLucia model.[112] As the SantaLucia model gives results at the base-pair step level, one can

only extract the average stacking strength for the two stacking interactions present in a given base-pair step. In particular, this means that in oxDNA the AA and TT stacking interactions are the same, whereas it is well known that the AA stacking interaction is significantly stronger than the TT interaction, an important property for DNA nanotechnology, for example, where poly-T single-stranded regions are often used as flexible linkers.[127;128] To remedy this we use new experimental data[89] to reparameterise the AA and TT stacking interactions in the model.

In the following sections we consider each change to the model in turn, and then we highlight some important aspects of the behaviour of the new model. It should be noted that for the development of the oxDNA2 model we mostly worked from the original, sequence-averaged parameterisation of the model rather than the sequence-dependent one. This is because it is more efficient to fit the thermodynamic parameters to sequence-averaged duplex melting temperatures as given by the SantaLucia model. The exception is the parameterisation of the AA and TT stacking strengths, which did use the sequence-dependent parameters from Ref. 97 as a starting point, to allow the best possible comparison between the model and the experimental results that were used for the fitting. After the parameters for oxDNA2 had been obtained, including new values for the sequence-averaged hydrogen bonding and stacking strengths, we then rescaled the sequence-dependent interaction strengths from Ref. 97 accordingly for use with the new model.

5.2 Introducing different widths for major and minor DNA grooves

B-DNA in the original oxDNA model has equal groove widths, while in reality DNA has a larger major groove and a smaller minor groove. Having realistic widths for the major

and minor grooves is equivalent to having appropriately positioned backbone sites in the model, an important feature for the physical properties of many DNA motifs. For example, in DNA origami antiparallel double helices are joined by crossovers, for which the position of the backbone has been shown to be crucial for origami structure.[8;9] Another example is anisotropic duplex bending: the duplex can be expected to bend more easily into the major groove than into the minor, if the groove widths are unequal.

The oxDNA nucleotide is composed of three interaction sites: the hydrogen-bonding, stacking and backbone sites. We introduce different groove widths by changing the position of the backbone site while keeping the duplex radius unchanged (Fig. 5.1), such that, rather than lying on a straight line, the three interaction sites lie in a plane. The new nucleotide shape introduces an additional parameter into the model, the angle γ between the line from the duplex centre to the backbone site and the line from the duplex centre to the stacking site (Fig. 5.1(a)). Given the coarse graining of the 18 atoms of the sugar-phosphate DNA backbone into a single interaction site, there is no definitive choice for the precise position of the backbone site and thus the value of the model parameter γ (Fig. 5.1(a)). We set $\gamma = 20^\circ$, a value which maps onto a full-atom representation of a DNA duplex well by visual inspection, although values of γ between 15° and 25° would give an equally satisfying visual match.

The backbone site is moved such that the duplex radius is unchanged, and we note that the modification has a negligible added computational cost when simulating the model. However, the thermodynamic and mechanical properties are slightly affected. For the thermodynamics, we found a change of 1-2 K in the duplex melting temperatures, and we modified the hydrogen-bonding and stacking strengths using the histogram reweighting method described in Section II C 2 of Ref. 90, so that the

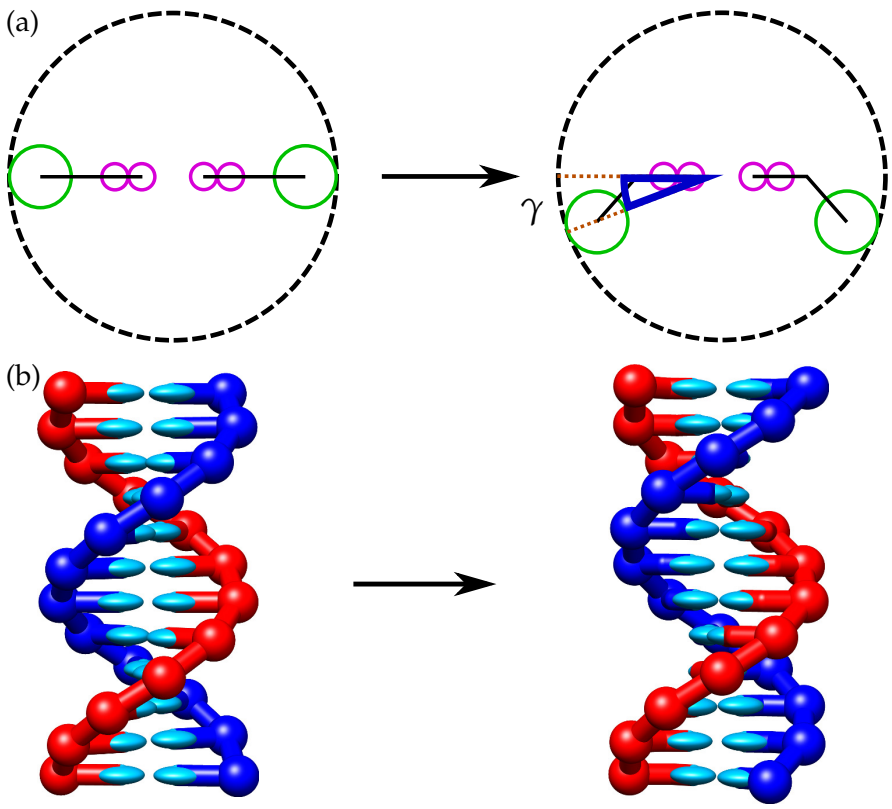

Figure 5.1: Schematics contrasting the original oxDNA model (left), with equal groove widths, with oxDNA2 (right), which has differentiated major and minor grooves. (a) A cross section of a duplex with one base pair displayed. The large dashed circle shows the helix radius, and each nucleotide is represented by three circles joined by a line; the large solid circles represent the backbone sites, while the small solid circles represent the stacking (closer to the backbone) and hydrogen-bonding (at the end of the nucleotide) sites. For oxDNA2, a value of 20° was chosen for the angle γ. (b) A representation of a DNA duplex for each model.

agreement with experimental melting temperatures was as good as for the original model. The mechanical properties of DNA are less well constrained experimentally and so were not refitted. The mechanical properties for the new model can be found in Section 5.7.

One illustration of the importance of the groove widths in oxDNA for the structural properties of DNA assemblies is provided by systems of 3-arm star tiles that are designed to form triangular prismatic polyhedra. We find that modifying the groove widths qualitatively changes the structure of trimers of these tiles (Fig. 5.2). Specifically, the body of the trimer defines a plane with two distinct faces. Zhang *et al.*[129] found that one of two possible isomers of the polyhedron preferentially formed, implying that the free arms of the trimer systematically pointed in the direction of one of these two faces. We find a consistent result when the groove widths specified by oxDNA2 are used. When equal-sized grooves are used (as in the original oxDNA model) the trimer arms point in the opposite direction.

5.3 Effective electrostatic interactions

Since the original goal of the oxDNA model was to simulate nanotechnology experiments, the thermodynamic and structural parameterisation was carried out at high salt. The very short electrostatic screening length at these conditions allows one to incorporate the electrostatics into a soft excluded volume, somewhat circumventing the necessity of a proper treatment. This was the approach taken for the original oxDNA model. The thermodynamics were then parameterised for 0.5 M [Na^+], a high enough value that further increasing it does not significantly change the physics at this level of coarse-graining.

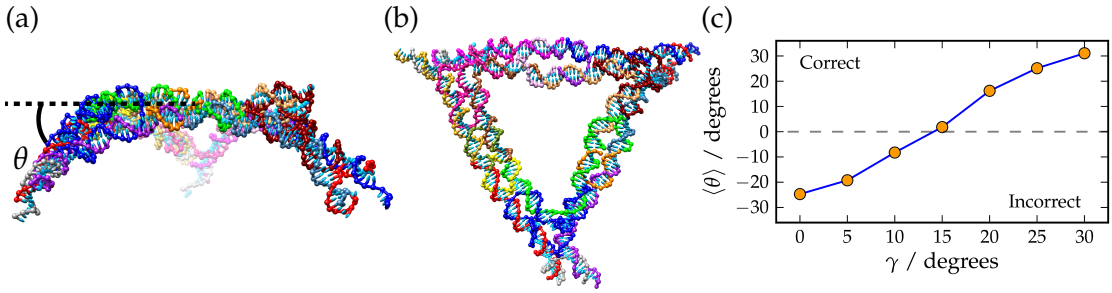

Figure 5.2: In oxDNA, the structure of the DNA trimer due to Zhang *et al.*[129] is found to qualitatively change as a function of duplex groove angle γ. A representative trimer configuration with $\gamma = 20^\circ$ is shown in (a) from the side and with θ, the angle between each arm and the plane of the main, triangular trimer section, displayed and in (b) from the top. (c) Shows the average value of θ as a function of γ, with $\gamma = 0^\circ$ corresponding to equal groove widths as for the original oxDNA model, and $\gamma = 20^\circ$ being chosen for oxDNA2.

The problem of treating electrostatics properly for DNA in solution is a complicated one. Perhaps the most evident issue is that the typical dimensions of nucleotides are comparable to the Debye length of the solution, so that a mean-field treatment is hard to justify. Also, in some cases the presence of salt ions affects the local structure of nucleic acids, by stabilizing some arrangements or destabilizing others. Ion condensation may also lead to stronger screening of the electrostatic interactions than otherwise expected – this has been incorporated into coarse-grained DNA models through partial effective charges.[65;130] Thus, in principle, many non-trivial effects must be taken into account when modelling the electrostatic interactions, and the debate on the best way to do so implicitly is still unresolved.[119;120]

Here, we choose a very simple treatment, based on the Debye-Hückel model for screened electrostatics. This approach has been used previously for other coarse-grained models,[60;65] and was first introduced into oxDNA by Wang and Pettitt.[78] We note that this treatment is consistent with the coarse-grained nature of the model, and we use the same top-down strategy that was used in the original parameterisation to design

the effective electrostatic interactions: our goal is to introduce a term in the potential that will reproduce the thermodynamic and mechanical effects of salt concentration on DNA, and thus should be regarded as an effective interaction rather than an attempt to rigorously model the local effects of charges. This should be kept in mind when interpreting the results obtained with oxDNA, in particular at low salt concentration.

Since the modelling of electrostatics is rather crude, we restrict our parameterisation to salt concentrations of 0.1 M of monovalent salt or greater. This restriction is also due to the fact that we parameterise our thermodynamics to the model of SantaLucia,[112] which was fitted in a similar salt regime. Importantly, physiological conditions fall within this range, which will allow quantitative comparison between simulations of DNA systems with our model and experiments at physiological conditions.

The detailed Debye-Hückel form which is added to the non-bonded interactions in the potential of Eq. 1.1 takes the following form:

$$V_{\mathrm{DH}}(T,I) = \sum_{ij} \frac{(q_{\mathrm{eff}}e)^2}{4\pi\epsilon_0\epsilon_{\mathrm{r}}} \frac{\exp\left\{-r_{ij}^{\text{b-b}}/\lambda_{\mathrm{DH}}(T,I)\right\}}{r_{ij}^{\text{b-b}}}, \tag{5.1}$$

where q_{eff} is the effective charge situated at the backbone site of each of the nucleotides, $r_{ij}^{\text{b-b}}$ is the magnitude of the distance between the backbone sites of nucleotides i and j, ϵ_0 is the permittivity of the vacuum, ϵ_r is the relative permittivity of water and e is the elementary charge. In principle, ϵ_r depends on r_{ij},[131] and weakly depends on temperature and salt concentration. However, for oxDNA2 we set ϵ_r to be a constant value, in keeping with the coarse-grained approach taken for the rest of the model. In particular, we choose $\epsilon_r = 80$, the standard value for water. In Eq. 5.1 we have stressed that the interaction depends on the temperature T and on the (monovalent)

salt concentration I through the Debye length $\lambda_{\mathrm{DH}}(T, I)$:

$$\lambda_{\mathrm{DH}}(T, I) = \sqrt{\frac{\epsilon_0 \epsilon_{\mathrm{r}} k_{\mathrm{B}} T}{2 N_{\mathrm{A}} e^2 I}}, \tag{5.2}$$

where N_{A} is Avogadro's number and k_{B} is Boltzmann's constant.

To improve the computational efficiency of simulating the new model, we set the interaction to zero at a finite distance; and to allow simulation with molecular dynamics, we introduce a quadratic smoothing potential so that the interaction goes to zero smoothly. The quadratic smoothing, the details of which are reported in Appendix A.1, is introduced after a cutoff r_{smooth}, which we choose to be $3\lambda_{\mathrm{DH}}$. This cutoff allows us to use all the standard techniques to improve the simulation efficiency via the use of Verlet lists and/or cells.[86] We have checked that introducing our chosen cutoff, $r_{\mathrm{smooth}} = 3\lambda_{\mathrm{DH}}$, has a negligible effect on the duplex thermodynamics results used to parameterise the interaction (Fig. 5.3(a)).

Our representation of DNA uses a single rigid body per nucleotide, and the best choice of where to put the charge is not obvious. All the atoms of the sugar and phosphate groups of the backbone are represented in a single interaction site, and it is thus natural to put the charge, which in real DNA is located on the phosphate, on that interaction site. Importantly, the backbone site of a nucleotide is placed almost in between the phosphate of that nucleotide and the phosphate of the neighbouring one, which could potentially lead to some unphysical effects. Also, we should stress that having a charge at each backbone site means that the DNA has as many charges as nucleotides, which is not always true in real systems: very often, the terminal phosphate at the $3'$ end is cut off, removing a charge. The absence of this charge can cause measurable effects on the thermodynamics, and indeed the SantaLucia model[112] requires the presence or absence of the terminal phosphates as an input parameter. In

keeping with our coarse-graining approach, we put a half effective charge on each of the terminal nucleotides to incorporate the fact that each charge should be halfway in between our backbone sites, emulating a system with the terminal charge removed, and parameterise to the SantaLucia model in a way consistent with this approach.

The parameter that we have tuned to reproduce the thermodynamics predicted by the model of SantaLucia is the effective charge q_{eff}. To do this, we used thermodynamic integration [132] to compute the melting temperatures of duplexes of length 5, 6, 7, 8, 10, 12 and 15 as a function of q_{eff} at several salt concentrations, and chose the value that best reproduced the melting temperatures predicted by SantaLucia's model.

The melting temperature of a duplex is defined as the temperature at which half of the strands in a stoichiometric bulk solution are in the duplex state. We cannot simulate the bulk system with oxDNA, as the computational cost of simulating such a large system is far too great. However, at the bulk melting temperature for a given concentration of strands, the bound and unbound free energies obtained for a system of two strands at the same concentration, F_{b} and F_{ub}, are related by:[133]

$$F_{\mathrm{b}}(T_{\mathrm{m}}, q_{\mathrm{eff}}) = F_{\mathrm{ub}}(T_{\mathrm{m}}, q_{\mathrm{eff}}) - k_{\mathrm{B}} T_{\mathrm{m}} \ln(2)\,, \tag{5.3}$$

where we have made explicit the dependence of F on the effective charge q_{eff}. The constant on the right hand side of Eq. 5.3 accounts for concentration fluctuations that are present in bulk but not for two strands in a periodic box.[133] The equation is exact in the dilute limit, which is indeed where the thermodynamics of duplex formation are usually studied and is the relevant limit for oxDNA.

We can take advantage of the relation of the free energies given by Eq. 5.3 when performing thermodynamic integration. For small changes in T and q_{eff}, it is possible

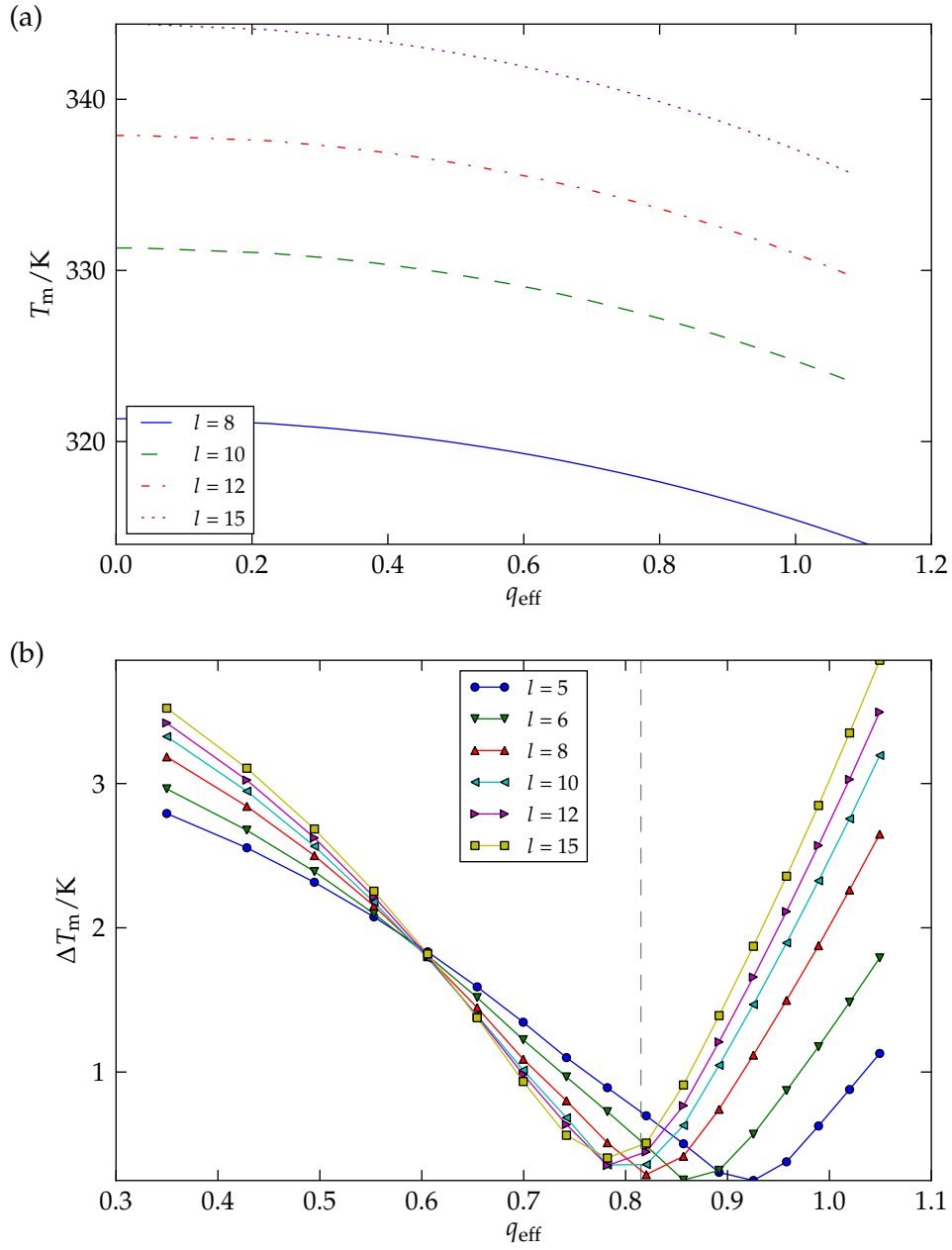

Figure 5.3: (a) Melting temperature of duplex DNA of different lengths l at [Na^+]=0.2 M as a function of the effective charge q_{eff} obtained with thermodynamic integration. (b) Average melting temperature difference between our model and the SantaLucia model[112] for different values of l as a function of q_{eff}. Each point corresponds to the average over [Na^+]=0.1, 0.2, 0.3, 0.4, 0.5 M of the magnitude of the difference in melting temperature between our model and the SantaLucia model. The plot suggests an optimal value for q_{eff} of 0.815, indicated by the vertical dashed line.

to write

$$F(T_m + dT, q_{eff} + dq_{eff}) = F(T_m, q_{eff}) + \frac{\partial F}{\partial T} dT + \frac{\partial F}{\partial q_{eff}} dq_{eff}$$
$$= F(T_m, q_{eff}) + \left(\frac{F - \langle V \rangle}{T_m} + \left\langle \frac{\partial V}{\partial T} \right\rangle \right) dT + 2 \frac{\langle V_{DH} \rangle}{q_{eff}} dq_{eff}. \tag{5.4}$$

The last step includes the unusual term $\langle \partial V / \partial T \rangle$. This is because our potential, like many coarse-grained models, depends explicitly on T through V_{stack} and V_{DH}.

To compute the change in T_m introduced by a small change in q_{eff}, one can impose the condition in Eq. 5.3 at the new T_m

$$F_b(T_m + dT, q_{eff} + dq_{eff}) = F_{ub}(T_m + dT, q_{eff} + dq_{eff}) - k_B(T_m + dT) \ln(2), \tag{5.5}$$

and, by using Eqs. 5.3, 5.4 and 5.5, one obtains

$$\frac{dT_m}{dq_{eff}} = \frac{(2/q_{eff})(\langle V_{DH} \rangle_{ub} - \langle V_{DH} \rangle_b)}{(\langle V_{ub} \rangle - \langle V_b \rangle)/T_m - (\langle \partial V_{ub}/\partial T \rangle - \langle \partial V_b / \partial T \rangle)}. \tag{5.6}$$

Eq. 5.6 is a differential equation that allows us to follow the change in melting temperature as q_{eff} is changed. We note that Eq. 5.6 is an extended Clausius-Clapeyron relation,[132] and the quantities on the right-hand side are readily accessible with separate simulations of the bound and unbound states. Since we are dealing with small systems, all these simulations are very quick and it is thus easy to achieve very accurate results. As a starting point for the thermodynamic integration we use the melting temperatures from SantaLucia, which our original model (equivalent to the current model at $q_{eff} = 0$) reproduces within tenths of a Kelvin. Since both V_{stack} and V_{DH} only depend weakly on temperature, we used Eq. 5.6 assuming $\langle \partial V_{ub}/\partial T \rangle - \langle \partial V_b / \partial T \rangle = 0$ to obtain an optimal value for q_{eff}, and checked *a posteriori* the results of thermodynamic integration with melting simulations. We have found that the melting temperature

differences computed with Eq. 5.6 are accurate to within 0.2 K.

For a range of salt concentrations and duplex lengths, we performed thermodynamic integration to find the melting temperature as a function of q_{eff}. Some of the results are depicted in Fig. 5.3. The melting temperature of the duplex decreases with increasing effective charge q_{eff}, but not dramatically so (Fig. 5.3(a)). This is because the increased interchain repulsion in the duplex state is partially balanced by the lower entropy of the single-stranded state (due to it adopting a slightly more extended state to reduce intrachain repulsion). The data in Fig. 5.3(b) show that the q_{eff} at which the difference in melting temperature between oxDNA and SantaLucia is minimised depends on the duplex length and lies in the range $0.75 < q_{\mathrm{eff}} < 0.95$. The best overall predictions are obtained for an effective charge $q_{\mathrm{eff}} = 0.815$. It is reassuring that this value does not deviate significantly from 1 (corresponding to the value given by Debye-Hückel theory), and that the best value for q_{eff} varies little with duplex length. In addition, it is not uncommon to use a value of $q_{\mathrm{eff}} < 1$ for coarse-grained DNA models.[65;130] One argument used to explain this is that ion condensation, which is known to occur for DNA, will screen the phosphate charges more strongly than expected from Debye-Hückel theory. This will lead to a lower effective charge when fitting a model using a Debye-Hückel treatment to experimental results, although such arguments should be applied with caution to a crude mean-field approach such as this one.

We note that introducing this explicit electrostatic term in our model potential will raise the computational expense of each simulation step compared to the original oxDNA model, as the electrostatic term will generally result in oxDNA2 having a larger interaction range than the original model has. This effect increases at lower salt, as the Debye-Hückel term becomes more long-ranged. For example, we find that simulating a

10-bp duplex with MD for a given number of steps takes 1.4× as long with oxDNA2 at [Na^+]=0.5 M and 1.7× as long with oxDNA2 at [Na^+]=0.1 M as it did for the original oxDNA.

5.4 Improving structure prediction for large-scale DNA objects

The model was originally parameterised for small single- and double-stranded DNA structures. If we wish to study large-scale structures, we need to ensure that we can reproduce the existing experimental data for these larger constructs. A good test case is provided by the work of Dietz *et al.*,[9] who measured the global twist of three different DNA origami structures, described as 10-by-6 helix bundles (Fig. 5.4). We denote the three origami structures as L-, N- and R-type. In the experiment, they were designed to impose a pitch of 10, 10.5, and 11 base pairs per turn (bp/turn) on the constituent DNA double helices and these different designs exhibited left-handed, no, and right-handed global twist, respectively, when multimerised to form ribbons and visualised with transmission electron microscopy. One might think that this result implies that DNA has a natural pitch of 10.5 bp/turn, since the design with that inherent periodicity did not result in a globally twisted system. However, it is not that straightforward.

	Global Twist/°	
Helix Bundle Type	oxDNA2	experiment
N	0.2	0
L	-22.2	-31 +/- 5
R	26.9	26 +/- 5

Table 5.1: Global twist in the DNA helix bundles of Dietz *et al.*[9] for simulations using oxDNA2 and determined from experiment. The oxDNA2 simulations were run with [Na^+]=0.5 M for 1.8×10^9 molecular dynamics (MD) steps for each design.

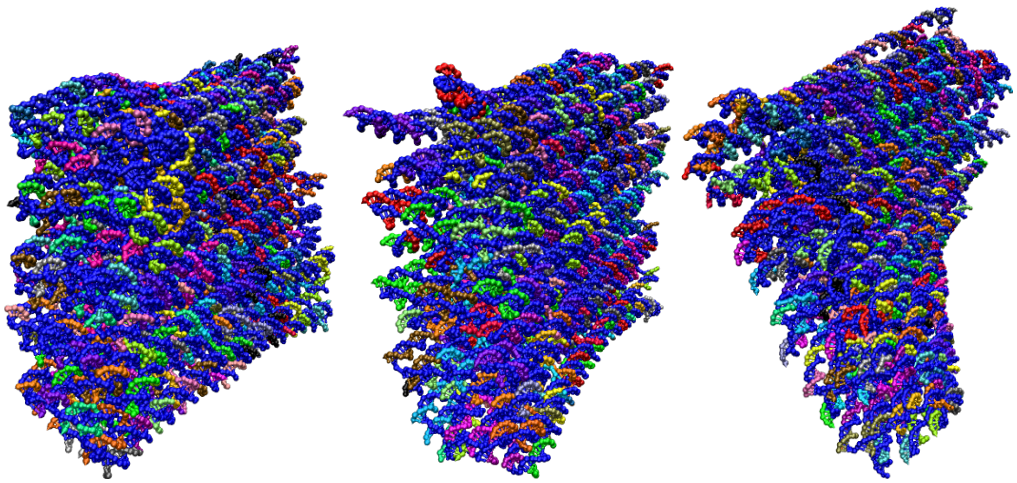

Figure 5.4: Simulation snapshots for the L- (left), N- (middle), and R- (right) type helix bundles, each composed of roughly 15,000 nucleotides. The designs are taken from Dietz et al.[9] In experiment and in oxDNA2, the N-type helix bundle exhibits close to zero global twist.

Unsurprisingly, when simulated with the original version of oxDNA, the N-type origami still showed a significant right-handed global twist, chiefly because the duplex pitch for the model was 10.36 bp/turn. However, even when simulated with a version of oxDNA modified so that the model duplex pitch was 10.5 bp/turn, the N-type origami displayed a right-handed twist. One reason for this is the following: in oxDNA, the helical twist across nicks and junctions is larger by about 3.9° and 2.5° degrees respectively than for a normal duplex step (we call this an overtwist), so that fractionally fewer base pairs are required for a helix turn than would otherwise be expected. Although these differences are small, they add up constructively to create a global twist on a structure that would otherwise have no such twist. For a nicked duplex in oxDNA, the overtwist occurs because, opposite the nick, the non-nicked strand prefers a larger twist than for duplex DNA (just as stacked single strands do in the model), and the nicked backbones lack the FENE spring that would usually oppose this tendency. A similar argument applies for strands with junctions.

Although there is no direct evidence available to show whether this overtwist is

physically realistic, there are multiple lines of experimental evidence that suggest that the pitch of duplex DNA is close to 10.5 bp/turn.[118;121] Given the evidence from Dietz *et al.*[9] that the effective pitch in DNA origami structures is also close to 10.5 bp/turn, we decided to reduce the overtwisting in oxDNA as much as is possible. To achieve this, we modify the coaxial stacking term of the potential, $V_{\text{coax stack}}$, so that the overtwist is 0° for a junction and 1.3° for a nick (see Appendix A.3 for details of the changes to the potential). We have verified that this change has very little effect on the other features of the model.

