Exploration of DNA systems under internal and external forcing using coarse-grained modelling



Megan Clare Engel

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2018

For Laurelin

Abstract

The profound simplicity and versatility of the molecule at the heart of all earth-bound life forms, DNA, continues to inspire new frontiers of scientific inquiry. Central to many of these, including the *de novo* design of novel DNA nanostructures and the use of DNA to probe the principles of biological self-assembly and the operation of cellular nanomachines, is the interaction of DNA with forces, both internal and external. This thesis comprises a survey of three key ways coarse-grained simulations using the oxDNA model can contribute to efforts to characterize these interactions. First, a non-equilibrium data analysis framework based on the Jarzynski equality from statistical physics is validated for use with oxDNA through the reconstruction of free energy landscapes for canonical DNA hairpin systems. We provide a framework for assessing errors in the method and apply it to study a system for which conventional equilibrium simulations would be impractical: DNA origami 'handles' proposed for use in force spectroscopy experiments. Next, we simulate the forcible unravelling of three DNA origami structures, the largest systems yet studied with simulated force spectroscopy. We combine these results with experimental AFM data to probe the mechanical response of origami in unprecedented detail, highlighting the effect of nanostructure design on unfolding behaviour. Lastly, we examine the validity of using widely-employed polymer elastic models to predict internal entropic forces in ssDNA. We develop a framework for measuring internal forces in the oxDNA coarse-grained model and apply it to analyze the pico-Newton range forces exerted by a recently proposed DNA origami force clamp, ultimately concluding that conventional means of estimating internal ssDNA forces are often inaccurate and should be supplemented with coarse-grained simulations. In addition to providing new insights about the DNA systems we present, our results highlight the significant fruits of complementing experimental studies with coarse-grained simulations.

Acknowledgements

No Quest, from the cosmically resonant to the relatively insignificant, can ever be completed by a person in isolation. Having been possessed of this wisdom since the first time I read *The Lord of the Rings*, it is curious indeed that I chose to put an ocean between myself and every soul I had ever loved to embark upon this academic endeavour. My need, however, for a Fellowship and community was answered in force.

Naturally, I owe a great debt to my academic supervisors, Jonathan Doye and Ard Louis, whose extensive knowledge, insight, aid – particularly in the writing of this thesis – and investment in me I appreciate deeply. Likewise, the generously-bestowed help from research group members, past and present, proved essential to my progress. In particular, I am thankful to Flavio Romano, John Schreck, Lorenzo Rovigatti, Majid Mosayebi, Ben Snodin, Garima Mishra, and those with whom sharing an office brightened the oft-grey Oxford days: Domen Prešern, Dan Tracey, Max Tortora, Ferdinando Randisi, and Chico Camargo. And to my academic mentor, Jim Williams: thank you for your help, without which I absolutely would not have made it this far.

I am also deeply grateful to the Rhodes Trust and to the Natural Sciences and Engineering Research Council of Canada for their financial support; their faith in me facilitated not only my studies, but all of the invaluable personal growth attendant upon them.

The remarkable friends I encountered in Oxford have left an indelible imprint upon me, and I could never have persevered in completing this document without their moral support, encouragement, inspiring example, and love. In no particular order – thank you to the 9PM Choir – especially Jean, John, and Paul – for the music, and for everything! And to the warm shepherds at the Chaplaincy who adopted my wandering soul – Andrew, Alex, Fr. Dushan, Fr. Matthew. To Eden & James, whose humour, compassion, and brilliance have forged moments that glow like a string of jewels in my memory. To Vinesh, to whom I am indebted on personal, academic, and spiritual levels for enlightening and challenging discourse and friendship, and to the beautiful Victoria: you showed me that Oxford could be just as wonderful as Home. To Michael Hawking for his conviction and unfailing ability to extract my smile even in circumstances of extreme duress. To Paul, whose

kindness moves me to tears and who has shown gratuitous love. To Madeleine, whose companionship has been constant succour. To Leena, whose generosity of spirit encompasses oceans. To Anne with an 'e'. To Aurora, who compelled me to be better than I am. To Lonn and Jenni, for bringing a piece of home to England and luxuriating in so many shared interests – and puns – with me. All of you have made 6 744 km away-from-home... home.

But this thesis would have been likewise impossible without the network of kindred spirits sending love from a distant shore. My bosom friend, Heather; Stephie and Krissie, whose faith in me never wavered; E3, without whom no homecoming was complete. Heidi, Brian, and Theresa – the parents and sister I am so fortunate to have gained and whose love and endearing packages have brightened our little Oxford Hobbit-hole. And my family: my cornerstone, because of whom leaving home was as violently painful as losing a limb. Thank you to Jen-Pen, my sweet sister, for our singular bond. To my Mom, for her unwavering lavishing of asymptotically-unconditional love, and for giving me everything I needed to make it to this point. And to my Dad: even when writing a physics thesis is the task in question, in the end, the most important thing you showed me wasn't that general relativity demo when I was 7! It was love, of course. Thank you for the lilacs & the bubble gum.

I am unequal to conveying the inexpressible: my gratitude to, and love for, my husband, Mitchell V. Engeldell, my vocation, dearest friend, sustenance, and truest Home. Thank you for accompanying me on this, and every, Adventure; "Frodo wouldn't have got far without Sam, would he, dad?" And to our sweet daughter: the long hours of thesis-writing were brightened incomparably by your dear little dances.

In truth, all thanks must lead to the source and summit of all my life and all that emanates therefrom: Christ, ever the end of my seeking.

After 1 648 days, 161 pages, and ∼208 Angrid Thai chilli chicken stir-fries (without onions), rather circularly, my main conclusion is something I already knew: "not all those who wander are lost."

Contents

List of Abbreviations **viii**

1 Introduction **1**
1.1 Modelling DNA 4
1.1.1 Single-nucleotide level coarse-grained DNA models 6
1.1.2 Modelling of DNA Nanotechnological Systems 11
1.1.3 The oxDNA Model 13
1.2 Thesis structure 15

2 Simulation Methods **16**
2.1 Simulation Methods 16
2.1.1 Molecular Dynamics Simulations 16
2.1.2 Monte Carlo Simulations 18
2.1.3 Umbrella Sampling 19
2.1.4 Weighted Histogram Analysis Method 20

3 Non-equilibrium bio-molecular unfolding under tension **23**
3.1 Introduction 23
3.2 Theory 28
3.3 Application of non-equilibrium landscape analysis to two DNA hairpins 34
3.3.1 Benchmarking with experimental results 34
3.3.2 Non-equilibrium landscape reconstructions 40
3.4 Application of non-equilibrium landscape analysis to a novel system: DNA origami handles 61
3.4.1 System description and methods 62
3.4.2 Results 65
3.5 Conclusions and outlook 71

4 Force-induced unravelling of DNA origami **73**
4.1 Introduction 73
4.2 Methods 77
4.3 Rothemund Tile 79
4.4 10-Helix Bundle 89

4.5 7-tile Biosensor 97
4.6 Conclusions 99

5 Measuring internal forces in single-stranded DNA 101
5.1 Introduction 101
5.1.1 Models of ssDNA Elasticity 103
5.2 Force propagation in the oxDNA model 110
5.2.1 Theory 113
5.2.2 Secondary structure 116
5.2.3 Method for calculating internal forces in ssDNA simulations 118
5.3 Application of internal force calculation to a single-molecule force clamp 121
5.3.1 Introduction and system description 121
5.3.2 Simulations 124
5.3.3 Results 130
5.4 Conclusion and outlook 136

6 Conclusions 138

Bibliography 142

Appendices

A Experimental methods and additional details supplementary to Chapter 4 159

List of Abbreviations

AFM . . . . . Atomic Force Microscope.

bp . . . . . Base pair.

DNA . . . . . Deoxyribonucleic acid.

dsDNA . . . . Double-stranded DNA.

exFJC . . . . . Extensible freely-jointed chain.

FEC . . . . . Force-extension curve.

FENE . . . . . Finitely extensible nonlinear elastic.

FJC . . . . . Freely-jointed chain.

LOT . . . . . Laser optical tweezers.

MC . . . . . Monte Carlo.

MD . . . . . Molecular dynamics.

nt . . . . . Nucleotide.

PMF . . . . . Potential of mean force.

PSF . . . . . Point spread function.

RNA . . . . . Ribonucleic acid.

SMFS . . . . . Single molecule force spectroscopy.

ssDNA . . . . Single-stranded DNA.

US . . . . . Umbrella sampling.

VMMC . . . . Virtual move Monte Carlo.

WHAM . . . . Weighted histogram analysis method.

WLC . . . . . Worm-like chain.

The Word is living, being, spirit, all verdant greening, all creativity. This Word manifests itself in every creature.

— St Hildegard von Bingen

1 Introduction

Contents

1.1 Modelling DNA 4
1.1.1 Single-nucleotide level coarse-grained DNA models . . . 6
1.1.2 Modelling of DNA Nanotechnological Systems 11
1.1.3 The oxDNA Model 13
1.2 Thesis structure 15

At least as deep-rooted in humanity as the curiosity that drove the ancients to study the heavens is the inward-facing, existential questioning that seeks to understand ourselves: life, and its unyielding forward motion. Inarguably a key to understanding life – both its primordial origins[213] and human biology[1] and evolutionary history[218] – is its fundamental building block, DNA, the information-carrying role of which was seminally elucidated by Watson & Crick[285] when they identified its iconic double-helical structure, shown in Figure 1.1. The outer backbone of the DNA helix is comprised of sugar and phosphate, and the inner rungs contain four types of nucleic acid bases – adenine (A), thymine (T), guanine (G), and cytosine (C) – that bind together predominantly in A-T, C-G pairings. It is this specificity of base pairing that is crucial to the transmission of genetic information: each strand of the double-helix bears a genetic 'blueprint' that is copied during the process of DNA replication[1]. The second pillar of DNA's enormous biological relevance is

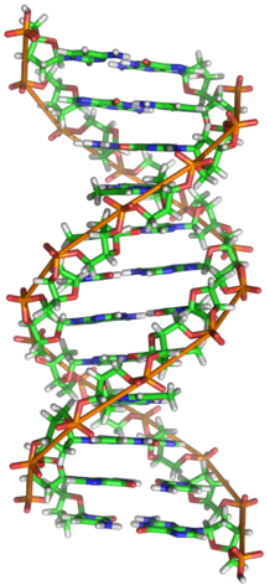


Figure 1.1: The iconic DNA double helix, proposed by Watson & Crick[285]. Sugar and phosphate molecules covalently link to form the backbone, and the rungs of the ladder are comprised of bases, which come in four flavours. Image from Ref. 38.

its role in the so-called 'central dogma of molecular biology'[44]: all proteins, the workhorses of cells, are encoded by DNA, as are RNA molecules – chemical relatives of DNA with diverse roles in protein construction and reaction catalysis. Current research tantalizingly suggests that even DNA sequences that are not involved in protein encoding may play fundamental, as-yet mysterious roles in the cell[56].

The same simplicity of structure and base-pairing specificity at the heart of DNA's biological significance allows for the possibility of rationally designing artificial DNA structures, as first suggested by Seeman[247] when he proposed that DNA could be manipulated to form self-assembling crystalline networks. This ingenious idea spawned the field of DNA nanotechnology, and from the first simple arrangements of two DNA helices linked by crossover junctions to impart enhanced mechanical stability[76], the complexity of architectures successfully constructed experimentally has soared. Not only has Seeman's vision of DNA crystals been realized in both two[293;286] and three[135] dimensions; a colourful array of nanoengineering elements have emerged including polyhedra[103], tensegrity structures[155], and nanotubes[174]. Capitalizing on duplex strand-replacement reactions[305], dynamic

DNA nanomachines have also been constructed, from DNA 'walkers' – motors that are able to navigate along a track[131;291] – to nanocages[57] designed for targeted drug delivery. The applications of such remarkable feats of nanoengineering are seemingly endless, spanning nanosensing[273], medicine[249], electronics[210], and even DNA-based computation[81].

The fact that DNA structure and thermodynamics are relatively well-understood renders it of great interest from a fundamental theoretical perspective: it represents a model system that can be used as a platform to explore the properties of single molecules and microscopic physics more generally. Many single-molecule techniques and theorems have been validated using simple DNA systems[96;63], and DNA is routinely employed as part of the instrumental apparatus that makes examination of other single biomolecules possible[303;209]. In particular, DNA has played a crucial role in the exploration of the 'folding problem': the quest to understand how the proteins encoded by a one-dimensional DNA sequence are able to self-assemble into three-dimensional functional structures so reliably[55]. By applying external forces to disassemble simple DNA molecules, their unfolding thermodynamics and kinetics can be repeatedly probed at the single molecule level, leading to insights about the out-of-equilibrium physics that governs biology.

The fruitful synergy between force and DNA doesn't end there. In biological contexts, the application of force to DNA is a central ingredient in DNA replication and transcription, as special molecules forcibly open the DNA helix to access its contents[1]. Mechanical force also mediates the extremely dense packaging that enables ∼metres of DNA to be stored in a ∼5 µm-diameter nucleus, and even plays a role in regulating genetic expression[276]. In nanotechnological contexts as well as biological ones, DNA's relationship with force is significant: nanostructure mechanical stability must be designable for applications involving force exposure, and the tensegrity structures mentioned above not only have to assemble against force, but feature constant tension throughout as part of their design[255]. The response of DNA to force is even relevant to mechanically-regulated cellular processes

involving molecules other than DNA – such as molecular motor transport, force-gated membrane channels, and intercellular communication facilitated by tension in cellular fibres – because the study of these processes is often carried out by DNA nanostructures, used to measure forces in[91;95] and apply forces to[193] other biological molecules.

Given its manifold applications, the study of DNA behaviour under force is important. While experimental work has contributed significantly to this end, as we shall see, computational DNA modelling offers unique, otherwise-inaccessible insights, supplying detailed microscopic and structural information that can complement – and guide – experimental observations.

1.1 Modelling DNA

The most detailed and realistic computational representations of DNA retain a quantum mechanical description of constituent subatomic charges, solving the Schroödinger equation to obtain electronic structures. These quantum chemical simulations are capable of reproducing quantitatively accurate descriptions of DNA systems – for example, base stacking and base pairing energies[126]. Quantum chemical computations have provided descriptions of backbone orientations[183] and aromatic base stacking[127] in DNA, and provide a route to capturing the molecular polarization underlying DNA-ion interactions[83]. While the quantum level of detail is the gold standard in computational DNA studies, intense computational requirements limit the applicability of quantum chemistry simulations to nearest-neighbour base pairs.

At the next finest level of description are classical atomistic models, which represent interactions between individual atoms using effective potentials implemented using force fields like AMBER[40] or CHARMM[24]. Such models can capture detailed microscopic interactions between DNA and explicit solvent atoms as well as other molecular species like proteins[148;162], and are able to access larger system sizes than quantum chemistry simulations. Atomistic molecular dynamics (MD) simulations have been profitably applied to the study of DNA behaviour under force, including stretching[99;250], twisting[216], and supercoiling[181;271], and have recently begun to

illuminate the mechanical behaviour of synthetic DNA nanostructures[302]. All-atom modelling has drawbacks, however: computational demands limit accessible timescales to ∼microseconds[148;164] and system sizes to a few thousand nucleotides at the maximum[302], which also renders sampling rare events difficult.

Lying at the opposite extreme are two popular simplified approaches. Continuum models that treat DNA as a semiflexible polymer using theoretical descriptions like the freely-jointed chain[229] or worm-like chain[197] models have successfully been applied to describe the large-scale behaviour of DNA under twist[191;93] and tension[259;169]. Additionally, thermodynamic models that are parametrized using empirical data to quantify the free energy of DNA based on contributions from each set of interacting nearest-neighbour nucleotides[235] are widely used to predict stable DNA structures and melting temperatures. Both of these schemes lack the ability to describe the microscopic dynamics of DNA, however, including all-important hybridization and unfolding processes.

A compromise between these two limits is afforded by coarse-grained (CG) DNA models. In such models, multiple atoms are grouped into particles that experience effective interactions, thus reducing the number of degrees of freedom, and solvent molecules are usually not modelled explicitly. While these simplifications greatly increase simulation efficiency, they also introduce a "representability problem": no CG model is able to accurately capture *all* aspects of system behaviour, and the type of model used must be chosen bearing in mind the specific requirements of a problem[160]. Simulating the interaction of DNA with force necessitates a model able to describe the mechanical properties of both single-stranded DNA (ssDNA) and double-stranded (dsDNA), as well as capture the hybridization reaction accurately, such that the breaking of individual base pairs in response to tension can be represented.

Among the most efficient CG models are finite-element-type models that represent DNA helices as elastic rods; most notably, the CanDo model of Kim et al.[137]. This model is able to capture equilibrium conformations and mechanical flexibility of DNA nanostructures constructed using a technique known as 'DNA

origami', in which a long 'scaffold' strand is arranged into designed structures with the aid of short 'staple' strands that bring distant regions of the scaffold into proximity[228]. DNA base pairs are represented as finite elements of continuous elastic rods, single-stranded regions are represented as springs, and connections between DNA helices are represented as rigid crossovers. While finite element methods work well for describing the expected structures and thermal fluctuations of fully formed DNA origami nanostructures, they cannot treat assembly or disassembly.

Similarly computationally efficient models coarse-grain at the level of multiple base pairs. For example, domain-level models[47] have as their fundamental units regions of contiguous origami staple base pairs, and these have been successfully applied to understand dynamic DNA nanostructure assembly[62]. A model recently introduced by Reshetnikov et al.[221] represents 7 base pairs by a single coarse-grained particle, and has been shown to reproduce DNA origami structures and dynamics. Even more detailed are models which assign one coarse-grained bead per base pair, such as the model of Mergell et al.[178]; a generalization of this model achieved reasonable agreement with experimentally measured twist and persistence length for DNA origami structures[11].

While the above strategies can successfully describe global configurations of large-scale DNA structures like origamis, models that coarse-grain at the level of one or more base pairs cannot capture the individual base pair breaking that is crucial for understanding tension-mediated unfolding. To access the microscopic details of DNA, models that coarse-grain at the level of single nucleotides are necessary. We provide a brief review of this most detailed level of coarse-grained DNA modelling below; a good recent review can also be found in Reference 49.

1.1.1 Single-nucleotide level coarse-grained DNA models

Single-nucleotide level models can broadly be classified according to a few main factors. First, the number of coarse-grained beads assigned to each nucleotide controls the level of realistic detail retained by the model and impacts its computational efficiency. Next, whether solvent particles and ions are represented

implicitly, using a mean-field description, or explicitly will govern how well the model can expect to capture electrostatics and salt-dependent dynamics. Another important consideration is whether the parametrization of the model is "top-down" or "bottom-up." Top-down descriptions obtain the effective potential parameters by fitting the model to empirical data, such as DNA melting temperatures, structural details, and mechanical properties. Bottom-up approaches attempt to fit the coarse-grained model predictions to results from finer-grained, usually atomistic models. One potential problem to bear in mind with bottom-up approaches is that the coarse-grained model inherits any deficiencies in the fine-grained description used for parametrization, and there is no guarantee that the finer-grained model – limited to small-scale investigations due to computational constraints – accurately captures the larger-scale properties the coarse-grained model is used to probe[59].

The simplest nucleotide-level models assign one bead to each nucleotide. Savelyev & Papoian[237] presented a bottom-up, 1-bead-per-nucleotide (1BPN) model with explicit ions that enabled them to reproduce the dependence of DNA persistence length on ionic strength[237;236], and it has been used to study DNA unwrapping from nucleosomes[30]. The model lacks sequence dependence, however. Another bottom-up 1BPN model that retains explicit ion representation is that of Naômé et al.[187], which also captures salt-dependent persistence lengths, and has correctly reproduced the topologies of DNA minicircles and the probability of closure of DNA rings[188]. The recently-proposed nCG-dsDNA model of Yagyu et al.[300] was parametrized using a top-down approach, but using only structural – not thermodynamic – data. While it retains sequence dependence, it has not yet been rigorously validated against experimental results. Furthermore, all three of the above-mentioned models are capable of describing DNA in its duplex state only, and cannot capture the hybridization dynamics crucial to investigations of force-induced dissasembly. One 1BPN model that *is* capable of describing hybridization is that of Araque et al.[9] (see also Reference 10), parametrized bottom-up, but from quantum chemical simulations. It neglects electrostatics entirely, however, is computationally expensive, and has only been used to treat short sequences.

Furthermore, it is not expected to reproduce melting temperatures are accurately as models based on thermodynamical parametrization[9].

Retaining more structural detail are 2-bead-per-nucleotide (2BPN) models, like that of Maffeo et al.[163], which represents solvent implicitly but can nonetheless capture salt-dependent effects. The bottom-up model was originally constructed to represent ssDNA only, and captures the force-extension behaviour of ssDNA well. It was recently extended to describe dsDNA and applied to the study of dsDNA translocation through nanopores[17;215]. Sequence dependence, hybridization and disassembly cannot be described by the model, however. The other major contender in the 2BPN category is oxDNA[203;281;261] , which features three interaction sites per nucleotide and is able to represent both ssDNA and dsDNA and interconversions between them. Basic DNA biophysical properties, including the elasticity of ssDNA and dsDNA as well as duplex melting behaviour, are well-described by oxDNA[203], as is force-induced duplex rupture behaviour[226;185]. Furthermore, as will be recounted in section 1.1.2, oxDNA has succeeded in modelling the assembly of large DNA nanostructures[262] and has been productively combined with experiments to investigate dynamical DNA nanomachines[202;136] and with atomistic simulations to elucidate the behaviour of highly stressed, supercoiled DNA minicircles[271].