Even when the overtwist at nicks and junctions has largely been removed, we require an intrinsic duplex pitch of 10.55 bp/turn in order for the oxDNA N-type helix bundle to have zero global twist. We set the pitch to this value (at $[Na^+]$=0.5 M) for oxDNA2 by modifying V_{backbone}. We found that this modification to the backbone potential changed the duplex melting temperatures by around 2 K, and we refitted the model's hydrogen bonding and stacking strengths to correct for this using the same method as was used for the thermodynamics refitting when implementing unequal helical groove widths. The requirement for a pitch of 10.55 bp/turn, rather than the 10.5 bp/turn that one might expect, is due to a few subtle effects that we have so far only partially investigated.

The global twist measured for the helix bundles with the new oxDNA2 model (Appendix B.4.1 describes how the global twist was measured from simulations) is compared to the experimental results in Table 5.1, while typical simulation snapshots are shown in Fig. 5.4. The slight modifications to V_{backbone} used to set the model pitch are given in Appendix A.2. We note that the modifications to V_{backbone} changed the duplex melting temperatures in the model by 1-2 K, which were then refitted using

histogram reweighting (as described in Section II C 2 of Ref. 90) to give an agreement with experimental melting temperatures that was as good as for the original model.

5.5 AA/TT sequence dependence

As a further improvement, the oxDNA2 model incorporates a more realistic sequence-dependent stacking interaction, which is achieved by differentiating between the AA and TT stacking interaction strengths. In the previous parameterisation of oxDNA,[97] the sequence-dependent base-pairing and stacking interaction strengths were obtained by fitting the oxDNA duplex melting temperatures to the SantaLucia model, a nearest-neighbour model that is able to predict experimental duplex melting temperatures very well.[112] The SantaLucia model is designed at the level of base-pair steps, where each base-pair step consists of four bases, with a free-energy difference between its single-stranded state and its duplex state. In total there are 10 unique base-pair steps in the SantaLucia model: AA/TT, AT/AT, TA/TA, GC/GC, CG/CG, GG/CC, GA/TC, AG/CT, TG/CA, GT/AC (for example, AG/CT refers to a base-pair step with complementary bases AG on one strand and CT on the other, both specified in the $3'$ to $5'$ direction). Therefore, fitting oxDNA to the SantaLucia model only allows one to find the sum of the strengths of the two stacking interactions between the nucleotides that are within a base-pair step. For example, it is only possible to find the average of the AA and TT stacking strengths from the SantaLucia model.

However, experimental evidence from sequence-dependent measurements of the mechanical properties of single-stranded DNA,[116;134] as well as hairpin stabilities and closing rates,[135;136] have revealed that sequences of A bases are much stiffer than equal length sequences of T bases. It has been argued that this is evidence that consecutive AA bases stack much more strongly than TT bases do. Here we use original

experimental data on the stabilities of hairpins with either a poly-A or a poly-T loop, to differentiate between the AA and TT stacking strengths. All hairpins have 6-bp stems and loops of either 21 or 31 bases. Details of the experimental setup and results can be found in the Supplementary Material of Ref. 89.

To parameterise our model to reproduce the experimental data, our procedure is to vary the AA stacking interaction strength, ϵ_{AA}, while fixing the sum of the AA and TT stacking interaction strengths to the value that is obtained by fitting oxDNA's duplex melting temperatures to the SantaLucia model, i.e.

$$\epsilon_{\mathrm{AA}} + \epsilon_{\mathrm{TT}} = 2\epsilon_{\mathrm{avg}}, \tag{5.7}$$

where ϵ_{TT} is the TT stacking strength and ϵ_{avg} is the strength for both AA and TT obtained from fitting oxDNA to the SantaLucia model. We vary the AA stacking interaction strength to match the experimental differences of the thermal stabilities of the poly-A-loop and poly-T-loop hairpins, specifically $\delta\Delta F = \Delta F_{(\mathrm{A-loop})} - \Delta F_{(\mathrm{T-loop})}$, where $\Delta F_{(\mathrm{X-loop})} = F_{\mathrm{b},(\mathrm{X-loop})} - F_{\mathrm{ub},(\mathrm{X-loop})}$ is the difference in free energy between the bound (hairpin) state and the unbound (open) state for a hairpin with a poly-X loop.

We obtain the bound and unbound free energies F_α, where $\alpha \in \{\mathrm{b}, \mathrm{ub}\}$, by using thermodynamic integration. We start by calculating $F_\alpha^{(0)}$, the free energy of a reference state for which the effective charge on the backbone site, q_{eff}, is set to 0, and $\epsilon_{\mathrm{AA}} = \epsilon_{\mathrm{TT}} = \epsilon_{\mathrm{avg}}$. $F_\alpha(I, x, q_{\mathrm{eff}})$, the free energy of the state with salt concentration I, stacking strengths ϵ_{AA} and ϵ_{TT}, and effective backbone charge q_{eff}, is then obtained by solving the following integrals

$$F_\alpha(I, x, q_{\mathrm{eff}}) = F_\alpha^{(0)} + \int_0^{q_{\mathrm{eff}}} \mathrm{d}q'_{\mathrm{eff}} \frac{\partial F_\alpha(I, 0, q'_{\mathrm{eff}})}{\partial q'_{\mathrm{eff}}} + \int_0^{x} \mathrm{d}x' \frac{\partial F_\alpha(I, x', q_{\mathrm{eff}})}{\partial x'} \tag{5.8}$$

where x measures the deviation of the AA and TT strengths from ϵ_{avg}, such that Eq. 5.7

is satisfied, i.e.

$$\epsilon_{\mathrm{AA}} = (1+x)\epsilon_{\mathrm{avg}}, \;\; \epsilon_{\mathrm{TT}} = (1-x)\epsilon_{\mathrm{avg}}\,. \tag{5.9}$$

Taking the derivatives of the free energy in Eq. 5.8 we obtain,

$$\begin{aligned} F_\alpha(I,x,q_{\mathrm{eff}}) =& F_\alpha^{(0)} + \int_0^{q_{\mathrm{eff}}} \mathrm{d}q'_{\mathrm{eff}} \frac{2\langle V_{\mathrm{DH}}(I,0,q'_{\mathrm{eff}})\rangle_\alpha}{q'_{\mathrm{eff}}} + \int_0^x \mathrm{d}x' \left[\frac{1}{1+x'} \langle V_{\mathrm{stack}}(I,x',q_{\mathrm{eff}})\rangle_\alpha^{\mathrm{AA}} \right. \\ & \left. - \frac{1}{1-x'} \langle V_{\mathrm{stack}}(I,x',q_{\mathrm{eff}})\rangle_\alpha^{\mathrm{TT}} \right], \end{aligned} \tag{5.10}$$

where $\langle V_{\mathrm{stack}}(I,x,q_{\mathrm{eff}})\rangle_\alpha^{\mathrm{AA(TT)}}$ is the average stacking energy of all AA(TT) nucleotides in a strand in state α with salt concentration I, x defined as in Eq. 5.9, and backbone charge q_{eff}. The terms appearing in the integrands of Eq. 5.10 can be obtained by running short simulations of bound and unbound states separately and therefore the free energies can be calculated in an efficient manner. The results for the difference in $\delta\Delta F$ between oxDNA and experiment as a function of ϵ_{AA} for the hairpins with 21- and 31-base loops are shown in Fig. 5.5; the value for the AA stacking strength that minimises this difference is found to be $\epsilon_{\mathrm{AA}} \approx 1.075\epsilon_{\mathrm{avg}}$ (corresponding to $\epsilon_{\mathrm{TT}} \approx 0.925\epsilon_{\mathrm{avg}}$), which is not too dissimilar to the preliminary value suggested in Ref. 97. We note that this value gives satisfactory predictions for $\delta\Delta F$ for a wide range of salt concentrations down to 0.05 M.

5.6 The oxDNA2 model

In summary, the oxDNA2 potential can be written as

$$\begin{aligned} V_{\mathrm{oxDNA2}} = & \sum_{\text{nearest neighbours}} (V^*_{\mathrm{backbone}} + V^*_{\mathrm{stack}} + V'_{\mathrm{exc}}) + \\ & \sum_{\text{other pairs}} (V^*_{\mathrm{HB}} + V_{\text{cross stack}} + V_{\mathrm{exc}} + V^*_{\text{coax stack}} + V^*_{\mathrm{DH}}), \end{aligned} \tag{5.11}$$

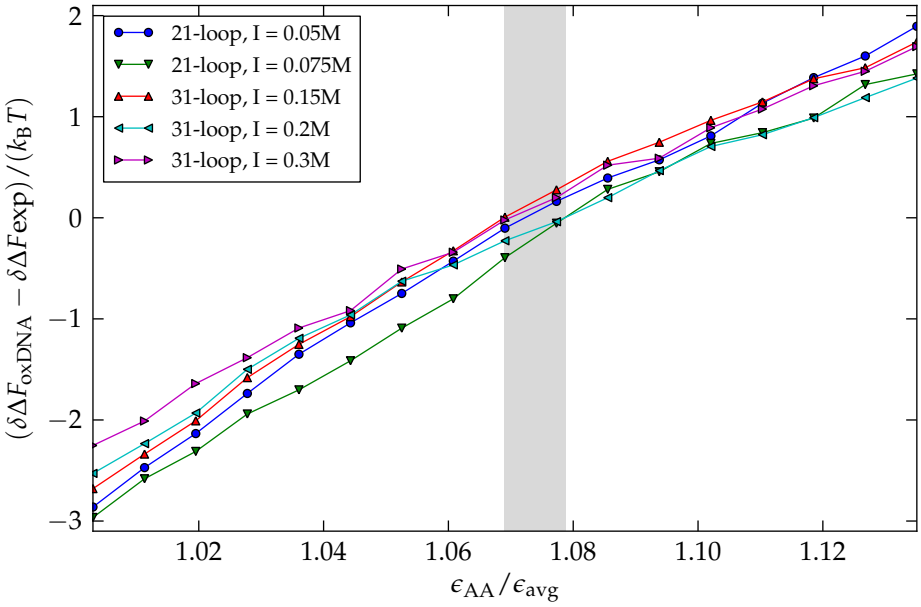

Figure 5.5: Deviation of oxDNA's value for $\delta\Delta F$ from experiment at 295.6K, as a function of $\epsilon_{\mathrm{AA}}/\epsilon_{\mathrm{avg}}$. The grey region contains the zero-deviation points for every curve. Details of the experimental setup and results can be found in the Supplementary Material of Ref. 89.

where a V_{x}^{*} indicates that the term is either modified for oxDNA2, or, in the case of V_{DH}^{*}, new in oxDNA2. The modified parameters for oxDNA2 are compared with those for oxDNA in Table A.1, and a full account of the changes is given in Appendix A. All other parameters remain the same as in the original model.

We emphasise that, after all the relevant changes to the potential were made, the hydrogen bonding and stacking parameters were modified to ensure that the close agreement to experimental duplex melting temperatures achieved for the original model was retained with oxDNA2 (this was done immediately after the changes to V_{backbone}, as described at the end of Section 5.4).

5.7 Physical properties of DNA in the oxDNA2 model

The structural, mechanical and thermodynamic properties of DNA for the new version of oxDNA presented in this chapter, which we call oxDNA2, are slightly different from

the properties for the original oxDNA model. We briefly highlight the most important of these changes in this section; and these properties are given as a function of salt concentration as this is now possible with the new model. We note that all of the results described in this section were computed using the final version of the oxDNA2 model as summarized in Section 5.6. The details of the simulations used to compute these results are given briefly in the figure captions and in detail in Appendix B.4.3, Table B.6 and Table B.7.

The structural properties of the new model, specifically the pitch and rise of double-stranded DNA, are presented in Fig. 5.6. As might be expected, the rise increases with decreasing salt concentration, due to the greater repulsion between backbone sites. The pitch also increases with decreasing salt, consistent with the measured increase in rise and slight decrease in neighbouring backbone-backbone distance measured as salt concentration is decreased (the duplex radius remains approximately constant). Although there is not much experimental evidence to compare this with, there is some indication that the pitch is roughly constant for low salt concentrations (0.162 M and below).[118] As described in Section 5.4, the pitch is chosen (by modifying the bonded neighbour backbone-backbone interaction) so that the global twist of origami structures agrees with experimental measurements. Specifically, we set the backbone-backbone interaction so that the helix bundle designed to have no global twist has no global twist in the model at $[Na^+]$=0.5 M. This results in a pitch of roughly 10.55 bp/turn at this salt concentration, compared to 10.36 bp/turn in the original model, and experimental values of around 10.45 suggested by cyclisation experiments,[118;121] albeit in the presence of some divalent salt.

The thermodynamics of duplex formation are shown in Fig. 5.7. The transition width

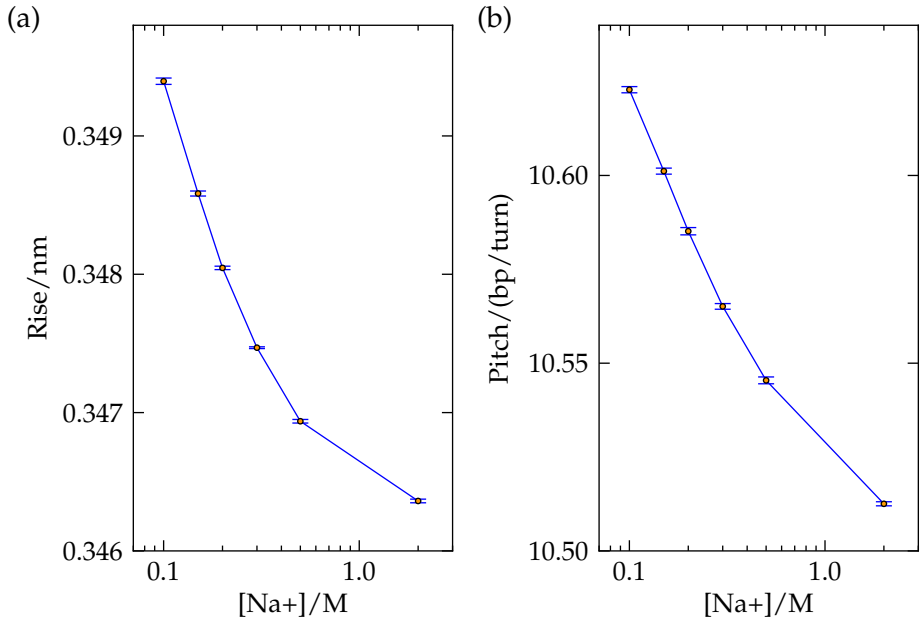

Figure 5.6: (a) Rise and (b) pitch of a 60-bp duplex as a function of salt concentration in oxDNA2. For each salt concentration the duplex was simulated for at least 3×10^9 MD steps. The error bars, which in some cases are narrower than the plot markers, show the standard error on the mean given by averaging over 10 independent estimates for each data point.

for the yield curve of a 10-bp duplex at 0.5 M [Na^+] in oxDNA2 is largely unchanged from the original oxDNA, and the transition widths depend weakly if at all on salt in oxDNA2. The free-energy profiles for duplex formation in oxDNA2 show that the free-energy cost of forming the first base pair decreases with increasing salt, presumably due to the reduced energetic cost of bringing the two single strands close together as the electrostatic repulsion between backbone sites becomes more short-ranged. The slope of the bound region of the free-energy profile also decreases with increasing salt, indicating a reduced free-energy cost of forming subsequent base pairs.

In Fig. 5.8, the duplex melting temperatures as a function of salt and for different duplex lengths in oxDNA2 are compared to the results from the SantaLucia model, to which oxDNA2 was parameterised. As expected, we find good agreement.

The mechanical properties of the model double-stranded DNA as a function of

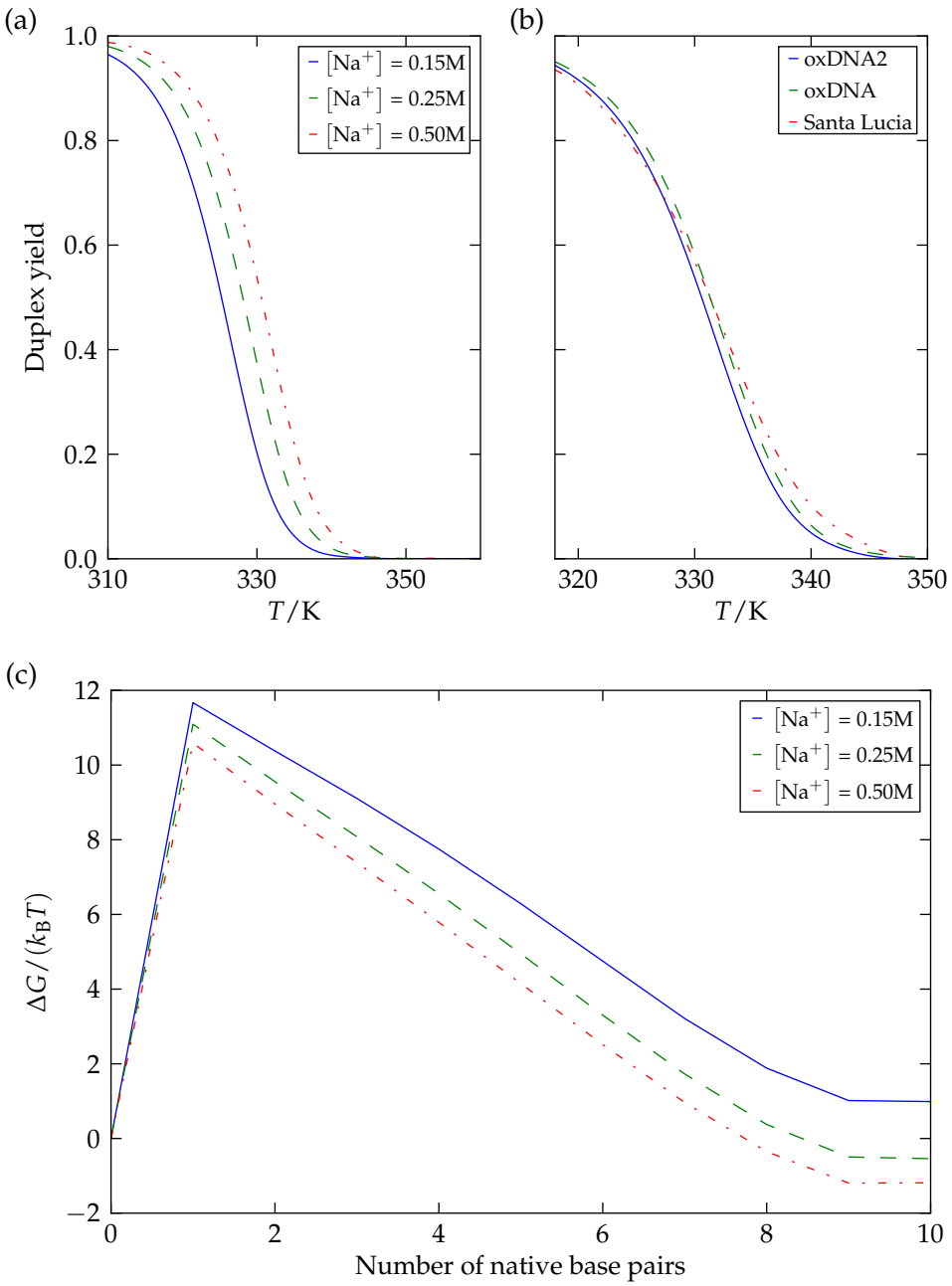

Figure 5.7: The thermodynamics of formation of a 10-bp duplex with oxDNA2. (a) Duplex yield for oxDNA2 at different salt concentrations. (b) Comparison of duplex yield profiles at $[Na^+]$=0.5 M as predicted by oxDNA2, the original oxDNA model and the SantaLucia model.[112] (c) The free-energy profile for duplex formation at different salt concentrations, at the melting temperature of a 10-bp duplex at $[Na^+]$=0.25 M. The duplex was simulated for at least 3×10^{10} virtual move Monte Carlo[137] (VMMC) steps at each salt concentration using umbrella sampling at a monomer concentration of 3.3×10^{-4} M.

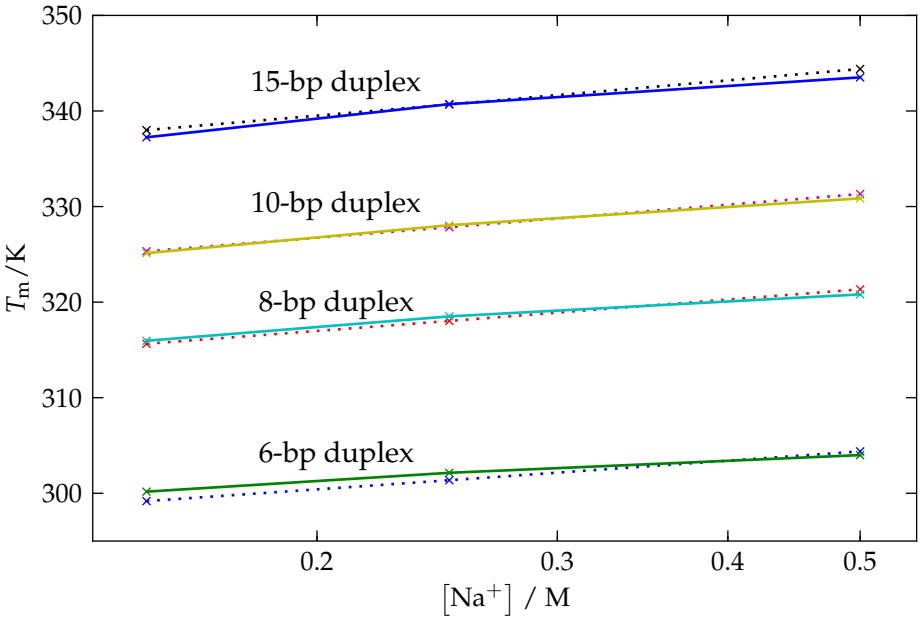

Figure 5.8: The melting temperature as a function of salt concentration for duplexes of different lengths simulated with oxDNA2 (crosses with solid lines) and as predicted by the SantaLucia model (crosses with dotted lines) for a DNA strand concentration of 3.3×10^{-4} M. Each duplex was simulated for roughly 4×10^9 VMMC steps at a salt concentration of 0.25 M, and the melting temperatures at 0.15 M and 0.5 M were computed by single histogram reweighting. [138]

salt, specifically the persistence length and torsional stiffness, are shown in Fig. 5.9. The persistence length was calculated by computing the correlation of the helix axis as described in Ref. 66. The effective torsional stiffness, C_{eff}, was computed from the linear regime of the torque response curve of a 60-bp duplex under tension (see Appendix B.4.2 for details). The simulations were carried out under a linear force of 30 pN, a high force regime where we expect that C_{eff} approximates the true torsional stiffness C_0.[79;139] The persistence length at [Na^+]=0.5 M, about 123 bp, is very slightly lower than for the original model, 125 bp, and the persistence length at different salt concentrations is consistent with the rather broad range of values suggested by experiment.[140;141] The decrease in persistence length with increasing salt is expected due to the decrease in repulsion between the duplex's backbone sites, which makes the duplex less stiff, in agreement with experiment.[140] The slight decrease in torsional stiffness with increasing salt can be rationalised in the same way. Regardless of the salt concentration, the torsional stiffness measured for oxDNA2 ($C_{\text{eff}} \approx 380 - 400$ fJfm) is lower than for the original model[79] ($C_0 \approx 473$ fJfm). Single-molecule twisting experiments give a value of $C_{\text{eff}} \approx 410$ fJfm for a pulling force of 3.5 pN[142] and a pulling force of 45 pN,[143] at around [Na^+]=0.1 M. There is limited experimental data on the salt dependence of torsional stiffness in DNA; however the torsional stiffness has been reported to be roughly constant in the range [Na^+]=0-0.162 M.[118]

The effects of various motifs on duplex melting temperatures in oxDNA2 are compared to results from the original oxDNA and the SantaLucia model in Table 5.2. These effects are either barely changed in oxDNA2, or are now closer to the SantaLucia values than the original oxDNA values were.

The melting temperatures of hairpins at 0.5 M [Na^+] are shown in Fig. 5.10. The

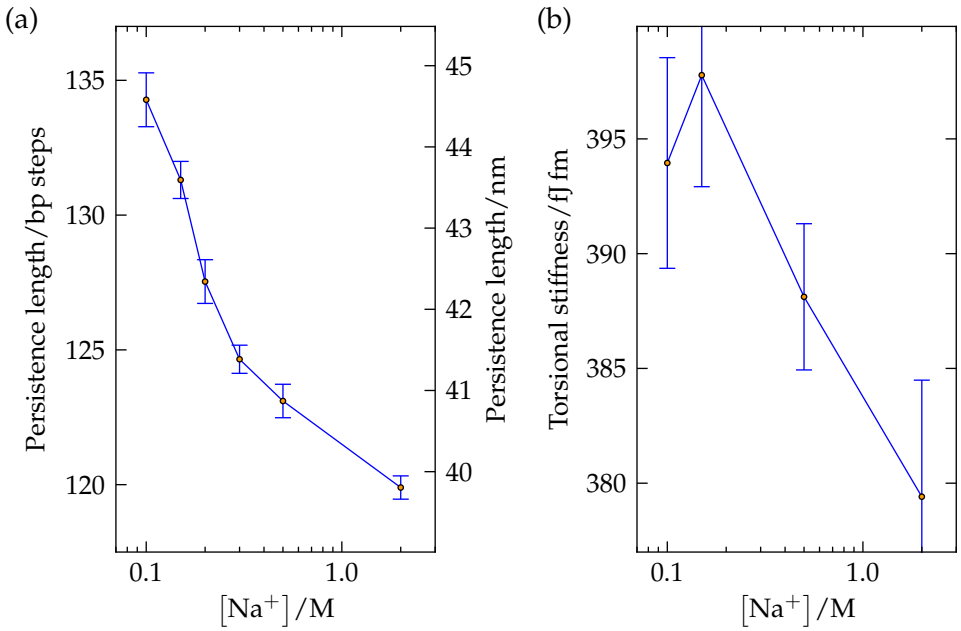

Figure 5.9: (a) The persistence length and (b) the torsional stiffness of duplex DNA in oxDNA2 as a function of salt concentration. For the persistence length, a 60-bp duplex was simulated for at least 3×10^9 MD steps for each salt concentration. The method for computing the torsional stiffness is described in Appendix B.4.2. Error bars for the simulation results show the standard error on the mean given by averaging over either 10 independent estimates (for the persistence length) or 5 independent estimates (for the torsional stiffness) for each data point.

	ΔT_m/K		
Motif	oxDNA2	oxDNA	SantaLucia
1-base bulge in 8-bp duplex	−18.7	−17.98	−23.19
2-base bulge in 8-bp duplex	−27.4	−23.92	−26.73
1-bp terminal mismatch in 5-bp duplex	+6.5	+6.71	+8.6
2-bp internal mismatch in 8-bp duplex	−15.9	−15.77	−13.99

Table 5.2: The effect of introducing various motifs on the duplex melting temperature at a monomer concentration of 3.3×10^{-4} M. ΔT_m is the difference between the melting temperature of the structure with the motif and a duplex consisting of the same number of complementary base pairs as the motif structure. For bulges and internal mismatches, the motif was placed at the centre of the duplex. The simulations of the bulges, terminal mismatches and internal mismatches were run for at least 6×10^{10}, 8.5×10^{10}, and 5.6×10^{10} VMMC steps respectively. All oxDNA2 results were obtained from simulations at a concentration of 4.2×10^{-5} M which were extrapolated to a concentration of 3.3×10^{-4} M.

hairpin melting temperatures are lower than for the original oxDNA by about 2 K, which in turn had hairpin melting temperatures about 3 K lower than those predicted by SantaLucia. The gap between the oxDNA2 hairpin melting temperatures and those given by the SantaLucia model widens somewhat as the salt concentration is lowered (Fig. 5.11), indicating a stronger salt dependence in oxDNA2, with a typical underestimate of around 7 K at a salt concentration of 100 mM for a relatively short loop length of 6 bases. For longer loops, the difference in predicted melting temperatures between oxDNA2 and SantaLucia widens further. This difference is unsurprising, as in oxDNA2 (and with physical DNA) ssDNA becomes stiffer at lower salt concentrations, making the formation of a hairpin less favourable,[80] whereas in the SantaLucia model the loops' contribution to hairpin stability is salt-independent.[112]

It seems plausible that oxDNA2's performance is better than implied by the salt-independent loop contribution in the SantaLucia model. In Fig. 5.12 we consider hairpins with a short 5-bp stem and a long 31-base loop, making the hairpin thermodynamics particularly sensitive to the change in loop stiffness with salt. The hairpin stability (i.e. the free energy difference between the bound and unbound state) as predicted by oxDNA2 and the SantaLucia model is compared to experimental data determined using FRET (experimental details are given in the Supplementary Material of Ref.[89]). As expected we find that the hairpin stabilities for oxDNA2 have a much stronger salt dependence (steeper gradient) than those of SantaLucia (shallow gradient). At higher salt the oxDNA2 results show a similar slope to the experimental curve, but at lower salt (0.2 M and below) oxDNA2 shows a stronger destabilization of the hairpins with decreasing salt (a steeper gradient) than experiment. Thus, although oxDNA2 does a reasonable job of capturing these physical effects, the comparison suggests that the single strands may be experiencing too much repulsion at low salt. We also note

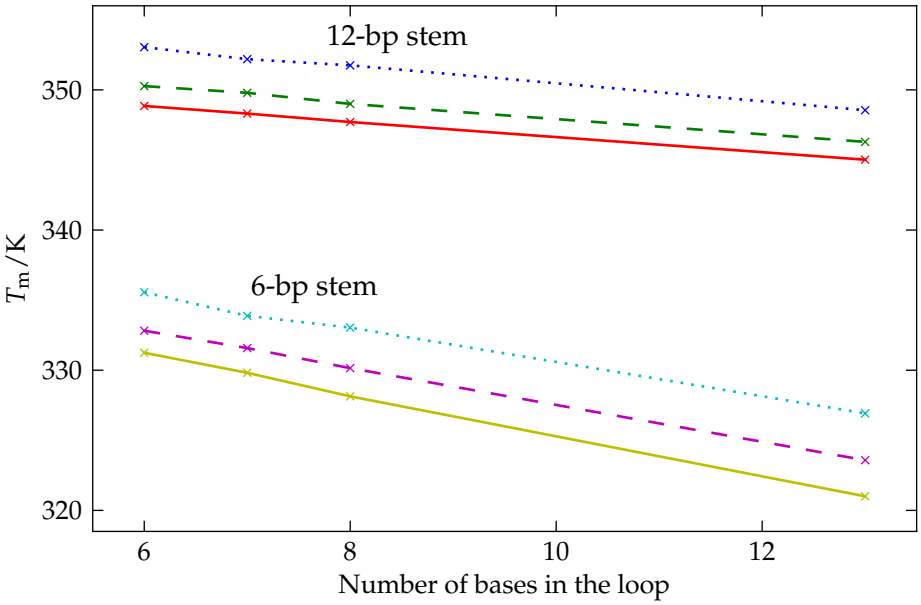

Figure 5.10: The melting temperature as a function of loop length for hairpins simulated with oxDNA2 (crosses with solid lines), the original oxDNA (crosses with dashed lines), and according to the SantaLucia model (crosses with dotted lines), at 0.5 M [Na^+]. For the oxDNA2 results, the hairpins were simulated at each salt concentration for at least 1.8×10^{10} VMMC steps.

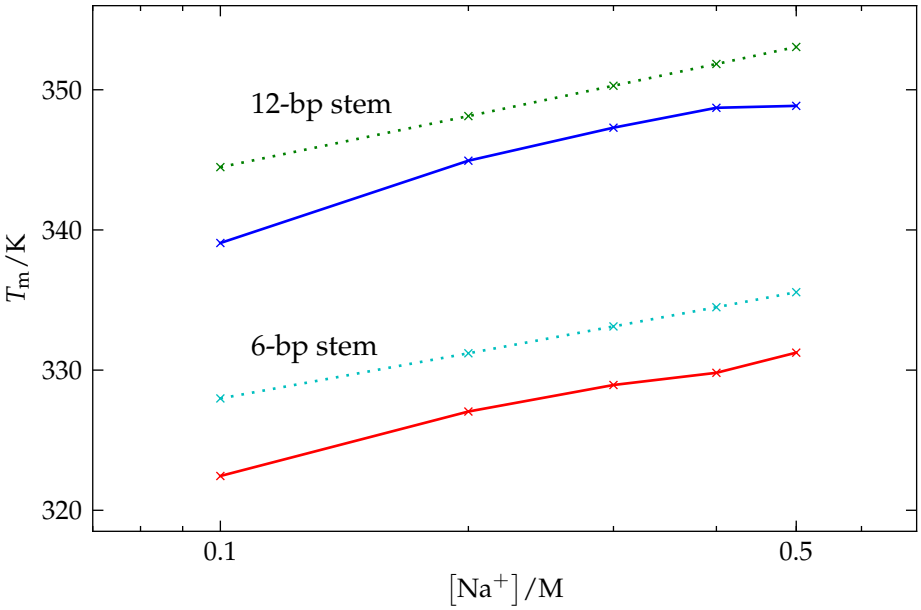

Figure 5.11: The melting temperature as a function of salt concentration of hairpins with 6-base loops simulated with oxDNA2 (crosses with solid lines) and according to the SantaLucia model (crosses with dotted lines). For the oxDNA2 results, the hairpins were simulated at each salt concentration for at least 1.5×10^{10} VMMC steps.

that the SantaLucia model is unable to predict the large difference in stability of the two hairpins due to their different loop sequences, because the SantaLucia model is insensitive to the loop sequence except for the two bases adjacent to the stem.

We examine the flexibility of single strands as a function of salt in oxDNA2 by measuring the radius of gyration, R_g, of a single strand of DNA at different salt concentrations, sequences and strand lengths, and comparing to experimental results (Fig. 5.13). The oxDNA2 model reproduces the overall trend of increasing R_g with salt, and is in agreement with the somewhat noisy experimental data. We note that oxDNA2 is able to capture two important effects: the greater stiffness (and hence larger R_g) of the more strongly-stacking poly-A strands; and the greater salt dependence of the poly-T strands' R_g compared to that of the poly-A strands, caused by the weaker stacking of the poly-T strands, which means that electrostatic repulsion makes a greater relative contribution to its stiffness.

5.8 Summary

The modifications and extensions to oxDNA presented in this section open up a variety of new potential applications for the model. With the introduction of an explicit salt-dependent term in the potential, the model can be used to simulate systems under physiological conditions, and to investigate the salt-dependent behaviour of DNA. Also, the parametrization presented here allows a quantitative comparison to experiments run in a wide range of salt concentrations rather than just in the high-salt limit. The introduction of major-minor grooving adds detail to the model, and, combined with other small modifications, allows the use of oxDNA2 to accurately characterise the structural properties of large DNA nanostructures, such as DNA origami. In particular the helical pitch and the twist angles at nicks and junctions were fine-tuned to obtain

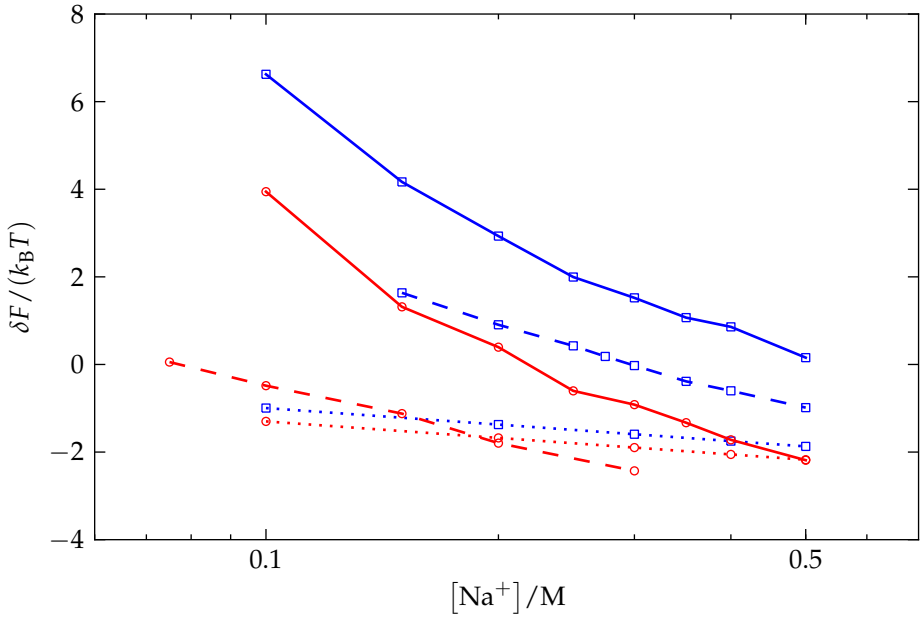

Figure 5.12: The free energy difference between the bound and unbound state for two hairpins as measured in experiment using FRET (dashed lines), and as predicted by oxDNA2 (solid lines) and the SantaLucia model (dotted lines). Results for two different hairpins are presented: the A_{31}-VW (red, circular markers) hairpin and the T_{31}-VW hairpin (blue, square markers). The oxDNA2 simulations were run for 10^{10} VMMC steps for each hairpin, with $q_{\text{eff}} = 0$ and $x = 0$ (with x defined in Eq. 5.9), and the results (computed for the standard oxDNA2 values for q_{eff} and x) were calculated using thermodynamic integration as described in Section 5.5 and Appendix B.4.3.

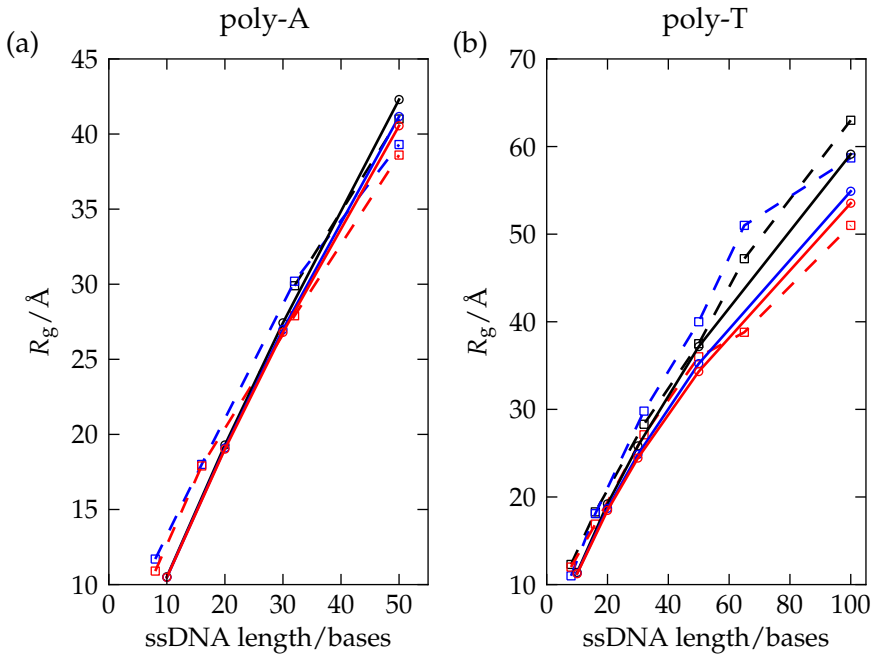

Figure 5.13: The radius of gyration (R_g) for ssDNA as a function of ssDNA length and for different salt concentrations, for (a) a poly-A ssDNA and (b) a poly-T ssDNA, from oxDNA2 (solid lines with circles) and from experiments due to Sim *et al.*[144] (dashed lines with squares). The salt concentrations shown are 0.225 M (blue lines), 0.525 M (red lines), 1.025 M (green lines). For the oxDNA2 results, the strands were simulated for between 2×10^9 and 5×10^{10} VMMC steps at each strand length and salt concentration.

a correct global twist for a test-case 3D origami structure. In the absence of definitive experimental values for these individual parameters we chose a combination that we deemed physically reasonable and that produces equilibrium structures that compare well with experimental ones. Finally, the sequence dependence in the model has been extended by introducing different interaction strengths for the AA and TT stacking. This change will be particularly useful for studying the effects of stacking in single strands or single-stranded sections (e.g. hairpin loops) as poly-A and poly-T sections are often used as paradigmatic examples of strongly and weakly stacking sequences, respectively, and for modelling DNA nanostructures where poly-T loops are often used as flexible linkers.

There are some limitations associated with the new features of the model. The Debye-Hückel treatment of the electrostatics is perhaps an oversimplification, and we should not expect it to capture all of the complexities associated with the electrostatics for DNA, but it is not straightforward to think of a different approach that would be consistent with our level of coarse graining. In particular, the fact that hairpin stability is reduced faster than in experiment suggests that single strands may experience too much repulsion in oxDNA2 at low salt, and care should be taken in making predictions based on our models at low salt concentration if the system under investigation depends crucially on the thermodynamics of long single-stranded sections.

In Chapter 6, we will carry out investigations on DNA origami structures using the improvements in the model's structural prediction for large-scale DNA nanostructures. Other projects using oxDNA2 to study large-scale DNA systems are also underway, including a study of structures composed of multi-arm tiles.[145] At the same time, the model is being used to study the salt-dependent thermodynamics of a diverse

set of systems, which would not have been possible without the introduction of salt dependence into the potential. The introduction of different strengths for the AA and TT stacking in the model allows us to better capture these effects, and in addition gives us more accurate sequence-dependent hairpin thermodynamics and kinetics, which are currently being exploited for further studies of hairpins.

A note regarding the credit for the work described in this chapter: The work was done in collaboration with many others. In particular, the work on modifying the AA and TT stacking strengths was done by Majid Mosayebi, using experimental data from Roman Tsukanov and Eyal Nir (and I did not contribute to it). The work on the salt-dependent interaction term was primarily done by Ferdinando Randisi under the supervision of Petr Šulc and Flavio Romano, and I contributed at a later stage. John Schreck provided Fig. 5.2, and otherwise the work on introducing different widths for the major and minor DNA grooves and improving the model's structural prediction for large nanostructures is my own, with Flavio Romano's supervision. Although I did not work directly on everything described in this chapter, I was responsible for testing and tweaking parameters to produce a final version of the new model. Moreover, it would seem odd to present an incomplete description of the new model merely for reasons of authorship.

6 Structural properties of DNA origami

In Chapter 5, we described an updated coarse-grained DNA model, oxDNA2. As noted in that chapter, the new model was fitted to better reproduce the structural properties of large-scale DNA nanostructures. In this chapter, we use the model to investigate the structural properties of 2D and 3D origamis, and of the individual Holliday junctions which are crucial for understanding origami structure. Where possible, we compare our results to experiment.

6.1 Previous studies of the structure of DNA objects

When origami structures are produced experimentally, they are typically visualised using AFM or (more often) TEM after attachment to a surface,[10;22;35;36] or sometimes with small-angle X-ray scattering (SAXS).[146] While sufficient to check that roughly the desired structure has been formed, these methods are not usually suitable for a more fine-grained structural analysis. Note that adsorption onto a surface may perturb the structure, especially for 2D origamis, which may be flattened and made to look more ordered – an advantage of SAXS is that is it is a solution measurement.

There have, however, been some attempts to experimentally characterise the structure of DNA origamis and other nanoscale structures in greater detail. Multiple studies have involved quantitative analyses using TEM imaging on a surface. One class of origamis was designed to multimerise to form twisted "ribbons," whose number of turns per unit distance when visualised on a surface could be used to infer the twist

angle of the corresponding monomers. We used this result when parameterising the oxDNA2 model in Chapter 5. In the same work, the distributions of bend angles were measured for a set of bent origami blocks, and the Castro group has made similar measurements of angles and distances for a variety of structures, such as bent bars, a "crank" which couples linear and angular motion, a nanoscale Bennett linkage, and a "scissor" structure.[147–149] The resolution of such measurements is usually relatively coarse, however.

In Ref. 126 and Ref. 39, the authors used cryo-EM imaging to characterise the structure of a DNA nanotetrahedron and a 3D DNA origami, respectively, and fitted all-atom DNA structures to the resulting electron density maps. This method may better represent the structure that the origami adopts in solution. Finally, fluorescence techniques may be used to characterise origami in solution or on a surface. FRET allows distances to be measured,[150] and a super-resolution imaging technique known as DNA PAINT[40;151;152] has recently been developed to charactise 3D DNA nanostructures.

Computer models have also been employed to study the structure of large DNA nanostructures. Notably, Yoo *et al.*[48] were able to use an all-atom model to simulate origamis consisting of thousands of base pairs (up to almost 3 million atoms including the solvent) for up to 140 ns, although it is not clear that the structures had completely relaxed into their equilibrium conformations in all cases. However, such simulations are extremely computationally intensive and cannot be performed routinely. More coarse-grained models have also been used to study origami structure: Arbona *et al.*[52] used a previously developed coarse-grained DNA model[153] to simulate small DNA-origami-style structures, computing their persistence lengths and examining the effect of changing B-DNA pitch on global twist. The relatively simple "stack of

plates" DNA model that they used treats base pairs as rigid bodies, so that base pairs cannot be broken, and the behaviour of the junctions, which are modelled as normal duplexes joined with additional elastic springs, is not well characterised. By far the most widely used model is CanDo,[12;53;54] a specialised model in which double helices are linear elastic rods, Holliday junctions are somewhat flexible connections between the rods, and single-stranded regions are freely-jointed chains. This relatively simplified representation allows equilibrium structural and mechanical properties to be computed in a matter of minutes, and produces structures which are generally in good agreement with experiment. However, the model treats only the elastic regime of local structural deformations, and does not allow base pairs to break.

Being a nucleotide-level model with realistic structural properties for single- and double-stranded DNA as well as motifs such as hairpins and bulges,[73] the oxDNA2 model can be used for a more detailed study of origami structure than CanDo. At the same time, it can simulate these structures for far longer times than all-atom simulations, so that they have a higher chance of adopting their fully equilibrated conformations during the simulations, which will be representative of experimental systems.

We note that for the work described in this chapter, we used the version of the oxDNA2 model with average-strength interactions, rather than sequence-dependent ones, as we were interested in generic structural properties. In addition, the latter model has only been directly fitted to reproduce the sequence dependence of the thermodynamics of melting, but with no attempt yet to reproduce the detailed sequence dependence of the structural or mechanical properties – this may be the subject of future work.

6.2 Characterising Holliday junction structure in oxDNA2

We begin our study of origami structure by considering the four-way junctions, also known as Holliday junctions, which are an essential feature of origami designs. These occur wherever two strands, usually staple strands,[1] cross from one double helix to another within the origami. Thus the junctions play the vital role of joining adjacent double helices together. Because each origami contains many junctions, their structural properties can potentially have major repercussions for the structure of the origami.

The structure of single isolated Holliday junctions has been characterised experimentally[154;155] through X-ray crystallography,[156;157] AFM[158] and FRET measurements.[159;160] Depending on the experimental conditions, a Holliday junction can exist in an open or stacked conformation, with the open conformation favoured in the absence of metal ions, and the stacked conformation adopted in the presence of divalent metal ions.[155] Because the latter conditions are used in the great majority of DNA nanotechnology studies, we consider only the stacked conformation here. This conformation is favoured by oxDNA for the salt conditions we consider (0.5 M).

In order to quantify the structure of the junction, we define two angles: ϕ, which measures the average angle between pairs of arms; and θ, which measures the average angle between the arms and the plane of the junction (see Fig. 6.1(a)).

6.2.0.1 Definitions of the ϕ and θ angles

To compute ϕ and θ, we first define a plane with normal $\boldsymbol{n}$, given by

$$\boldsymbol{n} = \boldsymbol{A} \times \boldsymbol{B} + \boldsymbol{B} \times \boldsymbol{C} + \boldsymbol{C} \times \boldsymbol{D} + \boldsymbol{D} \times \boldsymbol{A}, \tag{6.1}$$

[1] When the crossing strands are parts of the scaffold strand, the Holliday junctions, which usually line up along the structure, form what is known as the seam.

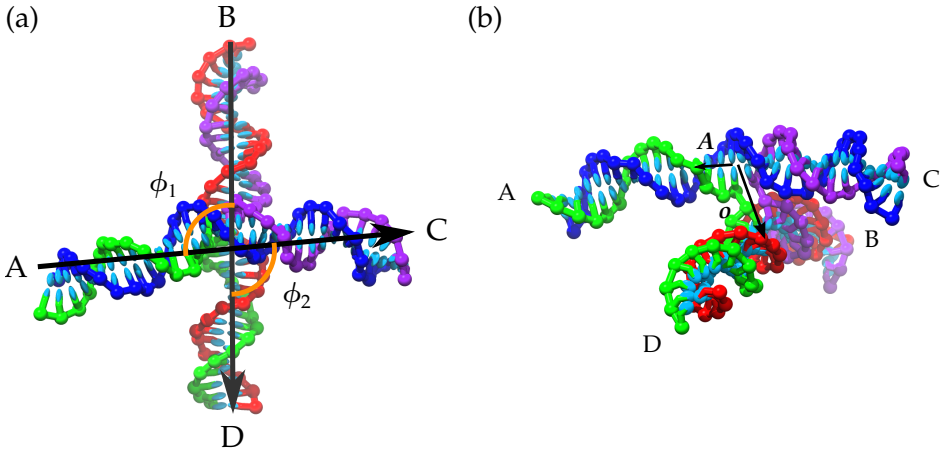

Figure 6.1: A typical configuration for a single Holliday junction simulated with oxDNA2 at 0.5 M salt. The green and purple strands cross from one double helix to another, while the red and blue strands occupy one double helix each. (a) A top-view from perpendicular to the plane of the junction. Arrows point along the helix axes in the the 3′ to 5′ direction of the red and blue strands. The four junction arms are labelled A, B, C and D. The angles ϕ_1 and ϕ_2 (loosely speaking, the angles between arms A, B and C, D respectively) are shown, where $\phi = (\phi_1 + \phi_2)/2$. θ is the average angle between each arm and the plane of the junction. (b) A view from the side. The arm vector $\boldsymbol{A}$ and orientation vector $\boldsymbol{o}$ are shown.

where we define $\boldsymbol{A}$ as the normalised vector pointing along the helix axis of arm A. To be exact, $\boldsymbol{A}$ points from the base-base midpoint of the base pair closest to the crossover to the 4th closest base pair's base-base midpoint, as shown in Fig. 6.1(b). This means that $\boldsymbol{A}$ spans three base-pair steps. $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ are defined in an analogous way. This definition is a compromise between not being too influenced by the motion of one or two base pairs and not being influenced by the motion of the part of the flexible arm far from the crossover.

In addition to this, we define an orientation vector $\boldsymbol{o}$ (Fig. 6.1(b))

$$\boldsymbol{o} = \frac{(\boldsymbol{b} + \boldsymbol{d})}{2} - \frac{(\boldsymbol{a} + \boldsymbol{c})}{2}, \tag{6.2}$$

where $\boldsymbol{a}$ is the position vector of the base-base midpoint closest to the crossover of arm A, and likewise for $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$. We will use $\boldsymbol{o}$ to define a ϕ angle that runs from 0° to

360°, using a triple scalar product.

To compute ϕ, we define two twist angles, ϕ_1 and ϕ_2, which are a measure of the twist of the Holliday junction arms around the axis of the crossover:

$$\phi_1 = \begin{cases} \arccos(\boldsymbol{A'} \cdot \boldsymbol{B'}) & \text{if } (\boldsymbol{A'} \times \boldsymbol{B'}) \cdot \boldsymbol{o} > 0 \\ 360^\circ - \arccos(\boldsymbol{A'} \cdot \boldsymbol{B'}) & \text{if } (\boldsymbol{A'} \times \boldsymbol{B'}) \cdot \boldsymbol{o} < 0 \end{cases},$$

and (6.3)

$$\phi_2 = \begin{cases} \arccos(\boldsymbol{C'} \cdot \boldsymbol{D'}) & \text{if } (\boldsymbol{C'} \times \boldsymbol{D'}) \cdot \boldsymbol{o} > 0 \\ 360^\circ - \arccos(\boldsymbol{C'} \cdot \boldsymbol{D'}) & \text{if } (\boldsymbol{C'} \times \boldsymbol{D'}) \cdot \boldsymbol{o} < 0 \end{cases},$$

where $\boldsymbol{A'} = (\boldsymbol{A} - (\hat{\boldsymbol{n}} \cdot \boldsymbol{A})\hat{\boldsymbol{n}})/|\boldsymbol{A} - (\hat{\boldsymbol{n}} \cdot \boldsymbol{A})\hat{\boldsymbol{n}}|$ is the normalised projection of the vector $\boldsymbol{A}$ into the plane perpendicular to $\boldsymbol{n}$, with similar definitions for $\boldsymbol{B'}$, $\boldsymbol{C'}$ and $\boldsymbol{D'}$. ϕ is then defined as the average of the two ϕ_i angles: $\phi = (\phi_1 + \phi_2)/2$. Note that experimentally determined crystal structures have $\phi \approx 240^\circ$, while the smallest angle between the double helical domains (which is the angle usually quoted) for those structures is around 60°.

To compute θ, we define angles θ_X, which are each related to the angle between the arm X and the plane of the Holliday junction:

$$\theta_A = \begin{cases} -|\arcsin(\boldsymbol{A} \cdot \hat{\boldsymbol{n}})| & \text{if } \boldsymbol{A} \cdot \boldsymbol{o} > 0 \\ |\arcsin(\boldsymbol{A} \cdot \hat{\boldsymbol{n}})| & \text{if } \boldsymbol{A} \cdot \boldsymbol{o} < 0 \end{cases},$$

$$\theta_B = \begin{cases} |\arcsin(\boldsymbol{B}\cdot\hat{\boldsymbol{n}})| & \text{if } \boldsymbol{B}\cdot\boldsymbol{o} > 0 \\ -|\arcsin(\boldsymbol{B}\cdot\hat{\boldsymbol{n}})| & \text{if } \boldsymbol{B}\cdot\boldsymbol{o} < 0 \end{cases},$$

$$\theta_C = \begin{cases} -|\arcsin(\boldsymbol{C}\cdot\hat{\boldsymbol{n}})| & \text{if } \boldsymbol{C}\cdot\boldsymbol{o} > 0 \\ |\arcsin(\boldsymbol{C}\cdot\hat{\boldsymbol{n}})| & \text{if } \boldsymbol{C}\cdot\boldsymbol{o} < 0 \end{cases}, \tag{6.4}$$

$$\theta_D = \begin{cases} |\arcsin(\boldsymbol{D}\cdot\hat{\boldsymbol{n}})| & \text{if } \boldsymbol{D}\cdot\boldsymbol{o} > 0 \\ -|\arcsin(\boldsymbol{D}\cdot\hat{\boldsymbol{n}})| & \text{if } \boldsymbol{D}\cdot\boldsymbol{o} < 0 \end{cases},$$

This definition ensures that an arm pointing away from the plane normal to $\boldsymbol{n}$ has a positive angle and an arm pointing into the plane has a negative one. Note that by definition $\theta_A \equiv \theta_C$ and $\theta_B \equiv \theta_D$.[2] We define θ as the average of the θ_X angles: $\theta = (\theta_A + \theta_B)/2$.

6.2.1 Comparison of an oxDNA2 junction with experimental results

We ran MD simulations of an isolated Holliday junction with arms 16 base pairs long at 0.5 M salt and room temperature (296.15K), binning the resulting configurations to find the free-energy landscape for the system as a function of ϕ and θ. We windowed the simulation in ϕ using a biasing potential (defined in Appendix B.5.1) to speed up sampling for values of ϕ around 180°, as these are the ϕ values relevant for origami structure. The sequence was chosen to prevent branch migration. The junction can adopt one of two equivalent isomers: in order to simplify the analysis, configurations

[2] This can be shown by considering each $\boldsymbol{X}\cdot\hat{\boldsymbol{n}}$ along with the definition of $\boldsymbol{n}$ given in Eq. 6.1

that were determined to be in the wrong isomer were discarded. Further details regarding the simulations, including how a configuration's isomer is determined, and DNA strand sequences are given in Appendix B.5.1.