There are several strong models at the 3-beads-per-nucleotide (3BPN) level. Among the earliest examples is the model of Morriss-Andrews et al.[184], which is parametrized from all-atom simulations and represents solvent ions implicitly. A unique strength of the model is the ellipsoidal representation of bases, which combined with anisotropic potentials allows realistic base-stacking interactions to be captured. The model describes temperature, salt, and sequence-dependent properties of ssDNA and dsDNA reasonably well, but produces persistence lengths that are a factor of $\sim$2 too low. The top-down parametrized model of Linak et al.[156] can describe Hoogsteen bonds, which allows it to capture more exotic DNA motifs, and has successfully reproduced experimental DNA hairpin melting temperatures. The more recent BioModi 3BPN model of Markegard et al.[166] has been optimized for large systems and the study of hybridization kinetics, and as

such is suitable for DNA nanotechnology studies. It has been used to study the strand concentration dependence of hybridization reactions[167], but its predicted duplex melting temperatures do not agree quantitatively with experimental results as well as those of oxDNA[166]. The above models have not yet been widely used or tested, nor have they been applied to the study of nanotechnological DNA systems.

Among the more extensively used single-nucleotide-level models is the top-down 3-site-per-nucleotide (3SPN) model of Knotts et al.[139]. Improvements made to the model over time[233;105;72] have resulted in a current version, 3SPN.2C, that incorporates anisotropic potentials including 3-body, as well as pairwise, interactions; captures major-minor grooving and sequence dependence; reproduces persistence lengths of ssDNA and dsDNA well; and achieves quantitative agreement with experimental DNA melting temperatures. It has been used to study nucleosome formation[73], tension-mediated nucleosome unwrapping[151], and DNA hybridization[106], and, impressively, has been adapted to explore composite DNA-protein[53] and DNA-nanoparticle[152] systems. One drawback of the model, however, is that it was originally parametrized for duplex DNA and consequently predicts ssDNA to be too stiff; while this issue has been mitigated in the latest version, it still exists. The 3-body interaction terms will also render it less efficient than models that include only pairwise interactions.

The Three-Interaction-Site DNA (TIS-DNA) model very recently proposed by Chakraborty et al.[32], parametrized top-down using DNA crystal structures and melting data, captures salt and sequence dependent effects, including salt dependent persistence lengths, and is suited to modelling hybridization. While promising, it has not yet been widely tested.

Finally, the most detailed single-nucleotide-level models assign 4 or more beads to each nucleotide. The bottom-up, 5-bead-per-nucleotide model of Korolev et al.[142] incorporates explicit solvent ions and reproduces salt-dependent persistence lengths of duplex DNA[143], but cannot treat ssDNA or, consequently, hybridization. Similarly unable to deal with hybridization reactions is the Martini-force-field-based model of Uusitalo et al.[277], which uses 6 or 7 beads-per-nucleotide (depending on the base)

and combines bottom-up and top-down parametrizations. It is well-suited, however, for multi-component simulations of DNA interacting with other biomolecules, as it can be easily interfaced with other coarse-grained models. The Martini model's accuracy in reproducing experimental persistence lengths and DNA structures is limited, however. Better structural agreement with nucleic acid data is achieved by the Nucleic Acid United Residue 2-Point (NARES-2P) model of He et al.[102], featuring 6-8 beads-per-nucleotide (BPN). The model exhibits poor agreement with experimental thermodynamic data, however, and thus is a poor choice to model hybridization and disassembly reactions.

The 6-8 BPN SIRAH model[50;48] can treat solvent explicitly and is able to capture DNA-ion binding interactions. It also lends itself to multiscale, multicomponent modelling, as it can be interfaced with atomistic simulations[161] and protein models[22]. It has not been optimized to reproduce the mechanical behaviour of ssDNA and dsDNA, however, which is fundamental to DNA-force studies, and has thus far only been used for very small systems (∼10s of nucleotides). Lastly, the 6-7 BPN High-Resolution DNA (HiReDNA) model of Cragnolini et al.[43] incorporates 3- and 4-body interaction terms, can model duplex hybridization, and is notable in its ability to capture non-canonical base pairing, allowing treatment of more exotic structures[42]. While its strength is its level of detail, this also renders it comparatively computationally inefficient, and simulating large nanostructures with HiReDNA is not feasible.

Of the coarse-grained models presented above, oxDNA, TIS-DNA, and the 3SPN.2C models all offer good mechanical descriptions of ssDNA and dsDNA, have been parametrized using experimental data and can represent hybridization and disassembly well, can capture salt- and sequence-dependent effects, and constitute a good compromise between computational efficiency and representational detail; they are thus reasonable choices for studies of DNA-force interactions. Furthermore, oxDNA and 3SPN.2C are among the most easily accessible of the aforementioned models, being freely available for download; oxDNA has recently even been incorporated into the widely-used LAMMPS MD code[104]. Only oxDNA, however, has been

extensively validated in the study of large DNA nanostructures[260;262;251;243]. In the current work, therefore, which seeks to describe the force response behaviour of DNA on a variety of scales, from short single strands to kilobase-long nanostructures, we employ the oxDNA model, which complements other modelling strategies that have been utilized in the study of DNA nanotechnological systems.

1.1.2 Modelling of DNA Nanotechnological Systems

Computational studies have contributed to the multifaceted development and understanding of DNA nanotechnology, from nanostructure mechanical properties to assembly processes. The aforementioned CanDo finite element model of Kim et al.[137], first presented in Reference 31, has been popularly and successfully applied to efficiently predict the global, minimum-free-energy conformations of nanostructures, and is freely available for download. In particular, the model has reproduced experimental 3D structures of a variety of DNA origami objects, from planar structures to crystalline tensegrity motifs[208], and been used to analyze origami mechanical deformations under a loading force[270]. CanDo has been extended to capture dynamics, as well – namely, thermal fluctuations about equilibrium conformations, and thus mechanical flexibility of nanostructures – and been applied to study conformational dynamics of DNA tweezers and various origami objects[246]. The early versions of the model lacked the ability to describe nicks (discontinuities in the phosphodiester backbone bonds in duplex regions) and bulges (regions of mismatched base pairs in a duplex), but recently the framework was parametrized to describe these features using atomistic simulations. This improved version of CanDo produced reasonably accurate predictions of the equilibrium structures of DNA polyhedra as compared with experimental data, but could not capture all features, such as the twisting of vertices in such structures, the description of which requires higher-resolution CG models[207]. Additionally, the model is limited by its inability to describe helix unzippering, secondary structures in single-stranded regions, and structures that have multiple stable states instead of a single free-energy

minimum. It also does not account for excluded volume interactions, which can lead to incorrect structural predictions for flexible origamis[251].

Other models that have been profitably applied to describe equilibrium structures and conformational dynamics of DNA origami structures were described in section 1.1.1. These include an extension to the 1-bead-per-base-pair CG model of Mergell et al.[178], which, as mentioned above, reproduced experimentally-observed flexibilities of origami 4-helix bundles[11], and the recently-proposed 1-bead-per-7-base-pairs model of Reshetnikov et al.[221].

Over the past few years, significant work at the all-atom level has begun to supplement the global picture provided by models like CanDo with local, microscopic details. The first atomistic origami simulations were performed by Yoo & Aksimentiev[302], and detailed local fluctuations for rod-like and bent origami objects. Not only have atomistic simulations succeeded in predicting experimentally-observed global origami structures[138], but their ability to capture DNA-ion interactions has enabled them to characterize the ionic conductivity of DNA origami placed over a nanopore – revealing that origami conductance is orientation-dependent[154] – and elucidate the microscopic conductance mechanisms of synthetic transmembrane ion channels constructed using DNA duplexes[88] and DNA origami[89]. Comparative atomistic studies of nanostructures built using DNA origami and those built using DNA bricks – short single strands of DNA that can be combined modularly to form larger structures – have also revealed that DNA-bricks-based assembly produces structures that are more compliant under external force than DNA origami[256]. While atomistic simulations have contributed greatly to DNA nanotechnology modelling, they are limited, as already discussed, in the time scales and system sizes that can be accessed.

To optimize the rational design of artificial nanostructures, an understanding of their self-assembly processes is necessary. Multiple modelling strategies have been applied to this end. The domain-level model of Dannenberg et al.[47] already mentioned has reproduced experimental results concerning the hysteresis between DNA origami annealing (formation) and melting curves, as well as the dependence

of annealing rates on staple concentration[47]. It has also been used to describe kinetically-steered origami assembly, in which different final structures are produced by applying small modifications to the staple set[62]. Self-assembly of a variety of 3D DNA brick nanostructures[220] has been studied using the lattice model of Reinhardt & Frenkel[219], in which DNA strands are portrayed by particles possessing rigid patches that represent binding domains. DNA brick self-assembly has also been studied using a recent theoretical kinetic model proposed by Fonseca et al.[69], parametrized using oxDNA, which reproduced experimentally-observed formation temperatures for 2D DNA brick structures.

Nucleotide-level coarse-grained models constitute a complementary simulation strategy to the above methods, offering a compromise between microscopic detail – particularly the ability to capture base pair breakage – and the efficiency necessary to simulate very large structures and probe rare events and assembly processes. As aforementioned, oxDNA is the only such model that has yet been widely applied to the description of nanotechnological systems. On the level of structural insights, oxDNA has been used to accurately predict bulge-size-dependent experimental conformations of DNA prisms[243] and the opening angles of DNA origami hinges[252;251]. It has also succeeded in simulating dynamical processes central to many DNA nanotechnological systems, including toehold-mediated strand displacement[265], the motion of DNA nanomachines that "walk" along tracks[202;269;136], and the operation of moveable crank-slider joints[251]. Finally, oxDNA has also provided insights on full self-assembly processes for DNA origami[262] and knotted DNA nanostructures[140]. Given its broad validation and utility in a variety of nanostructural contexts, in addition to the aforementioned ability to capture processes relevant to DNA-force interactions, simulations in this thesis were pursued with the oxDNA model.

1.1.3 The oxDNA Model

The oxDNA model has been extensively described previously[203;281;261]. Briefly, oxDNA represents nucleotides as rigid bodies featuring three interaction sites*; see

*Backbone, stacking, and hydrogen-bonding interactions are localized to distinct sites.

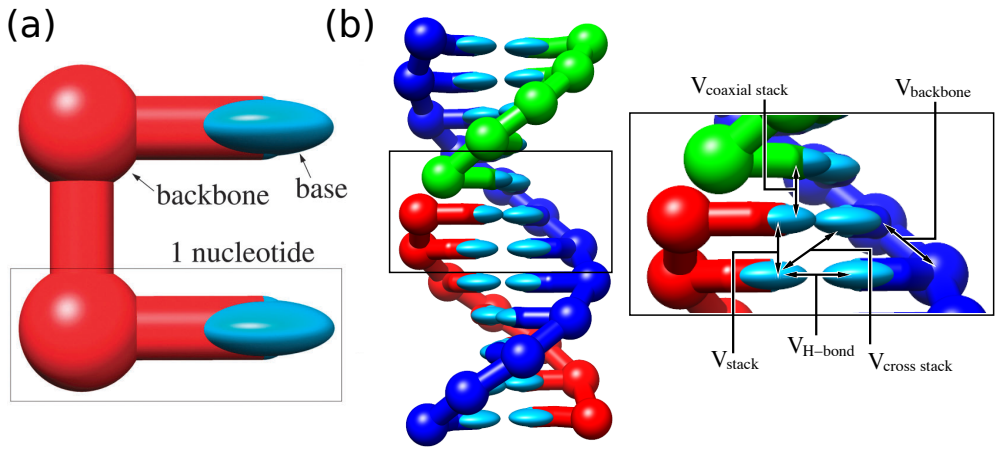


Figure 1.2: (a) oxDNA representation of a strand composed of two nucleotides and (b) a duplex featuring 3 strands. Each rigid-body nucleotide has three interaction sites: one on the backbone, and two on each base, one for stacking and one for hydrogen bonding interactions. The directions of coaxial stacking, hydrogen bonding, stacking, cross stacking, and backbone potentials are depicted. Figure adapted from Doye et al. [59]

Figure 1.2. Parameters in oxDNA were obtained using the "top-down" approach described above: the model was fitted to experimental results for structural, thermodynamic, and elastic properties of DNA. Nucleotides interact via pairwise potentials chosen to incorporate salient physical interactions, and the full potential energy function for the most recent version of the model, oxDNA2, is

$$V_{\text{oxDNA2}} = \sum_{\text{nearest neighbours}} (V_{\text{backbone}} + V_{\text{stack}} + V_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross stack}} + V_{\text{exc}} + V_{\text{coax stack}} + V_{\text{DH}}), \tag{1.1}$$

where V_{backbone}, V_{stack}, and V_{exc} represent backbone connectivity, stacking, and excluded volume interactions, respectively, and are summed over all pairs of nearest-neighbour nucleotides, and V_{HB}, $V_{\text{cross stack}}$, V_{exc}, $V_{\text{coax stack}}$, and V_{DH} are hydrogen bonding, cross stacking (i.e. stacking between bases on opposite duplex strands), excluded volume, coaxial stacking (i.e. stacking between contiguous bases on different strands, as across a nick), and electrostatic (according to Debye-Huckel theory) interactions, respectively. The detailed forms for each of these terms can be found in Chapter 2 of Reference 201 and Appendix A of Reference 260. While the model can capture salt, temperature, and sequence-dependence, it cannot represent

the detailed interactions of DNA with solvent ions or any non-canonical base pairing, though these shortcomings are not expected to affect the current work significantly.

1.2 Thesis structure

This thesis surveys three distinct ways in which coarse-grained modelling can contribute meaningful understanding to the study of DNA force response. Chapter 2 serves as a brief primer to the main simulation and analysis methods employed throughout the work. In Chapter 3, we'll explore how the force-mediated unfolding of simple DNA systems, despite occurring far from equilibrium, can reveal information about equilibrium free energies. While widely applied to experimental studies, the methods we adapt have heretofore not been explored in the context of a sophisticated coarse-grained DNA model like oxDNA. Next, Chapter 4 takes the same basic principle of unfolding DNA structures with force to an extreme, featuring the forcible unravelling of some of the largest DNA nanostructures ever simulated with a nucleotide-level coarse-grained model. The results presented shed light on the mechanical properties and unfolding pathways of DNA origami systems. Our third foray into force exploration, Chapter 5, focuses on the relationship between DNA and internal, rather than external, forces, specifically in the context of DNA nanostructures. We take a closer look at canonical models of the elastic behaviour of ssDNA that are widely applied in both experiments and simulations, cautioning that key details omitted by these models have led to inaccurate published estimations of force. In each of these three chapters, the potential fruits of using oxDNA simulations to supplement and refine experiments are highlighted; our main results are finally summarized in Chapter 6.

. . . No Doctor, I simply maintain that computers are more efficient than human beings, not better.

—Cdr. Spock

2 Simulation Methods

Contents

2.1 Simulation Methods 16
2.1.1 Molecular Dynamics Simulations 16
2.1.2 Monte Carlo Simulations 18
2.1.3 Umbrella Sampling 19
2.1.4 Weighted Histogram Analysis Method 20

The oxDNA code suite* is capable of running simulations in two modes: Monte Carlo (MC) and Molecular Dynamics (MD); each of these simulation techniques is introduced below.

2.1 Simulation Methods

2.1.1 Molecular Dynamics Simulations

In molecular dynamics (MD) simulations, Newton's equations of motion are numerically integrated for each coarse-grained particle to determine its trajectory. The forces in these equations are calculated using the derivatives of the potentials described in section 1.1.3. To integrate the equations, oxDNA employs the velocity Verlet algorithm[272], which is time-reversible and preserves the phase-space volume

*Available open source from https://sourceforge.net/projects/oxdna/files/

of the simulated system according to Hamiltonian evolution[74]; this leads to minimal long-term drift of system energy over the course of a simulation. For simulations in this thesis, a discrete integration time step of $\mathrm{d}t = 0.005$ simulation units, or 15.2fs, was used. This timestep provides a good compromise between simulation speed and accuracy.

The fact that solvent is implicit in oxDNA renders it necessary to enforce evolution in a canonical ensemble; oxDNA uses the NVT (constant particle number, volume, and temperature) ensemble. Coupling between the system and a heat bath is imitated using an Andersen-like thermostat[5;230], which generates realistic diffusive dynamics. In this scheme, the system is permitted to evolve in the microcanonical NVE (constant particle number, volume, and energy) ensemble for a certain number of steps, N_{Newt}, then a subset of particles are randomly chosen to have their velocities redrawn from a Maxwell-Boltzmann velocity distribution at the simulated temperature; this mimics stochastic collisions with the solvent bath. The probability of such a collision occurring is inversely proportional to the translational diffusion coefficient for each particle[230]; for simulations in this thesis, D=2.5 simulation units, or 6.0×10^{-7} m^2 s^{-1}, and $N_{Newt} = 103$, or about once every 1.55 ps. Our diffusion coefficient is chosen to deliberately accelerate diffusive dynamics, which in oxDNA are $\sim$2 orders of magnitude faster than in experiment[280], enabling our simulations to access long-timescale processes more efficiently. The choice of collisional frequency N_{Newt} reflects a balance between computational efficiency and ensuring that systems behave diffusively on relevant timescales. Andersen thermostats have been shown to reproduce a canonical distribution[74].

oxDNA also provides the option of using a Langevin thermostat, which mimics coupling to a thermal bath through the addition of a frictional damping term and a random forcing term to Newton's equations of motion[3]. The strength of the random forces is related to the magnitude of the frictional damping through the 'fluctuation-dissipation relation', which guarantees evolution in the canonical NVT ensemble[3]. Because oxDNA simulations are somewhat slower when the Langevin, rather than Andersen-like, thermostat is used, it makes only a brief appearance in Chapter 5.

oxDNA MD simulations can be run on CPUs or GPUs, which offer a speed-up for systems larger than a few hundred nucleotides. All simulations of large DNA origami nanostructures in this thesis were run on GPUs.

2.1.2 Monte Carlo Simulations

When equilibrium values of observables, rather than full dynamical trajectories, are sought, Monte Carlo (MC) simulation can be applied[†]. In MC simulations, system configurations are sampled according to their Boltzmann weights; that is, over the course of a simulation, the amount of time the N-particle system spends in a configuration $\mathbf{r}^N$ is proportional to the Boltzmann factor, $e^{-\beta U(\mathbf{r}^N)}$, with $U(\mathbf{r}^N)$ the energy of the configuration and β the usual inverse thermal energy. This 'importance sampling' allows the simulation to focus on regions of phase space that contribute most significantly to averages, boosting efficiency. Equilibrium Boltzmann averages can be computed simply by sampling the system at each step. To achieve this in practice, an algorithm known as the Metropolis Monte Carlo method is used[179]:

1. Initialize the system in some configuration $\mathbf{r}_0^N$.
2. Generate a proposed new configuration, $\mathbf{r}_n^N$.
3. If the new configuration has a lower energy $U(n)$, accept the move $0 \to n$. If it has a higher energy, accept the move with probability $P_{acc} = e^{-\beta(U(n)-U(0))}$:
   - generate a random number R between 0 and 1
   - accept the move if $R < P_{acc}$, and reject otherwise
4. Sample observables.
5. Repeat.

The Metropolis algorithm fulfils detailed balance; that is, in equilibrium, the number of moves that lead to state i from state j is equal, on average, to the number of moves that lead to state j from state i:

$$e^{-\beta U(i)}\pi(i \to j) = e^{-\beta U(j)}\pi(j \to i) \tag{2.1}$$

[†]MC schemes for simulating dynamics do exist, however; see Ref. 68

where $\pi(i \rightarrow j)$ is the transition probability from state i to state j. There are a number of possible choices in regards to how proposed moves are generated. For systems of strongly interacting particles, if only single-particle moves are attempted, the bond breakage resulting from a proposed diffusive movement will make the move energetically unfavourable and unlikely to be accepted. Furthermore, for a cluster of interacting particles to diffuse, separate moves must be attempted for each particle in the same direction. oxDNA therefore employs the Virtual Move Monte Carlo (VMMC) scheme[289;290], in which moves are attempted for clusters of, rather than individual, particles. A particle is randomly selected, and a cluster comprised of particles it interacts with is generated. A move – either translational or rotational – is then attempted for the whole cluster of interacting particles, enabling strong pairwise interactions to be preserved and permitting efficient diffusion of the system. While multiple cluster-move algorithms exist, the VMMC algorithm in particular offers an additional advantage beyond accelerating cluster diffusion: clusters are chosen based on particle interactions both before and after a proposed move, which allows the system to move along potential energy gradients as it diffuses. This facilitates more efficient relaxation of polymers, rendering VMMC an ideal choice for use with oxDNA. In practice, the Metropolis acceptance probability given above must be modified slightly to preserve detailed balance in VMMC[289;290], but this added complexity is compensated for by the great enhancements in diffusion and relaxation efficiency.

2.1.3 Umbrella Sampling

Even with the added efficiency afforded by the VMMC algorithm, simulating systems for which states of interest are separated by large free energy barriers can be difficult. To boost the sampling of regions of phase space that are rarely visited under the traditional VMMC algorithm, the umbrella sampling technique[275] can be used. Umbrella sampling acquires its name from the fact that it involves the application of a biasing potential energy to the system which serves to 'bridge' the gap between

energetically separated regions of phase space. In effect, this bias lowers free energy barriers to render them surmountable in a reasonable amount of time.