The results of the simulations are shown in Fig. 6.2. The free-energy minimum for the oxDNA2 junction is at $(\phi, \theta) = (93.5^\circ, 2.5^\circ)$, while integrating over θ gives a preferred ϕ angle of 90.5°, with a mean value of 91.8°; such junction configurations are said to be left-handed. However, the junctions observed in crystal structures are usually right-handed with 240° being a typical value for ϕ.[155] The preferred value of ϕ that we see with oxDNA2 can be understood as a way of aligning the backbone sites of the two double helices at the junction so that the distance between the excluded volumes of nearby backbones is maximized while the bonded backbone sites are separated by their preferred distance, as seen in Fig. 6.3(a). Configurations with ϕ corresponding to that of the crystal structure require the backbones to cross at the junction (Fig. 6.3(b)), which in oxDNA2 probably involves destabilizing the stacking and/or base-pairing interactions for the central base pairs, in order to keep the excluded volumes of the four central backbones far enough apart and the bonded backbones at their preferred distance. While there is a hint of a local minimum for these configurations in oxDNA2, they are disfavoured. Although oxDNA2 does use a simplified representation of the DNA backbone, we are not aware of any result definitively determining the handedness of the junction in solution. We also note that a crystalised left-handed junction has been reported[161] for an RNA-DNA complex, and that both chiral forms have been seen as local minima for a junction in solution in all-atom simulations.[162]

That oxDNA2 is unable to reproduce the experimental junction crystal structures' preference to be right-handed is, fortunately, not particularly detrimental to modelling

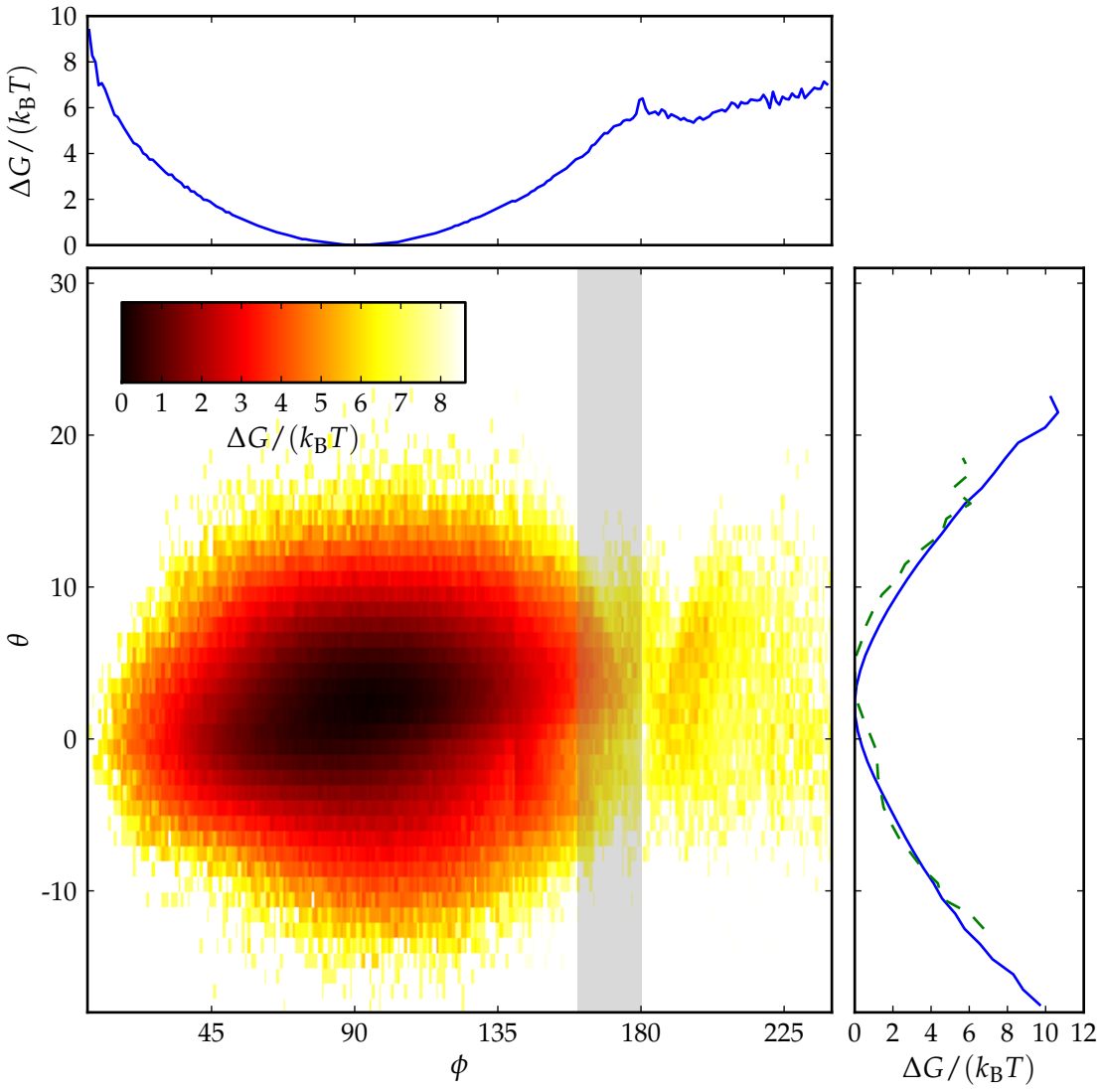

Figure 6.2: The free-energy landscape for a Holliday junction in oxDNA2 at a temperature of 296.15 K. Free-energy profiles for ϕ and θ are also shown. The full free-energy profile for θ is shown as a solid blue line, while the green dashed line shows the subset $160^\circ \leq \phi \leq 180^\circ$, shaded grey in the free-energy landscape.

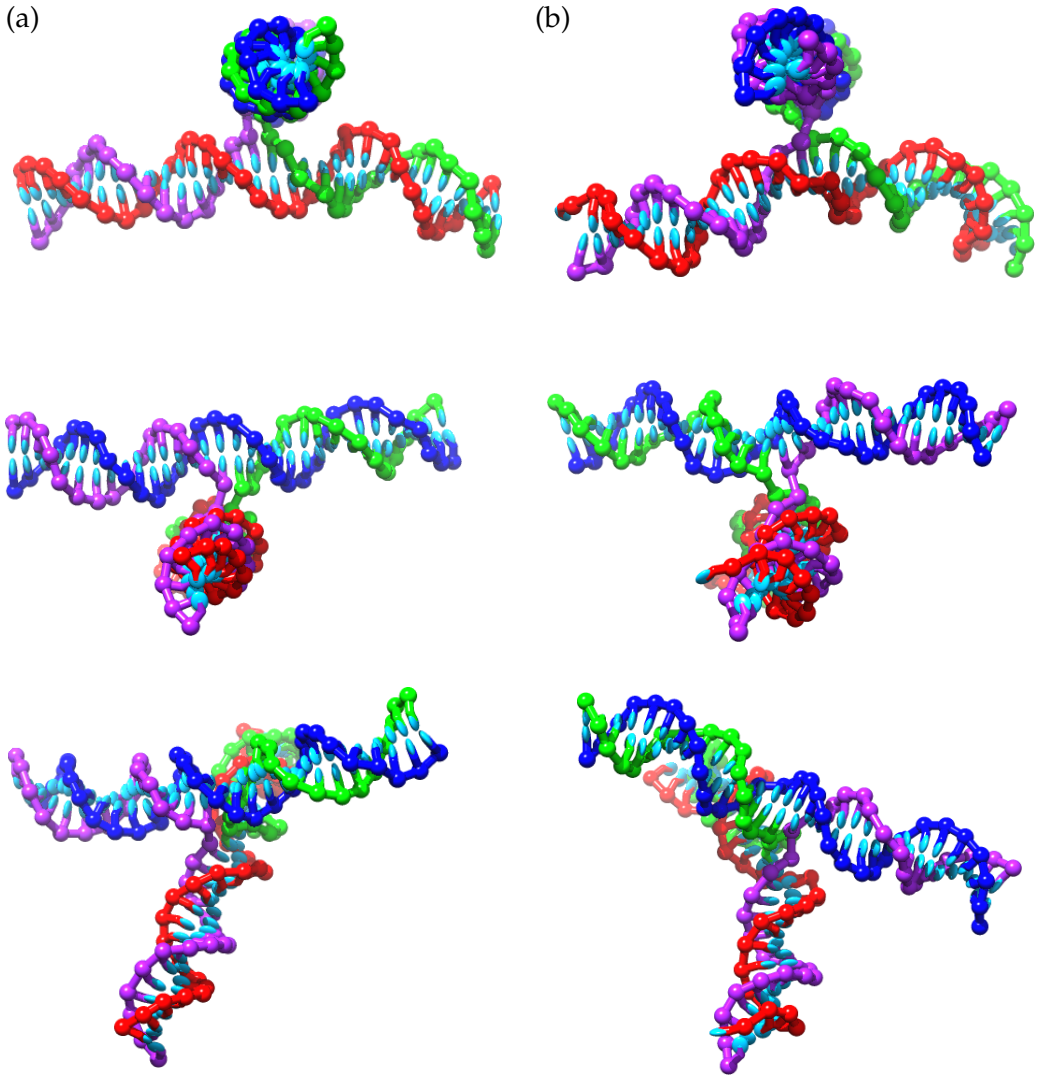

Figure 6.3: Holliday junction geometry in oxDNA2 shown with three different views for (a) the ϕ-θ minimum in oxDNA2 ($\phi \sim 93.5^{\circ}$) and (b) a configuration with a ϕ angle similar to the angle determined for Holliday junction crystal structures ($\phi \sim 240^{\circ}$), which is not a favourable state in oxDNA2. Note that the backbone bonds must cross for (b) but not for (a), which is particularly evident in the top panel, and that the middle panel in (b) shows a disruption of the top duplex's geometry at the junction.

origami structure with oxDNA2 for the following reasons. Firstly, the junction is very flexible in ϕ so this preference is not too strong (although clearly the oxDNA2 junction will very rarely adopt a configuration with $\phi \sim 240^\circ$). Secondly, in an origami the junctions are constrained to ϕ values close to 180°, i.e. 90° from the preferred junction angle for oxDNA2 and 60° for real DNA, so the junctions in both cases will have a somewhat similar level of stress. Finally, junctions which have been experimentally resolved within origamis have been found to exhibit ϕ angles slightly *below* 180° [39], a counter-intuitive result because one would probably expect deviations to take the junction towards (not away from) the preferred geometry. However, this is good news for oxDNA2, because, as we will see, the origami junctions in oxDNA2 do twist towards their preferred left-handed orientation, thus leading to qualitatively correct ϕ angles within origamis. This tendency to have a slight left-handed twist has also been observed in atomistic simulations of origamis.[48] The cause of this twist is not completely clear, but is probably due to a rougher free-energy landscape (as a function of ϕ) than predicted by oxDNA2, caused by the details of the interactions between the DNA, solvent and ions around the junction point.

The free-energy profile for θ in Fig. 6.2 shows a slight preference for a positive θ, with a free-energy minimum at $\theta = 2.5^\circ$, corresponding to a tendency for the helix arms to point slightly out of the plane of the junction. The effect is greater, with the minimum at $\theta = 4.5^\circ$, for a subset of the data for which $160^\circ \leq \phi \leq 180^\circ$, the region likely to be relevant within origamis. We will discuss the relevance of this finding to DNA origami structure in the next section. A simple argument explaining the behaviour of θ for the free junction is that negative values of θ will cause the arms to bump into each other more often, and this becomes more likely when the arms are approximately aligned, as for $\phi \approx 180^\circ$. Intriguingly, however, this argument would lead us to expect

a similar effect as we move towards $\phi \approx 0^\circ$, and this effect is not evident from the free-energy landscape. This suggests that the details of what are happening at the junction itself (i.e. the centre of the DNA complex) are also important, with the geometry at the junction for $\phi \approx 0^\circ$ being very different to that for $\phi \approx 180^\circ$. As $\phi \approx 0^\circ$ configurations are unimportant for junctions within DNA origamis we do not investigate the origin of this effect further here.

We also note a slight second minimum in the ϕ free-energy profile for a right-handed junction with $\phi \sim 200^\circ$, with a small barrier of perhaps $1 k_B T$. Although this apparent slight barrier may be mostly due to noise, there is a clear change of gradient after $\sim 180^\circ$, indicating qualitatively different behaviour.

6.3 Structural properties of a 2D origami

The coupling between many Holliday junctions present in a DNA origami generates a rich set of structural properties. In this section we consider 2D origamis, which consist of a single "sheet" of (anti-)parallel DNA helices joined by crossovers. In the following section we address 3D origamis.

We focus on a particular, fairly generic, 2D origami design for our study of 2D origami structure in this section. A simulation snapshot showing the equilibrated structure in solution is shown in Fig. 6.4 (see Fig. B.3 for the cadnano representation of the structure). In this section we refer to this design as "the 2D tile." We choose to study the 2D tile as it is fairly homogeneous and roughly flat, due to strategic positioning of the junctions designed to match the pitch of the DNA double helix, which is desirable as excessive curvature or twist can add unnecessary complications to structural analysis.

We can see some of the basic features of origami structure in Fig. 6.4. The origami

sheet is not noticeably twisted, but there is some curvature. We expect the structure to be flattened out when adsorbed onto a surface. The structure does fluctuate considerably in solution, but we focus on the average structure for this study.

We focus our study on two structural phenomena: the "weave pattern" and "corrugation."

6.3.1 The weave pattern

The weave pattern in 2D origamis, where adjacent double helices tend to push apart away from the junctions (as can be seen in Fig. 6.4), has been well known since the DNA origami method was originally devised, and is clearly visible in experimental images.[3] There are a few possible causes for this behaviour: (i) electrostatic repulsion between the negatively charged helices; (ii) an entropic effect due to the increased conformational space available when adjacent double helices are not perfectly parallel; or (iii) the local structure around the junctions favouring the arms pointing slightly away from each other (favouring a positive θ angle in the junction measurement scheme outlined in the previous section). For simulations with oxDNA2, we will see that all three of these effects play a role in the weave pattern of 2D origamis.

We quantify the weave pattern of the 2D tile by measuring the distance between the helix axes (defined by the midpoint between the bases for each base pair) for adjacent double helices (see Fig. 6.5). The results shown for the tile at a temperature of 300 K and $[Na^+] = 0.5\,M$ are plotted in Fig. 6.6.[3] Because of the pattern of junction placement in the origami's design (Fig. B.3), there are two obvious groups that the

[3] To simplify the appearance of the plot we omit some inter-helix distances. Namely, those involving the double helices at the top and bottom edges of the origami, as these are only constrained on one side and so exhibit slightly different behaviour; a few affected by branch migration which resulted in spurious results near the affected junction; and one affected by a partially melted staple which caused artificially enhanced flexibility.

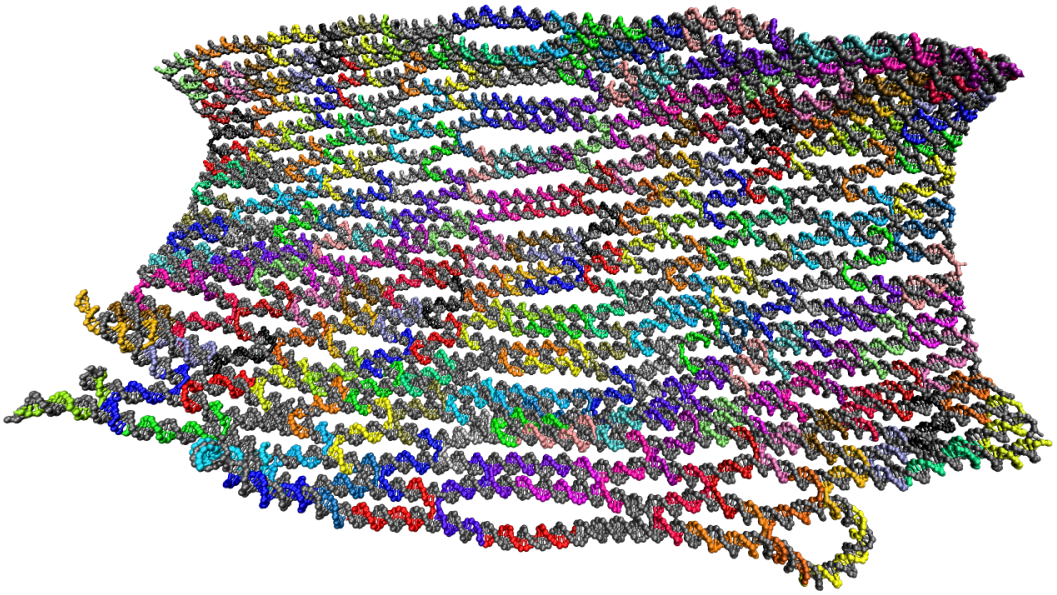

Figure 6.4: A simulation snapshot showing the 2D tile after equilibration with oxDNA2.

pairs of double helices can be divided into. In this plot one is shown as solid lines and the other as dotted lines. Each group exhibits a wave-like pattern with troughs at the crossovers, where the double helices are brought closest together, and peaks away from the crossovers, normally at a position which is both midway between the junctions and where the adjacent pair of helices have a crossover. This pattern has a periodicity of about 32 base-pair steps, corresponding to the periodic junction placement in the origami. In the middle of the plot, (around base-pair index 150), a different pattern is evident. This is due to the presence of the origami's seam (a series of junctions where the scaffold strand is exchanged, rather than staples strands as is more usual), which runs along the middle of the tile. In this region, one group of double helix pairs has a particularly large gap without any junctions and so opens up to the largest extent here, as is also very clear from Fig. 6.4; the modulations in the distance in the middle of this region reflect the presence of junctions on adjacent pairs of helices. By contrast, the other group of double helix pairs has a shorter distance between junctions due to the extra scaffold crossovers.

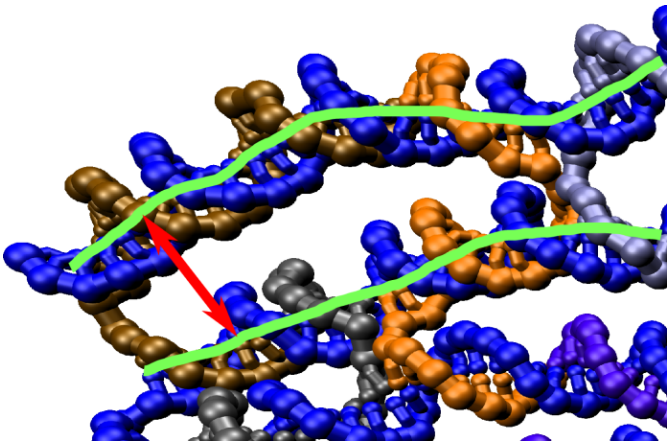

Figure 6.5: An illustration of the measurement of the weave pattern in oxDNA2. The green lines show the helix axis and the red line with arrow heads shows the distance between the axes for a particular pair of base pairs.

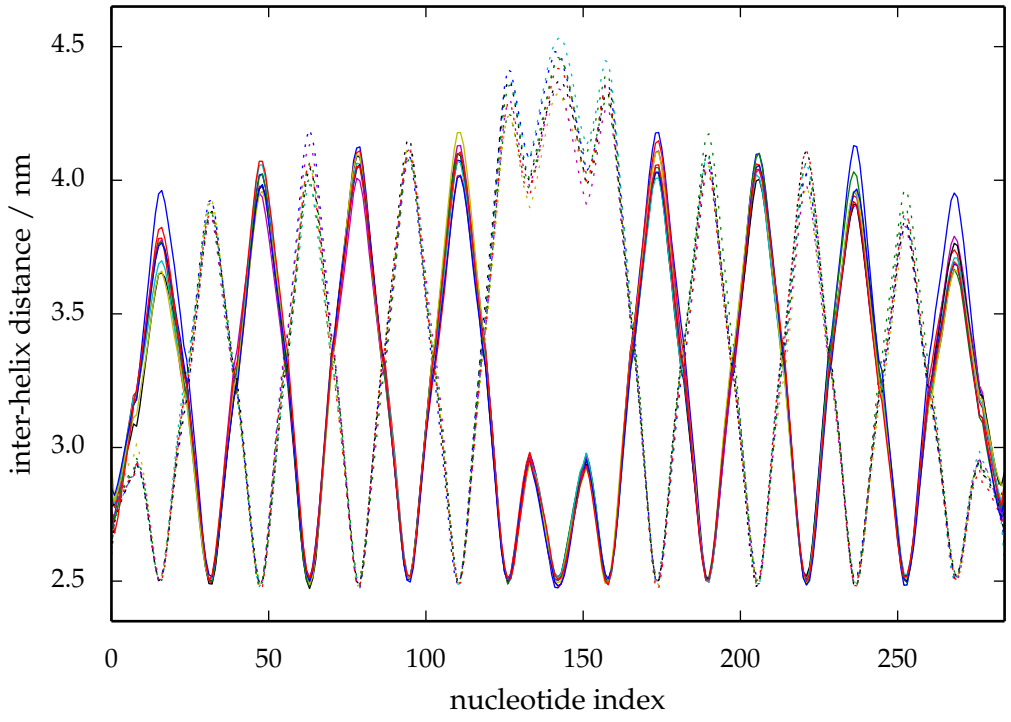

Figure 6.6: The weave pattern for the 2D tile at 300 K and $[Na^+] = 0.5$ M, quantified as the inter-helix distance as a function of base-pair index along the origami. Each line corresponds to a different pair of adjacent double helices on the origami. Some pairs have been omitted for clarity (see main text). The symmetry of the design is such that the pairs may be split into two groups: here one group is plotted with solid lines and one with dotted lines.

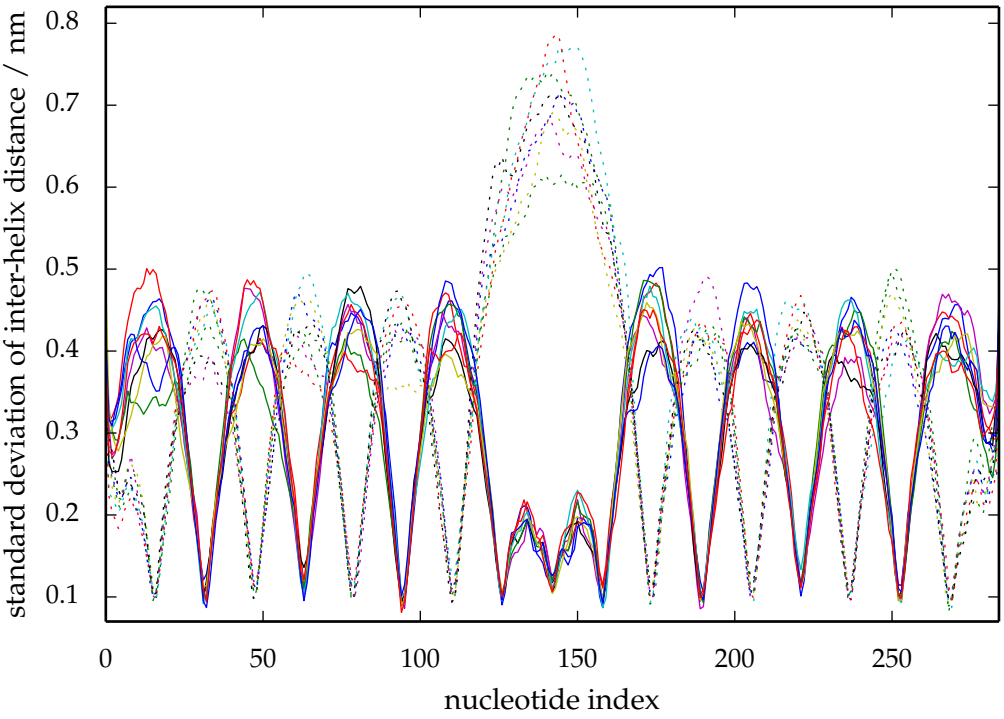

Figure 6.7: The standard deviation of the weave pattern for the 2D tile at 300 K and $[Na^+] = 0.5\,M$. The size of the standard deviation is significantly smaller than the size of the weave pattern itself. Data shown in an analogous way to Fig. 6.6.

6.3.1.1 Investigating the cause of the weave pattern in oxDNA2

From the free-energy landscape and profiles shown in Fig. 6.2 we can see that the free oxDNA2 junction is fairly flexible in θ, with the free-energy minimum occurring where θ is slightly above 0°, meaning that the junction arms tend to slightly point away from each other. This preference for a positive θ will, all else being equal, create a weave pattern, and in fact the preference is slightly stronger for values of ϕ around 180°, the set of conformations likely to be relevant for origami structure. In fact, the bending that creates the weave pattern is mostly localized at the junctions (see Fig. 6.6 and Fig. 6.8), probably due to the stiffness of the duplex sections between junctions, which being only 16 base pairs long are much shorter than the duplex persistence length (123 base pairs in oxDNA2 under these conditions), and relative flexibility of the junctions.

To assess whether the weave pattern is at least in part a static structural preference, and not purely caused by fluctuations which will tend to favour a pushing apart of the helices on average, we can examine the fluctuations in the weave pattern. These fluctuations (defined as the standard deviation of the inter-helix distance) are shown in Fig. 6.7. The plot shows that the fluctuations, which are in the range of 0.1 to 1 nm (and are smallest at the junctions and largest at the midpoints between the junctions), are significantly smaller in magnitude than the weave pattern itself, which ranges between 2.5 and 4.5 nm. Thus the junctions in the origami are very unlikely to have a value of θ near to 0, and have a static structural preference for finite θ.

In order to investigate the effect of electrostatic repulsion, we repeated the simulations of the 2D tile, this time using such a high salt concentration that the explicit electrostatic repulsion term in the oxDNA2 model is effectively removed. The result is shown in Fig. 6.9. We found that the weave pattern remained, albeit with a reduced

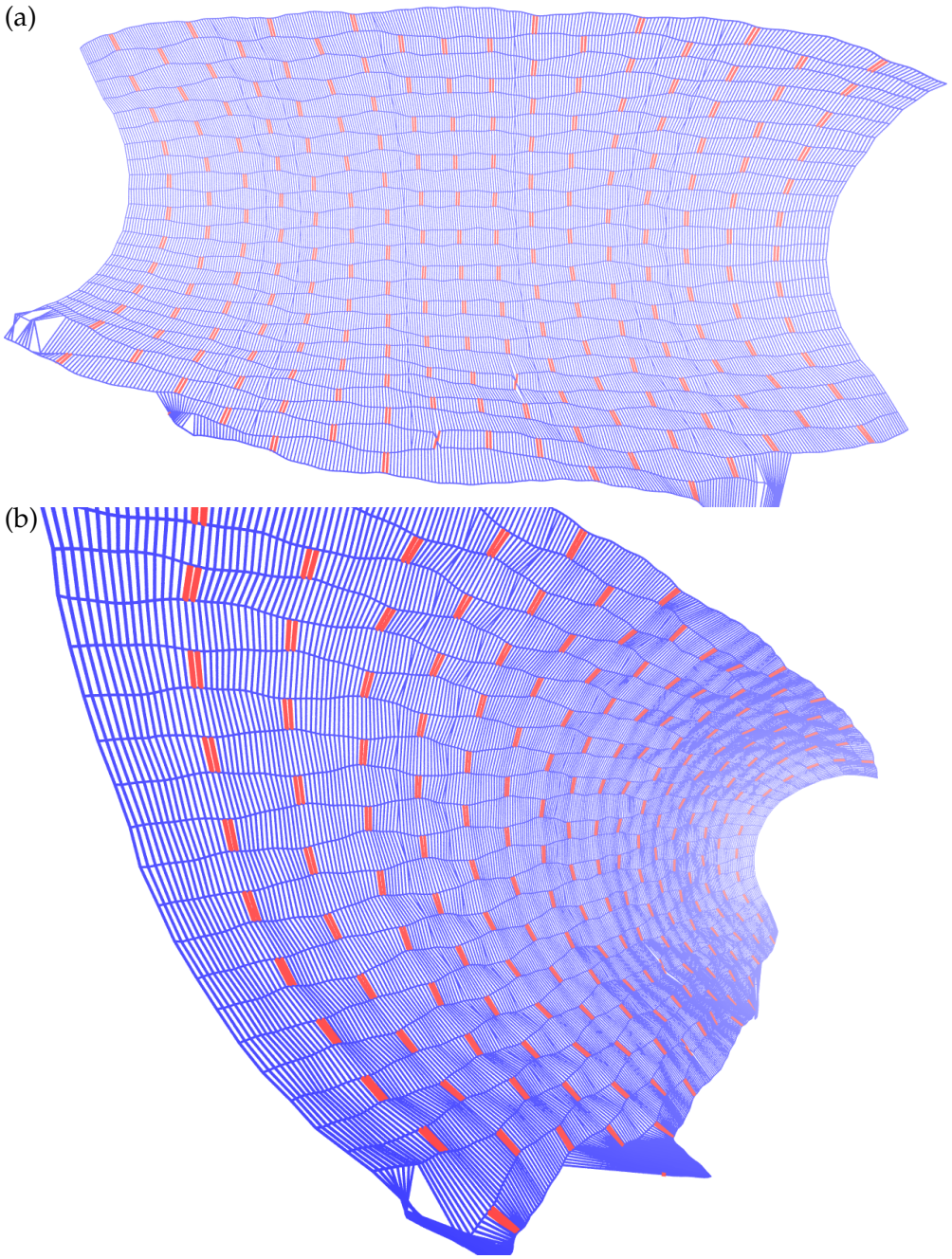

Figure 6.8: A "chickenwire" representation of an average-structure configuration for the 2D tile (see Appendix B.5.3 for details of how the average structure was computed). The tile is shown (a) from the front, as in Fig. 6.4, so that the weave pattern can be clearly seen; and (b) at an angle, to show the corrugation pattern on the tile's surface. Lines running horizontally along the origami show the central axes of the double helices that make up the origami. Pairs of red vertical lines represent double crossovers, while blue vertical lines represent the inter-helix vectors used for the quantitative analysis of the weave and corrugation, and are shown primarily to give shape to the surface. Deviations from the typical structure such as that seen in the bottom left corner of the origami are caused by staple melting or branch migration.

magnitude of the oscillations constituting the pattern. Under realistic salt conditions (typical of applications in DNA nanotechnology), $[Na^+] = 0.5\,M$, and at 300 K, the maximum separation is just above 4.5 nm, while without the electrostatic term it is just above 4.0 nm. This indicates that, although electrostatic repulsion enhances the weave pattern in oxDNA2, it is not the sole cause.