First, one or more 'order parameters', or reaction coordinates, ξ, are chosen which delineate degrees of freedom in phase space along which free energy barriers exist. For example, in this work, the order parameters employed relate to the unfolding reaction of DNA hairpins: the number of base pairs formed and the end-to-end distance of the terminal hairpin nucleotides. Sampling of phase space proceeds according to *weighted* Boltzmann factors:

$$\begin{aligned} P_{bias}(\xi(\mathbf{r}_i^N)) &= e^{-\beta\left(U(\mathbf{r}_i^N)+V_{bias}(\xi(\mathbf{r}_i^N))\right)} \\ &= w(\xi_i)e^{-\beta U(\mathbf{r}_i^N)} \end{aligned} \tag{2.2}$$

where $\mathbf{r}_i^N$ is a configuration of the N-particle system, $P_{bias}(\xi(\mathbf{r}_i^N))$ is the (unnormalized) probability of observing order parameter $\xi(\mathbf{r}_i^N)$, $V_{bias}(\xi(\mathbf{r}_i^N))$ is the umbrella biasing potential, and $w(\xi(\mathbf{r}_i^N))$ is the effective weighting function. To obtain unbiased equilibrium probabilities at the end of the simulation, the measured probabilities are *un*weighted accordingly:

$$P^0(\xi_i) = \frac{P_{bias}(\xi_i)}{w(\xi_i)} \tag{2.3}$$

The ideal weighting function to use is one which renders $P_{bias}(\xi_i)$ equal for all states, so all states are well-sampled: $w(\xi_i) = \frac{1}{P^0(\xi_i)}$. Unfortunately, this requires knowledge of the equilibrium Boltzmann probability distribution $P^0(\xi(\mathbf{r}_i^N))$, which is unknown *a priori*. One way forward is to choose an initial guess for the $w(\xi_i)$ based on experience and intuition regarding which regions of phase space will be difficult to sample, then run preliminary simulations to obtain an idea of $P^0(\xi_i)$ and adjust the weights to be inversely proportional to this distribution. The process can then be iterated until even sampling across order parameter values is achieved.

2.1.4 Weighted Histogram Analysis Method

For some systems, it may inefficient to use umbrella sampling alone. In such cases, the system's phase space can be divided into 'windows' and several parallel MC

simulations can be run, in each of which the system's configuration is restricted to – or a least concentrated in – a particular window. For example, to sample the unfolding transition for the DNA hairpins featured in Chapter 3, the end-to-end terminal nucleotide distance was restricted to be < 4.3 nm in one window, > 1.7 nm and < 6.0 nm in a second window, and so on. Another common practice is to use a harmonic biasing potential $V_{bias}(\xi) \propto (\xi_i - \xi_0)^2$, with each window centred on a different equilibrium order parameter ξ_0. Within each of the K windows, umbrella sampling is carried out as described above, yielding K independent estimates of the (unnormalized) Boltzmann probability distribution for the system – though each of these estimates will only be reliable over the limited phase space region sampled in the kth window. Further, since only relative occupation probabilities are computed in each window (the full partition function is not known), results from different windows will need to be shifted and 'stitched' together; overlap between windows in phase space is therefore essential.

While various schemes could be used to combine the results from each window, this thesis employs the Weighted Histogram Analysis Method (WHAM) of Kumar et al.[147]. Consider K simulations carried out in different windows, with n_k samples accumulated in each simulation. Phase space is divided into M bins corresponding to different values of the order parameter, ξ_i. The optimal estimate of the overall, unbiased probability distribution is given by:

$$P_i^0 = \frac{\sum_{k=1}^{K} N_{i,k}}{\sum_{k=1}^{K} n_k w_{i,k} \exp\{f_k\}} \tag{2.4}$$

where the sums are taken over all simulation windows, $N_{i,k}$ is the number of counts obtained in the ith order parameter bin in the kth simulation window, n_k is the total number of samples collected in the kth simulation window, and f_k, given by

$$f_k = -\ln \sum_{i=1}^{M} w_{i,k} P_i^0, \tag{2.5}$$

is the free energy of the kth simulation. Here, the sum is taken over all order parameter bins, $w_{i,k}$ is the weight for the ith order parameter bin in the kth

simulation window, and P_i^0 is the optimal estimate of the intrinsic probability distribution given by Equation 2.4.

Equations 2.4 and 2.5 constitute the WHAM equations, and they are solved iteratively until the f_k converge to within a chosen error tolerance. The equations are derived by writing P_i^0 as a weighted sum over the K estimates for P_i^0 from each window, $\Omega_{j,k}$:

$$P_i^0 = \sum_{k=1}^{K} c_{i,k}\Omega_{i,k} \tag{2.6}$$

and choosing the weights $c_{i,k}$ to minimize the variance in P_i^0. The interested reader can find good derivations of the WHAM equations in References 6 and 147.

It is a part of probability that many improbabilities will happen.

— Agathon, as quoted in Aristotle's *Poetics 9*.

3

Non-equilibrium bio-molecular unfolding under tension

Contents

3.1 Introduction 23
3.2 Theory 28
3.3 Application of non-equilibrium landscape analysis to two DNA hairpins 34
3.3.1 Benchmarking with experimental results 34
3.3.2 Non-equilibrium landscape reconstructions 40
3.4 Application of non-equilibrium landscape analysis to a novel system: DNA origami handles 61
3.4.1 System description and methods 62
3.4.2 Results 65
3.5 Conclusions and outlook 71

3.1 Introduction

In the 1960s and 70s, seminal experiments confirmed that all the information about a protein's 'native' functional three-dimensional structure is encoded in its one-dimensional sequence of amino acids[8;7]. Subsequently, what became known as the 'protein folding problem' – grasping the mechanism of this 1D to 3D mapping – enticed scientists across multiple disciplines. The principles that

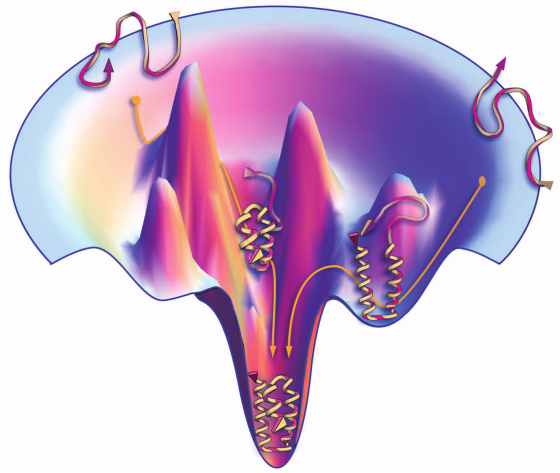


Figure 3.1: Illustration of a generic free energy landscape, with the 'native' folded state, the global free energy minimum, at the bottom of the funnel. The 'rim' of the funnel represents the ensemble of unfolded conformations; its width is indicative of the relatively higher entropy of the unfolded state. Local free energy minima represent folding intermediates, and several barriers – which slow diffusion on the landscape – are visible. Taken from Dill & MacCallum[55].

govern protein folding apply more broadly to other biomolecular structures, as well – DNA and RNA, for example – rendering the significance of unlocking the 'folding problem' far-reaching. Understanding the biological mechanism by which sequences of biomolecular monomers are mapped to 3D structures carries monumental implications for treating human diseases that result from improper molecular folding[70], including Alzheimer's and Parkinson's, as well as for *de novo* protein and DNA design, which has applications ranging from nanomedicine[109;301] to nanoelectronics[210]. Initially puzzling to scientists was the timescale on which folding occurred: milliseconds to seconds. As famously pointed out in 'Levinthal's paradox', this is far too short a window for a random search through the space of possible molecular conformations[153;25]. In the 1980s, physicists offered a solution to the puzzle by viewing the problem through the lens of statistical mechanics[25]: folding occurs along an *ensemble* of pathways. Biomolecular folding can be viewed as a diffusive search over a funnel-shaped, hyper-dimensional free energy landscape[198]; see Figure 3.1. Conformations closer to the native state have lower free energy, resulting in molecules being funnelled preferentially toward a global minimum.

Roughness, in the form of local peaks and troughs on the landscape, determines the folding diffusion constant, and thus folding rates. Local minima represent folding intermediates. The folding landscape offers a comprehensive picture of molecular folding, and indeed contains much of the information relevant to the folding process[212;303], enabling prediction of folding behaviour[298]. Thus, once the energy landscape framework emerged, scientists turned their efforts to characterising these landscapes for biomolecules.

Traditional experimental bulk measurement methods are inadequate for characterising folding energy landscapes because averaging over the behaviour of many molecules obscures rare trajectories, intermediates, alternate conformations, and misfolded minima: essential parts of the energy landscape[309;92]. To access all relevant regions of the landscape, individual molecular folding trajectories must be monitored. A particularly fruitful means of achieving this is via the application of external force to unfold molecules. Not only does tension-induced unfolding enable the study of individual unfolding trajectories, but force biases the unfolding pathway and can thus be used to explore transiently-occupied states that would rarely be observed under unbiased conditions[92].

The application of force to study biomolecular folding landscapes has been widely and fruitfully applied experimentally, mainly through atomic force microscope (AFM) or laser optical tweezer (LOT) set-ups[92;298]. In both AFM and LOT setups, either one or both ends of a molecule are suspended in harmonic potentials that are then moved to effect unfolding. These so-called single-molecule force spectroscopy (SMFS) experiments have been used to reconstruct energy landscapes for proteins[98;79] and nucleic acids[297;63] and validate fundamental physical theorems[157;37]. It is important to note that, to date, SMFS experiments can only succeed in illuminating one-dimensional projections – or the potential of mean force (PMF) – of full, hyper-dimensional energy landscapes. This is because technical difficulties limit such experiments to monitoring a single degree of freedom, or reaction coordinate.

The simplest SMFS modality involves application of a constant tension in equilibrium experiments. Extracting landscapes from equilibrium data can be

accomplished by collecting the probability distribution of a molecule's end-to-end extension and finding the free energy as a function of extension by inverting the Boltzmann relation; this technique has been validated experimentally for proteins and nucleic acids[296;79]. When the barrier to unfolding is high, however, the time needed for a biomolecule to unfold may render such equilibrium experiments impractical due to requirements on instrumental stability[97;298]. In such cases, alternative solutions are necessary.

Turn-of-the century theoretical developments in non-equilibrium physics opened up the possibility of extracting free energy landscapes for biomolecules even when equilibrium studies are intractable. In 1997, an equality due to Christopher Jarzynski[123] enabled for the first time the extraction of *equilbrium* free energies from *non-equilibrium* processes. The Jarzynski equality is part of a class of related identities developed around the same time known as non-equilibrium fluctuation theorems[45], and states that the equilibrium free energy difference between two states, ΔG, can be extracted from an ensemble of realizations of the process A $\rightarrow$ B by taking a weighted average of the work ΔW performed in each realization:

$$\langle e^{-\beta \Delta W} \rangle = e^{-\beta \Delta G}, \tag{3.1}$$

even when the process occurs arbitrarily far from equilibrium. Such fluctuation theorems are profound and remarkable because they provide a direct link between an equilibrium quantity and a path-dependent non-equilibrium quantity. On the heels of these theoretical advances, methodologies for applying them to out-of-equilibrium SMFS studies emerged[112;115] and have been experimentally validated in experiments on DNA hairpins[96;63]; experimental applications of these methods have become routine[98;303].

Concurrently with their experimental implementation, non-equilibrium theories like the Jarzynski equality also seeded new possibilities in molecular simulation. On one hand, whether non-equilibrium methods can perform more efficiently than traditional free-energy calculation techniques is an open question that has received much attention[51]. Debate continues on this front, with some earlier studies

concluding non-equilibrium methods offer no practical advantage over conventional simulation techniques[196], though this principle is not without exception[150]. More recent suggestions of combining Jarzynski-based methods with importance sampling have revived the possibility that non-equilibrium simulations can offer an advantage[195;278;33;12], and research in this area is ongoing.

More generally, non-equilibrium simulation techniques can be enormously beneficial for studying the dynamics of biomolecular response to force. In 'steered molecular dynamics' (SMD) simulations, force is applied to bias a molecule along a pathway, either to study particular regions of the unfolding landscape or to explore mechanical response to external forces. As with experimental non-equilibrium techniques, SMD offers the ability to investigate rarely-occupied states[125], and simulated SMFS can complement experiments by offering descriptions of force-mediated processes in experimentally-inaccessible detail. Biomolecular SMFS simulations have been profitably applied to reveal insights about dynamic processes of streptavidin-biotin rupture[94], fibronectin unfolding[204], and stability of RNA kissing complexes under force[34], among others[264;122;120], and have revealed unanticipated intermediate states[234] and multiple folding pathways[101].

Despite the widespread use of *in silico* SMFS, instances where simulated full free-energy landscapes have been reconstructed with Jarzynski-based techniques are lacking. Some examples do exist, but these either use simplified models that lack molecular detail[23;118;117;60] – often for the purposes of validating proposed analysis schemes – or atomistic simulations[125;182;36], which are limited in the timescales that can be accessed. Notably scarce are applications of Jarzynski-based methods to SMD of nucleic acids, though such reconstructions can offer great value, particularly as complements to experiments.

As aforementioned, experiments can typically monitor only one reaction coordinate, effectively projecting the hyper-dimensional energy landscape into a 1D PMF. Some choices of reaction coordinate, however, are poor in the sense that interesting features on the landscape – barriers to unfolding and intermediate states – may be obscured by the projection[60]. Choosing different pulling directions can

also reveal previously hidden pathways and intermediates[54]. For some systems, reactions can only be captured accurately using multiple order parameters[265;294]. Finally, experiments are plagued by artefacts such as potential surface effects and the convolution of molecular landscapes with point spread functions (PSF) associated with the instrument and the long molecular linker handles that are typically employed[298]. *In silico* SMFS can alleviate these issues by illuminating good choices for reaction coordinates and investigating instrument-free landscapes. They also offer the fundamental advantage of being able to correlate features in the landscape with structural dynamics of the unfolding process. In light of these potential benefits, we here investigate the application of Jarzynski-based methods to coarse grained nucleic acid models using oxDNA[203;281], a sophisticated, nucleotide-level model of DNA that has enjoyed broad success in reproducing experimental thermodynamic and kinetic data[261;243;226;185], as detailed in Chapter **??**.

The remainder of this chapter begins with a brief derivation of the relevant theoretical framework, following closely the treatment of Hummer & Szabo[112]. The main results of applying these formulae to oxDNA simulations are then presented, followed by an analysis of errors in the results. The chapter closes with an application of this method to a large DNA origami system for which equilibrium simulations would have great difficulty extracting free energy landscapes.

3.2 Theory

Consider a system evolving according to Liouville's equation in phase space:

$$\frac{\partial f(x,t)}{\partial t} = \mathcal{L}_t f(x,t) \tag{3.2}$$

where $\mathcal{L}_t$ is the time-dependent Liouville evolution operator and $f(x,t)$ is the phase space distribution of the system. The Boltzmann distribution

$$p(x,t) = \frac{e^{-\beta H(x,t)}}{\int e^{-\beta H(x',0)}\,\mathrm{d}x'}, \tag{3.3}$$

where $\beta = 1/k_{\mathrm{B}}T$ is the inverse thermal energy and $H(x,t)$ is a time-dependent Hamiltonian, is a stationary solution to Eq. 3.2:

$$\frac{\partial p(x,t)}{\partial t} = \mathcal{L}_t p(x,t) = 0. \tag{3.4}$$

Further, taking the time derivative of the Boltzmann distribution yields $\partial p/\partial t = -\beta(\partial H/\partial t)p$, which we can combine with Eq. 3.4 to derive a parabolic partial differential equation for which Eq. 3.3 is a solution:

$$\frac{\partial p(x,t)}{\partial t} = \mathcal{L}_t p(x,t) - \beta \frac{\partial H}{\partial t} p(x,t). \tag{3.5}$$

We can find a *second* representation for the solution to Eq. 3.5, however, using the Feynman-Kac theorem from quantum mechanics, which states that a functional of a stochastic trajectory $y(t)$, $u(y(t),t)$, that obeys a parabolic partial differential equation

$$\frac{\partial u}{\partial t} + A(y,t)\frac{\partial u}{\partial y}(y,t) + B(y,t)\frac{\partial^2 u}{\partial y^2}(y,t) + C(y,t)u(y,t) = 0 \tag{3.6}$$

and which takes on the known form $u(y,0) = \Psi(y)$ at initial time $t = 0$ is given by the solution

$$u(x,\tau) = \left\langle \exp\Big\{ -\int_0^\tau C[y(t'),t']\,\mathrm{d}t' \Big\} \Psi(y(0)) \right\rangle_{\text{conditional}}, \tag{3.7}$$

provided $y(t)$ evolves according to a Gaussian random process[245]. Here, $\Psi(y(0))$ is evaluated on the initial $t = 0$ distribution, and the angled brackets indicate a *conditional* expectation value: the average is taken over all possible paths of stochastic evolution $y(t)$ that terminate with the random variable value $y = x$ at time $t = \tau$. Eq. 3.7 is the Feynman-Kac formula. We can re-write the conditional expectation value by inserting a Dirac delta function that 'picks out' the stochastic trajectories that terminate as required at $y(\tau) = x$:

$$u(x,\tau) = \left\langle \delta(x - y(\tau)) \exp\Big\{ -\int_0^\tau C[y(t'),t']\,\mathrm{d}t' \Big\} \Psi(y(0)) \right\rangle. \tag{3.8}$$

Eq. 3.5 is analogous to Eq. 3.6 with $C(y,t) = -\beta\frac{\partial H}{\partial t}$. We take $p(y,0) = \Psi(y)$ to be an initial equilibrium distribution, such that $\Psi(y) = 1$ if y is in the initial distribution and 0 otherwise. We then simply take the expectation value in Eq. 3.8 over all stochastic trajectories that begin in the equilibrium distribution and terminate at $y(\tau) = x$:

$$p(x,\tau) = \Big\langle \delta(x - y(\tau)) \exp\Big\{ -\int_0^\tau \beta \frac{\partial H}{\partial t}(y(t'),t')\,\mathrm{d}t' \Big\} \Big\rangle. \tag{3.9}$$

Equating the two different solutions to Eq. 3.5, relabelling the dummy variables $y \to x$ and $\tau \to t$ for convenience, and generalising to a multidimensional phase space coordinate $\mathbf{x}$ yields

$$\frac{e^{-\beta H(\mathbf{x},t)}}{\int e^{-\beta H(\mathbf{x}',0)}\,\mathrm{d}\mathbf{x}'} = \Big\langle \delta(\mathbf{x} - \mathbf{x}(t)) \exp\Big\{ -\int_0^t \beta \frac{\partial H}{\partial t'}(\mathbf{x}(t'),t')\,\mathrm{d}t' \Big\} \Big\rangle. \tag{3.10}$$

The significance of this expression lies in the fact that an equilibrium Boltzmann distribution (left) is given in terms of a weighted average of non-equilibrium trajectories (right). It therefore enables computation of equilibrium quantities from nonequilibrium data. The famous Jarzynski equality[123] is obtained from Eq. 3.10 by integrating it over all $\mathbf{x}$:

$$\langle e^{-\beta \Delta W} \rangle = e^{-\beta \Delta G} \tag{3.11}$$

where the definition of free energy G in terms of partition functions $Z = \int e^{-\beta H(\mathbf{x},t)}\,\mathrm{d}\mathbf{x}$, $\Delta G = -\frac{1}{\beta}\log(Z/Z_0)$, is used[†], and where the work performed during a non-equilibrium trajectory $\mathbf{x}(t)$ is $W(t) = \int_0^t \frac{\partial H}{\partial t'}(\mathbf{x}(t'),t')\,\mathrm{d}t'$. Our expression for work is born of the fact that the Hamiltonian represents the total energy of the system, and its derivative is thus the rate of change of energy. Integrating $\partial H/\partial t'$ over time gives the total change in energy, equivalent to the external work performed on the system.

The Jarzynski equality can be interpreted physically as follows. When changes in a thermodynamic system state are effected *quasistatically*, and thus infinitely

[†]Here and throughout the chapter, G has been used to denote free energy. While Jarzynski originally proved his equality using Helmholtz free energy[123], it is valid for both Helmholtz and Gibbs free energies[46], and for this chapter, G has been used instead of F to prevent confusion with 'force'.