Finally, we simulated the tile at different temperatures (with $[Na^+] = 0.5\,M$) to obtain further insight into whether the weave pattern has a partly entropic, as well as energetic, origin. Fig. 6.10 shows the weave pattern at 270 K and 330 K. Together with the weave pattern at 300 K (Fig. 6.6), the plots indicate that the magnitude of the oscillations characterising the weave pattern increase somewhat with increasing temperature. Thus, thermal fluctuations play a role in determining the magnitude of the weave pattern, with this entropic component favouring a more pronounced weave pattern.

6.3.2 Corrugation in 2D origamis

A second structural property seen in 2D origamis is what we term corrugation, where the origami displays a systematic, out-of-plane bending of the double helices, so that the junctions have a ϕ angle that is not exactly 180°, as would be the case for an origami with perfectly antiparallel double helices. This results in a wave-like pattern on the surface of the origami visible for average-structure configurations, as shown in Fig. 6.8. The appearance of this effect, which is much smaller in magnitude than the weave pattern, is not surprising given the preference for the free Holliday junction to have a ϕ angle lower than 180°.

Our approach to measure the corrugation is to follow how the orientations of the

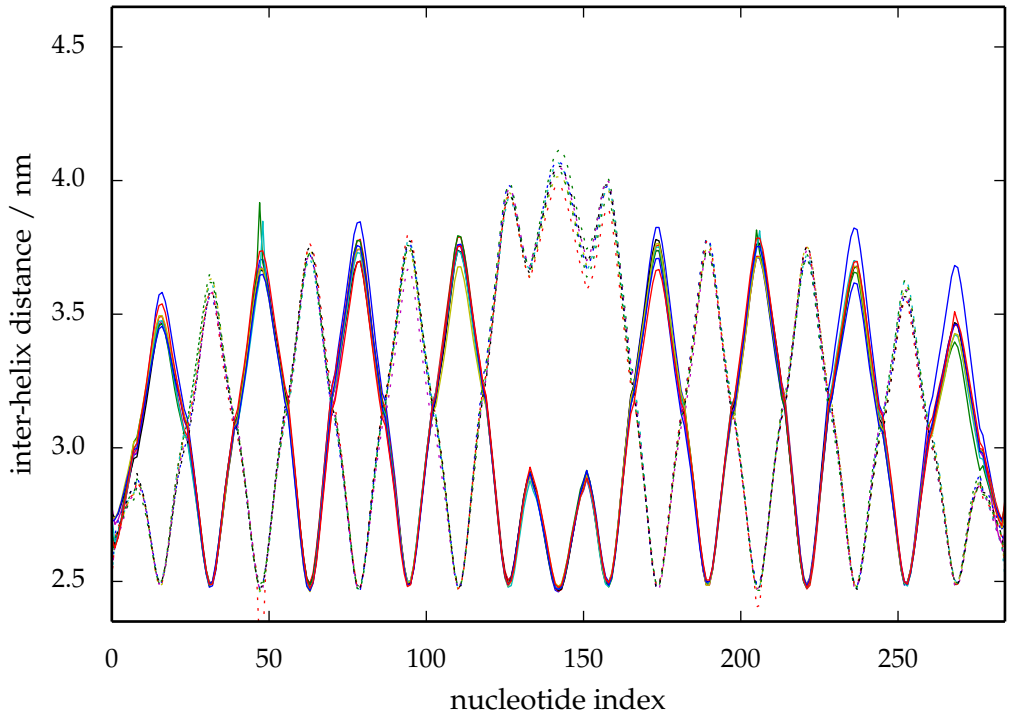

Figure 6.9: The weave pattern for the 2D tile at very high salt, which in oxDNA2 effectively corresponds to removing the electrostatic interaction. The pattern has a lower amplitude than for the $[Na^+] = 0.5\,M$ case, indicating that electrostatic repulsion enhances the weave pattern in oxDNA2.

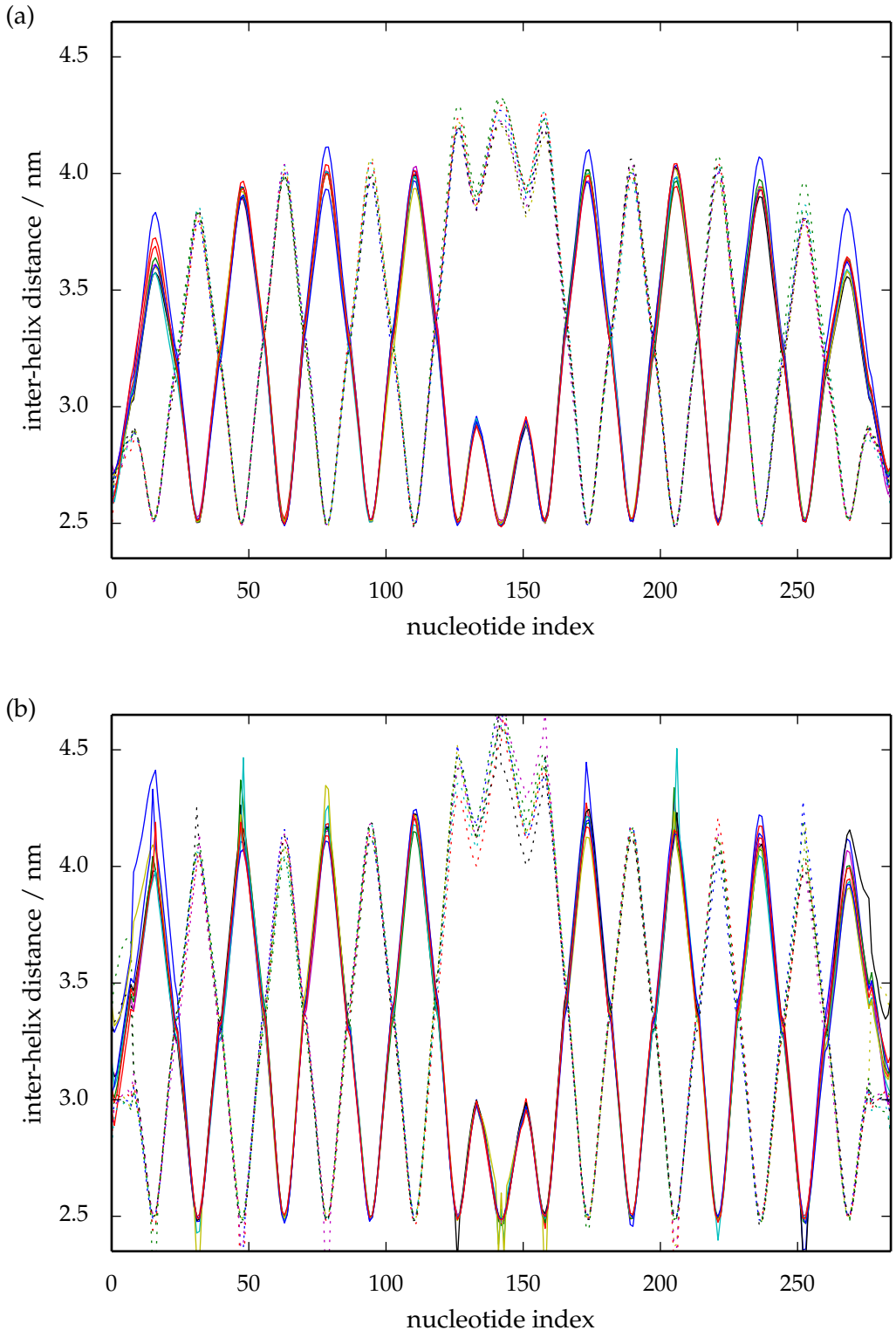

Figure 6.10: The weave pattern for the 2D tile at $[Na^+] = 0.5\,M$ at a temperature of (a) 270 K and (b) 330 K. We find that the amplitude of the pattern increases with increasing temperature. Data shown in an analogous way to Fig. 6.6.

inter-helix vectors vary as one moves away from a junction. Specifically, for every such inter-helix vector we measure the angle between the inter-helix vector and the average inter-helix vector at the nearest junction between that pair of helices when projected onto the plane perpendicular to the average helix axis at that junction. The sign of the angle is determined from the sign of the scalar triple product of the two projected inter-helix vectors with the average helix axis, in order to distinguish clockwise and anticlockwise twisting. Thus, this measure quantifies the amount of twisting, in the plane perpendicular to the helix axis, between adjacent double helices near to junctions.

The corrugation for the 2D tile as measured with this method is plotted in Fig. 6.11.[4] The plot shows the tendency for the double helices to come slightly out of the plane of the tile. They are systematically rotated one way on one side of the junction (base pair index less than 0) and the opposite way on the other (base pair index greater than 1). This corresponds to a ϕ angle of less than 180° for each junction, which is as expected from our study of the free junction in Section 6.2.1 and agrees with the findings for the structure determined in Ref. 39 with cryo-EM, which we consider in the next section. Although each junction shows qualitatively the same behaviour, there is clearly a wide variation in the magnitude of each curve. Although this will be partly due to noise from the fluctuations of the origami, which will decrease with more sampling, it seems likely that it is also due to the inhomogeneous average properties of the structure, which means that the corrugation angles are on average different around different junctions. The general shape of the curves can be understood by considering the junctions connecting a particular double helix to both of its neighbours. These are spaced approximately 16 base pairs apart, and their chiral twist causes the sign of the in- or out-of-plane modulation to change at every junction (with the magnitude passing

[4] As for the weave pattern plots, some junctions have been omitted for clarity, including the outermost junctions on the tile, and the junctions next to the scaffold seam as well as the seam itself.

through zero at the junction), as shown in Fig. 6.12.

Although, to our knowledge, this corrugation effect has not been reported in any experimental studies of 2D origamis, this is perhaps not surprising because the effect is small in magnitude and would tend to be reduced or removed when the structure is placed on a surface to be visualised, as is usually the case for experiments. A similar effect has however been observed with the CanDo structural model[54] as "out of plane bending" in a series of planar "ring" origamis. Presumably, just as with our findings, this is due to the preference of the CanDo junctions to have a ϕ angle other than 180 °.

6.4 Structure of the 3D "pointer" origami

In Chapter 5, we showed that oxDNA2 reproduces the global twist of three 3D origami structures fairly well (Table 5.1), having been fitted to the global twist of one of them. In this section, we study a 3D origami that has been well characterised experimentally, which we will refer to as the "pointer" structure, due to its resemblance, from some angles, to a pointing hand. This structure was designed, experimentally assembled, and characterised, by Bai *et al.*[39] The design uses a square lattice for the double helices, which means that the crossovers are spaced 32 base pair steps apart, and we expect a slight global twist in the structure. The authors determined the structure of the origami using cryo-EM imaging, reconstructing an electron density map and using it to fit an all-atom average structure of the origami.

The average structure computed from oxDNA2 simulations (see Appendix B.5.3 for details of the averaging procedure) is compared to the average structure calculated from cryo-EM in Fig. 6.13. The structure was simulated at [Na^+]=0.5 M, a salt concentration which roughly emulates the experimental conditions of 20 mM Mg^{2+}, 5 mM Na^+, and

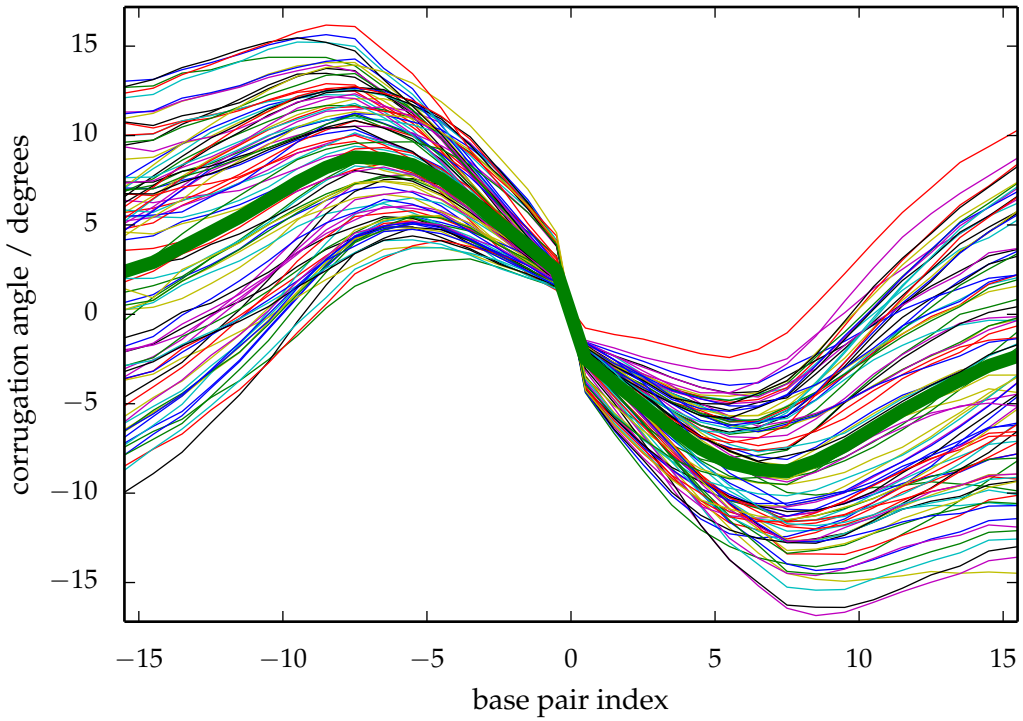

Figure 6.11: The corrugation pattern for the 2D tile. The angle between inter-helix vectors is is shown on the y-axis (see the main text for details of the definition). The x-axis shows the location of the inter-helix vector, in base-pair steps relative to the midpoint of the junction. Each line corresponds to a different junction, with the thick green line being an average over the data. Some junctions have been omitted for clarity (see main text).

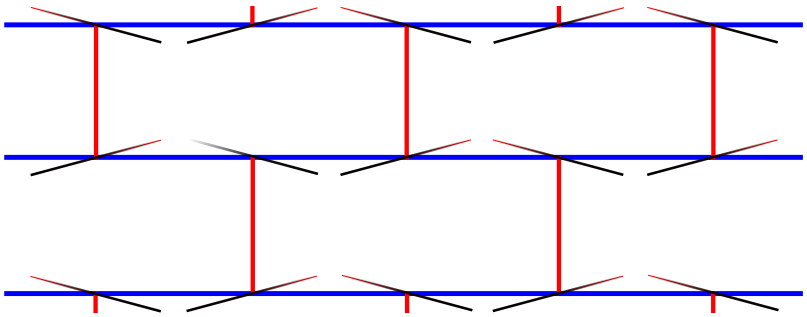

Figure 6.12: A schematic illustrating the origin of the form of the corrugation pattern. A small section of a 2D origami is shown, with blue lines representing the double helices (shown as straight for simplicity) and red lines representing the locations of the crossovers. Diagonal black lines show the preferred chirality of the junctions, with the faded end pointing into the page, and the solid end coming out of the page. In the absence of an overall curvature of the origami, the modulations in and out of the plane of the page should balance out, so that the junctions line up and the corrugation goes to zero every 16 base pairs, the number of base pairs between each junction. This is the approximate pattern seen in Fig. 6.11, albeit with a wide spread.

1 mM EDTA. By eye, there is a good agreement between simulation and experiment.

We attempt to quantify the level of agreement between our simulations and the cryo-EM structure by calculating the square root of the mean squared displacement (RMSD) between the simulated structure and the experimentally determined one. Details of the RMSD calculation are given in Appendix B.5.4. We find that the RMSD is 0.89 nm, a reasonably good agreement with experiment. A graphical comparison is shown in Fig. 6.14. In this figure, is clear that the overall size of the cross-sectional lattice computed with oxDNA2 is very close to the experimentally determined one, indicating that the size of the weave pattern and double helical radii match experiment well. In addition, the overall twist of the structure is reproduced. The majority of the contribution to the RMSD is due to the double helices towards the outside of the structure, which are more clearly displaced from the experimentally determined structure. One potential reason for this disparity is that our average structure includes the effects of thermal fluctuations at room temperature, and it is not clear to what extent these fluctuations will be frozen in during the cryo-EM process. We also note that the estimated resolution of the cryo-EM characterisation of the pointer is reported as 0.97 nm at the core and 1.4 nm at the periphery, comparable to the RMSD we have found.

We find that the 3D pointer's double helix axes trace out a left-handed helix with a period of approximately 32 base-pair steps per turn, which corresponds to the spacing of junctions between each adjacent double helix pair in this design. We can loosely think of the origin of this effect as follows: as we move along the double helix, its base pairs are pulled closer to each of its four neighbouring double helices in turn, as each junction is encountered, with a junction between our double helix and a particular neighbour occurring once every 32 base-pair steps. Oscillations periodic in the junction spacing

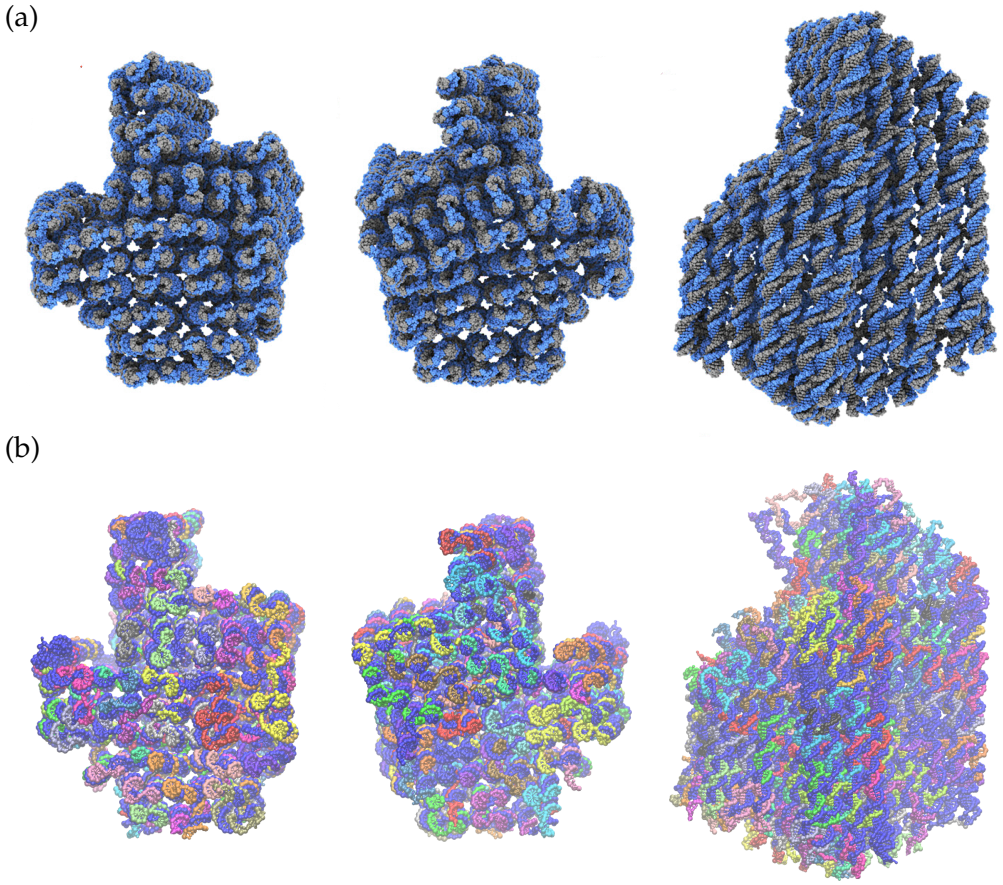

Figure 6.13: Three different views of the pointer structure determined by (a) fitting an atomistic model to cryo-EM data (reproduced with permission from Ref. 39, copyright © 2015 National Academy of Sciences) and (b) the oxDNA2 model.

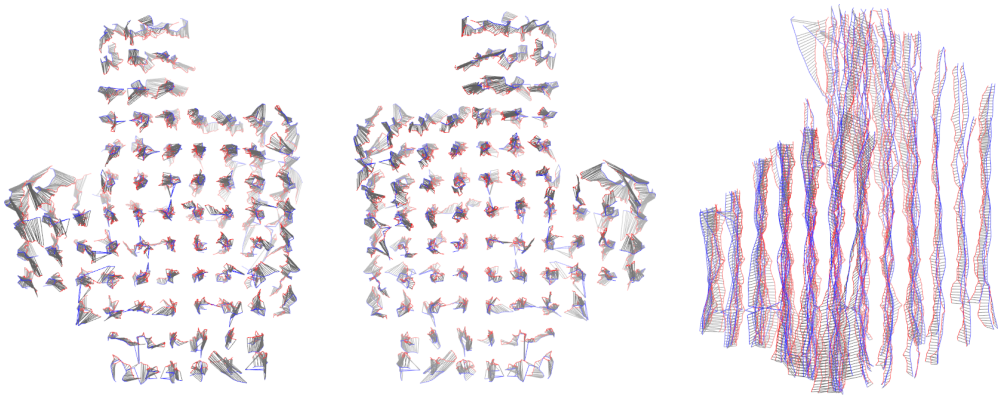

Figure 6.14: The helix axes used for the RMSD calculation are shown for three different views of the structure. The simulated structure is in blue, and the experimental one is in red. Grey lines show the displacement vectors used in the RMSD calculation.

were also seen for some structures in the atomistic origami structure study of Ref. 48.

Because the cryo-EM technique is a promising experimental method for precise structural characterisation, we may wish to improve our method of mimicking the cryo-EM procedure with oxDNA2. One way of doing this might be to add a "freezing" step to our protocol for comparing to these experimentally determined structures.

6.5 Summary

In this chapter we have characterised some of the basic properties of DNA origami structure with oxDNA2. We have seen that Holliday junction structure within origamis, a vital determinant of origami structure, is reproduced well by oxDNA2. Taking advantage of this capability, we have investigated origami structure with simulations of full-scale 2D and 3D origamis with the model. We characterised the weave pattern for a 2D origami and determined its origin within the model, and quantified the subtle corrugation effect. For the 3D pointer origami, we noted the left-handed helix formed by the axes of the double helices, and quantified the agreement between oxDNA2's average structure for the pointer with the experimental structure determined with cryo-EM.

Having completed this basic investigation of origami structural properties within the model, the good agreement with experiment that we have found provides us with confidence to use the model to study the structural properties of other large DNA nanostructures, and origamis in particular. Structures that have been carefully characterised experimentally will give further opportunities to test and refine the structural predictions of the model, while for nanostructures that have only been visualised using low-resolution methods, such as TEM on a surface, oxDNA2 has the potential to

provide more detailed structural insights. The model will be of particular use when a more detailed view than CanDo can provide is needed, for structures with complex features that cannot easily be handled by CanDo, or when internal stresses might lead to broken base pairing.

7 Conclusion

In this work we have used the coarse-grained DNA model oxDNA to study the highly successful DNA origami self-assembly method and, having introduced an enhanced version of the model, the structural properties of the example two- and three-dimensional origamis.

We studied the self-assembly of a 384-base-pair origami from a solution of single-stranded scaffold and staple strands using direct MD simulations. We saw that staple strands bound to the scaffold in parallel, with cooperativity evident for the binding of the second staple domain if the adjacent staple was already partially formed. We were able to simulate a complete assembly of the origami for a system with exactly one copy of each strand. This assembly occurred in an intermediate temperature window: at low temperatures assembly was prevented by misbonding while at high temperatures the free-energy barriers to assembly became too large for our simulation time scales. When we attempted an assembly with a large staple excess and a high strand concentration, completion of the assembly was prevented because of kinetic traps caused by multiple copies of the same staple strand binding simultaneously to different domains on the scaffold, preventing each other from completing their binding. This last effect helped us to rationalise some successful origami design strategies.

We then studied the self-assembly of a similar origami with oxDNA using FFS, a rare-event simulation technique. Using FFS made simulations possible at experimental concentrations, which are three orders of magnitude lower than the concentrations that

we used in our direct simulations, and at higher temperatures where binding rates are lower and staple melting is enhanced. At these concentrations we found a clear time scale separation between the binding rate for the initial domain of each staple and the binding rate for its subsequent domains, although this separation may be reduced for more complex origami designs, so that the staple blocking effect identified for high-concentration systems with direct simulation may still come into play. As with the direct assembly simulations, we found the strongest cooperative effect for the binding of the fourth staple domain at a Holliday junction. We also found evidence of reduced domain formation and melting rates for transitions which required a staple strand tail to be threaded through a small loop, while the potentially topologically difficult problem of binding a middle staple domain when the end domains were already bound surprisingly occurred perhaps even faster than the typical rate for binding a third staple domain. When a staple had bound an end domain, subsequently binding the non-adjacent domain at the other end of the staple proved to be significantly slower than binding the middle domain, which is probably the main reason why binding the second staple domain was faster when the first domain that bound was the middle domain rather than an end domain. All of these findings provide hints of how rational design can guide origami self-assembly along desirable pathways, avoiding potential strand threading or wrapping problems which can drastically reduce yields.

We also introduced oxDNA2, an enhanced version of the oxDNA model. The model was previously parameterised to one salt concentration, a constraint which we relaxed by adding a salt-dependent term allowing treatment of salt conditions of $[Na^+] = 0.1$ to 2 M. We used new experimental data to add additional sequence dependence into the model's thermodynamics, introducing different interaction strengths for AA stacking and TT stacking, an important distinction for DNA biophysics and nanotechnology.

Additionally, we improved the model's treatment of large-scale DNA nanostructures by introducing differentiated widths for the major and minor DNA double helical grooves, and modifying the coaxial stacking and backbone-backbone interactions to fit the global twist of an origami to an experimentally determined value.

We used the improvements introduced with oxDNA2 to study general structural properties of 2D and 3D origamis, comparing to experiment where possible. We investigated Holliday junctions, an important element of origami structure, in some detail, finding that the behaviour of oxDNA2 junctions within origamis agrees with that determined by experiment. We quantified the weave pattern in 2D origamis, which in the model results from the preferred local structure of the Holliday junctions, electrostatic repulsion between backbones, and thermal fluctuations. We also characterised the subtle corrugation effect in 2D origamis. We studied a 3D origami which has been characterised in detail experimentally, finding that each double helix axis formed a left-handed helix due to the regular pattern of junction placement. We found a good agreement with experiment in the overall size and shape of the structure, with an RMSD between the simulated and experimental structure of 0.89 nm.

Future simulations with oxDNA, with the help of FFS or perhaps a different rare-event method, have the potential to yield further insights into origami assembly. The studies we have conducted here have highlighted likely fruitful areas for further investigation, such as potential problems with strand threading and staple blocking, as well the implications of cooperative effects during Holliday junction formation. Detailed studies of smaller model systems, for which accurate statistics can be more easily obtained, will be particularly important in this regard.

The simple treatment of electrostatic interactions we have used for the oxDNA2

model leads to a disagreement with experimental hairpin melting data at low salt, suggesting that the electrostatic repulsion between DNA single strands at low salt may be too strong. This is a potential area for future improvement of the model.

Having taken advantage of the improved structural prediction capabilities introduced with oxDNA2 to conduct an investigation into the basic structural properties of origamis, we can now use the model to gain structural insights into a wide range of origamis and other DNA nanostructures. There is much interest in this area, with significant opportunity for collaborations with experimental groups. The insights that oxDNA can provide will be particularly useful for nanostructures that have a significant range of dynamical motion and that are harder to characterise experimentally. At the same time, further detailed experimental characterisation of DNA nanostructures will increase our confidence in the model's predictions and allow us to make additional improvements to the model in the future.

References

[1] N C Seeman, 'Nucleic acid junctions and lattices', *J Theor Biol*, **99**, 237–247, 1982.

[2] V Linko and H Dietz, 'The enabled state of DNA nanotechnology', *Curr Opin Biotech*, **24**, 555–561, 2013.

[3] P W K Rothemund, 'Folding DNA to create nanoscale shapes and patterns.', *Nature*, **440**, 297–302, 2006.

[4] Y Ke, L L Ong, W M Shih and P Yin, 'Three-dimensional structures self-assembled from DNA bricks', *Science*, **338**, 1177–1183, 2012.

[5] E Benson, A Mohammed, J Gardell, S Masich, E Czeizler, P Orponen and B Högberg, 'DNA rendering of polyhedral meshes at the nanoscale', *Nature*, **523**, 441–444, 2015.

[6] M Matthies and T L Schmidt, 'Design and synthesis of triangulated origami trusses', in review.