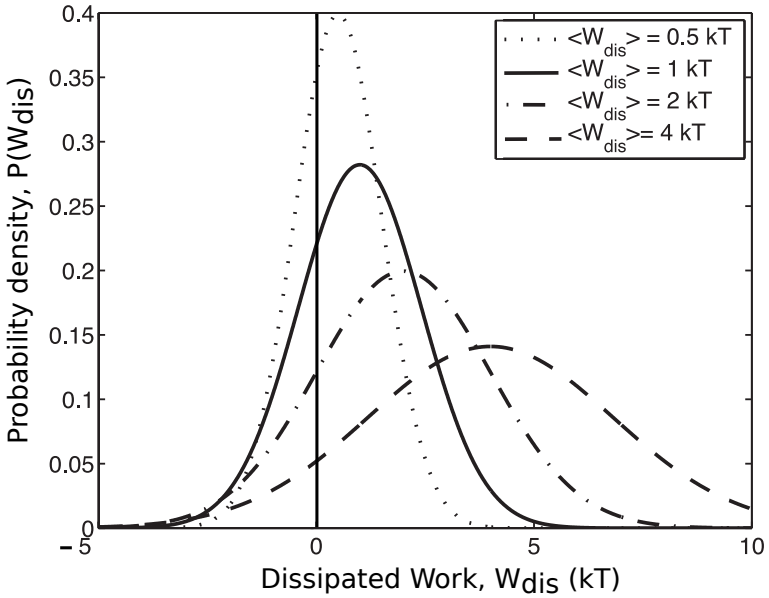


Figure 3.2: Illustration of increasingly far-from-equilibrium work distributions. Here, the abscissa is dissipated work W_{dis}, which is the difference between W, the work applied to effect a change in system state, and the free energy difference ΔG between the initial and final states: $W_{dis} = W - \Delta G$. The ordinate indicates the likelihood of observing a value of W_{dis} in any given realization of a process. For a process performed quasistatically, $W_{dis} = \langle W_{dis} \rangle = 0$, indicated as a delta function in the plot. As the process is performed farther and farther from equilibrium, the average dissipated work and distribution width increase, but a 'negative dissipated work' tail below $W_{dis} = 0$ remains. Adapted from Gore et al.[90].

slowly, the work required is identical to the free energy difference between the initial and final state, ΔG; no work is dissipated, or lost – for example, to friction[123]. For real processes performed in finite time, each realisation, characterized by external work W, will have non-zero dissipated work, $W_{dis} = W - \Delta G$. A W_{dis} probability distribution can be constructed from an ensemble of such realisations; see Fig. 3.2. In the quasistatic limit, $W_{dis} = 0$. If the process is performed nearly quasistatically, the dissipated work values will have an average close to zero and be Gaussian distributed[90]. As the change in system state is effected farther and farther from equilibrium, both the average dissipated work and the width of the work distribution grow larger, and the distribution may deviate from a purely Gaussian form[90;124]. Crucially, however, one tail of the distribution always extends into *negative* W_{dis} values, corresponding to instances where the system required *less* work than ΔG to

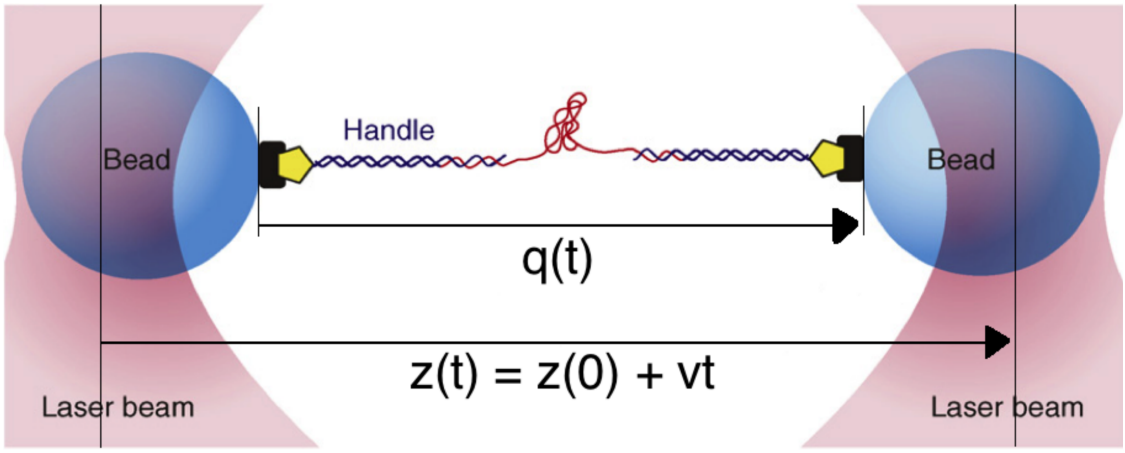


Figure 3.3: Typical experimental setup for single-molecule pulling experiments. A molecule is attached to long, dsDNA linker handles to eliminate potential cross-talk between lasers (in the case of laser optical tweezer (LOT) setups) or interactions with the surface slide to which one end of a molecule is affixed (in the case of AFM pulling experiments). These handles are attached to beads (for LOT setup) or to a glass slide and cantilever (for AFM). One or both endpoints are subjected to moving harmonic potentials, which forces the molecule (red) to unfold. The energy landscape with respect to the internal molecular coordinate, q, is of interest, but Jarzynski's equality refers to the external control parameter, z. Adapted from Woodside et al. [299].

change its state[123]; effectively, the second law of thermodynamics is only true *on average*. Microscopically, individual violations of the law are possible. Physically, one can imagine such realisations occurring when the system harnesses thermal 'kicks' from the environment, reducing the amount of external work required to alter its state. By weighting these 'negative W_{dis}' trajectories more heavily in an average, the equilibrium free energy can still be recovered from non-equilibrium distributions. Jarzynski's equality establishes that the correct weight to use is the exponential of the non-equilibrium work.

As can be seen in Fig. 3.2, the farther a process is from equilibrium, the smaller the probability that realisations in the tail of the work distribution will be observed. Eqs. 3.10 and 3.11 are only valid in the limit that averages $\langle \ldots \rangle$ are taken over all possible realisations of a process. The validity of results derived using fluctuation theorems will depend on whether or not the tail of the work distribution, which contributes most significantly to the average in Fig. 3.2, is adequately sampled in any real experiment or simulation, characterised by a finite number of trajectories.

Unfortunately, free energy landscapes of biomolecules cannot be extracted from the pulling experiments discussed in section 3.1 through straightforward application of Eq. 3.11; Figure 3.3 illustrates the problem. We are typically interested in the free energy landscape as a function of molecular extension q, $G(q)$. Jarzynski's equality, however, relates ΔG to the work performed by continuously changing a control parameter according to a well-defined protocol[123;114]: in the case of pulling experiments, this is the distance between the harmonic traps in which the ends of a molecule are held, z, which is varied in an identical manner from trajectory to trajectory. The end-to-end extension, meanwhile, changes stochastically. Following the formalism developed by Hummer & Szabo[112], $G(q)$ can be extracted from Eq. 3.10 as follows. First, take the relevant Hamiltonian for molecular pulling in harmonic traps:

$$H(\mathbf{x},t) = H_0(\mathbf{x}) + k(z(0) + vt - q(t))^2/2 \tag{3.12}$$

where $H_0(\mathbf{x})$ is the Hamiltonian of the unperturbed system corresponding to phase space coordinates $\mathbf{x}$; $q(t)$ is the end-to-end molecular extension at time t; $z(0)$ is the initial separation of the harmonic traps; v is the speed at which the harmonic potentials are moved apart; and k is the spring constant of the effective potential. Note that $q = q(\mathbf{x})$. For this Hamiltonian, the integral $W(t) = \int_0^t \frac{\partial H}{\partial t'}(\mathbf{x}(t'),t')\,\mathrm{d}t'$ becomes:

$$W(t) = kv\Big(z(0)t + vt^2/2 - \int_0^t q(t')\,\mathrm{d}t'\Big). \tag{3.13}$$

Substituting Eq. 3.13 into Eq. 3.10, multiplying both sides by $\delta(q - q(\mathbf{x}))$ to 'pick out' a particular molecular end-to-end extension from a phase space point $\mathbf{x}$, integrating over all $\mathbf{x}$, and taking the logarithm of both sides, we arrive at[112;114]:

$$G_0(q) = -\beta^{-1}\log\langle\delta(q - q(t))e^{-\beta\Delta W(t)}\rangle \tag{3.14}$$

where $\Delta W(t) = W(t) - k(z(0) + vt - q(t))^2/2$, the external work performed in time t minus the instantaneous bias energy stored in the harmonic potential,

and we have defined the unperturbed (zero force) free energy as a function of molecular coordinate thus:

$$G_0(q) = -\beta^{-1}\log\frac{\int \delta(q-q(\mathbf{x}))e^{-\beta H_0(\mathbf{x})}\,\mathrm{d}\mathbf{x}}{\int e^{-\beta H(\mathbf{x},0)}\,\mathrm{d}\mathbf{x}}. \tag{3.15}$$

Given a very large number of pulling trajectories, one could determine the entire free energy surface by performing the trajectory average in Eq. 3.14 at a single moment in time. In practice, only a small window in q, near the separation of the pulling spring $z(0)+vt$, will be sampled adequately at each time, so it is desirable to perform an average over all times in the pulling trajectory. This is analogous to umbrella sampling[275] in several windows, as described in section 2.1.4, where in this case the data must be unbiased for the biasing potential energy at time t as well as the non-equilibrium work. We can apply the weighted histogram analysis method[147], developed for optimally combining data from multiple umbrella sampling windows (see section 2.1.4), to calculate Eq. 3.14 from pulling trajectories:

$$G_0(q) = -\beta^{-1}\log\frac{\sum_t \frac{\langle\delta(q-q(t))e^{-\beta W(t)}\rangle}{\langle e^{-\beta W(t)}\rangle}}{\sum_t \frac{e^{-\beta V(q,t)}}{\langle e^{-\beta W(t)}\rangle}} \tag{3.16}$$

where $\Delta W(t)$ has been re-labelled $W(t)$ for ease of notation. Here, the angular brackets denote an average to be taken over all trajectories, and the sums are taken over N discrete time slices that map the realization of the process: $\{t_0, t_1, \ldots t_N\}$; normally $N = \text{sampling rate}^{-1}$. The $V(q,t)$ term in the denominator is the biasing potential effecting the change in system state – here, a harmonic spring potential.

3.3 Application of non-equilibrium landscape analysis to two DNA hairpins

3.3.1 Benchmarking with experimental results

To validate oxDNA's ability to capture the relevant features of DNA folding reactions, we first used equilibrium simulations to calculate the landscapes for three hairpins – one 6bp stem and two 20bp stem – for which experimental data were available. The sequence of the 20bp stem hairpin 20TS06T4 (where the nomenclature indicates

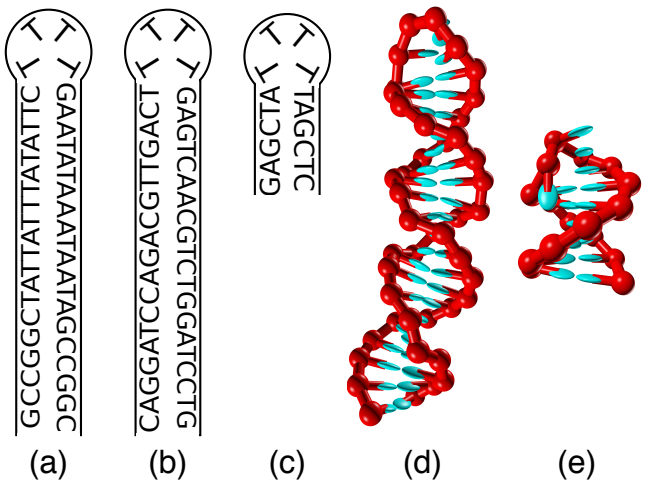


Figure 3.4: Sequences of three hairpins for which experimental landscape data were available: (a) hairpin 20TS06T4 from Woodside et al.[297], Manuel et al.[165], and Neupane et al.[192]; (b) hairpin 20R55T4 from Neupane et al.[192] and Pfitzner et al.[209]; and (c) hairpin 06R50T4 from Pfitzner et al.[209]. oxDNA representation of hairpins with a 4 base pair loop and (d) a 20 base pair stem and (e) a 6 base pair stem.

the hairpin has a 20bp stem, a transition state 6bp from the stem opening, and a 4 thymine loop) studied in Woodside et al.[297], Manuel et al.[165], and Neupane et al.[192] is illustrated in Figure 3.4(a). Fig. 3.4(b) contains the sequence of hairpin 20R55T4 (where the nomenclature indicates the hairpin has a 20bp stem, a sequence with 55% guanine-cytosine content, and a 4 thymine loop), studied in Neupane et al.[192] and Pfitzner et al.[209]. The sequence of hairpin 06R50T4 (where the nomenclature indicates the hairpin has a 6bp stem, a sequence with 50% guanine-cytosine content, and a 4 thymine loop), studied by Pfitzner et al.[209], is shown in Fig. 3.4(c). Figs. 3.4(d) and (c) depict oxDNA representations of generic 20bp stem, 4 base loop and 6bp stem, 4 base loop hairpins, respectively.

All of the experiments mentioned above were LOT studies, and all save that of Pfitzner et al.[209] utilized long dsDNA handles to connect the hairpin of interest to the trapped beads. In the case of Pfitzner et al.[209], DNA origami handles were used. The experiments were carried out at various temperatures and salt concentrations: 30 °C and $[Na^+] = 0.46$ M in Pfitzner et al.[209]; 23 °C and $[K^+] = 0.2$ M in Woodside et al.[297]; and 20 °C and $[K^+] \sim 0.2$ M in Neupane et al.[192] and Manuel et al.[165].

The salt concentration in the latter case is not exact; an additional, unquantified amount of [Na^+] was added by experimentalists to establish pH = 7.0. In all cases, the hairpins were held at a constant tension; these were equilibrium experiments.

Given the range of experimental temperatures and salt concentrations used, Virtual Move Monte Carlo (VMMC) simulations were performed at standard room temperature, 25 °C, and a salt concentration of [Na^+] = 0.5 M, using oxDNA2.0 with sequence dependence incorporated. The choice of salt concentration was in part motivated by efficiency: oxDNA simulations are much faster at higher salt. In order to achieve adequate sampling along both reaction coordinates, (a) # hydrogen bonds and (b) end-to-end distance, two-dimensional umbrella sampling (US)[275] was employed. Further, achieving good sampling necessitated splitting the US simulations into windows, and results from different windows were combined using the weighted histogram analysis method (WHAM)[147]; see Chapter 2 for further details of these techniques. To compare the results with the experimental landscapes, which give free energy as a function of extension *along the direction of force*, first, the simulated probability distributions $P(R)$, with R the end-to-end extension, were projected into $P(q)$ distributions, with q the end-to-end extension along one axis. The probability distributions were then inverted according to

$$G(q) = -\mathrm{k_B T} \ln P(q) \tag{3.17}$$

to yield one-dimensional landscapes.

Figures 3.5-3.7 contain the simulation results, compared to previously published experimental data. For each hairpin, multiple replicas of the US simulations were performed; landscape errors are standard deviations across these replicas. The simulated landscapes have been 'tilted' by applying a hypothetical constant force, $F_{1/2}$, such that the folded and unfolded wells have equal free energy; this allows comparison with the experimental landscapes, which are not zero-force landscapes as experiments were performed at constant tension. The forces used for tilting were $F_{1/2}$=10.2 pN, $F_{1/2}$=11.9 pN, and $F_{1/2}$=7.0 pN for the 20TS06T4, 20R55T4, and 06R50T4 simulated landscapes, respectively. These values are satisfyingly close to

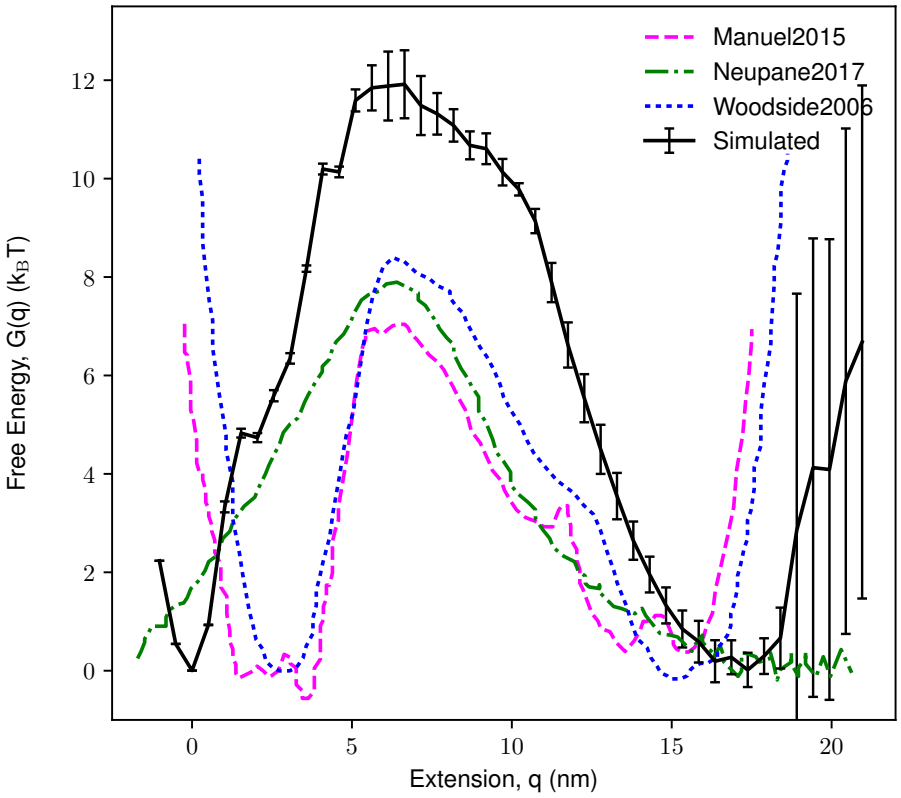


Figure 3.5: Comparison of oxDNA hairpin landscape, tilted using a constant force of $F_{1/2}$=10.2 pN, with previous experimental results from Woodside et al.[297], Manuel et al.[165], and Neupane et al.[192] for hairpin 20TS06T4. Discrepancies between experimental and simulated landscapes are attributable to convolution effects and differences in the salt concentrations at which simulations and experiments were performed.

the experimentally reported constant tensions used: $F_{1/2}$=11-14 pN for 20TS06T4 from Woodside et al.[297]*, $F_{1/2}$=14 pN for 20R55T4 from Pfitzner et al.[209], and $F_{1/2}$=6.5 pN for 06R50T4 from Pfitzner et al.[209]. To aid the eye, the barrier peaks of all landscapes have been aligned. Ordinates are in units of k_BT at 25 °C.

The agreement between simulation and experiment for all three hairpins is more than sufficient to demonstrate oxDNA's capability to reproduce landscapes correctly, particularly in light of oxDNA's previous success describing hairpin thermodynamics[203]. In all cases, the simulated landscapes have higher free energy barriers (or similar, as in Fig. 3.6). Experimental LOT landscapes are plagued by instrumental artefacts of the elastic response of the beads and linker handles, with

*Neupane et al.[192] and Manuel et al.[165] do not specify the tension used.