[7] F Zhang, S Jiang, S Wu, Y Li, C Mao, Y Liu and H Yan, 'Complex wireframe DNA origami nanostructures with multi-arm junction vertices', *Nat Nano*, **10**, 779–784, 2015.

[8] S M Douglas, H Dietz, T Liedl, B Högberg, F Graf and W M Shih, 'Self-assembly of DNA into nanoscale three-dimensional shapes', *Nature*, **459**, 414–418, 2009.

[9] H Dietz, S M Douglas and W M Shih, 'Folding DNA into twisted and curved nanoscale shapes', *Science*, **325**, 725–730, 2009.

[10] D Han, S Pal, J Nangreave, Z Deng, Y Liu and H Yan, 'DNA origami with complex curvatures in three-dimensional space', *Science*, **332**, 342–346, 2011.

[11] T Liedl, B Högberg, J Tytell, D E Ingbe and W M Shih, 'Self-assembly of three-dimensional prestressed tensegrity structures from DNA', *Nature Nanotech*, **5**, 520–524, 2010.

[12] C E Castro, F Kilchherr, D-N Kim, E L Shiao, T Wauer, P Wortmann, M Bathe and H Dietz, 'A primer to scaffolded DNA origami', *Nat Methods*, **8**, 221–229, 2011.

[13] A V Pinheiro, D Han, W M Shih and H Yan, 'Challenges and opportunities for structural DNA nanotechnology', *Nat Nano*, **6**, 763–772, 2011.

[14] Y-J Chen, B Groves, R A Muscat and G Seelig, 'DNA nanotechnology from the test tube to the cell', *Nat Nano*, **10**, 748–760, 2015.

[15] A Rajendran, M Endo and H Sugiyama, 'Single-molecule analysis using DNA origami', *Angew Chem Int Edit*, **51**, 874–890, 2012.

[16] R Tsukanov, T E Tomov, R Masoud, H Drory, N Plavner, M Liber and E Nir, 'Detailed study of DNA hairpin dynamics using single-molecule fluorescence assisted by DNA origami', *J Phys Chem B*, **117**, 11932–11942, 2013.

[17] S F J Wickham, M Endo, Y Katsuda, K Hidaka, J Bath, H Sugiyama and A J Turberfield, 'Direct observation of stepwise movement of a synthetic molecular transporter', *Nature Nanotech*, **6**, 166–169, 2011.

[18] Q Jiang, C Song, J Nangreave, X Liu, L Lin, D Qiu, Z-G Wang, G Zou, X Liang, H Yan and B Ding, 'DNA origami as a carrier for circumvention of drug resistance', *J Am Chem Soc*, **134**, 13396–13403, 2012.

[19] S M Douglas, I Bachelet and G M Church, 'A logic-gated nanorobot for targeted transport of molecular payloads', *Science*, **335**, 831–834, 2012.

[20] M Langecker, V Arnaut, T G Martin, J List, S Renner, M Mayer, H Dietz and F C Simmel, 'Synthetic lipid membrane channels formed by designed DNA nanostructures', *Science*, **338**, 932–936, 2012.

[21] J R Burns, K GÃűpfrich, J W Wood, V V Thacker, E Stulz, U F Keyser and S Howorka, 'Lipid-bilayer-spanning DNA nanopores with a bifunctional porphyrin anchor', *Angew Chem Int Edit*, **52**, 12069–12072, 2013.

[22] N A W Bell, C R Engst, M Ablay, G Divitini, C Ducati, T Liedl and U F Keyser, 'DNA origami nanopores', *Nano Lett*, **12**, 512–517, 2012.

[23] E Pfitzner, C Wachauf, F Kilchherr, B Pelz, W M Shih, M Rief and H Dietz, 'Rigid DNA beams for high-resolution single-molecule mechanics', *Angew Chem Int Edit*, **52**, 7766–7771, 2013.

[24] J Cheng, S Sreelatha, R Hou, A Efremov, R Liu, J R C van der Maarel and Z Wang, 'Bipedal nanowalker by pure physical mechanisms', *Phys Rev Lett*, **109**, 238104, 2012.

[25] T E Tomov, R Tsukanov, M Liber, R Masoud, N Plavner and E Nir, 'Rational design of DNA motors: Fuel optimization through single-molecule fluorescence', *J Am Chem Soc*, **135**, 11935–11941, 2013.

[26] L Adleman, 'Molecular computation of solutions to combinatorial problems', *Science*, **266**, 1021–1024, 1994.

[27] L Qian, E Winfree and J Bruck, 'Neural network computation with DNA strand displacement cascades', *Nature*, **475**, 368–372, 2011.

[28] T Ran, S Kaplan and E Shapiro, 'Molecular implementation of simple logic programs', *Nat Nano*, **4**, 642–648, 2009.

[29] N Goldman, P Bertone, S Chen, C Dessimoz, E M LeProust, B Sipos and E Birney, 'Towards practical, high-capacity, low-maintenance information storage in synthesized DNA', *Nature*, **494**, 77–80, 2013.

[30] B Yurke, A J Turberfield, A P Mills, F C Simmel and J Neumann, 'A DNA-fuelled molecular machine made of DNA', *Nature*, **406**, 605–608, 2000.

[31] A S Walsh, H Yin, C M Erben, M J A Wood and A J Turberfield, 'DNA cage delivery to mammalian cells', *ACS Nano*, **5**, 5427–5432, 2011.

[32] H Gradišar, S Božič, T Doles, D Vengust, I Hafner-Bratkovič, A Mertelj, B Webb, A Šali, S Klavžar and R Jerala, 'Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments', *Nat Chem Biol*, **9**, 362–366, 2013.

[33] C Tian, C Zhang, X Li, Y Li, G Wang and C Mao, 'Artificial, parallel, left-handed DNA helices', *J Am Chem Soc*, **134**, 20273–20275, 2012.

[34] J-P J Sobczak, T G Martin, T Gerling and H Dietz, 'Rapid folding of DNA into nanoscale shapes at constant temperature', *Science*, **338**, 1458–1461, 2012.

[35] Y Ke, G Bellot, N V Voigt, E Fradkov and W M Shih, 'Two design strategies for enhancement of multilayer–DNA-origami folding: underwinding for specific intercalator rescue and staple-break positioning', *Chem Sci*, **3**, 2587–2597, 2012.

[36] K F Wagenbauer, C H Wachauf and H Dietz, 'Quantifying quality in DNA self-assembly', *Nat Commun*, **5**, 3691, 2014.

[37] X Wei, J Nangreave and Y Liu, 'Uncovering the self-assembly of DNA nanostructures by thermodynamics and kinetics', *Acc Chem Res*, **47**, 1861–1870, 2014.

[38] K E Dunn, F Dannenberg, T E Ouldridge, M Kwiatkowska, A J Turberfield and J Bath, 'Guiding the folding pathway of DNA origami', *Nature*, **525**, 82–86, 2015.

[39] X-C Bai, T G Martin, S H W Scheres and H Dietz, 'Cryo-EM structure of a 3D DNA-origami object', *Proc Nat Acad Sci USA*, **109**, 20012–20017, 2012.

[40] R Jungmann, M S Avendano, J B Woehrstein, M Dai, W M Shih and P Yin, 'Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT', *Nat Methods*, **11**, 313–318, 2014.

[41] D Svozil, P Hobza and J Šponer, 'Comparison of intrinsic stacking energies of ten unique dinucleotide steps in A-RNA and B-DNA duplexes. Can we determine correct order of stability by quantum-chemical calculations?', *J Phys Chem B*, **114**, 1191–1203, 2010.

[42] A Mladek, M Krepl, D Svozil, P Cech, M Otyepka, P Banas, M Zgarbova, P Jurecka and J Šponer, 'Benchmark quantum-chemical calculations on a complete set of rotameric families of the DNA sugar-phosphate backbone and their comparison with modern density functional theory', *Phys Chem Chem Phys*, **15**, 7295–7310, 2013.

[43] J Šponer, J E Šponer, A Mládek, P Banáš, P Jurečka and M Otyepka, 'How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists', *Methods*, **64**, 3 – 11, 2013.

[44] C A Laughton and S A Harris, 'The atomistic simulation of DNA', *Wiley Interdisc Rev Comput Mol Sc*, **1**, 590–600, 2011.

[45] A Pérez, F J Luque and M Orozco, 'Frontiers in molecular dynamics simulations of DNA', *Acc Chem Res*, **45**, 196–205, 2012.

[46] F Oteri, M Falconi, G Chillemi, F F Andersen, C L Oliveira, J S Pedersen, B R Knudsen and A Desideri, 'Simulative analysis of a truncated octahedral DNA nanocage family indicates the single-stranded thymidine linkers as the major player for the conformational variability', *J Phys Chem C*, **115**, 16819–16827, 2011.

[47] C Maffeo, B Luan and A Aksimentiev, 'End-to-end attraction of duplex DNA', *Nucleic Acids Res*, **40**, 3812–3821, 2012.

[48] J Yoo and A Aksimentiev, 'In situ structure and dynamics of DNA origami determined through molecular dynamics simulations', *Proc Natl Acad Sci USA*, **110**, 20099–20104, 2013.

[49] C Maffeo, J Yoo, J Comer, D B Wells, B Luan and A Aksimentiev, 'Close encounters with DNA', *Journal of Physics: Condensed Matter*, **26**, 413101, 2014.

[50] C Bustamante, J F Marko, E D Siggia and S Smith, 'Entropic elasticity of λ-phage DNA', *Science*, **265**, 1599–1600, 1994.

[51] J M Arbona, J Elezgaray and J-P Aimé, 'Modelling the folding of DNA origami', *EPL*, **100**, 28006, 2012.

[52] J M Arbona, J-P Aimé and J Elezgaray, 'Modeling the mechanical properties of DNA nanostructures', *Phys Rev E*, **86**, 051912, 2012.

[53] D-N Kim, F Kilchherr, H Dietz and M Bathe, 'Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures', *Nucleic Acids Res*, **40**, 2862–2868, 2012.

[54] K Pan, D-N Kim, F Zhang, M R Adendorff, H Yan and M Bathe, 'Lattice-free prediction of three-dimensional structure of programmed DNA assemblies', *Nat Comm*, **5**, 5578, 2014.

[55] A Savelyev and G A Papoian, 'Chemically accurate coarse graining of double-stranded DNA', *Proc Natl Acad Sci USA*, **107**, 20340–20345, 2010.

[56] Q Cao, C Zuo, Y Ma, L Li and Z Zhang, 'Interaction of double-stranded DNA with a nanosphere: a coarse-grained molecular dynamics simulation study', *Soft Matter*, **7**, 506–514, 2011.

[57] N Korolev, D Luo, A P Lyubartsev and L Nordenskiöld, 'A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo', *Polymers*, **6**, 1655–1675, 2014.

[58] M C Linak, R Tourdot and K D Dorfman, 'Moving beyond Watson–Crick models of coarse grained DNA dynamics', *J Chem Phys*, **135**, 205102, 2011.

[59] J C Araque, A Z Panagiotopoulos and M A Robert, 'Lattice model of oligonucleotide hybridization in solution. I. Model and thermodynamics', *J Chem Phys*, **134**, 165103, 2011.

[60] A Morriss-Andrews, J Rottler and S S Plotkin, 'A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist and chirality', *J Chem Phys*, **132**, 035105, 2010.

[61] C Maffeo, T T M Ngo, T Ha and A Aksimentiev, 'A coarse-grained model of unstructured single-stranded DNA derived from atomistic simulation and single-molecule experiment', *J Chem Theory Comput*, **10**, 2891–2896, 2014.

[62] T Cragnolini, P Derreumaux and S Pasquali, 'Coarse-grained simulations of RNA and DNA duplexes', *J Phys Chem B*, **117**, 8047–8060, 2013.

[63] D M Hinckley, G S Freeman, J K Whitmer and J J de Pablo, 'An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization', *J Chem Phys*, **139**, 144903, 2013.

[64] V Ortiz and J J de Pablo, 'Molecular origins of DNA flexibility: Sequence effects on conformational and mechanical properties', *Phys Rev Lett*, **106**, 238107, 2011.

[65] D M Hinckley, J P Lequieu and J J de Pablo, 'Coarse-grained modeling of DNA oligomer hybridization: Length, sequence, and salt effects', *J Chem Phys*, **141**, 035102, 2014.

[66] T E Ouldridge, A A Louis and J P K Doye, 'Structural, mechanical and thermodynamic properties of a coarse-grained model of DNA', *J Chem Phys*, **134**, 085101, 2011.

[67] T E Ouldridge, 'Coarse-grained modelling of DNA and DNA self-assembly', *PhD Thesis, University of Oxford*, 2011.

[68] T E Ouldridge, A A Louis and J P K Doye, 'DNA nanotweezers studied with a coarse-grained model of DNA', *Phys Rev Lett*, **104**, 178101, 2010.

[69] P Šulc, T E Ouldridge, F Romano, J P K Doye and A A Louis, 'Simulating a burnt-bridges DNA motor with a coarse-grained DNA model', *Nat Comput*, **13**, 535–547, 2014.

[70] T E Ouldridge, P Šulc, F Romano, J P K Doye and A A Louis, 'DNA hybridization kinetics: zippering, internal displacement and sequence dependence', *Nucleic Acids Res*, **41**, 8886–8895, 2013.

[71] T E Ouldridge, R L Hoare, A A Louis, J P K Doye, J Bath and A J Turberfield, 'Optimizing DNA nanotechnology through coarse-grained modeling: A two-footed DNA walker', *ACS Nano*, **7**, 2479–2490, 2013.

[72] L Rovigatti, F Smallenburg, F Romano and F Sciortino, 'Gels of DNA nanostars never crystallize', *ACS Nano*, **8**, 3567–3574, 2014.

[73] J S Schreck, T E Ouldridge, F Romano, A A Louis and J P K Doye, 'Characterizing the bending and flexibility induced by bulges in DNA duplexes', *J Chem Phys*, **142**, 165101, 2015.

[74] J S Schreck, T E Ouldridge, F Romano, P Šulc, L Shaw, A A Louis and J P K Doye, 'DNA hairpins destabilize duplexes primarily by promoting melting rather than by inhibiting hybridization', *Nucleic Acids Res*, **43**, 6181–6190, 2015.

[75] N Srinivas, T E Ouldridge, P Šulc, J M Schaeffer, B Yurke, A A Louis, J P K Doye and E Winfree, 'On the biophysics and kinetics of toehold-mediated DNA strand displacement', *Nucleic Acids Res*, **41**, 10641–10658, 2013.

[76] M Mosayebi, A A Louis, J P K Doye and T E Ouldridge, 'Force-induced rupture of a DNA duplex', *arXiv:150203623*, 2015.

[77] C Matek, T E Ouldridge, A Levy, J P K Doye and A A Louis, 'DNA cruciform arms nucleate through a correlated but asynchronous cooperative mechanism', *J Phys Chem B*, **116**, 11616–11625, 2012.

[78] Q Wang and B M Pettitt, 'Modeling DNA thermodynamics under torsional stress', *Biophys J*, **106**, 1182–1193, 2014.

[79] C Matek, T E Ouldridge, J P K Doye and A A Louis, 'Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA', *Sci Rep*, **5**, 7655, 2015.

[80] M Mosayebi, F Romano, T E Ouldridge, A A Louis and J P K Doye, 'The role of loop stacking in the dynamics of DNA hairpin formation', *J Phys Chem B*, **118**, 14326–14335, 2014.

[81] F Romano, D Chakraborty, J P K Doye, T E Ouldridge and A A Louis, 'Coarse-grained simulations of DNA overstretching', *J Chem Phys*, **138**, 085101, 2013.

[82] J P K Doye, T E Ouldridge, A A Louis, F Romano, P Šulc, C Matek, B E K Snodin, L Rovigatti, J S Schreck, R M Harrison and W P J Smith, 'Coarse-graining DNA for simulations of DNA nanotechnology', *Phys Chem Chem Phys*, **15**, 20395–20414, 2013.

[83] R R F Machinek, T E Ouldridge, N E C Haley, J Bath and A J Turberfield, 'Programmable energy landscapes for kinetic control of DNA strand displacement', *Nat Comm*, **5**, 5324, 2014.

[84] L Rovigatti, P Šulc, I Z Reguly and F Romano, 'A comparison between parallelization approaches in molecular dynamics simulations on GPUs', *J Comput Chem*, **36**, 1–8, 2015.

[85] E F Pettersen, T D Goddard, C C Huang, G S Couch, D M Greenblatt, E C Meng and T E Ferrin, 'UCSF Chimera – A visualization system for exploratory research and analysis', *J Comput Chem*, **25**, 1605–1612, 2004.

[86] D Frenkel and B Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 1996.

[87] S Whitelam and P L Geissler, 'Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles', *J Chem Phys*, **127**, 154101, 2007.

[88] S Whitelam and P L Geissler, 'Erratum: "Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles" [J. Chem. Phys.127, 154101 (2007)]', *J Chem Phys*, **128**, 219901, 2008.

[89] B E K Snodin, F Randisi, M Mosayebi, P Šulc, J S Schreck, F Romano, T E Ouldridge, R Tsukanov, E Nir, A A Louis and J P K Doye, 'Introducing improved structural properties and salt dependence into a coarse-grained model of DNA', *J Chem Phys*, **142**, 234901, 2015.

[90] P Šulc, F Romano, T E Ouldridge, J P K Doye and A A Louis, 'A nucleotide-level coarse-grained model of RNA', *J Chem Phys*, **140**, 235102, 2014.

[91] B E K Snodin, F Romano, L Rovigatti, T E Ouldridge, A A Louis and J P K Doye, 'Direct simulation of the self-assembly of a small DNA origami', *ACS Nano*, **10**, 1724–1737, 2016.

[92] W C Swope, H C Andersen, P H Berens and K R Wilson, 'A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters', *J Chem Phys*, **76**, 637–649, 1982.

[93] J A Anderson, E Jankowski, T L Grubb, M Engel and S C Glotzer, 'Massively parallel Monte Carlo for many-particle simulations on GPUs', *J Comput Phys*, **254**, 27–38, 2013.

[94] N Metropolis, A W Rosenbluth, M N Rosenbluth, A H Teller and E Teller, 'Equation of state calculation by fast computing machines.', *J Chem Phys*, **21**, 1087–1092, 1953.

[95] R J Allen, C Valeriani and P R ten Wolde, 'Forward flux sampling for rare event simulations', *J Phys: Condens Matter*, **21**, 463102, 2009.

[96] L Cademartiri and K J M Bishop, 'Programmable self-assembly', *Nature Materials*, **14**, 2–9, 2015.

[97] P Šulc, F Romano, T E Ouldridge, L Rovigatti, J P K Doye and A A Louis, 'Sequence-dependent thermodynamics of a coarse-grained DNA model', *J Chem Phys*, **137**, 135101, 2012.

[98] T Murtola, A Bunker, I Vattulainen, M Deserno and M Karttunen, 'Multiscale modeling of emergent materials: biological and soft matter', *Phys Chem Chem Phys*, **11**, 1869–1892, 2009.

[99] Y Gao, L K Wolf and R M Georgiadis, 'Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison', *Nucleic Acids Res*, **34**, 3370–3377, 2006.

[100] L E Morrison and L M Stols, 'Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution', *Biochemistry*, **32**, 3095–3104, 1993.

[101] D Y Zhang and E Winfree, 'Control of DNA strand displacement kinetics using toehold

exchange', *J Am Chem Soc*, **131**, 17303–17314, 2009.

[102] E M Craig, M D Crothers and P Doty, 'Relaxation kinetics of dimer formation by self complementary oligonucleotides', *J Mol Biol*, **62**, 383–392, 1971.

[103] D Pörschke and M Eigen, 'Co-operative non-enzymatic base recognition III. Kinetics of the helix-coil transition of the oligoribouridylic – oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH', *J Mol Biol*, **62**, 361–364, 1971.

[104] D Pörschke, O C Uhlenbeck and F H Martin, 'Thermodynamics and kinetics of the helix-coil transition of oligomers containing GC base pairs', *Biopolymers*, **12**, 1313–1335, 1973.

[105] C Chen, W Wang, Z Wang, F Wei and X S Zhao, 'Influence of secondary structure on kinetics and reaction mechanism of DNA hybridization', *Nucleic Acids Res*, **35**, 2875–2884, 2007.

[106] F Dannenberg, K E Dunn, J Bath, M Kwiatkowska, A J Turberfield and T E Ouldridge, 'Modelling DNA origami self-assembly at the domain level', *arXiv:150903066*, 2015.

[107] A J Genot, D Y Zhang, J Bath and A J Turberfield, 'Remote toehold: a mechanism for flexible control of DNA hybridization kinetics', *J Am Chem Soc*, **133**, 2177–2182, 2011.

[108] F Romano, A Hudson, J P K Doye, T E Ouldridge and A A Louis, 'The effect of topology on the structure and free energy landscape of DNA kissing complexes', *J Chem Phys*, **136**, 215102, 2012.

[109] B K Müller, A Reuter, F C Simmel and D C Lamb, 'Single-pair FRET characterization of DNA tweezers', *Nano Lett*, **6**, 2814–2820, 2006.

[110] W Bae, K Kim, D Min, J-K Ryu, C Hyeon and T-Y Yoon, 'Programmed folding of DNA origami structures through single-molecule force control', *Nat Comm*, **5**, 5654, 2014.

[111] S M Douglas, A H Marblestone, S Teerapittayanon, A Vazquez, G M Church and W M Shih, 'Rapid prototyping of 3D DNA-origami shapes with caDNAno', *Nucleic Acids Res*, **37**, 5001–5006, 2009.

[112] J SantaLucia, Jr and D Hicks, 'The thermodynamics of DNA structural motifs', *Annu Rev Biophys Biomol Struct*, **33**, 415–40, 2004.

[113] A Shapiro, A Hozeh, O Girshevitz, A Abu-Horowitz and I Bachelet, 'Cooperativity-based modeling of heterotypic DNA nanostructure assembly', *Nucleic Acids Res*, **43**, 6587–6595, 2015.

[114] V Kočar, J S Schreck, S Čeru, H Gradišar, N Bašić, T Pisanski, J P K Doye and R Jerala, 'Design principles for folding modular single-chain DNA into nanostructures', submitted.

[115] R Tsukanov, T E Tomov, Y Berger, M Liber and E Nir, 'Conformational dynamics of DNA hairpins at millisecond resolution obtained from analysis of single-molecule FRET histograms', *J Phys Chem B*, **117**, 16105–16109, 2013.

[116] A Bosco, J Camunas-Soler and F Ritort, 'Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions', *Nucleic Acids Res*, **42**, 2064–2074, 2013.

[117] M E Polinkovsky, Y Gambin, P R Banerjee, M J Erickstad, A Groisman and A A Deniz, 'Ultrafast cooling reveals microsecond-scale biomolecular dynamics', *Nat Comm*, **5**, 5737, 2014.

[118] W H Taylor and P J Hagerman, 'Application of the method of phage T4 DNA ligase-catalyzed ring-closure to the study of DNA structure: II. NaCl-dependence of DNA flexibility and helical repeat', *J Mol Biol*, **212**, 363 – 376, 1990.

[119] F Mocci and A Laaksonen, 'Insight into nucleic acid counterion interactions from inside molecular dynamics simulations is "worth its salt"', *Soft Matter*, **8**, 9268–9284, 2012.

[120] D A Potoyan, A Savelyev and G A Papoian, 'Recent successes in coarse-grained modeling of DNA', *WIREs Comput Mol Sci*, **3**, 69–83, 2013.

[121] D Shore and R L Baldwin, 'Energetics of DNA twisting: II. Topoisomer analysis', *J Mol Biol*, **170**, 983 – 1007, 1983.

[122] A Dawid, F Guillemot, C Brème, V Croquette and F Heslot, 'Mechanically controlled DNA extrusion from a palindromic sequence by single molecule micromanipulation', *Phys Rev Lett*, **96**, 188102, 2006.

[123] S Ido, K Kimura, N Oyabu, K Kobayashi, M Tsukada, K Matsushige and H Yamada, 'Beyond the helix pitch: Direct visualization of native DNA in aqueous solution', *ACS Nano*, **7**, 1817–1822, 2013.

[124] Q Du, C Smith, N Shiffeldrim, M Vologodskaia and A Vologodskii, 'Cyclization of short DNA fragments and bending fluctuations of the double helix', *Proc Nat Acad Sci USA*, **102**, 5397–5402, 2005.

[125] J C Wang, 'Helical repeat of DNA in solution', *Proc Nat Acad Sci USA*, **76**, 200–203, 1979.

[126] T Kato, R P Goodman, C M Erben, A J Turberfield and K Namba, 'High-resolution structural analysis of a DNA nanostructure by cryoEM', *Nano Lett*, **9**, 2747–2750, 2009.

[127] F F Andersen, B Knudsen, C L P Oliveira, R F Frohlich, D Kruger, J Bungert, M Agbandje-McKenna, R McKenna, S Juul, C Veigaard, J Koch, J L Rubinstein, B Guldbrandtsen, M S Hede, G Karlsson, A H Andersen, J S Pedersen and B R Knudsen, 'Assembly and structural analysis of a covalently closed nanoscale DNA cage', *Nucleic Acids Res*, **36**, 1113–1119, 2008.

[128] C Tian, C Zhang, X Li, C Hao, S Ye and C Mao, 'Approaching the limit: Can one DNA strand assemble into defined nanostructures?', *Langmuir*, **30**, 5859–5862, 2013.

[129] C Zhang, W Wu, X Li, C Tian, H Qian, G Wang, W Jiang and C Mao, 'Controlling the chirality of DNA nanocages', *Angew Chem Int Edit*, **51**, 7999–8002, 2012.

[130] C Maffeo, R Schöpflin, H Brutzer, R Stehr, A Aksimentiev, G Wedemann and R Seidel, 'DNA-DNA interactions in tight supercoils are described by a small effective charge density', *Phys Rev Lett*, **105**, 158101, 2010.

[131] J Mazur and R L Jernigan, 'Distance-dependent dielectric constants and their application to double-helical DNA', *Biopolymers*, **31**, 1615–1629, 1991.

[132] C Vega, E Sanz, J Abascal and E Noya, 'Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins', *J Phys: Condens Matter*, **20**, 153101, 2008.

[133] T E Ouldridge, A A Louis and J P K Doye, 'Extracting bulk properties of self-assembling systems from small simulations', *J Phys: Condens Matter*, **22**, 104102, 2010.

[134] T Uzawa, T Isoshima, Y Ito, K Ishimori, D E Makarov and K W Plaxco, 'Sequence and temperature dependence of the end-to-end collision dynamics of single-stranded DNA', *Biophys J*, **104**, 2485–2492, 2013.

[135] N L Goddard, G Bonnet, O Krichevsky and A Libchaber, 'Sequence dependent rigidity of single stranded DNA', *Phys Rev Lett*, **85**, 2400–2403, 2000.

[136] G Bonnet, O Krichevsky and A Libchaber, 'Kinetics of conformational fluctuations in DNA hairpin-loops', *Proc Nat Acad Sci USA*, **95**, 8602–8606, 1998.

[137] S Whitelam, E H Feng, M F Hagan and P L Geissler, 'The role of collective motion in examples of coarsening and self-assembly', *Soft Matter*, **5**, 1251–1262, 2009.

[138] S Kumar, J M Rosenberg, D Bouzida, R H Swendsen and P A Kollman, 'The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method', *J Comput Chem*, **13**, 1011–1021, 1992.

[139] J D Moroz and P Nelson, 'Torsional directed walks, entropic elasticity, and DNA twist stiffness', *Proc Nat Acad Sci USA*, **94**, 14418–14422, 1997.

[140] A Savelyev, 'Do monovalent mobile ions affect DNA's flexibility at high salt content?', *Phys Chem Chem Phys*, **14**, 2250–2254, 2012.

[141] E Herrero-Galán, M E Fuentes-Perez, C Carrasco, J M Valpuesta, J L Carrascosa, F Moreno-Herrero and J R Arias-Gonzalez, 'Mechanical identities of RNA and DNA double helices unveiled at the single-molecule level', *J Am Chem Soc*, **135**, 122–131, 2013.

[142] X J A Janssen, J Lipfert, T Jager, R Daudey, J Beekman and N H Dekker, 'Electromagnetic torque tweezers: A versatile approach for measurement of single-molecule twist and torque', *Nano Lett*, **12**, 3634–3639, 2012.

[143] Z Bryant, M D Stone, J Gore, S B Smith, N R Cozzarelli and C Bustamante, 'Structural transitions and elasticity from torque measurements on DNA', *Nature*, **424**, 338–341, 2003.

[144] A Y L Sim, J Lipfert, D Herschlag and S Doniach, 'Salt dependence of the radius of gyration and flexibility of single-stranded DNA in solution probed by small-angle X-ray scattering', *Phys Rev E*, **86**, 021901, 2012.