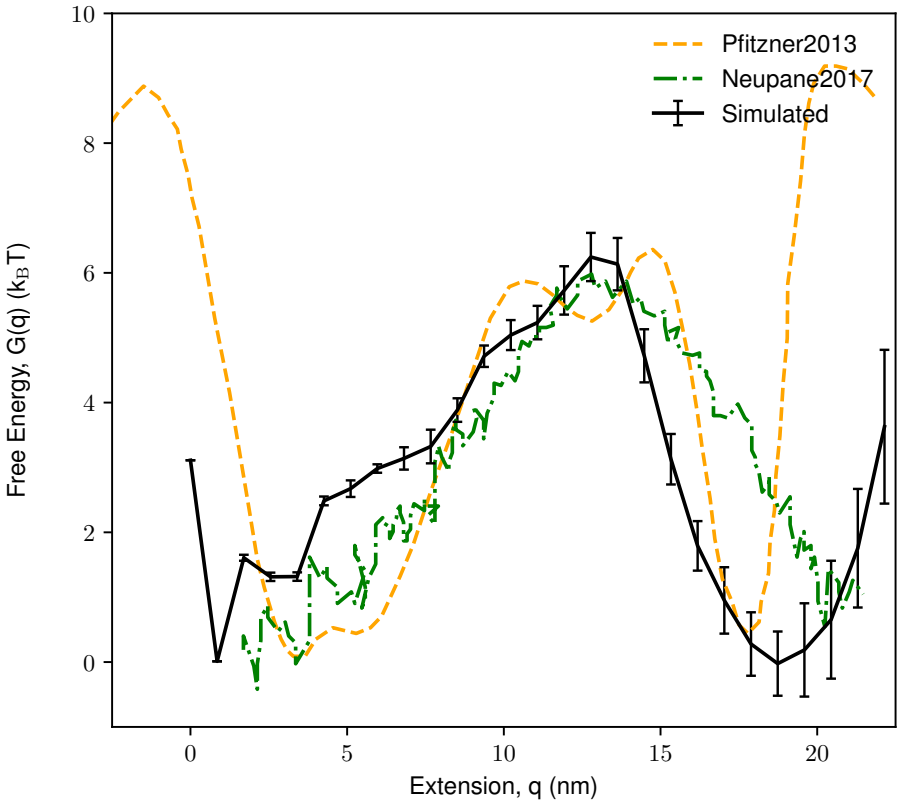


Figure 3.6: Comparison of oxDNA hairpin landscape, tilted using a constant force of $F_{1/2}$=11.9 pN, with previous experimental results from Neupane et al. [192] and Pfitzner et al. [209] for hairpin 20R55T4.

which the hairpin landscapes are convolved, and these effects are known to broaden and soften landscapes[298;107]. For the 20bp stem hairpins, Woodside et al. [297], Pfitzner et al. [209], and Manuel et al. [165] did attempt to deconvolve their data to remove instrumental artefacts, and Neupane et al. [192] employed a landscape reconstruction method claimed to obviate the requirement for deconvolution. Nevertheless, such procedures are necessarily imperfect, and the fact that simulated landscapes possess higher barriers than even deconvolved experimental landscapes is in fact expected. Additional factors underlying the discrepancies are differences in temperatures and salt concentrations; at lower salt, increased electrostatic repulsion will destabilize the hairpins and lower unfolding barriers somewhat. The landscape for hairpin 20R55T4 (Fig. 3.6) agrees more closely with experimental results than does the landscape for 20TS06T4 (Fig. 3.5); while the better agreement with the data

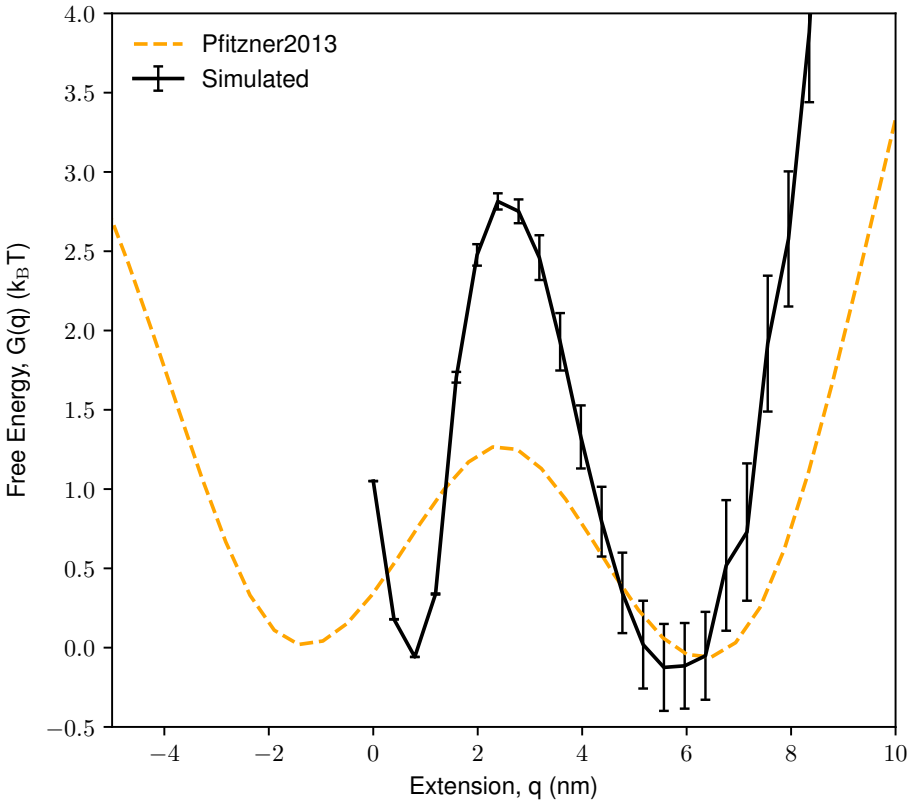


Figure 3.7: Comparison of oxDNA hairpin landscape, tilted using a constant force of $F_{1/2}$=7.0 pN, with previous experimental results from Pfitzner et al.[209] for hairpin 06R50T4. The experimental landscape is much broader than the simulated one due to instrumental convolution effects.

of Pfitzner et al.[209] can be attributed to the fact that their landscape was collected under a salt concentration ($[Na^+]$ = 0.46 M) close to the simulated one ($[Na^+]$ = 0.5 M), it is unclear why the Neupane et al.[192] landscape also agrees more closely with simulated results in this case. One possible explanation is that the Neupane et al.[192] landscape in Fig. 3.5 is based on only about 1/3 as many data points as the Neupane et al.[192] landscape in Fig. 3.6. Furthermore, given the fact that only the barrier regions were reconstructed by Neupane et al.[192], there is some ambiguity in how best to align these landscapes with simulated ones.

The disagreement between oxDNA results and experiment is the most egregious for the 6bp hairpin (Fig. 3.7). This is to be expected: the effect of experimental noise on measurements is more severe for shorter molecular time and length scales,

and motions faster than a limiting instrumental response time cannot be resolved at all[107]. Because the 6bp hairpin transitions are much more rapid and occur on smaller length scales than those of longer hairpins, they are correspondingly more difficult to detect above instrumental noise. We thus expect convolution with the instrument response to distort the 6bp landscape more than the 20bp ones, leading to the greater discrepancy observed. In light of these considerations, we can conclude that oxDNA captures the relevant physics well and that equilibrium landscapes produced to validate the procedures discussed in subsequent sections can be trusted.

3.3.2 Non-equilibrium landscape reconstructions

Generation of non-equilibrium landscapes

To simulate non-equilibrium single-molecule pulling experiments on DNA hairpins, the centres-of-mass of nucleotides at either end of each hairpin were subjected to 3D harmonic potentials. The long, dsDNA linker handles typically used in experimental SMFS setups were omitted, as they only degrade the data extracted by convolving the hairpin unfolding with a point spread function (PSF) associated with the elasticity of the handles[107]. The initial separation of the harmonic traps was set to be the average equilibrium end-to-end distance of each hairpin. Because Eq. 3.10 is predicated upon trajectories starting in an equilibrium Boltzmann distribution[112], hairpins were allowed to equilibrate in the harmonic traps for $\sim 5\times10^7$ MD steps ($\sim$0.76 µs in simulation time and $\sim$ 100 minutes on CPU) prior to beginning the pulling simulations[†]. To unfold the hairpins, one harmonic trap was moved relative to the other at a constant speed. Multiplying this speed by the trap stiffness yields the average 'force loading rate' for the simulation.

All trajectories at a given pulling rate were run for the same number of MD steps, long enough to ensure the hairpins reached the unfolded state in each case. This number varied from 1×10^7 steps (0.15 µs) for the fastest rate and the smaller hairpin to 3×10^9 steps (45.5 µs) for the slowest rate and the larger hairpin. All pulling simulations were performed at 30°C and 0.5 M NaCl, using oxDNA 1.0

[†]Arrival in equilibrium was confirmed by ensuring the system energy had reached a stable average value.

(oxDNA 2.0 had not yet been released when the simulations commenced). Average, sequence-independent stacking and hydrogen bonding strengths were used, as subtleties of hairpin sequences beyond the canonical Watson-Crick base pairing are not thought to be relevant to this study.

To reconstruct free energy landscapes from these non-equilibrium simulations, first, force-extension curves (FECs) were computed using the end-to-end molecular extensions, recorded throughout the simulation, and the instantaneous forces corresponding to those extensions. To calculate the forces, the extension $q(t)$ was subtracted from the known separation between the centres of the harmonic traps $z(t)$. This displacement was then multiplied by an effective trap stiffness, computed by combining the two 3D traps' stiffnesses in series: $\frac{1}{k_{\text{eff}}} = \frac{1}{k_1} + \frac{1}{k_2} = \frac{2}{k}$, to determine $F(q(t)) = -k_{\text{eff}}(q(t) - z(t))$. Illustrative results of this procedure are depicted in Figure 3.8; left and right panels show FECs for the 6bp and 20bp hairpin, respectively. Typically, the force increases with very little extension gain up to some critical value, at which a "rip" occurs. The sudden decrease in force and increase in extension signify the unfolding of the hairpin. Note the stochasticity of the process, evident in the distribution of unfolding forces.

From the FECs, $G_0(q)$ can be computed according to Eq. 3.16 as follows. First, the non-equilibrium work corresponding to the ith time bin and the kth pulling trajectory is computed through numerical integration:

$$W_{i,k} = kv\Big[\frac{vt_i^2}{2} + t_i z_0 - \frac{1}{2}\sum_{j=1}^{i}(t_j - t_{j-1})(q_{j,k} + q_{j-1,k})\Big], \tag{3.18}$$

with z_0 the initial trap separation. Then, $\langle e^{-\beta W(t)} \rangle$ is approximated for the ith time bin:

$$\langle e^{-\beta W(t)} \rangle \approx \frac{1}{N}\sum_{k=1}^{N} e^{-\beta W_{i,k}}, \tag{3.19}$$

where N is the total number of pulling curves acquired. The top numerator of Eq. 3.16 is then approximated:

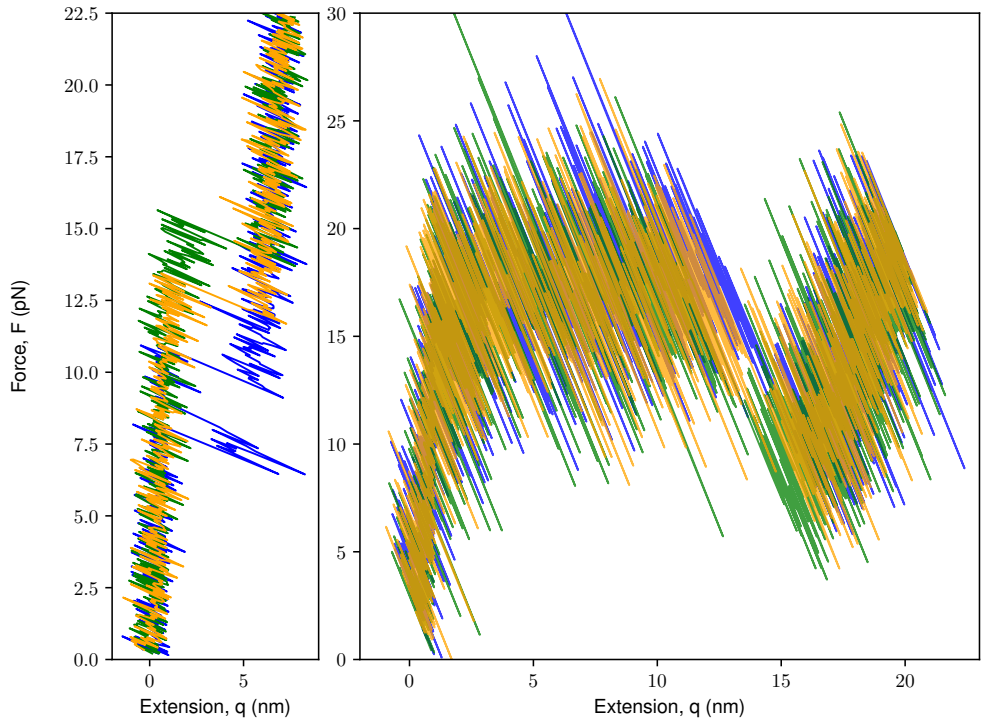


Figure 3.8: Representative simulated force-extension curves for the 6bp, 06R50T4 (left; trap stiffness 0.57 pN/nm and trap rate 5.62×10^6 nm/s) and 20bp, 20R55T4 (right; trap stiffness 5.71 pN/nm and trap rate 5.62×10^5 nm/s) hairpins.

$$\langle \delta(q_l - q(t)) e^{-\beta W(t)} \rangle \approx \frac{1}{N} \sum_{k=1}^{N} e^{-\beta W_{i,k}} \Theta_l(q_{i,k}), \tag{3.20}$$

where q_l is the midpoint of the lth extension bin, i indexes the time bin, and we compute a sum over all N pulling trajectory curves. Θ is a Heaviside function that is 1 if $q_{i,k}$ falls in the lth extension bin and 0 otherwise. The top denominator of Eq. 3.16 is computed using the spring potential energy of the traps:

$$V(q_l, t) = \frac{k(z_i - q_l)^2}{2}, \tag{3.21}$$

where z_i is the controlled separation of harmonic potentials in the ith time bin and k is the effective stiffness of the combined two-trap system. The numerator and denominator of Eq. 3.16 are then computed by taking the sum over all time

bins to yield the free energy corresponding to q_l and the process is repeated for all extension bins to yield a complete free energy landscape.

To explore the effect of pulling protocol on errors in the reconstructed landscapes, simulations were also performed using a 'sawtooth' force ramp. Instead of being subjected to harmonic traps, a constant tension (starting from 0) was applied to hairpin end nucleotides for a 'wait time' of τ_{inc} seconds before being instantaneously ramped up by increment ΔF_{inc} pN. Note that such a protocol is not feasible experimentally[306]. To compute energy landscapes for these simulations, the above procedure was adopted with the following modifications: the work in the ith time bin for the kth trajectory is given by:

$$W_{i,k} = -\sum_{j=1}^{i} q_{j,k} \Delta F_{inc}, \tag{3.22}$$

where the sampling rate is set to be equal to the inverse wait time, $1/\tau_{inc}$, and the instantaneous bias energy in the top denominator of Eq. 3.16 is

$$V(q_l, t_i) = -F_i q_l. \tag{3.23}$$

Results

The results of applying the procedure presented above are shown in Figures 3.9 and 3.10 for several representative simulation parameter sets. Alongside the reconstructed non-equilibrium landscapes is an equilibrium landscape for each hairpin, derived according to the methods described in section 3.3.1. To impart a sense of the variability in the trajectory averages upon which Eqs. 3.16 and 3.11 rely, 10 sets of 100 pulling curves were randomly drawn from a pool of $\geq$500. The curves in Figs. 3.9 and 3.10 are the averages of these ten landscapes, along with bounding curves representing the 'best' and 'worst' landscape realisations. In this context, the proximity of a non-equilibrium landscape to the true, equilibrium result is governed by how well the negative tail of the W_{dis} distribution is sampled by the 100 chosen trajectories. An exploration of how this sampling varies with simulation protocol, and how to determine *a priori* whether a reconstructed landscape is reliable without recourse to the equilibrium result, is the subject of the subsequent section.

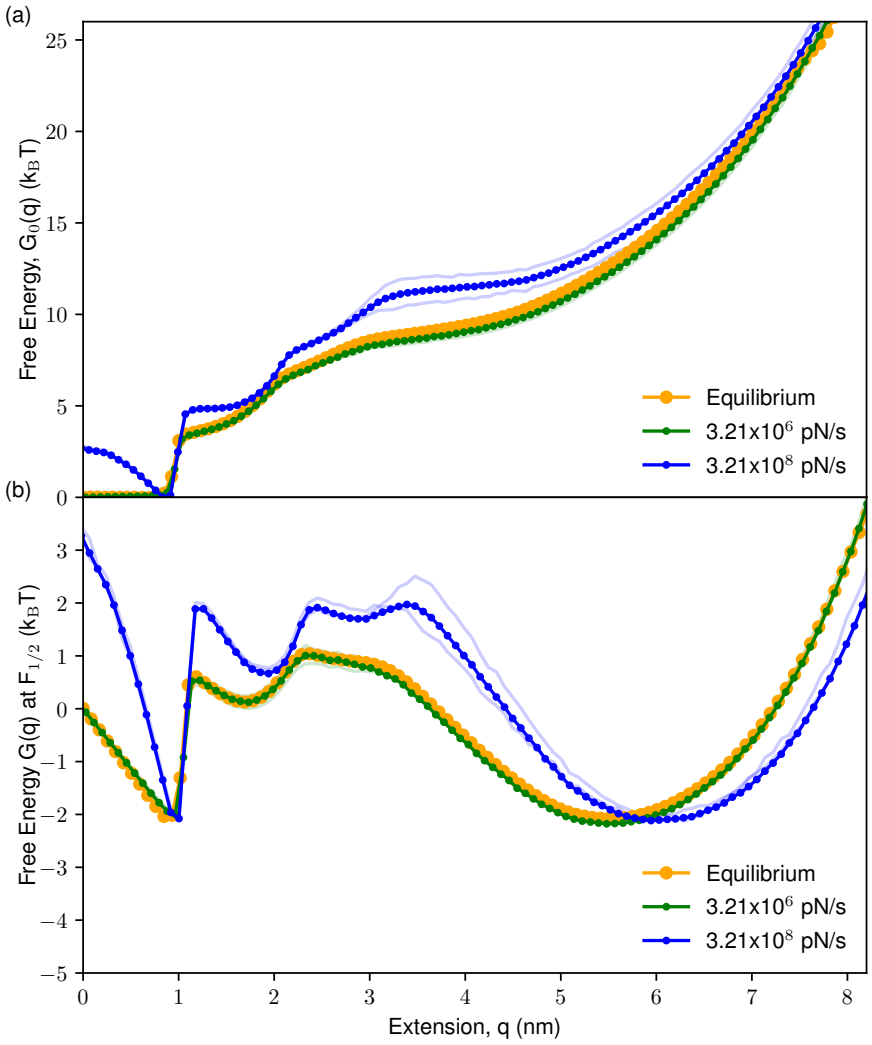


Figure 3.9: (a) Representative non-equilibrium landscape reconstructions for the 6 base pair hairpin 06R50T4, compared to the equilibrium result derived using VMMC simulations. Both the 3.21×10^6 pN/s and 3.21×10^8 pN/s curves are the result of 100 trajectories collected using a trap movement rate of 5.62×10^6 nm/s, but using trap stiffnesses of 0.57 pN/nm and 57 pN/nm, respectively. Lighter, bounding curves are the 'best' and 'worst' of a set of 10 reconstructions for which 100 curves were randomly chosen from sets of 500 (for the 0.57 pN/nm data) and 5000 (for the 57 pN/nm data). (b) Same landscapes as in (a), but tilted to the constant force at which the folded and unfolded wells have the same free energy: 7.3 pN for the equilibrium landscape and 7.3 pN and 8.3 pN for the green and blue landscapes, respectively.

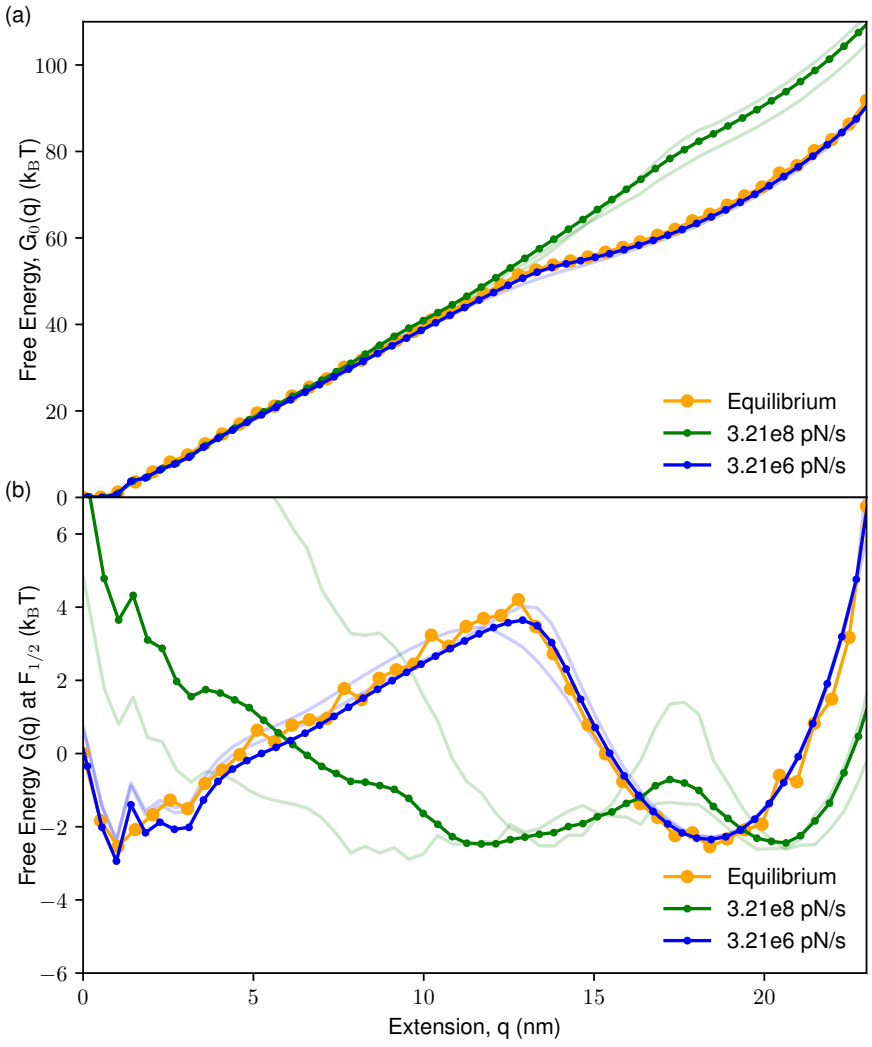


Figure 3.10: (a) Representative non-equilibrium landscape reconstructions for the 20 base pair hairpin 20R55T4, compared to the equilibrium result derived using VMMC simulations. Both the 3.21×10^{6} pN/s and 3.21×10^{8} pN/s curves are the result of 100 trajectories collected using trap stiffness 5.71 pN/nm, but using trap movement rates of 5.62×10^{5} nm/s and 5.62×10^{7} nm/s, respectively. Lighter, bounding curves are the 'best' and 'worst' of a set of 10 reconstructions for which 100 curves were randomly chosen from a set of 5000. (b) Same landscapes as in (a), but tilted to the constant force at which the folded and unfolded wells have the same free energy: 11.2 pN for the equilibrium landscape and 15.3 pN and 11.1 pN for the green and blue landscapes, respectively.

Free energy landscapes in the literature are often presented 'tilted' to the constant force $F_{1/2}$ at which the folded and unfolded wells are at the same free energy, to more clearly visualise the barrier[297;96;209]. The bottom panels of Figs. 3.9 and 3.10 adopt this convention, and show $G_{tilt} = G(q) - F_{1/2}q$. Very close agreement with the equilibrium landscapes is achieved for both systems at the slower force loading rate, $\dot{F} = 3.21 \times 10^6$ pN/s, even when only 100 curves are used to compute the average. Deviations from equilibrium landscapes, observable for both systems at the faster force loading rate, $\dot{F} = 3.21 \times 10^8$ pN/s, manifest in two ways: overall free energy differences are too large, and unfolding barriers are shifted to too-high extensions, as has been observed elsewhere[63]. However, at the slower $\dot{F}$, corresponding to non-prohibitive total simulation times of 220 CPU hrs (6bp) and 600 CPU hrs (20bp), we have demonstrated our central result: the coarse-grained simulator oxDNA can successfully be combined with non-equilibrium analysis techniques to reconstruct full free energy landscapes for DNA molecules. Furthermore, the equilibrium simulations underlying the landscapes in section 3.3.1 took ∼70-300 days of CPU time[‡], rendering the non-equilibrium simulations a much more efficient means for generating landscapes in this case.