[145] C Zhang, Y He, M Su, S H Ko, T Ye, Y Leng, X Sun, A E Ribbe, W Jiang and C Mao, 'DNA self-assembly: from 2D to 3D', *Faraday Discuss*, **143**, 221–233, 2009.

[146] E S Andersen, M Dong, M M Nielsen, K Jahn, R Subramani, W Mamdouh, M M Golas, B Sander, H Stark, C L P Oliveira, J S Pedersen, V Birkedal, F Besenbacher, K V Gothelf and J Kjems, 'Self-assembly of a nanoscale DNA box with a controllable lid', *Nature*, **459**, 73–76, 2009.

[147] L Zhou, A E Marras, H-J Su and C E Castro, 'DNA origami compliant nanostructures with tunable mechanical properties', *ACS Nano*, **8**, 27–34, 2014.

[148] A E Marras, L Zhou, H-J Su and C E Castro, 'Programmable motion of DNA origami mechanisms', **112**, 713–718, 2015.

[149] C E Castro, H-J Su, A E Marras, L Zhou and J Johnson, 'Mechanical design of DNA nanostructures', *Nanoscale*, **7**, 5913–5921, 2015.

[150] I H Stein, V Schüller, P Böhm, P Tinnefeld and T Liedl, 'Single-molecule FRET ruler based on rigid DNA origami blocks', *ChemPhysChem*, **12**, 689–695, 2011.

[151] R Iinuma, Y Ke, R Jungmann, T Schlichthaerle, J B Woehrstein and P Yin, 'Polyhedra

self-assembled from DNA tripods and characterized with 3D DNA-PAINT', *Science*, **344**, 65–69, 2014.

[152] M Raab, J J Schmied, I Jusuk, C Forthmann and P Tinnefeld, 'Fluorescence microscopy with 6 nm resolution on DNA origami', *ChemPhysChem*, **15**, 2431–2435, 2014.

[153] B Mergell, M R Ejtehadi and R Everaers, 'Modeling DNA structure, elasticity, and deformations at the base-pair level', *Phys Rev E*, **68**, 021911, 2003.

[154] D M J Lilley, 'Structures of helical junctions in nucleic acids', *Quart Rev Biophys*, **33**, 109–159, 2000.

[155] D M J Lilley, 'Four-way helical junctions in DNA molecules', in C J Benham, S Harvey, W K Olson, D W Sumners and D Swigon, editors, *Mathematics of DNA structure, function and interactions*, volume 150 of *The IMA Volumes in Mathematics and its Applications*, 213–224. Springer New York, 2009.

[156] M Ortiz-Lombardia, A Gonzalez, R Eritja, J Aymami, F Azorin and M Coll, 'Crystal structure of a DNA holliday junction', *Nat Struct Biol*, **6**, 913–917, 1999.

[157] B F Eichman, J M Vargason, B H M Mooers and P S Ho, 'The Holliday junction in an inverted repeat DNA sequence: Sequence effects on the structure of four-way junctions', *Proc Nat Acad Sci USA*, **97**, 3971–3976, 2000.

[158] C Mao, W Sun and N C Seeman, 'Designed two-dimensional DNA holliday junction arrays visualized by atomic force microscopy', *J Am Chem Soc*, **121**, 5437–5443, 1999.

[159] P S Eis and D P Millar, 'Conformational distributions of a four-way DNA junction revealed by time-resolved fluorescence resonance energy transfer', *Biochemistry*, **32**, 13852–13860, 1993.

[160] S Hohng, R Zhou, M K Nahas, J Yu, K Schulten, D M J Lilley and T Ha, 'Fluorescence-force spectroscopy maps two-dimensional reaction landscape of the holliday junction', *Science*, **318**, 279–283, 2007.

[161] J Nowakowski, P J Shim, C Stout and G F Joyce, 'Alternative conformations of a nucleic acid four-way junction', *J Mol Biol*, **300**, 93–102, 2000.

[162] J Yu, T Ha and K Schulten, 'Conformational model of the Holliday junction transition deduced from molecular dynamics simulations', *Nucleic Acids Res*, **32**, 6683–6695, 2004.

[163] C De Michele, L Rovigatti, T Bellini and F Sciortino, 'Self-assembly of short DNA duplexes: from a coarse-grained model to experiments through a theoretical link', *Soft Matter*, **8**, 8388–8398, 2012.

[164] P Yakovchuk, E Protozanova and M D Frank-Kamenetskii, 'Base-stacking and base-pairing contributions into thermal stability of the DNA double helix', *Nucleic Acids Res*, **34**, 564–574, 2006.

Appendices

The following chapters contain extra details to supplement the material described in the main part of this thesis. Appendix A provides details regarding the potential energy function introduced for oxDNA2 in Chapter 5, while Appendix B provides further supplementary information for Chapters 3, 4, 5 and 6.

A The oxDNA2 potential

A.1 Introducing an electrostatic term into the model potential

We introduce an explicit electrostatic term into the model using Debye-Hückel theory. To aid computational efficiency when simulating the model, we ensure that the interaction goes to zero within a finite distance $r_{\text{cut,DH}}$, and we introduce a quadratic smoothing function so that the interaction goes to zero smoothly. The form of the electrostatic term is then

$$V_{\text{electrostatic}}(r^{\text{b-b}}, T, I) = \begin{cases} V_{\text{DH}}(r^{\text{b-b}}, T, I) & \text{if } r_{\text{smooth,DH}} > r^{\text{b-b}}, \\ V_{\text{smooth}}(r^{\text{b-b}}, T, I) & \text{if } r_{\text{cut,DH}} > r^{\text{b-b}} \geq r_{\text{smooth,DH}}, \\ 0 & \text{otherwise,} \end{cases} \tag{A.1}$$

where T is the temperature, I is the salt concentration, and $r_{\text{smooth,DH}}$, the distance beyond which smoothing is introduced, is chosen to be equal to $3\lambda_{\text{DH}}$ (λ_{DH} is defined below). The Debye-Hückel-like term, $V_{\text{DH}}(r^{\text{b-b}}, T, I)$, is given by

$$V_{\text{DH}}(r^{\text{b-b}}, T, I) = \frac{(q_{\text{eff}}e)^2}{4\pi\epsilon_0\epsilon_{\text{r}}} \frac{\exp\left\{-r^{\text{b-b}}/\lambda_{\text{DH}}(T, I)\right\}}{r^{\text{b-b}}}, \tag{A.2}$$

where q_{eff} is the effective charge (which is equal to 1 in Debye-Hückel theory but which we choose to be 0.815 after fitting to experimental data), e is the charge of an electron, ϵ_0 is the permittivity of a vacuum, ϵ_r is the relative permittivity of water (which we choose to be 80), and $r^{\text{b-b}}$ is the distance between the pair of interacting backbone sites.

λ_{DH} is the Debye screening length, and is given by

$$\lambda_{\mathrm{DH}}(T,I) = \sqrt{\frac{\epsilon_0 \epsilon_{\mathrm{r}} k_{\mathrm{B}} T}{2 N_{\mathrm{A}} e^2 I}}\,, \tag{A.3}$$

where k_{B} is Boltzmann's constant and N_{A} is Avogadro's number. The quadratic smoothing term, $V_{\mathrm{smooth}}(r^{\mathrm{b\text{-}b}}, T, I)$, is given by

$$V_{\mathrm{smooth}}(r^{\mathrm{b\text{-}b}}, T, I) = B(r^{\mathrm{b\text{-}b}} - r_{\mathrm{cut,DH}})^2, \tag{A.4}$$

and B and $r_{\mathrm{cut,DH}}$ are constants chosen such that $V_{\mathrm{electrostatic}}$ is smooth and differentiable. Note that the above definition implies that the cutoff radius for the Debye-Hückel term, $r_{\mathrm{cut,DH}}$, depends on the salt concentration I and the temperature T.

A.2 Modifying the average pitch of the model duplex

The model duplex pitch can be controlled by modifying the position of the minimum in the FENE bonded backbone potential, $\delta r^0_{\mathrm{backbone}}$, which has a small effect on the thermodynamics (which was refitted using the histogram reweighting method described in Section II C 2 of Ref. 90) and little effect on the other properties of the model DNA. For oxDNA2, $\delta r^0_{\mathrm{backbone}}$ was set to give a global twist of zero for the N-type helix bundle due to Dietz *et al.*;[9] the new and old values for $\delta r^0_{\mathrm{backbone}}$ are given in Table A.1.

A.3 Modifying the average coaxial stacking term in the oxDNA potential

We modified the coaxial stacking term in the oxDNA potential in order to change the twist angle across a nick and across an origami junction, while changing the thermodynamic properties of a nicked duplex and other motifs as little as possible. This was achieved by moving the position of the minimum with respect to the angle θ_1 (shown in Fig. A.1) of the coaxial stacking term, $\theta^0_{\mathrm{coax},1}$, and by increasing the force constant, k_{coax},

to make the minimum more energetically favourable. Increasing the force constant was necessary because it is less favourable for a pair of coaxially stacking nucleotides in a nicked duplex to stack with this tighter angle; a stronger overall coaxial stacking interaction compensates for this effect. The changes to the model are shown in Table A.1.

The altered optimal angle $\theta^0_{\text{coax},1}$ requires an alteration of the modulation function involving θ_1 for coaxial stacking. In the original oxDNA, the θ_1 modulation is given by $f_4(\theta_1) + f_4(2\pi - \theta_1)$, where f_4 is defined in Ref. 67. The second term was introduced to ensure that the potential is differentiable at $\theta_1 = \pi$, where the gradients of the two terms cancel. In oxDNA, this additional term had a very small effect on the numerical value of the potential for $\theta_1 < \pi$. Using the same trick in oxDNA2, however, is problematic, because $\theta^0_{\text{coax},1}$ is much closer to π. The second term, $f_4(2\pi - \theta_1)$, would then be substantial for $\theta_1 < \pi$ and would distort the modulating function in an undesired fashion. To avoid this, we replace the second term, $f_4(2\pi - \theta_1)$, with a new function $f_6(\theta_1)$:

$$f_6(\theta) = \begin{cases} \frac{A}{2}(\theta - B)^2. & \text{if } \theta \geq B, \\ 0 & \text{otherwise.} \end{cases} \tag{A.5}$$

The $f_4(\theta_1) + f_4(2\pi - \theta_1)$ factor in the coaxial stacking term is then replaced by $f_4(\theta_1) + f_6(\theta_1)$. The values for $A_{\text{coax},1}$ and $B_{\text{coax},1}$ are chosen to ensure that the coaxial stacking potential is smooth and differentiable; they are specified in Table A.1.

For oxDNA2, we made an additional change to the coaxial stacking potential: we removed the $f_5(\cos(\phi_3), a_{\text{coax},3'}, \cos(\phi_3)^*_{\text{coax}})$ and $f_5(\cos(\phi_4), a_{\text{coax},4'}, \cos(\phi_4)^*_{\text{coax}})$ terms. These terms had allowed only right-handed blunt-ended stacking, disallowing left-handed blunt-ended stacking; however, since the development of the original oxDNA

Parameter	oxDNA	oxDNA2
γ/degrees	0	20
q_{eff}	0	0.815
ϵ_{r}	-	80
k_{coax}	46	58.5
$\theta^0_{\text{coax,1}}$	$\pi - 0.60$	$\pi - 0.25$
$A_{\text{coax,1}}$	-	40
$B_{\text{coax,1}}$	-	$\pi - 0.025$
ϵ_{HB}	1.077	1.0678
ϵ_{stack}	$1.3448 + 2.6568 k_{\text{B}}T$	$1.3523 + 2.6717 k_{\text{B}}T$
$\delta r^0_{\text{backbone}}$	0.7525	0.7564

Table A.1: Comparison of model parameters between oxDNA and oxDNA2. All other parameters are unchanged in oxDNA2. All angles except γ are given in radians, all lengths are defined with respect to a reduced length scale (1 unit = 8.518 Å) and all energies are defined with respect to a reduced temperature ($k_{\text{B}}T$ = 0.1 corresponding to 300K). γ is the angle between the line from the helix axis to the backbone site and the line from the helix axis to the stacking site (see Fig. 5.1), which defines the backbone site position in oxDNA2. q_{eff} is the effective charge used for the Debye-Hückel treatment introduced in the new model for the electrostatic interactions due to the negatively-charged phosphate groups in the DNA backbone. ϵ_{r} is the relative permittivity of water which is used for the Debye-Hückel treatment. k_{coax} determines the overall strength of the coaxial stacking interaction while $\theta^0_{\text{coax,1}}$ specifies the position of the minimum with respect to the angle θ_1 of the coaxial stacking potential. The values for $A_{\text{coax,1}}$ and $B_{\text{coax,1}}$ are chosen to ensure that the coaxial stacking potential is smooth and differentiable. ϵ_{HB} and ϵ_{stack} determine the well depths for the hydrogen bonding and stacking interactions respectively. $\delta r^0_{\text{backbone}}$ gives the position of the minimum of the FENE bonded backbone potential.

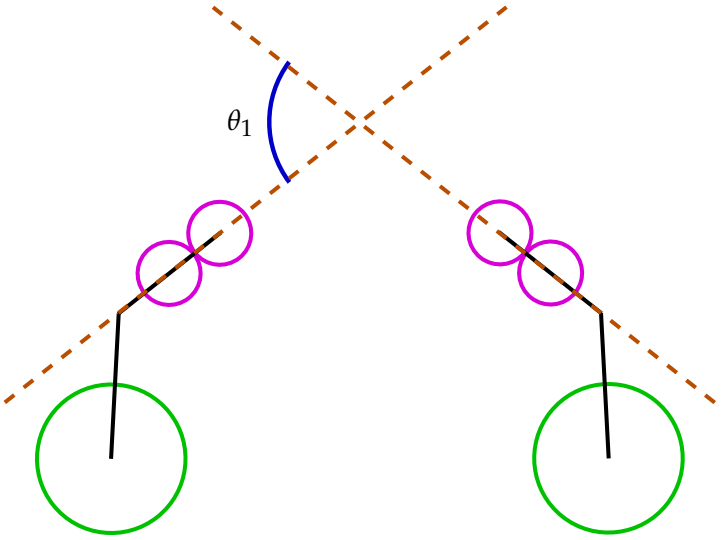

Figure A.1: Schematic showing the angle θ_1 between two nucleotides. Each nucleotide is represented by three circles joined by a line; the large solid circles represent the backbone sites, while the small solid circles represent the stacking (closer to the backbone) and hydrogen-bonding (at the end of the nucleotide) sites. In updating the oxDNA model we modify the term dependent on θ_1 for the coaxial stacking interaction in the oxDNA model, in order to reduce the overtwist across nicks and junctions.

model, this constraint has been deemed unnecessary, as the blunt-ended coaxial stacking interaction was found to be likely to be too weak compared to experiment,[163] and there is no experimental evidence indicating that left-handed blunt-ended coaxial stacking is not possible.

Incorporating the changes described above, the final form for the oxDNA2 coaxial stacking term is

$$\begin{aligned} V_{\text{coax\_stack}} = {} & f_2(\delta r_{\text{stack}}, k_{\text{coax}}, \delta r^0_{\text{coax}}, \delta^{c,\text{low}}_{\text{coax}}, \delta r^{c,\text{high}}_{\text{coax}}, \delta r^{\text{low}}_{\text{coax}}, \delta r^{\text{high}}_{\text{coax}}) f_4(\theta_4, a_{\text{coax},4}, \theta^0_{\text{coax},4}, \Delta\theta^*_{\text{coax},4}) \\ & \times (f_4(\theta_1, a_{\text{coax},1}, \theta^0_{\text{coax},1}, \Delta\theta^*_{\text{coax},1}) + f_6(\theta_1, A_{\text{coax},1}, B_{\text{coax},1})) \\ & \times (f_4(\theta_5, a_{\text{coax},5}, \theta^0_{\text{coax},5}, \Delta\theta^*_{\text{coax},5}) + f_4(\pi - \theta_5, a_{\text{coax},5}, \theta^0_{\text{coax},5}, \Delta\theta^*_{\text{coax},5})) \end{aligned}$$

$$\times \left(f_4(\theta_6, a_{\text{coax},6}, \theta^0_{\text{coax},6}, \Delta\theta^*_{\text{coax},6}) + f_4(\pi - \theta_6, a_{\text{coax},6}, \theta^0_{\text{coax},6}, \Delta\theta^*_{\text{coax},6})\right). \tag{A.6}$$

Table A.2 gives a comparison of the thermodynamic properties of the motifs relevant to the coaxial stacking term in oxDNA and oxDNA2. The free-energy change upon stacking across a nick, $\Delta G_{\text{nick\_stack}}$, is slightly more negative in oxDNA2, indicating that a nicked duplex stacks slightly more strongly across the nick in oxDNA2 than in the original oxDNA. In both cases the stacking is stronger than is seen in experiment ($-2.62 k_B T$ in Ref. 164 – a discussion comparing the value of $\Delta G_{\text{nick\_stack}}$ measured for the original oxDNA with experimental results is given in Ref. 67). The stabilising effect of a hairpin stem on duplex hybridisation (see Fig. A.2), $\Delta T_{\text{m,hairpin\_stem,Xmer}}$, is found to be slightly stronger than for the original oxDNA, and is on average in better agreement with the stability predicted by the SantaLucia model. Finally, dimerisation of two duplexes through blunt-ended coaxial stacking is found to be less favourable in oxDNA2 than in oxDNA. This is despite the mitigating effect of removing the f_5 terms in the coaxial stacking interaction, which strengthens blunt-ended stacking. A previous study[163] has shown that oxDNA already underestimated the stability of blunt-ended coaxial stacking.

A.4 Modifying the hydrogen bonding and stacking interaction strengths

Model duplexes were found to fray slightly less often when different widths for the major and minor duplex grooves were introduced compared to the equal-groove-width case, which led to a slight raising of the duplex melting temperatures. In addition, the change to the position of the minimum in the FENE bonded backbone potential, $\delta r^0_{\text{backbone}}$, was found to destabilise the duplex state, slightly lowering the duplex

Measurement	oxDNA	oxDNA2	SantaLucia
$\Delta G_{\text{nick\_stack}}$	$-4.3 k_{\text{B}}T$	$-4.4 k_{\text{B}}T$	-
$\Delta T_{\text{m,hairpin\_stem,6mer}}$	+9.9K	+10.6K	+11.7K
$\Delta T_{\text{m,hairpin\_stem,7mer}}$	+7.7K	+8.3K	+8.5K
$\Delta T_{\text{m,hairpin\_stem,8mer}}$	+6.3K	+6.9K	+6.4K
$\Delta G_{\text{blunt\_ended\_stacking}}$	$+4.9 k_{\text{B}}T$	$+5.3 k_{\text{B}}T$	-

Table A.2: Comparison of some thermodynamic quantities that depend on the coaxial stacking interaction in oxDNA and oxDNA2. $\Delta G_{\text{nick\_stack}} = \Delta G_{\text{stacked}} - \Delta G_{\text{unstacked}}$ is the free-energy change upon stacking across a nick for a 20-bp duplex with a nick at the centre at 37° C. $\Delta T_{\text{m,hairpin\_stem,6mer}} = T_{\text{m,hairpin\_stem,6mer}} - T_{\text{m,6mer}}$ gives the difference in melting temperature at a concentration of 3.3×10^{-4} M between a 6-bp duplex adjacent to a hairpin and a 6-bp duplex (see Fig. A.2), and $\Delta T_{\text{m,hairpin\_stem,7mer}}$ and $\Delta T_{\text{m,hairpin\_stem,8mer}}$ have analogous definitions. $\Delta G_{\text{blunt\_ended\_stacking}}$ is the free-energy change upon dimerisation of two 6-bp duplexes through blunt-ended coaxial stacking at a monomer concentration of 5.37 mM and at 19.85° C. $\Delta G_{\text{nick\_stack}}$ was computed from a simulation of 1.9×10^{11} VMMC steps, each $\Delta T_{\text{m,hairpin\_stem,Xmer}}$ from simulations of 8×10^{10} VMMC steps, and $\Delta G_{\text{blunt\_ended\_stacking}}$ from a simulation of 2.4×10^{11} VMMC steps.

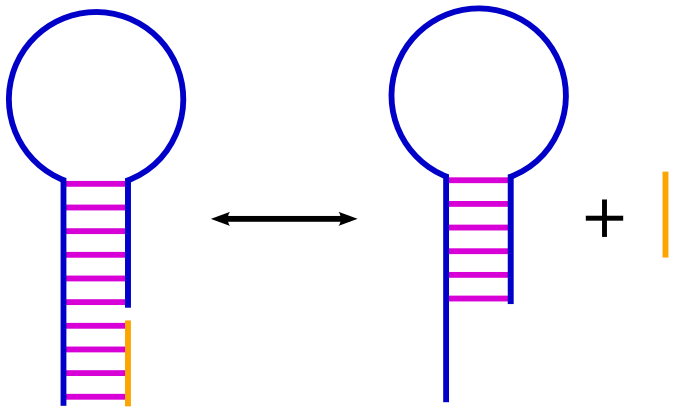

Figure A.2: A schematic showing the transition important for the calculation of $\Delta T_{\text{m,hairpin\_stem,Xmer}}$ (see Table A.2), which is given by $\Delta T_{\text{m,hairpin\_stem,Xmer}} = T_{\text{m,hairpin\_stem,Xmer}} - T_{\text{m,Xmer}}$, where $T_{\text{m,hairpin\_stem,Xmer}}$ is the melting temperature of a X-bp duplex adjacent to a hairpin (the schematic shows this system for the X = 4 case) and $T_{\text{m,Xmer}}$ is the melting temperature of a X-bp duplex.

melting temperatures. In the process of the reparameterisation of oxDNA to incorporate these two changes, the duplex melting temperatures were reset on the introduction of each change by making small modifications to the stacking and hydrogen bonding strengths using the histogram reweighting method described in Section II C 2 of Ref. 90, giving an agreement with experimental melting temperatures that was just as good as before. The new parameters are specified, and compared to the old parameters, in Table A.1, while Fig. 5.7 gives a comparison of the duplex thermodynamics with the old and new versions of the model.

B Supplementary information

B.1 Standard MD simulation parameters

The standard input parameters for the MD simulations used to generate the results described in this work are given in Table B.1.

B.2 DNA strand sequences used for the direct simulation of origami assembly

The DNA strand sequences used for the twelve staple strands and scaffold strand for the directly simulated origami assembly described in Chapter 3 are given in Table B.2.

B.3 Supplementary information for Chapter 4

B.3.1 Strand sequences used for the simulation of origami assembly using FFS

The DNA strand sequences used for the origami assembly studied with FFS as described in Chapter 4 are given in Table B.3.

Parameter	Value
newtonian_steps	103
diff_coef	2.5
thermostat	brownian
dt	0.005
verlet_skin	0.05

Table B.1: The simulation parameters used in the oxDNA code's input file for the MD simulations run for this work.

Strand	Scaffold Domains	Sequence
scaffold	-	ATGCAAAGATACGGAAAAGGGAGAGAAAGAAGAG
		GGACACGGGAAAGGGCAAAAAAACAAAGAATGGG
		AGTAGGAAGCGAGAAAATAACGGGCAGCTATAAA
		GAAAAAAAGGGAAAGATAGAACAGGGAGGAGGAA
		AGAGACAGTAGAGTGGTGATAGGGAGAAGAAAAG
		AAGAAAAAGAGGAGAGCAAGAAACGGGATAGAGA
		AGAGGAACAGCAAAAGACAAAGAAGAGGGGCAAG
		AGAGGATAGAAGTACGAGATAAGGGGTAGAAGCG
		AAAAGCAAAAAATAAGGGATAAGAGACGGGGAGA
		AGTGAAGAAAGAAGGAGGATAAAAAGACAGATGG
		ACACGAAAGGAAGATAGGAGGGGGAGCGACGAAT
		AAAAGGAAAC
staple 1	1, 2	CTTCTTTCTCTCCCTTTTCCGTATCTTTGCAT
staple 2	3, 4	CCTTTCCCGTGTCCCTTTCTTTGTTTTTTTGC
staple 3	5, 8	CCCTGTTCTATCTTTCTCGCTTCCTACTCCCA
staple 4	6, 7	GCTGCCCGTTATTTTCCCTTTTTTTCTTTATA
staple 5	9, 10	CTGTCTCTTTCCTCCTCCCTATCACCACTCTA
staple 6	11, 14	TTGTCTTTTGCTGTTCTTTCTTCTTTTCTTCT
staple 7	12, 13	TTCTTGCTCTCCTCTTCTCTTCTCTATCCCGT
staple 8	15, 16	CTCTTGCCCCTCTTCTCTCGTACTTCTATCCT
staple 9	17, 20	CTTCTTTCTTCACTTCCGCTTCTACCCCTTAT
staple 10	18, 19	CTTATTTTTTGCTTTTTCCCCGTCTCTTATCC
staple 11	21, 22	CTGTCTTTTTATCCTCTTCCTTTCGTGTCCAT
staple 12	23, 24	GTTTCCTTTTATTCGTCGCTCCCCCTCCTATC

Table B.2: DNA strand sequences used for the assembly simulations described in Chapter 3. Sequences were obtained with NUPACK with a search designed to minimize secondary structure and intra-staple binding. The "Scaffold Domain" column indicates which scaffold domains the staple strands bind to (see Fig. 3.1).

Strand	Sequence
scaffold	CTTGTTGAACGTGCGCTCATCCTAGATCATCCTGACTACGACACGGAG
	GACAACAGGATCCCAACAAATTAGGGCGGGAGCGATACCTAGGCCAAA
	ATCGGTTGAATGTATCTTCAACTCCGGACGTCATGGATTTGGTCCTGA
	GGTGAGTTCCAAGGCATCACAACCCAAACCTCTGTTAGTAAACTTGGT
	GGGGATAAGCGATGTCACATGTTAGGCCAGTTATGCTCATAATAATTT
	CATTAGCGCGTAGGATGTTAGGAGCCAGCCAGTATGTGGCAACTGTAA
	TTATTAGTCCGGTATCATCGGTGAGTCAAGGTGGCACGAGTCGAATGC
	TGGCCTTTCATAGTACGCCTTTCTCTGAACGAAAAGATTGCCACGGCG
staple Aa	TGCCTTGGAACTCACCAAATTATTATGAGCATATCCTACGCGCTAATG
staple Ba	CTGGCTGGCTCCTAACAACTGGCCTAACATGTGAGGTTTGGGTTGTGA
staple Ca	ACCAAGTTTACTAACAGACATCGCTTATCCCCTTACAGTTGCCACATA
staple Da	GATACCGGACTAATAAACCTTGACTCACCGATGCATTCGACTCGTGCC
staple Ab	GATACATTCAACCGATGACGTCCGGAGTTGAATCAGGACCAAATCCAT
staple Bb	CGCCGTGGCAATCTTTGCGCACGTTCAACAAGTTTGGCCTAGGTATCG
staple Cb	CTCCCGCCCTAATTTGGGATGATCTAGGATGATCGTTCAGAGAAAGGC
staple Db	GTACTATGAAAGGCCACTCCGTGTCGTAGTCATTGGGATCCTGTTGTC

Table B.3: The random DNA strand sequences used for the assembly simulations described in Chapter 4. Fig. 4.1 shows the location of the staples.

B.3.2 Number of possible substates for an assembly state

Under the scheme laid out for studying origami assembly using FFS in Chapter 4, each assembly state Sn, except for the first and last states S0 and S24, has more than one possible combination of scaffold domains occupied by staples. As there are 24 scaffold domains, the number of combinations for assembly state Sn is given by

$$\binom{24}{n} = \frac{24!}{n!(24-n)!}. \tag{B.1}$$

This leads to 735 471 combinations for S8, as stated in the main text.

We can reduce this number by assuming that, for each staple, once the first staple domain has bound to the scaffold, the staple binds the rest of its domains before a new staple binds somewhere else on the scaffold. If the assumption always holds, our states Sn will only have between 0 and 8 fully bound staples and up to 1 partially bound staple (it turns out that this is true in practice in our simulations for the vast majority of cases).

Call the number of fully bound staples $N = \text{floor}(n/3)$ and the number of partially bound domains $k = n \bmod 3$ (these definitions follow from the fact that there are three domains per staple). Then the number of combinations C is given by

$$C = \begin{cases} \binom{8}{N} & k = 0, \\ 3(8-N)\binom{8}{N} & k = 1, 2. \end{cases} \tag{B.2}$$

On top of this, we can reduce the effective number of combinations by one half by treating the origami as symmetric under a rotation of 180° in the plane of the origami, as this rotation leaves the complete origami changed only in sequence. Returning to our

previous example, there are 252 possible combinations for S8 under these assumptions.