Bias in non-equilibrium landscapes

Since equation 3.11 is only valid in the limit of infinite trajectories, one can define a *bias* in any finite-trajectory Jarzynski average:

$$B_N = G_N - \Delta G, \tag{3.24}$$

where the subscript N indicates the number of trajectories included in the average, G_N is the finite-N Jarzynski estimate of free energy, and ΔG is the 'true', intrinsic free energy difference being probed. Because of its exponential nature, the average in Eq. 3.11 is often dominated by very rare trajectories corresponding to low work values. Analyses based on the Jarzynski equality therefore exhibit poor convergence

[‡]An iterative process was used to identify suitable umbrella sampling weights and thus achieve sufficient sampling, which led to lengthy simulation times.

with N[124]. Work has been done to characterize this convergence behaviour, both theoretically[124;304;225] and experimentally[90;206;97], as well as to investigate optimal protocols for minimising the bias B_N in free energy estimates[274;80]. In these studies, the focus is on extreme optimisation for the purpose of streamlining free-energy extraction. Though some have suggested optimal force protocols contain some discontinuous jumps in force[80;240], we have simulated an experimental harmonic pulling protocol, since our aim with oxDNA is to provide a platform that complements experimental setups closely. For our practical purposes, it is desirable to have (i) an idea of which elements of our molecular pulling protocol have the greatest impact on the bias in reconstructed landscapes; and (ii) a rough estimate, obtainable without access to equilibrium data, of whether errors lie in an acceptable range for a given set of simulations.

I. Impact of parameters on bias B_N

Non-equilibrium hairpin pulling simulations were repeated using various simulation parameters to probe their effect on the bias B_N. Since full landscape reconstructions effectively compute a Jarzynski average at every extension value, and the bias as defined by Eq. 3.24 is for a single free energy difference, one must adopt a working definition of bias 'in a landscape'. The most interesting facet of the unfolding landscape is usually the disparity between folded and unfolded free energies, which is also the part most susceptible to bias because it is the transition farthest out of equilibrium. Even landscapes that resolve the pre-barrier region well may accumulate bias for extensions near and beyond the barrier to unfolding. Consequently, taking an 'average' bias across all extensions will tend to overestimate how well the landscape captures the feature of interest. We therefore take B_N to be the *largest* difference between the equilibrium and non-equilibrium curves across all extensions.

Our results yield insight into which parameters have the greatest impact on error in reconstructed landscapes. Comprehensive results are reported in Tables 3.1-3.3, while Figures 3.11-3.14 summarise the salient points. Where errors are listed, they are the result of taking ten random samples of N curves from a larger pool,

Table 3.1: Summary of the effect of various parameters on the bias B_N (Eq. 3.24) in the reconstructed landscapes for the 6bp hairpin; harmonic protocol. Yellow rows in this and Table 3.2 highlight the beneficial effect of choosing a harmonic, rather than sawtooth protocol; green rows show that few slow pulls should be prioritized over many fast ones. Purple rows reveal no clear advantage to preferencing smaller stiffnesses.

Harmonic protocol					
Stiffness (pN/nm)	Speed (nm/s)	Force loading rate (pN/s)	N	Bias (k_BT)	CPU hours
0.57	5.62×10^6	3.21×10^6	500	-0.48	1.1×10^4
			100	$-0.51^{+0.34}_{-0.32}$	2200
2.86	5.62×10^6	1.60×10^7	500	0.86	2200
			100	$0.88^{+0.68}_{-1.71}$	444
5.71	5.62×10^7	3.21×10^8	5000	1.69	1111
			1000	$2.02^{+1.70}_{-2.35}$	222
			500	$2.08^{+1.78}_{-2.98}$	111
			100	$2.77^{+4.15}_{-5.00}$	22
	5.62×10^6	3.21×10^7	5000	0.71	1.1×10^4
			1000	$0.86^{+0.77}_{-1.82}$	2200
			500	$0.86^{+0.63}_{-0.47}$	1100
			100	$1.35^{+1.25}_{-3.43}$	220
	5.62×10^5	3.21×10^6	5000	0.65	1.1×10^5
			1000	$0.59^{+0.26}_{-0.19}$	2.2×10^4
			500	$0.68^{+0.35}_{-0.24}$	1.1×10^4
			100	$0.62^{+1.76}_{-1.54}$	2200
14.27	5.62×10^6	8.02×10^7	500	1.13	890
			100	$1.25^{+0.78}_{-2.55}$	178
28.55	5.62×10^6	1.60×10^8	500	2.18	780
			100	$2.14^{+0.22}_{-0.27}$	156
42.82	5.62×10^6	2.41×10^8	500	2.27	665
			100	$2.26^{+0.38}_{-0.43}$	133
57.09	5.62×10^7	3.21×10^9	500	5.99	67
			100	$6.59^{+1.93}_{-0.97}$	13
	5.62×10^6	3.21×10^8	5000	2.21	6650
			1000	$2.19^{+0.20}_{-0.10}$	1330
			500	$2.25^{+0.29}_{-0.38}$	665
			100	$2.26^{+0.79}_{-0.68}$	133
	5.62×10^5	3.21×10^7	500	0.95	6650
			100	$0.62^{+0.24}_{-1.13}$	1330

Table 3.2: Summary of the effect of various parameters on the bias B_N (Eq. 3.24) in the reconstructed landscapes for the 6bp hairpin; sawtooth protocol. Yellow rows in this and Table 3.2 highlight the beneficial effect of choosing a harmonic, rather than sawtooth protocol.

Sawtooth protocol					
Wait time (ps)	Increment (pN)	Force loading rate (pN/s)	N	Bias (k_BT)	CPU hours
15.2	0.0049	3.21×10^8	4990	1.99	887
			1000	$5.77^{+3.48}_{-4.15}$	178
			500	$6.87^{+3.62}_{-2.29}$	89
			100	$9.03^{+4.79}_{-4.32}$	18

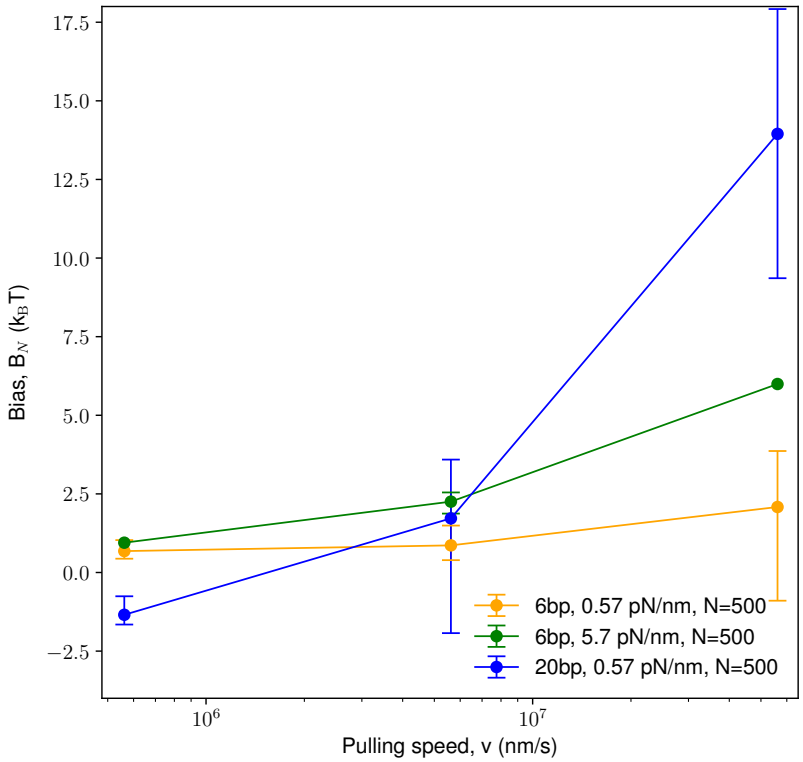


Figure 3.11: Effect of pulling speed on bias B_N (Eq. 3.24) in landscapes. Where shown, errors are the 'best' and 'worst' of a set of 10 reconstructions of $N = 500$ curves randomly chosen from a larger set. To minimize B_N, one should collect fewer slower trajectories rather than many faster ones.

Table 3.3: Summary of the effect of various parameters on the bias B_N (Eq. 3.24) in the reconstructed landscapes for the 20bp hairpin. Yellow rows highlight the beneficial effect of choosing a harmonic, rather than sawtooth protocol; green rows show that few slow pulls should be prioritized over many fast ones.

Harmonic protocol					
Stiffness (pN/nm)	Speed (nm/s)	Force loading rate (pN/s)	N	Bias (k_BT)	CPU hours
2.86	5.62×10^6	1.60×10^7	500	1.04	3750
			100	$3.25^{+1.89}_{-4.68}$	750
5.71	5.62×10^7	3.21×10^8	5000	11.24	3000
			1000	$12.44^{+4.10}_{-2.53}$	600
			500	$13.95^{+3.97}_{-4.59}$	300
			100	$18.86^{+2.01}_{-4.27}$	60
	5.62×10^6	3.21×10^7	5000	0.73	3×10^4
			1000	$0.78^{+1.93}_{-2.17}$	6000
			500	$1.72^{+1.87}_{-5.52}$	3000
			100	$3.26^{+0.60}_{-1.15}$	600
	5.62×10^5	3.21×10^6	1000	-1.38	6×10^4
			500	$-1.34^{+0.59}_{-0.31}$	3×10^4
			100	$-1.43^{+0.65}_{-1.10}$	6000
14.27	5.62×10^6	8.02×10^7	500	-1.73	2500
			100	$1.35^{+2.56}_{-4.58}$	500
28.55	5.62×10^6	1.60×10^8	500	2.35	2100
			100	$2.51^{+1.23}_{-1.00}$	420
42.82	5.62×10^6	2.41×10^8	500	0.80	1900
			100	$1.07^{+2.27}_{-2.03}$	380
57.09	5.62×10^6	3.21×10^8	500	2.00	1800
			100	$2.44^{+1.69}_{-0.76}$	360
Sawtooth protocol					
Wait time (ps)	Increment (pN)	Force loading rate (pN/s)	N	Bias (k_BT)	CPU hours
15.2	0.00049	3.21×10^8	500	22.17	6889
			100	$23.49^{+3.36}_{-2.86}$	1378

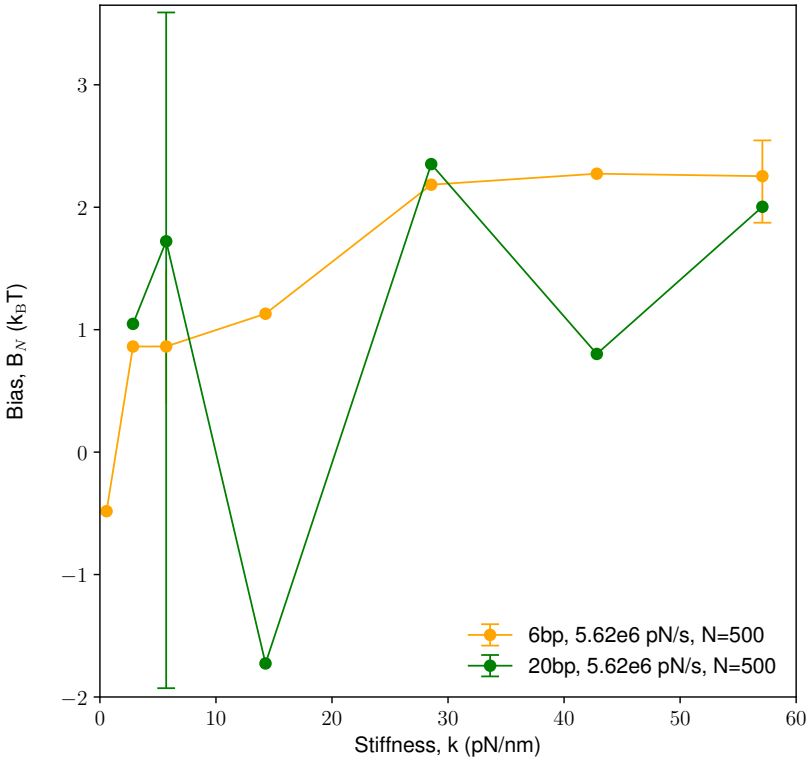


Figure 3.12: Effect of trap stiffness on bias B_N (Eq. 3.24) in landscapes. Where shown, errors are the 'best' and 'worst' of a set of 10 reconstructions of $N = 500$ curves randomly chosen from a larger set. Optimizing trap stiffnesses does not appear to be as effective in reducing B_N as optimizing protocol and pulling speed.

computing the average of all ten landscapes, and taking the landscapes with the largest and smallest biases as upper and lower error limits, respectively. Clear from Figs. 3.11-3.14 is the fact that errors grow the farther from equilibrium a simulation is performed; this is expected given that non-equilibrium averages are increasingly sensitive to rare events with increasing work distribution width. Where biases are negative, rare, low-work trajectories happened to be randomly sampled before higher-work segments of the work distribution, leading to underestimates of the free energy difference.

Our results show that the landscape bias decreases with decreasing pulling rate (Fig. 3.11); decreasing trap stiffness (Fig. 3.12); increasing number of tra-

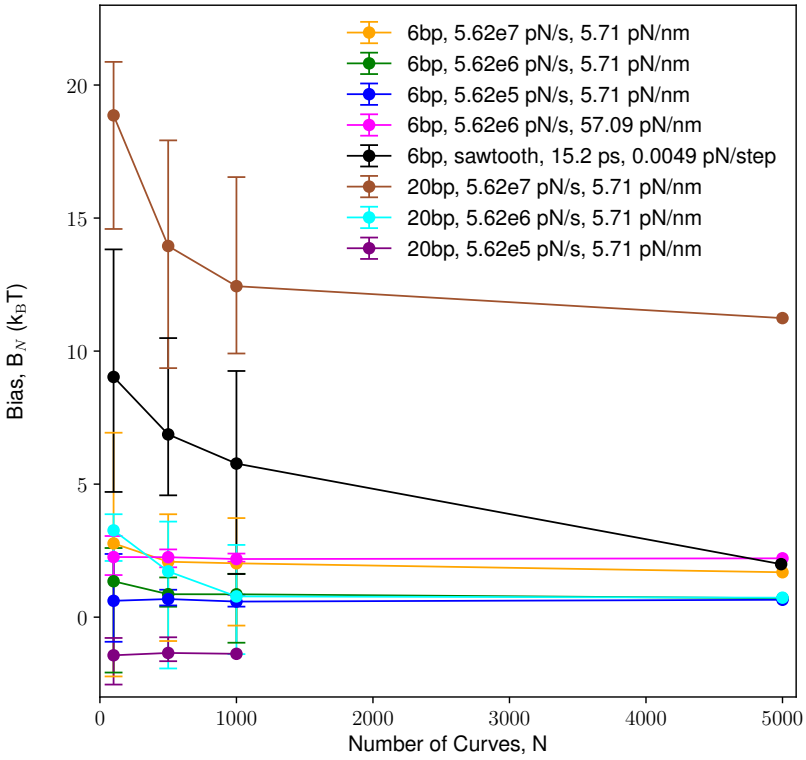


Figure 3.13: Effect of number of trajectories on bias B_N (Eq. 3.24) in landscapes. Where shown, errors are the 'best' and 'worst' of a set of 10 reconstructions of N curves randomly chosen from a larger set. For a given set of pulling parameters, more trajectories will reduce B_N, but the convergence to $B_N = 0$ is slow. Biases < 0 correspond to simulations in which rare, low-work trajectories happened to be randomly sampled before higher-work segments of the work distribution, leading to underestimates of the free energy difference.

jectories (Fig. 3.13); and a harmonic, compared to a sawtooth, pulling protocol (Fig. 3.14). For a given number of CPU hours, however, which of these four factors should be optimised?

The largest gains are to be made by choosing a good protocol. Consider the yellow highlighted rows in Tables 3.1, 3.2, and 3.3. For the 6bp hairpin, 178 CPU hours using a sawtooth protocol at a force loading rate of $\dot{F} = 3.21 \times 10^8$ pN/s yields an average bias of 5.77 $k_\mathrm{B}T$. A harmonic protocol at the same loading rate leads to average biases of 2.26 $k_\mathrm{B}T$ for 133 CPU hours and 2.08 $k_\mathrm{B}T$ for 111 CPU hours. For the 20 bp hairpin at $\dot{F} = 3.21 \times 10^8$ pN/s, roughly 1300 CPU

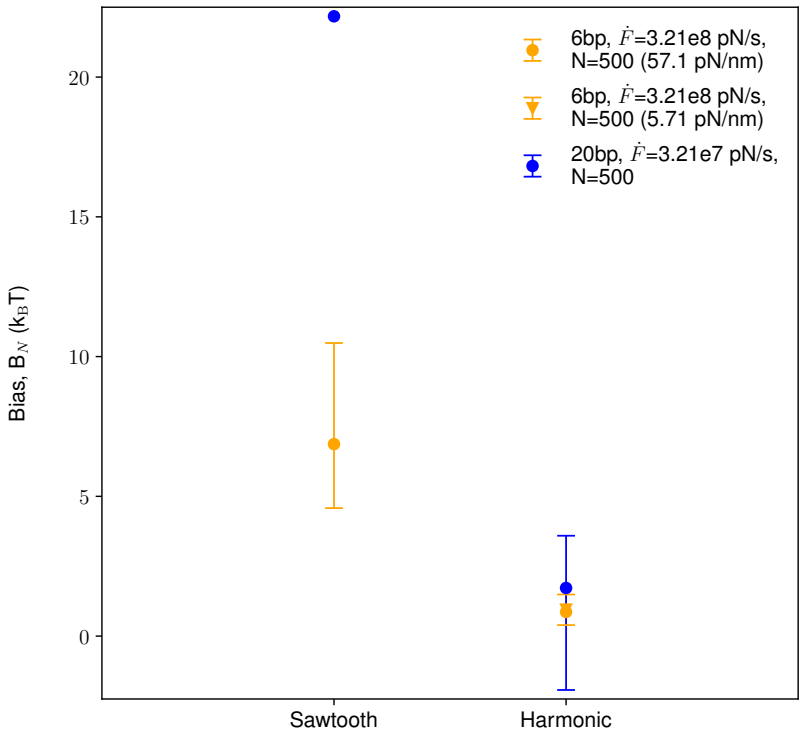


Figure 3.14: Effect of pulling protocol (for a given force loading rate) on bias B_N (Eq. 3.24) in landscapes. Where shown, errors are the 'best' and 'worst' of a set of 10 reconstructions of $N = 500$ curves randomly chosen from a larger set. Choice of protocol is extremely important in minimizing B_N, and a harmonic protocol performs much better than a sawtooth one.

hours using a sawtooth protocol yields a bias of 24.3 k_BT, while smaller biases of 12.44 k_BT and 2.44 k_BT are possible with less than half that simulation time (600 and 360 CPU hours, respectively) using a harmonic protocol[†]. One possible explanation for this is that sawtooth pulling is characterized by monotonically increasing external force that instantaneously pushes the system out of equilibrium, while harmonic pulling provides *continuous* opportunities for the system to relax somewhat, lowering the force it experiences and moving the system marginally closer to equilibrium. Interestingly, some have suggested that using discontinuous

[†]That 360 hours yield a smaller bias than 600 hours for the harmonic protocol is a result of choice of pulling speed, discussed in the text.

jumps of a carefully-chosen magnitude at the beginning and end of a protocol can optimize the performance of the Jarzynski equality[240]; repeated instantaneous forcing, however, is clearly non-optimal.

Next, one should prioritise fewer slow simulations over many fast simulations; this result has been noted elsewhere[112;51]. Consider the green highlighted rows in Tables 3.1 and 3.3. For the 6 bp hairpin pulled with 57.1 pN/nm traps, 1330 CPU hours can yield a bias of 2.19 $k_{\mathrm{B}}T$ for 1000 trajectories collected with pulling speed 5.62×10^6 pN/s, or 0.62 $k_{\mathrm{B}}T$ for 100 trajectories collected with pulling speed 5.62×10^5 pN/s. Similarly, with traps of stiffness 5.71 pN/nm, ∼1100 CPU hours produce biases of 1.69 $k_{\mathrm{B}}T$ for 5000 trajectories collected at 5.62×10^7 pN/s and 0.86 $k_{\mathrm{B}}T$ for only 500 trajectories collected at 5.62×10^6 pN/s. The same trends are clear for the 20 bp hairpin: for example, 3000 CPU hours can produce 5000 trajectories collected at 5.62×10^7 pN/s for a bias of 11.24 $k_{\mathrm{B}}T$ or 500 trajectories collected at 5.62×10^6 pN/s for a bias of 1.72 $k_{\mathrm{B}}T$ (both for trap stiffness 5.71 pN/nm). Due to the poor convergence with N, fewer slower trials are broadly preferable.