B.3.3 Order parameters and simulation parameters for the origami assembly FFS simulations

The definitions for the discrete order parameters λ used for the FFS assembly simulations of Chapter 4 depend on the states Q of the pairs of staple and scaffold domains. The definitions for Q, which were different for transitions between different assembly states, are given in Tables B.4 and B.5. In those definitions, d denotes the smallest of the distances, in simulation length units, between pairs of bases that are designed to be base-paired in the final structure (we call these "target base pairs"); n denotes the number of target base pairs for which the hydrogen-bonding interaction energy term is less than -0.1 simulation energy units; and n_{near} denotes the number of target base pairs for which the following is true: all of the factors in the hydrogen-bonding interaction term are non-zero, or all but one of them are non-zero, but the hydrogen-bonding interaction energy is greater than -0.1 simulation energy units.

For a forwards transition Si $\rightarrow$S$i+1$, the A-state contains i bound domains and 120-i unbound domains (due to the staple excess, there are 120 pairs of staple and scaffold domains that could potentially bind, although a maximum of 24 can bind simultaneously), and the B-state contains $i+1$ bound domains and $120-(i+1)$ unbound domains. In order to monitor the progress of the assembly, we define λ so that: $\lambda = 0$ is defined by 120-i domains satisfying $Q = 0$; $\lambda = 1, .., n-1$ are defined by at least $i+1$ domains satisfying $Q \geq \lambda$ and no more than i domains satisfying $Q \geq \lambda + 1$; and $\lambda = n$ is defined by at least $i+1$ domains satisfying $Q = n$.

For example: for the S10$\rightarrow$S11 transition, which used order parameter definition C (Table B.4), $\lambda = 0$ when 110 domains have $Q = 0$ (i.e. $d > 1$), $\lambda = 1$ when at least

Q	Definition A	Definition B	Definition C	Definition D
0	$d > 8$	$d > 8$	$d > 1$	$d > 1$
1	$8 \geq d > 6$	$8 \geq d > 6$	$d \leq 1$ & $n_{\text{near}} = 0$	$d \leq 1$ & $n_{\text{near}} = 0$
2	$6 \geq d > 3$	$6 \geq d > 3$	$n_{\text{near}} = 1$ & $n = 0$	$n_{\text{near}} = 1$ & $n = 0$
3	$3 \geq d > 0.8$	$3 \geq d > 0.8$	$n = 1$	$n = 1$
4	$d \leq 0.8$ & $n_{\text{near}} = 0$	$d \leq 0.8$ & $n_{\text{near}} = 0$	$n = 2$	$n = 2$
5	$n_{\text{near}} = 1$ & $n = 0$	$n_{\text{near}} = 1$ & $n = 0$	$3 \leq n < 13$	$3 \leq n < 14$
6	$n = 1$	$n = 1$	$n = 13$	$n \geq 14$
7	$n = 2$	$n = 2$	$n \geq 14$	-
8	$n = 3$	$n = 3$	-	-
9	$4 \leq n < 7$	$4 \leq n < 14$	-	-
10	$7 \leq n < 14$	$n \geq 14$	-	-
11	$n \geq 14$	-	-	-

Table B.4: The four Q definitions used for the FFS simulations in Chapter 4 for the forwards transitions. Definition A was used for the S0→S1 transition; definition B was used for the S3→S4, S6→S7, S9→S10 and S12→S13 transitions; definition C was used for the S1→S2, S4→S5, S7→S8, S8→S9, S10→S11 and S11→S12 transitions; and definition D was used for the S2→S3, S5→S6, S13→S14, S14→S15 transitions. d, n_{near} and n are defined in the main text.

Q	Definition E	Definition F	Definition G
0	$n > 13$	$n > 13$	$n > 13$
1	$13 \geq n > 10$	$13 \geq n > 10$	$13 \geq n > 10$
2	$10 \geq n > 4$	$n = 10$	$n = 10$
3	$n = 4$	$n = 9$	$n = 9$
4	$n = 3$	$n = 8$	$n = 8$
5	$n = 2$	$n = 7$	$n = 7$
6	$n \leq 1$ & $d < 1$	$n = 6$	$n = 6$
7	$5 > d \geq 1$	$n = 5$	$n = 5$
8	$d \geq 5$	$n = 4$	$n = 4$
9	-	$n = 3$	$n = 3$
10	-	$n = 2$	$n = 2$
11	-	$n \leq 1$ & $d < 1$	$n \leq 1$ & $d < 5$
12	-	$5 > d \geq 1$	$d \geq 5$
13	-	$d \geq 5$	-

Table B.5: The three Q definitions used for the FFS simulations in Chapter 4 for the backwards transitions. Definition E was used for the S3→S2 transition, definition F was used for the S2→S1 transition, and Definition G was used for the S1→S0 transition. d, n_{near} and n are defined in the main text.

11 domains satisfy $Q \geq 1$ (i.e. $d \leq 1$) and no more than 10 domains satisfy $Q \geq 2$ (i.e. $n_{\text{near}} = 1$ or $n \geq 1$), $\lambda = 2$ when at least 11 domains satisfy $Q \geq 2$ (i.e. $n_{\text{near}} = 1$ or $n \geq 1$) and no more than 10 domains satisfy $Q \geq 3$ (i.e. $n \geq 1$), and so on. Note that the conditions are defined so as to avoid ambiguity regarding the value of λ. The B-state is then defined as at least 11 domains satisfying $Q = 7$, or $n \geq 14$.

For the backwards calculations, we have a transition Si $\rightarrow$S$i-1$. In this case we define λ so that: $\lambda = 0$ corresponds to i domains satisfying $Q = 0$; $\lambda = 1, .., n-1$ are defined by $120 - i + 1$ domains satisfying $Q \geq \lambda$ and no more than $120 - i$ domains satisfying $Q \geq \lambda + 1$; and $\lambda = n$ is defined by at least $120 - i + 1$ domains satisfying $Q = n$.

The forwards FFS calculations were run using flux scheme 1 and the backwards ones were run using flux scheme 2, both of which are described in Section 2.4. The MD parameters used in the input file for the oxDNA flux and shooting simulations are given in Table B.1.

For each transition between assembly states, usually at least 25000 configurations were generated during the flux simulations and at least 5000 configurations were generated during each set of shooting simulations, with the exception of the S3$\rightarrow$S2 transition, for which a factor of 10 fewer configurations was generated at each stage. In addition, only 1000 configurations were saved at λ_n for most of the rate calculations, with the exception of the S0$\rightarrow$S1, S1$\rightarrow$S2, S2$\rightarrow$S3, S3$\rightarrow$S4, S4$\rightarrow$S5 and S5$\rightarrow$S6 transitions. It can be argued that it is reasonable to save on computational effort in this way because, while earlier configurations are needed to launch future simulations without any equilibration, the configurations generated at λ_n will be equilibrated before they are (potentially) used for a rate calculation for a different transition, so it is not necessary to generate

as diverse a set of configurations. With the exception of the S0→S1 transition, the calculations were split up into 4 separate, independent FFS calculations, so that an approximate estimate of the error on the rates could be determined.

The variety of different definitions for Q shown in Tables B.4 and B.5 partly reflects the different requirements for efficient sampling of different types of transitions, as well as tinkering and improvements based on experience. For example, the $Q = 10$ condition for definition A was found to be unnecessary.

B.3.4 Assembly substate breakdown for backwards transitions not shown in Chapter 4

A breakdown of the backwards transitions by substate is shown for the S1→S0 transition in Fig. B.1 and for the S3→S2 transition in Fig. B.2. The rates are approximately uniform (the small error bars on the C0→S0 and D1→S0 points are misleading as these are poorly sampled, with only 20 and 34 simulations reaching the B-state, respectively).

B.3.5 Creating the average rate plots

The average forwards rates (Fig. 4.5) were calculated as the average rate from all of the transitions between substates within each run, with the result being averaged over the four runs (except for the S0→S1 transition, for which there was only one run).

The average backwards rates (Fig. 4.6) were calculated in the same way as the average forwards rates, except that the flux and success probability data for the transitions between substates was first combined for substates that were related by the rotational symmetry of the origami.

The rates shown in most of the figures examining the order of staple domain binding (Figs. 4.7, 4.8, 4.12 and 4.14) were calculated in the same way as the average forwards

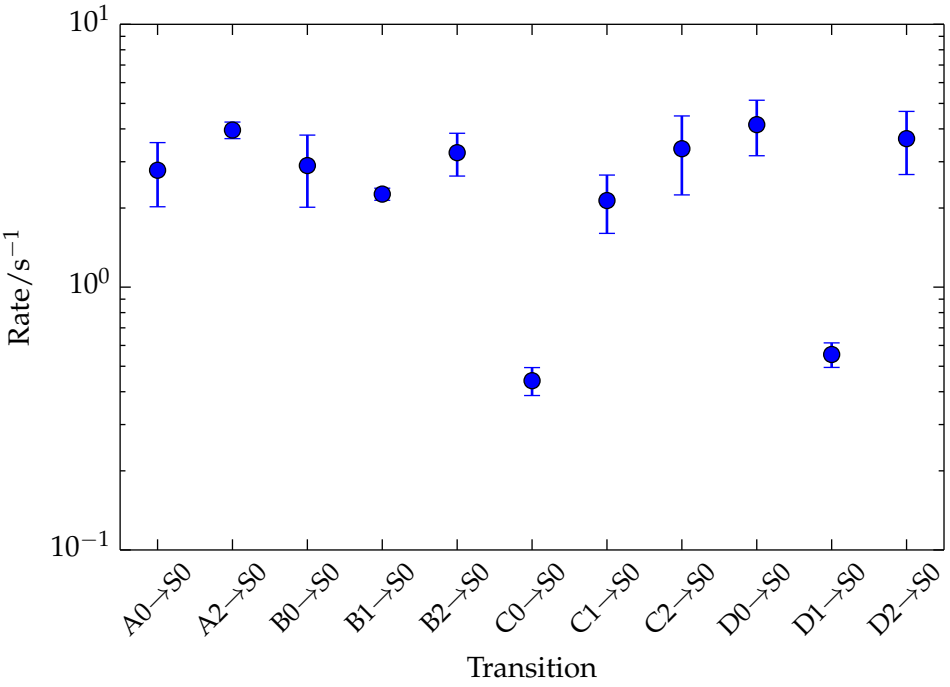

Figure B.1: The rates for the transitions from each of the substates of assembly state S1 to assembly state S0. Error bars show the standard error on the mean calculated from up to four independent FFS calculations.

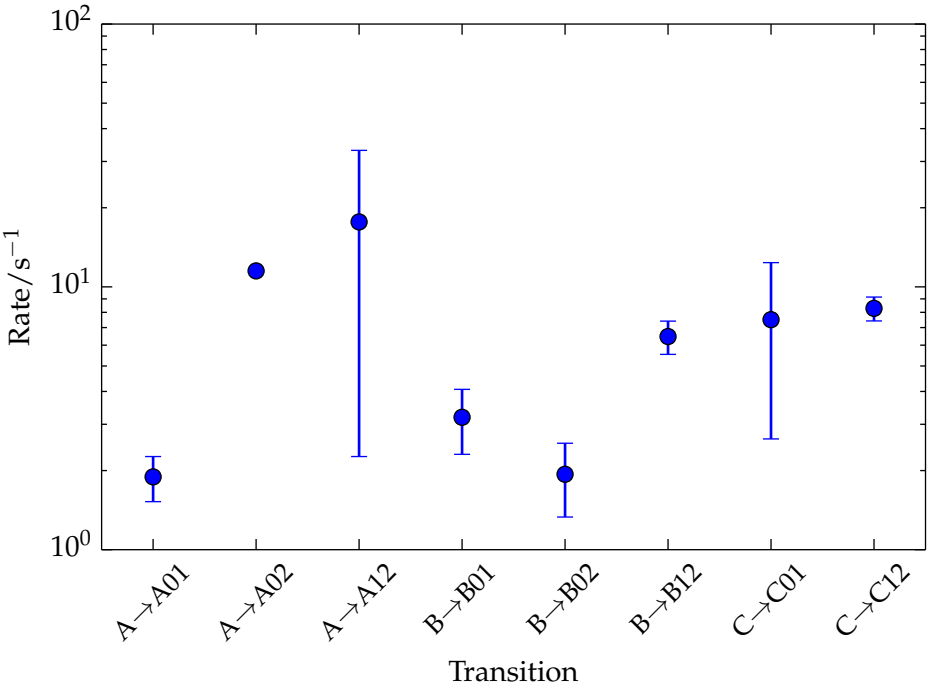

Figure B.2: The rates for the transitions from each of the substates of assembly state S3 to the substates of assembly state S2. Error bars show the standard error on the mean calculated from up to four independent FFS calculations.

rates, except that the transitions between substates were split into two groups, as shown in the figures.

The figures demonstrating threading effects (Figs. 4.9 and 4.11) and cooperativity (Figs. 4.15, 4.17, 4.18 and 4.19) present rates for each of the transitions between substates, each of which is an average of all of the rates calculated from the runs that sampled that transition (up to four).

B.4 Supplementary information for Chapter 5

B.4.1 Measuring the global twist of an origami

We measure the global twist of the origami due to Dietz *et al.*[9] in the following way: The origami (for example, the L-type, although this applies equally well to all three types) can be thought of as a set of antiparallel double helices on a hexagonal lattice joined together by crossovers. The overall shape of this structure is roughly that of a (possibly twisted) rectangular prism. To measure the twist, we consider the two faces of the rectangular prism which are opposite each other and approximately normal to the axes of the double helices. For each face, we define a vector along its long edge, and we superimpose these two vectors into a plane perpendicular to the average of the helix axes. We then define the global twist as the angle between these two vectors.

To avoid end effects, the top and bottom rows of helices (3 helices in the top row and 3 helices in the bottom row) were excluded for this analysis, as were approximately 22 base pairs from each end of every double helix.

B.4.2 Measuring the torsional stiffness of a duplex in the oxDNA model

The torsional stiffness of a duplex in oxDNA2 was measured according to the scheme outlined by Matek *et al.*[79] In summary, this method involves pulling a duplex while also twisting it, and measuring the torque required to impose a given twist. In this case, a 60-bp duplex was pulled at a constant force of 30 pN and held with virtual traps which imposed a helical pitch different from the equilibrium pitch. For each salt concentration, simulations were run with imposed pitches of 9.8, 10.1, 10.4, 10.7, 11.0, and 11.3 bp/turn, with each simulation run for at least 2.6×10^9 MD steps. The torque observed was plotted as a function of imposed twist (this is called a torque response curve), and the gradient of the linear region used to approximate the effective torsional stiffness.

The torsional stiffness C_{eff} is computed as

$$C_{\text{eff}} = \frac{\Delta\Gamma}{\Delta\sigma}\frac{a_0}{\theta_0}, \tag{B.3}$$

where $\Delta\Gamma$ is the change in torque, and $\Delta\sigma$ the change in superhelical twist density, measured in the linear regime of the torque response curve, a_0 is the double helical rise for a relaxed duplex and θ_0 is the twist angle across a base-pair step for a relaxed duplex, measured in radians. According to Moroz and Nelson,[139] the effective torsional stiffness C_{eff} measured by twisting a duplex under tension can be related to the true torsional stiffness C_0 by

$$C_{\text{eff}} = C_0 \left\{ 1 - \frac{C_0}{4B_0}\sqrt{\frac{k_{\text{B}}T}{B_0 F}} \right\}, \tag{B.4}$$

where F is the linear force and B_0 is the bending stiffness. Our simulations are in the high-force regime, where C_{eff} should be a good approximation for C_0.

B.4.3 Details of the simulation protocols

The simulation results presented in Chapter 5 were obtained using either VMMC or MD with a Brownian thermostat. Note that, where VMMC steps are reported in the main text, they refer to the number of *accepted* moves during the simulation. Details of the simulation parameters used are given in Table B.6 and Table B.7.

For the duplex free-energy profiles, yields and melting temperatures (Fig. 5.7 and Fig. 5.8) we used umbrella sampling with an order parameter defined in terms of the number of native base pairs (0 to n where n is the number of base pairs in a fully formed duplex). A base pair was defined as being formed if the potential's hydrogen-bonding energy term for that pair was lower than $-0.596\,\mathrm{kcal mol}^{-1}$. For the hairpin melting temperature calculations (Fig. 5.10, Fig. 5.11 and Fig. 5.12) we used an analogous order parameter for the native base pairs in the hairpin stem.

The oxDNA2 results reported in Fig. 5.12 were computed using thermodynamic integration as described in Section 5.5. For each oxDNA2 data point shown in the figure, first an umbrella sampling simulation of each hairpin was carried out for a version of the model with $x = 0$, where x is given by Eq. 5.9, and $q_{\mathrm{eff}} = 0$ in order to find the reference free energy ($F_{\alpha}^{(0)}$ in Eq. 5.10). The simulations were run for 6×10^{11} VMMC steps and used the parameters specified in Table B.7. The integrals in Eq. 5.10 were discretised into 15 gridpoints for the integral in q'_{eff} and 10 gridpoints for the integral in x. Then simulations to compute these integrals were run for roughly 6×10^{7} VMMC steps for each grid point, so that each data point in the figure required 25 grid point simulations, with one umbrella sampling simulation per hairpin.

	Simulation description				
Parameter	Standard	Bulge	Nick stack.	Hairpin stab.	Blunt coax
Algorithm	VMMC	VMMC	VMMC	VMMC	VMMC
δ_{trans}	0.11	0.1	0.03	0.15	0.22
δ_{rot}	0.22	0.2	0.15	0.15	0.22
max. cluster size	-	-	-	-	12
Verlet skin	1	1	1	1	1
T	T_{m}	T_{m}	310.15K	T_{m}	293K

Table B.6: The results in Chapter 5 were computed using simulations of oxDNA2 with these parameters (see also Table B.7). The "standard" parameters were used for the following calculations: ΔT_{m} for a terminal mismatch and ΔT_{m} for an internal mismatch (Table 5.2); all hairpin T_{m} calculations (Fig. 5.10 and Fig. 5.11); and all duplex T_{m} calculations (Fig. 5.7). The "bulge" parameters were used for the ΔT_{m} calculations for a bulge motif (Table 5.2), the "nick stack." parameters were used for the $\Delta G_{\text{nick\_stack}}$ calculation (Table A.2), the "hairpin stab." parameters were used for the $\Delta T_{\text{m,hairpin\_stem,Xmer}}$ calculations (Table A.2), and the "blunt coax" parameters were used for the calculation of $\Delta G_{\text{blunt\_ended\_stacking}}$ (Table A.2). δ_{trans} and δ_{rot} are the standard deviation of the normally distributed move size for the translational and rotational VMMC moves respectively, max. cluster size is the size of the biggest cluster that VMMC is permitted to choose for a trial move, Verlet skin is the thickness of the Verlet skin used for Verlet lists, and T is the temperature of the heat bath (T_{m} indicates that simulations were run at or close to the melting temperature for that system). All numbers are given in oxDNA units (1 oxDNA length unit = 0.8518 nm, 1 oxDNA time unit = 3.03 ps), unless given explicitly (angles are quoted in radians).

	Simulation description			
Parameter	Hairpin δF	$C_{\text{torsional}}$	l_{p}	Helix bundle twist
Algorithm	VMMC	MD	MD	MD
δ_{trans}	0.11	-	-	-
δ_{rot}	0.22	-	-	-
max. cluster size	-	-	-	-
Verlet skin	1	0.05	0.1	0.05
δt_{step}	-	0.005	0.005	0.005
diffusion coefficient	-	2.5	2.5	2.5
T	295.6K	296.15K	296.15K	296.15K

Table B.7: The results in Chapter 5 were computed using simulations of oxDNA2 with these parameters (see also Table B.6). The "hairpin δF" parameters were used for the initial free energy calculation for the thermodynamic integration to find the relative stability of the hairpins shown in Fig. 5.12, and to find the radius of gyration (R_{g}) of ssDNA (Fig. 5.13), the "$C_{\text{torsional}}$" parameters were used for the calculation of the torsional stiffness (Fig. 5.9(b)), the "l_{p}" parameters were used for the calculation of the persistence length (Fig. 5.9(a)) and the pitch and rise (Fig. 5.6), and the "helix bundle twist" parameters were used for the calculation of the global twist of the helix bundle origami structures (Table 5.1). δ_{trans} and δ_{rot} are the standard deviation of the normally distributed move size for the translational and rotational VMMC moves respectively, max. cluster size is the size of the biggest cluster that VMMC is permitted to choose for a trial move, Verlet skin is the thickness of the Verlet skin used for Verlet lists, δt_{step} is the size of the simulation time step in MD, diffusion coefficient gives the diffusion coefficient used for a single nucleotide, and T is the temperature of the heat bath (T_{m} indicates that simulations were run at or close to the melting temperature for that system). All numbers are given in oxDNA units (1 oxDNA length unit = 0.8518 nm, 1 oxDNA time unit = 3.03 ps), unless given explicitly (angles are quoted in radians).

B.5 Supplementary information for Chapter 6

B.5.1 Holliday junction free-energy landscape

We simulated the free energy landscape of the junction described in Section 6.2.1 using MD simulations with an Andersen thermostat with the parameters given in Table B.1. The sequences of the strands are given in Table B.8. We windowed the simulations using a biasing potential defined as

$$V(\phi) = \begin{cases} k(\phi - (\phi_0 - \Delta))^4 & \text{if } \phi \leq \phi_0 - \Delta, \\ 0 & \text{if } \phi_0 - \Delta < \phi \leq \phi_0 + \Delta, \\ k(\phi - (\phi_0 + \Delta))^4 & \text{otherwise,} \end{cases} \tag{B.5}$$

where k, ϕ_0 and Δ are constants. The relevant derivatives of this potential, requiring derivatives of ϕ with respect to the positions and orientations of the relevant nucleotides, was obtained so that the bias could be applied to MD simulations. Table B.9 shows which constants were used for each window, and how long the simulations were run for in each window. For the biased simulations, only data from the flat region of the biasing potential, $\phi_0 - \Delta < \phi \leq \phi_0 + \Delta$, was used. The free-energy landscapes computed for each window were combined using the weighted histogram analysis method (WHAM).[138]

strand	sequence
1	AGCAACTTGACGGTAAGAAGCAGAACTCTCCT
2	TCACTGCTAGGTGATCTTTGTTAAGGGTCTCA
3	TGAGACCCTTAACAAATTACCGTCAAGTTGCT
4	AGGAGAGTTCTGCTTCGATCACCTAGCAGTGA

Table B.8: The sequences used for the Holliday junction for which the free-energy landscape was computed.

parameter	window 1	window 2	window 3	window 4
ϕ_0	-	160°	180°	200°
Δ	-	40°	40°	40°
k	-	2	2	2
simulation steps	8.4×10^{10}	2×10^9	2.7×10^{10}	4×10^9

Table B.9: The parameters used for the different windows when computing the Holliday junction's free energy. Window 1 did not use any biasing potential.

It is possible for a free Holliday junction to form either one of two equivalent isomers. The isomers are defined by which two strands are the exchanging strands, and which two stack through the junction. For these simulations, configurations where the Holliday junction had switched to the other, unintended isomer were discarded. These configurations are likely to be inappropriate because the biasing potential is only effective for the intended isomer.

The nature of the isomer was determined by monitoring the distances between the four central base pairs at the centre of the junction. We denote these base pairs A, B, C and D, such that, in the intended isomer, each of the A bases usually stacks (or coaxially stacks) with each of the B bases, and likewise for the C and D bases. In addition, in the other isomer, the A bases will usually (coaxially) stack with the C bases, and the B bases with the D bases. We label the distance between the midpoints of the base pairs as d_{XY}, with d_{AB} denoting the distance between the A and B base-pair midpoints, for example. We then define a configuration to correspond to the intended isomer when d_{AB} and d_{CD} are both shorter than either of d_{AC} and d_{BD}. This definition correctly identifies isomer flips, and was found to exclude only $\sim$1 in 1000 configurations, presumably due to extreme fluctuations, in trajectories where an isomer flip did not occur.

B.5.2 Simulating 2D origami structure

The simulations to compute the weave pattern in Section 6.3 were run using MD with the standard parameters given in Table B.1. Unless otherwise stated, the simulations were run at a salt concentration of 0.5 M and a temperature of 300 K. All four sets of simulations mentioned in the main text (one at a high enough salt concentration that the model's electrostatic interaction is cut off within the excluded volume of the backbone and is effectively removed, and three at 0.5 M at three different temperatures) were run for 2×10^9 simulation steps each. The design of the "2D tile" used as the model system for the investigation of the weave pattern and other structural properties of 2D origamis is shown in Fig. B.3

B.5.3 Calculating an average structure

We used a relatively simple method to calculate an "average structure" of an origami from a simulation trajectory (i.e. a series of configurations/simulation snapshots) from an origami simulation.

This method has two steps: Firstly, local axes are computed for the origami for each configuration. This is done as follows: the x-axis is defined as the average direction of the vectors pointing along the double helix axes. For 2D origamis, a vector $\mathbf{v}$ is defined as the normal of the plane of the origami, which itself is given by the cross product between the average direction of the vectors pointing along the double helices and the average inter-helix vector between the top and bottom double helices. For 3D origamis, $\mathbf{v}$ is the average inter-helix vector between helices that are in the same row in the origami's cadnano[111] design. The z-axis is defined as the direction of the cross product between a vector pointing along the x-axis and $\mathbf{v}$, and the y-axis is defined

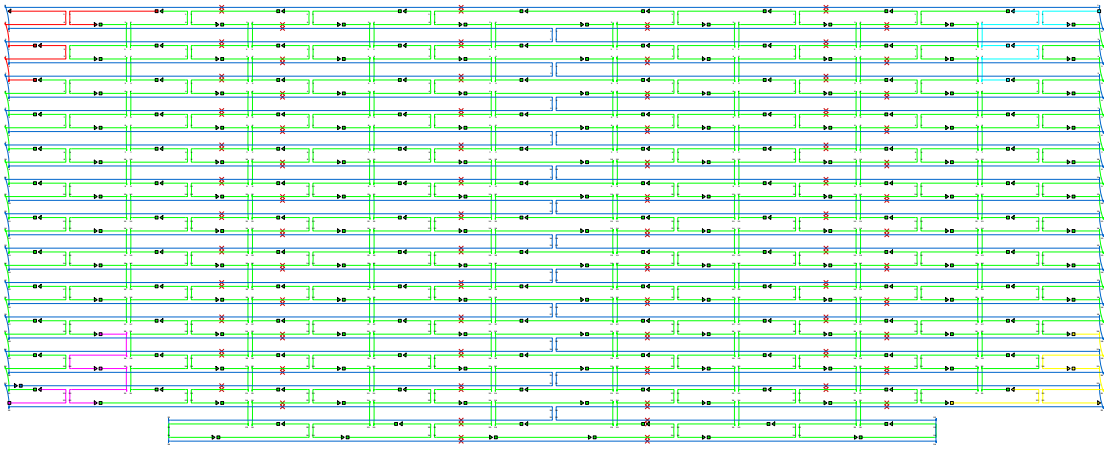

Figure B.3: The cadnano[111] design for the 2D tile which was used as a model system for the investigation of the structural properties of 2D origamis in Section 6.3. The scaffold strand is shown in blue, the staples are in green, and the red crosses indicate deletions (bases that have been removed from the design). Pairs of horizontal lines represent double helices and vertical lines represent junctions between them. Squares indicate the $5'$ end and triangles indicate the $3'$ end of each staple.

as the direction of the cross product between a vector pointing along the z-axis and a vector pointing along the x-axis. Secondly, for each configuration the centre of mass is placed at the origin, and the positions and orientations of the nucleotides in this local coordinate system are computed. These are then added to the running total, which is normalised at the end of the procedure.

Thus, the average structure is an arithmetic mean of the coordinates and orientations of the nucleotides, in a reference frame which moves with the origami. This simple scheme is adequate for relatively stiff structures, but tends to give less reliable results for structures with significant flexibility, as the average positions of the nucleotides in these flexible parts will tend to shrink towards the centre.

B.5.4 Comparison to experiment for the 3D "pointer" origami structure

For the pointer, we quantify the level of agreement between the oxDNA2 model's average structure and the experimentally determined one by finding the RMSD between the two structures. The RMSD was calculated at the level of base pairs. That is, we measured one distance for each base pair in the structure, namely the distance between the base-pair midpoint in the simulated structure and that in the experimentally determined structure. We then took the average of the sum of the squares of these distances, and took the square root of the result.

The structures were superimposed on each other to facilitate the RMSD calculation as follows: First, the structures were translated so that the centres of mass were both at the origin. Then a vector pointing along a certain reference double helix was found for each structure, and the experimental structure was rotated so that these vectors were aligned. Using this as a starting point, a simple zero-temperature Monte Carlo algorithm was used to find the orientation of the experimental structure that minimised the RMSD. Trial moves were rotations with a random angle and rotation axis, and moves were accepted only if they lowered the RMSD. This procedure was repeated several times to check the result was robust and not just a local minimum.