Finally, while lower stiffnesses uniformly reduce biases for the 6 bp hairpin (see Fig. 3.12), the trend is not clear for the 20bp hairpin. Further, even for the 6bp hairpin, for a given number of available CPU hours, fewer curves collected at lower stiffnesses do not always yield an advantage over more curves collected at higher stiffnesses: see purple rows in Table 3.1. In particular, note that 2200 CPU hours can produce 1000 curves for a 5.71 pN/nm trap, 500 curves for a 2.86 pN/nm trap, or 100 curves for a 0.57 pN/nm trap, corresponding to biases of 0.86 $k_{\mathrm{B}}T$, 0.86 $k_{\mathrm{B}}T$, and -0.51 $k_{\mathrm{B}}T$, respectively. It is not obvious that decreasing the stiffness at the cost of increased simulation time is worthwhile, at least over this range of trap stiffnesses.

II. Framework for characterising non-equilibrium bias

To develop a rough 'rule-of-thumb' for determining the required number of trajectories under a given simulation protocol, a framework developed by Palassini & Ritort[206] was adapted and supplemented with insight from Jarzynski[124]. The difficulty of treating bias in the Jarzynski equality stems from the exponential

nature of the estimator. Even if work values are normally distributed – which is not the case in general[304] – there is no closed-form solution for the error in exponential estimators[227]. A further complication is the fact that the form of a work distribution is protocol and system dependent[206], and unknown in advance. Various approximations for the error in the Jarzynski estimator have been presented[295;310;90], but the most comprehensive and recent of these is that of Palassini & Ritort[206]. In developing an expression for bias in finite-N Jarzynski estimates of free energy, they propose fitting the tail of the empirical work distribution with the following form[206]:

$$p(W) \sim q\frac{\Omega^{\alpha-1}}{|W-W_c|^{\alpha}}\exp\Big(-\frac{|W-W_c|^{\delta}}{\Omega^{\delta}}\Big), \tag{3.25}$$

where W_c is a characteristic work value, normally taken to be the peak of the distribution; Ω measures the width of the tail; δ and α are shape parameters; and q is a normalization constant. In order for the Jarzynski equality to hold on this distribution, $\delta > 1$, and since $|W - W_c|^{\alpha}$ is slowly varying compared to the exponential, α can be set to 0 for empirical fits. For a Gaussian work distribution, $\langle W\rangle = W_c$, $\sigma^2 = \Omega^2/2$, $\delta = 2$, and $q = \pi^{-1/2}$. Palassini & Ritort[206] note that in practice, their method does not work well for δ close to 1, as the work distribution may not be well-modeled by Eq. 3.25 in that case; they suggest employing it only for distributions such that $\delta > 1.4$[§]. As the work distributions for the hairpins discussed below all have $1.8 \lesssim \delta \lesssim 2.5$, this is not an issue in the present work.

The authors proceed to derive approximations for the bias B_N in finite-N Jarzynski estimates in three regimes, parametrized by $\lambda = \frac{\log N}{\Omega^{\delta/(\delta-1)}}\delta^{\delta/(\delta-1)}$, which indicates the relative magnitudes of Ω and N:

$$B_N = D_c + \log N - \Omega(\log N)^{1/\delta} - \lambda^{(1-\delta)/\delta}\Big[\gamma_E + \\ + \frac{1-\alpha-\delta}{\delta}\log\log N + \log(q/\delta)\Big], \lambda \ll 1 \tag{3.26}$$

[§]This is because the fitted work distributions may fail to satisfy the Crooks fluctuation theorem[45] as $\delta \to 1$; see the supplementary information of Ref. 206.

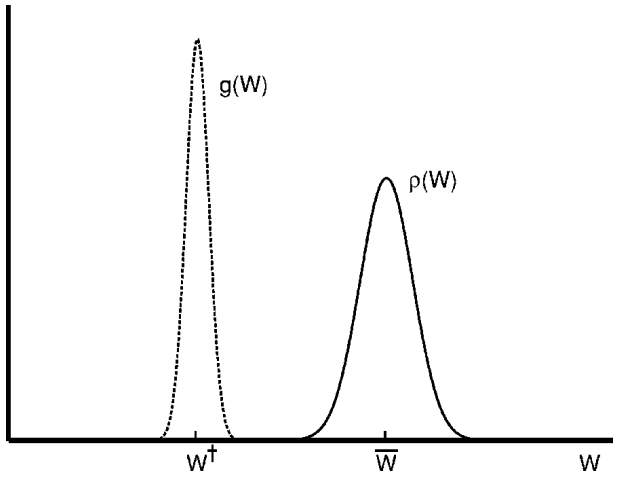


Figure 3.15: Work distributions $p(W)$ and $g(W) = e^{-\beta W}p(W)$ for a generic thermodynamic process. The ordinate is the probability of a work value being observed in a given realization of a process; $\bar{W}$ is the value of work most likely to be observed, and $\mathrm{W}^{\dagger}$ is the work value that contributes most to the Jarzynski average. Adapted from Jarzynski [124].

$$B_N = D_c + \log N - \Omega(\log N)^{1/\delta} + \gamma_E - \lambda^{(1-\delta)/\delta}\Big[\gamma_E + \frac{1-\alpha-\delta}{\delta}\mathrm{loglog}N + \\ + \frac{1}{2}\log\frac{\pi q^2}{2\delta(\delta-1)} + \theta^2 + \log\mathrm{erfc}(\theta)\Big], \lambda \simeq 1 \quad (3.27)$$

Here, erfc is the complementary error function, $\theta = (\lambda^{(1-\delta)/\delta}-1)\sqrt{\delta \log N}/\sqrt{2(\delta-1)}$, and $D_c = W_c - \Delta G$ is the dissipated work at the critical work value, where ΔG is the 'true' free energy difference. The result for the $\lambda \gg 1$ regime is not shown; for practical applications, one normally cares about the bias in cases of limited N compared to work distribution width, Ω. The 'true' free energy difference ΔG is not known *a priori*, but a saddle point calculation over distribution 3.25 yields $D_c \simeq (\delta-1)\left(\frac{\Omega}{\delta}\right)^{\delta/(\delta-1)}$ to leading order (the equivalence is exact for Gaussian work distributions), with errors typically less than 0.5 - 2 $k_{\mathrm{B}}T$. This justifies approximating D_c using fit parameters[206].

One can also use fits of $p(W)$ to estimate the probability of observing *dominant* work values, which contribute most substantially to the Jarzynski average. A generic distribution of dominant work values, $g(W) = e^{-\beta W}p(W)$, is depicted in Figure 3.15; while the most probable work value to be observed is the peak of $p(W)$, $\bar{W}$, the most dominant value in an exponentially-weighted average is

$W^\dagger = e^{-\beta \bar{W}}$. For wider $p(W)$ distributions, corresponding to processes with greater average dissipated work $W_{dis} = \bar{W} - \Delta G$, the peak of $g(W)$ is found at increasingly improbable work values[124].

Figure 3.16 shows work distributions from hairpin pulling simulations (orange), collected at the final time sampled, along with fits to Eq. 3.25 (black), for some representative simulation parameter sets. Only the tails of the distributions – primary determinants of error in free energy estimates – were fit, according to the method of Palassini & Ritort[206]. Also shown are the 'best' fits to the distribution, obtained using the maximum number of trajectories simulated with given parameters (black, dashed). Fit parameters were used to estimate bias according to Eqs. 3.26 or 3.27 (depending on the value of λ); these 'Ritort bias' results appear, alongside the empirical bias B_N calculated from the landscapes, in Table 3.4, and are accurate within an order of magnitude.

Empirical $g(W)$ distributions – computed from the orange work distribution histograms – are shown as purple histograms in Fig. 3.16, alongside blue curves corresponding to the $g(W)$ obtained from the work distribution fits. The degree to which the purple histograms sample the blue curves provides a rough ruler for gauging the convergence of a landscape. An estimate of the average number of trajectories needed to sample the peak of $g(W)$ can be obtained by by integrating $p(W)$ on $(-\infty, W^\dagger]$:

$$N_{est,WD} \approx \frac{1}{\int_{-\infty}^{W^\dagger} p(W)\,\mathrm{d}W}. \tag{3.28}$$

$N_{est,WD}$ estimates are included in Table 3.4. Our results suggest that if the $g(W)$ distribution is sampled near its peak, one can expect the resulting landscape to have a small ($\mathcal{O}(1 k_\mathrm{B}T)$) bias.

An important caveat must be borne in mind when attempting to characterise the bias in a particular simulation setup using fits of $p(W)$: such fits are highly susceptible to statistical sampling errors. To illustrate this point, consider the dashed lines in Fig. 3.16 I and II, which more closely approximate the 'true' $p(W)$ as they are fits of Eq. 3.25 to the largest number of trajectories collected for a given

Figure 3.16a: Empirical work probability distributions (orange) and the same distributions weighted by $e^{-\beta W}$ (purple) collected at the final time in pulling simulations. Black curves are $p(W)$ fits according to Eq. 3.25, and blue curves are $g(W) = e^{-\beta W} p(W)$. The extent to which the purple histograms sample the blue curves is indicative of how well-converged the free energy estimate is. I and II correspond respectively to the 'worst' and 'best' of 10 landscapes of where 100 curves were randomly selected from a larger set, shown in III. Dashed black and blue curves in I and II represent the fits in III, and are thus more accurate estimates of $p(W)$ and $g(W)$. Differences between the dashed and solid curves indicate potential pitfalls of using a small number of curves to estimate $p(W)$. (a) 6bp hairpin, stiffness 0.57 pN/nm and pulling rate 5.62×10^6 pN/s. (b) 6bp hairpin, stiffness 57 pN/nm and pulling rate 5.62×10^6 pN/s. (c) 6bp hairpin, sawtooth protocol with force loading rate $\dot{F} = 3.21 \times 10^8$ pN/s.

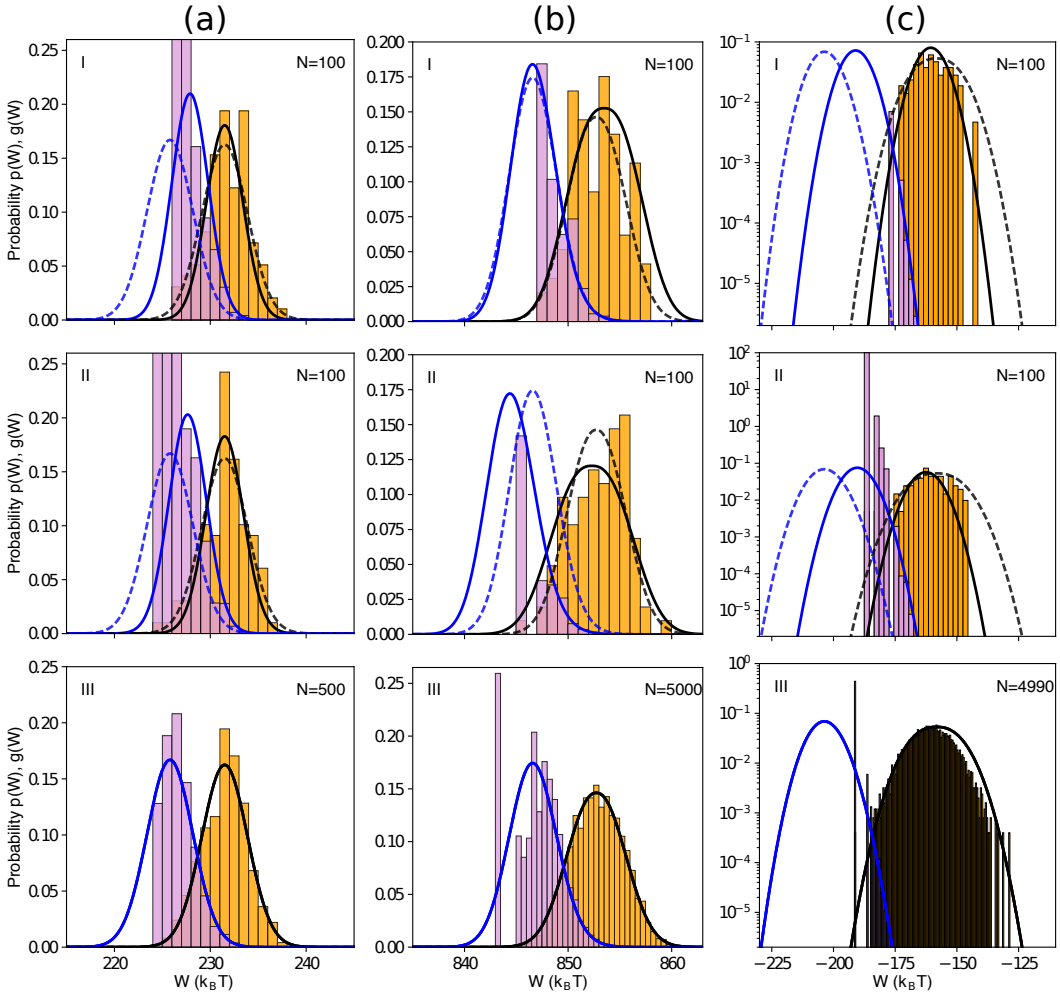
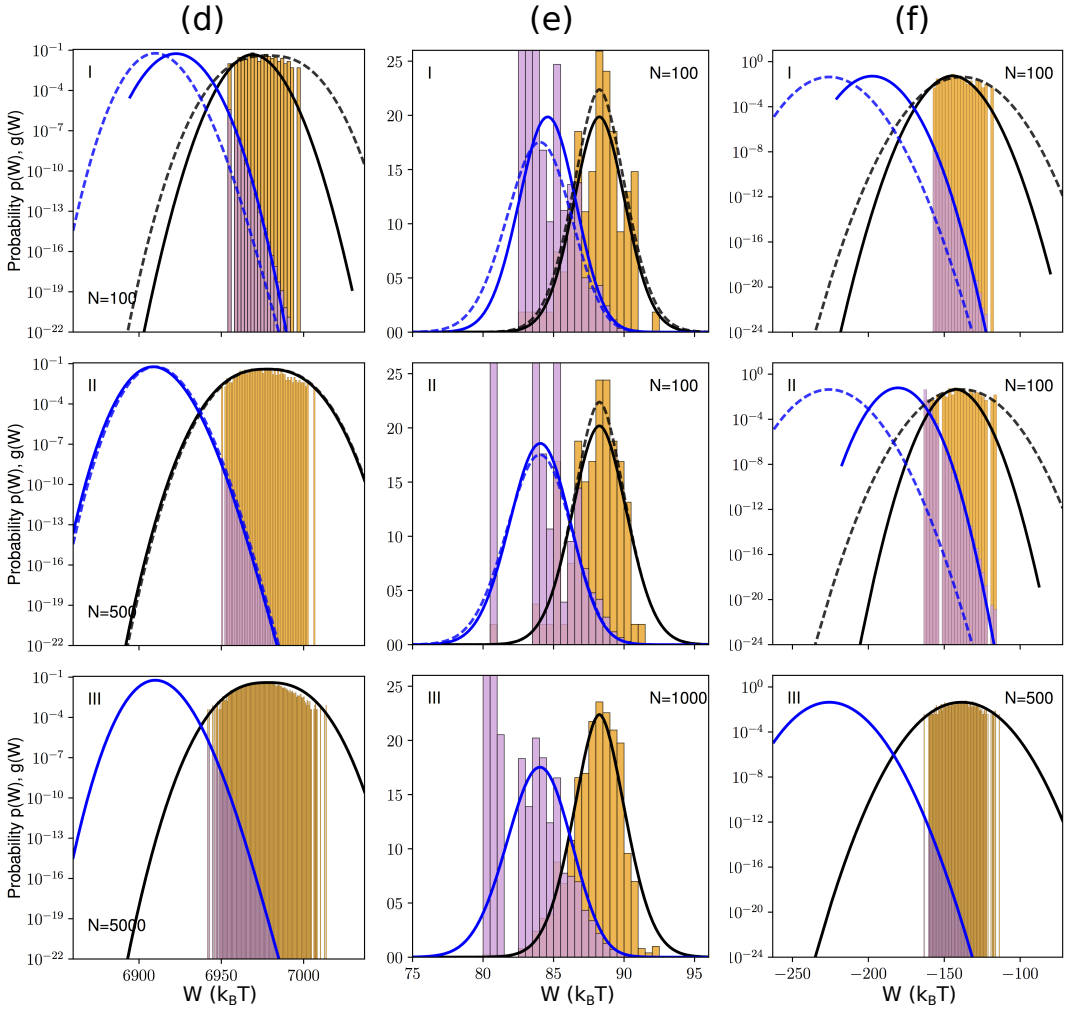


Table 3.4: Results of applying the bias characterization frameworks discussed in the text to representative simulation sets for the hairpins; corresponding work distributions are shown in Fig. 3.16. B_N is the value measured from reconstructed landscapes; Ritort Bias is an estimate of the bias computed with Eqs. 3.26 or 3.27, using parameters from a fit of the empirical work distribution $p(W)$ to Eq. 3.25; $N_{est,WD}$ and $N_{est,Jar}$ are estimates of the average number of trajectories required to sample the peak of the $g(W)$ distribution, and thus at least approach a converged landscape, using Eqs. 3.28 and 3.30, respectively. Average dissipated work W_{dis} is estimated according to Eq. 3.29.

Harmonic protocol, 6bp							
Parameters	Fig. 3.16	N	B_N (k_BT)	Ritort Bias (k_BT)	$N_{est,WD}$	$N_{est,Jar}$	W_{diss} (k_BT)
0.57 pN/nm 5.62×10^6 nm/s	(a)	500	-0.48	0.51	140	13	2.57
		100	-0.17	0.64	44	10	2.27
		100	-0.83	0.48	49	20	2.95
57.1 pN/nm 5.62×10^6 nm/s	(b)	5000	2.21	0.98	142	23	3.15
		100	3.05	0.51	137	9	2.18
		100	1.58	0.62	240	31	3.45
Sawtooth protocol, 6bp							
15.2 ps 0.0049 pN	(c)	4990	1.99	4.63	5.15×10^9	2.41×10^{10}	24.9
		100	13.82	4.61	6.05×10^7	1.57×10^5	12.0
		100	4.70	4.60	2.32×10^7	4.46×10^8	19.9
Harmonic protocol, 20bp							
5.71 pN/nm 5.62×10^7 nm/s	(d)	5000	11.24	9.42	2.18×10^{13}	1.31×10^{11}	25.6
		500	17.92	12.92	2.72×10^{13}	1.34×10^8	18.71
		100	20.88	10.10	9.73×10^{11}	2.9×10^6	14.89
5.71 pN/nm 5.62×10^5 nm/s	(e)	1000	-1.38	0.46	58	10	2.26
		100	-0.78	0.38	46	7	1.96
		100	-2.53	0.29	55	30	3.37
Sawtooth protocol, 20bp							
15.2 ps 0.00049 pN	(f)	500	22.17	17.45	6.27×10^{19}	1.59×10^8	18.9
		100	26.8	10.51	2.78×10^{13}	39341	11.0
		100	20.62	7.89	5.07×10^{12}	8.03×10^7	18.2

Figure 3.16b: (continued) (d) 20bp hairpin, stiffness 5.71 pN/nm and pulling rate 5.62×10^7 pN/s. In this case, instead of the 'best' set of $N = 100$ curves, II shows the 'worst' set of $N = 500$ curves pulled from the larger set in III. (e) 20bp hairpin, stiffness 5.71 pN/nm and pulling rate 5.62×10^5 pN/s. (f) 20bp hairpin, sawtooth protocol with force loading rate $\dot{F} = 3.21 \times 10^7$ pN/s.



set of parameters (Fig. 3.16 III shows this largest set of trajectories). These dashed curves differ from the solid lines – the 'first pass' fits of $p(W)$ using fewer curves – and slight differences in the $p(W)$ fit parameters lead to significant differences in $g(W)$ (and thus $N_{est,WD}$). This problem is exacerbated farther from equilibrium, for wider $p(W)$. Nevertheless, as is evident from Table 3.4, applying the aforementioned techniques can supply a reasonable estimate of the order of magnitude of trajectories that would, on average, be necessary to reduce biases to $\mathcal{O}(1k_{\mathrm{B}}T)$.

Also included in Table 3.4 are two additional characterisations of bias from the literature. Firstly, the average dissipated work, W_{diss}, is approximated using

$$W_{diss} = \bar{W} - \Delta G \approx \bar{W} - \Delta G_N, \tag{3.29}$$

where ΔG_N is the finite-N Jarzynski free energy estimate. Hummer[111] has suggested W_{diss} should be $\approx 1k_{\mathrm{B}}T$ for optimal results, though this approximate rule was derived for Gaussian work distributions only. Secondly, an estimate of the number of trials required for 'reasonable convergence' of the Jarzynski average was provided by Jarzynski[124;304]:

$$N_{est,Jar} \approx e^{-\beta W_{diss}}. \tag{3.30}$$

This estimate differs from $N_{est,WD}$, especially for simulations far from equilibrium (see Table 3.4), but combining both estimates gives a sense of whether $\mathcal{O}(100)$ or $\mathcal{O}(> 1 \times 10^6)$ trials are required. We note that it does not appear to be necessary for $W_{diss} \approx 1k_{\mathrm{B}}T$ to obtain biases $< 1k_{\mathrm{B}}T$.

Our recommendation is to employ all the aforementioned measures in concert to impart a rough sense of how reliable a given landscape reconstruction is. Landscapes that are well-converged are seen in Fig. 3.16 and Table 3.4 to have well-sampled $g(W)$ and estimates for $N < 100$. Landscapes that are very poorly converged have correspondingly sparsely-sampled $g(W)$ and estimates for $N > 1 \times 10^6$.

3.4 Application of non-equilibrium landscape analysis to a novel system: DNA origami handles

In addition to supplementing experimental studies, another context where the above-validated techniques are particularly useful is in the simulation of large systems for which conventional, Monte Carlo equilibrium simulation techniques are impractically slow. Here, we apply non-equilibrium analysis to explore the use of DNA origami tubes as a substitute for the dsDNA handles traditionally used in SMFS experiments.

As already mentioned, the elastic response of the long dsDNA linker handles in SMFS experiments gets convolved with the energy landscape of the molecule under

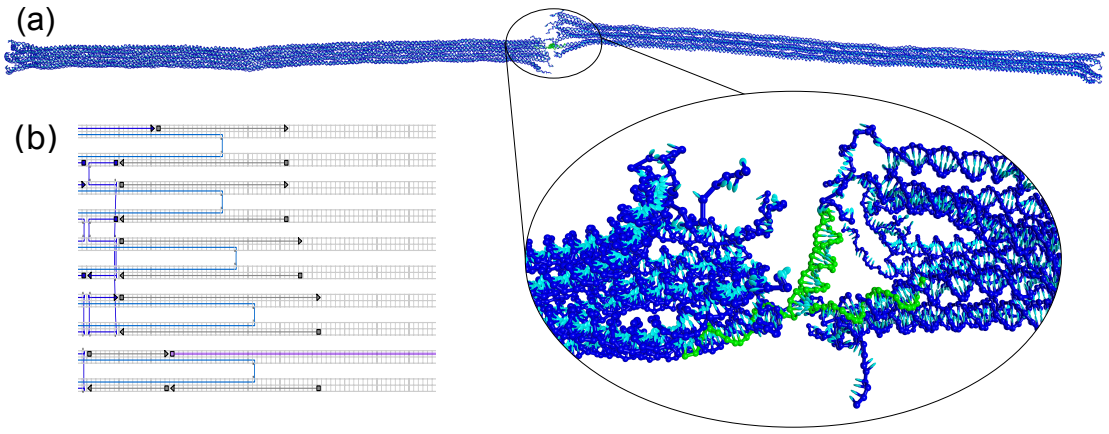


Figure 3.17: (a) oxDNA representation of the DNA origami handles of Pfitzner et al.[209]. Hairpin 20R55T4 is suspended between the handles as part of a bridging staple; see blow-up. (b) Fragment of the innermost part of the *left* handle's Cadnano[58] design file, showing detail of the overhanging portions of staples (grey), which are poly-T. The purple strand contains the hairpin.

investigation, 'smearing out' features of interest. Pfitzner et al.[209] suggested using DNA origami[228] tubes to mitigate this issue, as DNA origami tubes are much stiffer than dsDNA, and thus convolve the molecule's landscape with a much narrower point spread function (PSF)[298]. It is an open question whether such origami handles *alter* the intrinsic landscape of the molecule of interest – for instance through steric interactions – and origami handles have not yet seen widespread adoption in the SMFS experimental community. By replicating the pulling experiments of Pfitzner et al.[209] and comparing hairpin landscapes extracted with and without the presence of origami handles, we shed light on this question. Because the origami handles used are so large, conventional equilibrium simulations are unfeasible for this system.

3.4.1 System description and methods

Pfitzner et al.[209] perform constant-force LOT experiments on a 6bp and 20bp hairpin (06R50T4 and 20R55T4 as described in section 3.3.1 and shown in Fig. 3.4) to extract free energy landscapes, using origami 10-helix bundles (10HB) as linker handles. Figure 3.17 shows an oxDNA representation of the handles, connected by a bridging staple that contains the 20bp hairpin. The left and right handles contain 15 275 nt (15 247 nt in the case of the 6bp design) and 15 189 nt, respectively. Each

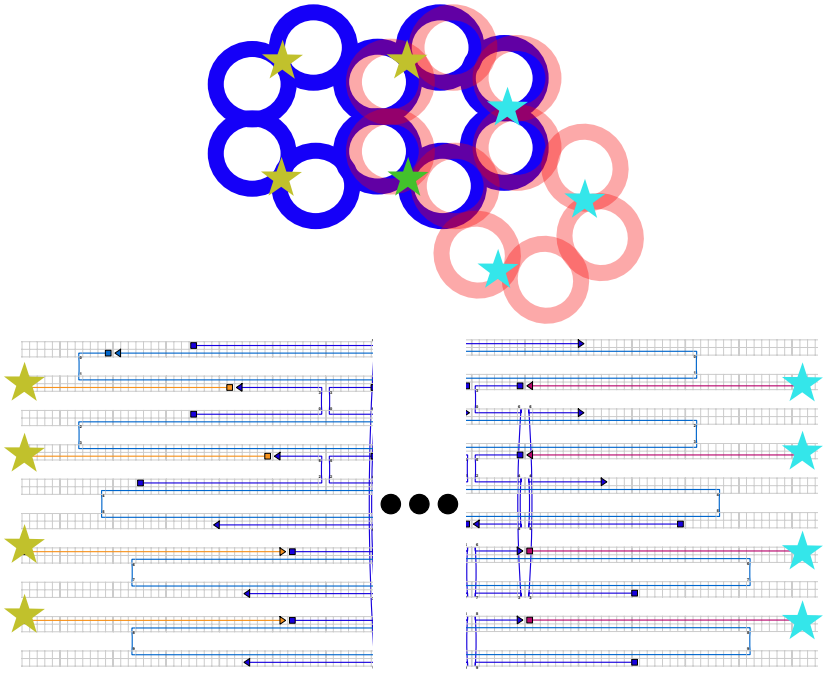


Figure 3.18: Top: Cross section of the 10 helix bundles. Blue is the 'left' handle (as depicted in Fig. 3.17), and red is the 'right' handle. The red handle has been slightly offset as a visual aid, but six helices of each handle are directly aligned in the plane perpendicular to trap movement. Bottom: Cadnano snapshots of the handle ends, depicting the experimental bead attachment points and the simulated location of 3D harmonic traps. Yellow stars are traps on the left handle, blue are traps on the right handle, and the green star indicates where left and right handle traps share the same position in the plane perpendicular to trap motion.

250 nm-long handle contains overhanging portions of staples that crowd the vicinity of the hairpin, potentially altering the measured landscape.

In the LOT experiments, forces were applied to four staples at the ends of each origami handle, which were functionalized with molecules that enabled linkage to laser-trapped dielectric beads. To mimic this as closely as possible, in simulations, 3D harmonic traps of stiffness 5.71 pN/nm were applied to the centres of mass of the nucleotides at the ends of these extended staples; these traps are indicated by stars in Figure 3.18. The positions of the traps were chosen such that the helices containing the hairpin staple were aligned; see Figure 3.18 for the configuration. The traps on the right handle were then moved at constant rate away from the left handle.

Simulations were carried out for the 6bp hairpin at a pulling rate of 2.8×10^{7} nm/s and the 20bp hairpin at pulling rates of 5.6×10^{7} nm/s and 5.6×10^{6} nm/s, respec-

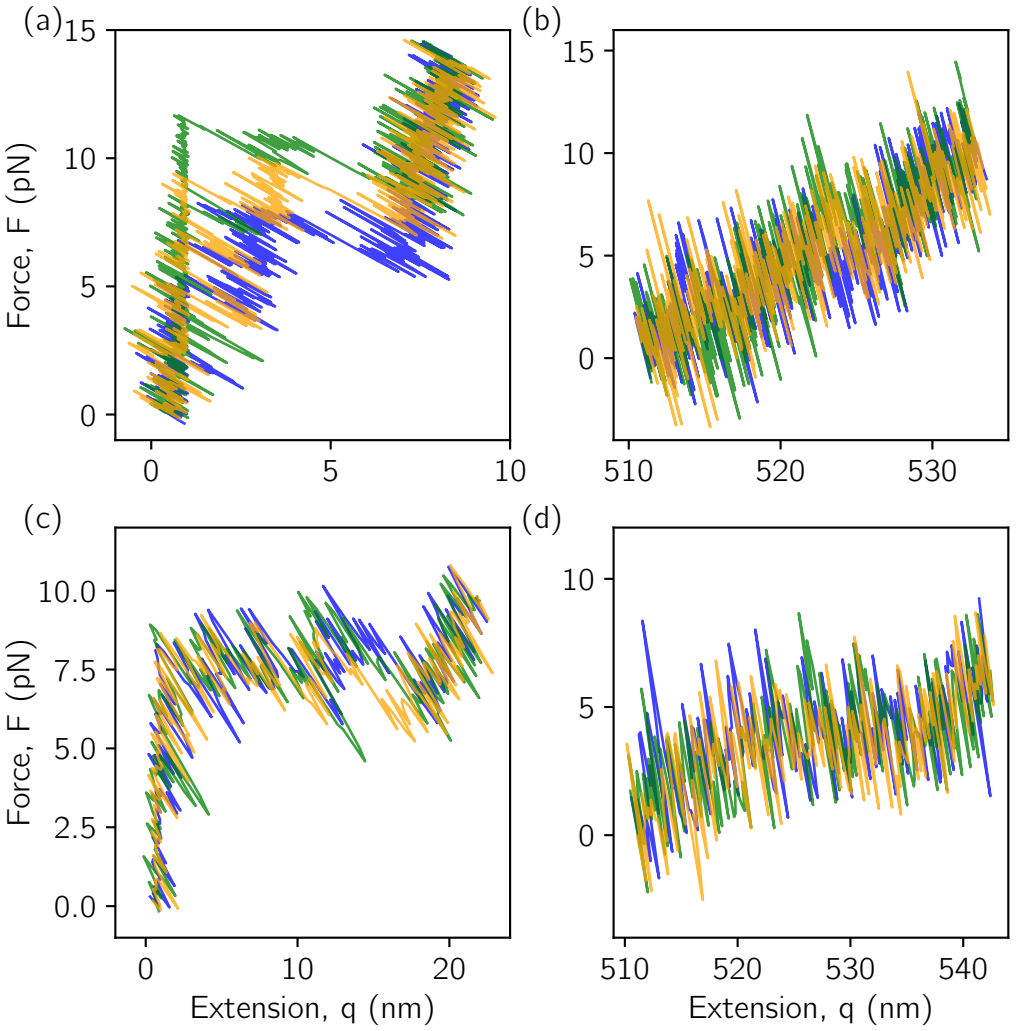


Figure 3.19: Sample force-extension curves (FECs) for the handle-hairpin setup. (a) FECs for the 6bp hairpin unfolding between the handles at pulling rate 2.8×10^7 nm/s. (b) FECs for the 6bp hairpin pulls in (a), but using the end-to-end distance of the handles themselves. (c) FECs for the 20bp hairpin unfolding between the handles at pulling rate 5.6×10^6 nm/s. (d) FECs for the 20bp hairpin pulls in (c), but using the end-to-end distance of the handles themselves.

tively. In total, 500 trajectories for the 6bp hairpin (301 days on GPU), 300 for the 20bp hairpin at the faster rate (154 days on GPU), and 64 for the 20bp hairpin at the slower rate (288 days on GPU) were collected.

3.4.2 Results

During simulations, the average separation of the traps along the pulling axis and the end-to-end extension of the hairpin were measured; these were used to construct FECs, shown in Figure 3.19. Because the force of the harmonic traps is not communicated directly to the hairpin, but rather is convolved with the elastic response of the handles, in order to arrive at reasonable estimates for the forces in the hairpin FECs, a more realistic value for the stiffness actually experienced by the hairpin was needed. To estimate this, a histogram of end-to-end distances along the axis of constraint, q_{tot}, for the fully unbound hairpin attached to the trapped handles was fit to a Gaussian. The variance of q_{tot} was assumed to be equal to the sum of the individual variances of each hairpin end nucleotide in a hypothetical 3D harmonic trap: $\sigma^2_{q_{tot}} = \sigma^2_{q_1} + \sigma^2_{q_2} = \frac{1}{\beta k} + \frac{1}{\beta k}$, with the individual variances $\sigma^2_{q_i}$ equal and determined using the equipartition theorem: $\sigma^2_{q_i} = k_{\mathrm{B}}T/k$. This procedure yielded a stiffness of $k = 1.58$ pN/nm.

The results of applying the landscape reconstruction procedure detailed in section 3.3.2 are shown in Figures 3.20-3.22. Landscapes were reconstructed using the end-to-end distances of both the handles (green) and the hairpins (orange). For comparison, the equilibrium landscape (blue) and a non-equilibrium landscape (black) collected at similar stiffness and pulling speed values to the orange curves, but *without* origami handles, are also shown.

In all cases, the FECs of the handles (Fig. 3.19 (b) and (d)) are insufficient to resolve much of the unfolding barrier at all; see green curves in Figs. 3.20-3.22. As described in section 3.3.1, this is an effect of convolution of the hairpin's unfolding with the elastic response of the handles. This convolution also affects the orange landscapes, as the hairpins don't directly experience the harmonic traps. Furthermore, the convolution effect is more significant here than in the experimental work of Pfitzner et al.[209], as their experiments were carried out at constant force; non-equilibrium studies suffer from more complex, force-dependent handle elastic response. While deconvolution has been attempted for non-equilibrium data[303], the

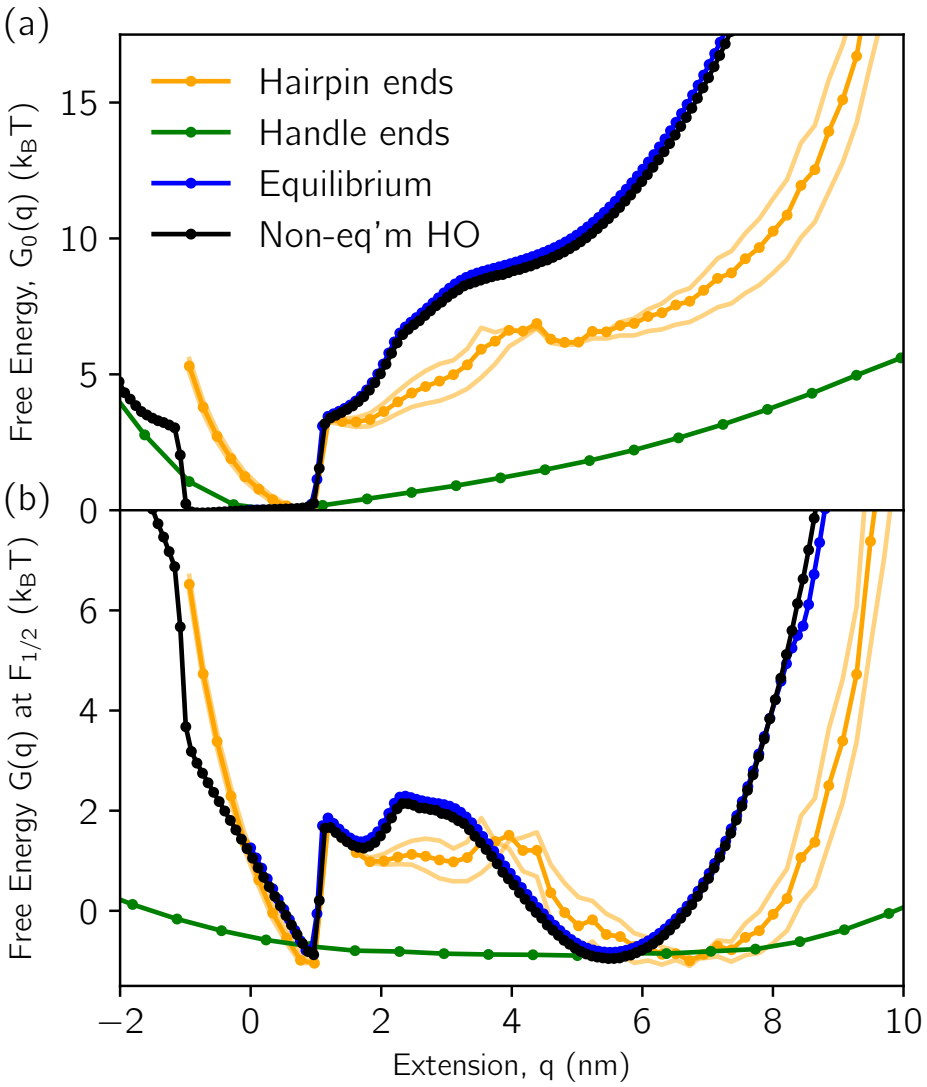


Figure 3.20: Free energy landscape reconstructions from 500 trajectories of the 6bp hairpin pulled at 2.8×10^7 nm/s. Orange landscapes are the result of using FECs collected using hairpin end-to-end extension directly; green landscapes use the end-to-end extension of the handles, and are thus 'smeared out'. For comparison, the equilibrium (blue) landscape is shown, as well as a 'hairpin-only' (HO) non-equilibrium landscape (black) collected under similar pulling conditions: 500 curves at 5.6×10^6 nm/s and 0.57 pN/nm. (a) Zero-force energy landscapes; (b) landscapes tilted to the forces at which the folded and unfolded wells have the same free energy (3.9 pN for the orange curve). The light orange bounding curves represent landscapes using 1.58 pN/nm±25% stiffness to construct the hairpin landscapes; that they retain the discrepancy in shape with equilibrium result suggests the origami handles are intrinsically altering the landscape.

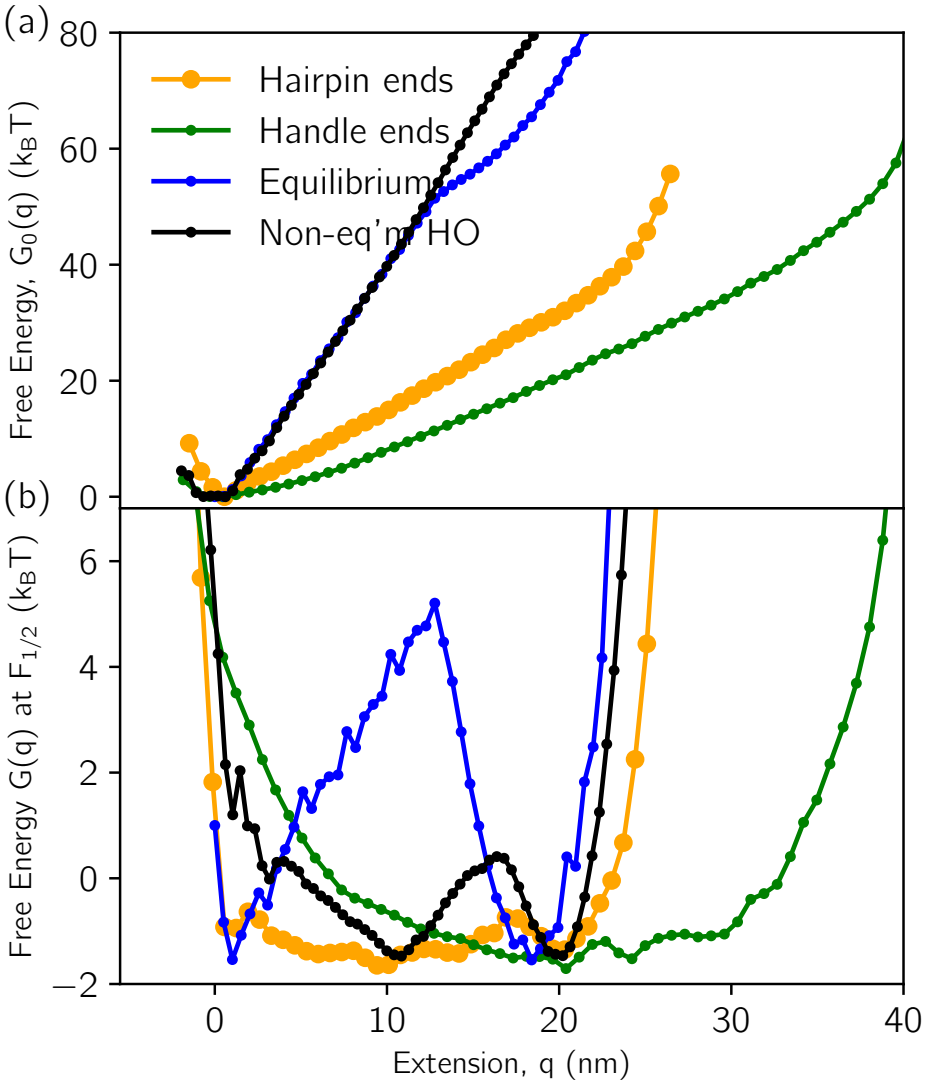


Figure 3.21: Free energy landscape reconstructions from 300 trajectories of the 20bp hairpin pulled at 5.6×10^7 nm/s. Orange landscapes are the result of using FECs collected using hairpin end-to-end extension directly; green landscapes use the end-to-end extension of the handles, and are thus 'smeared out'. For comparison, the equilibrium (blue) landscape is shown, as well as a 'hairpin-only' (HO) non-equilibrium landscape (black) collected under similar pulling conditions: 500 curves at 5.6×10^7 nm/s and 5.71 pN/nm. (a) Zero-force energy landscapes; (b) landscapes tilted to the forces at which the folded and unfolded wells have the same free energy (5.0 pN for the orange curve). Note that for these rates even the hairpin only simulations show significant differences from equilibrium curves.

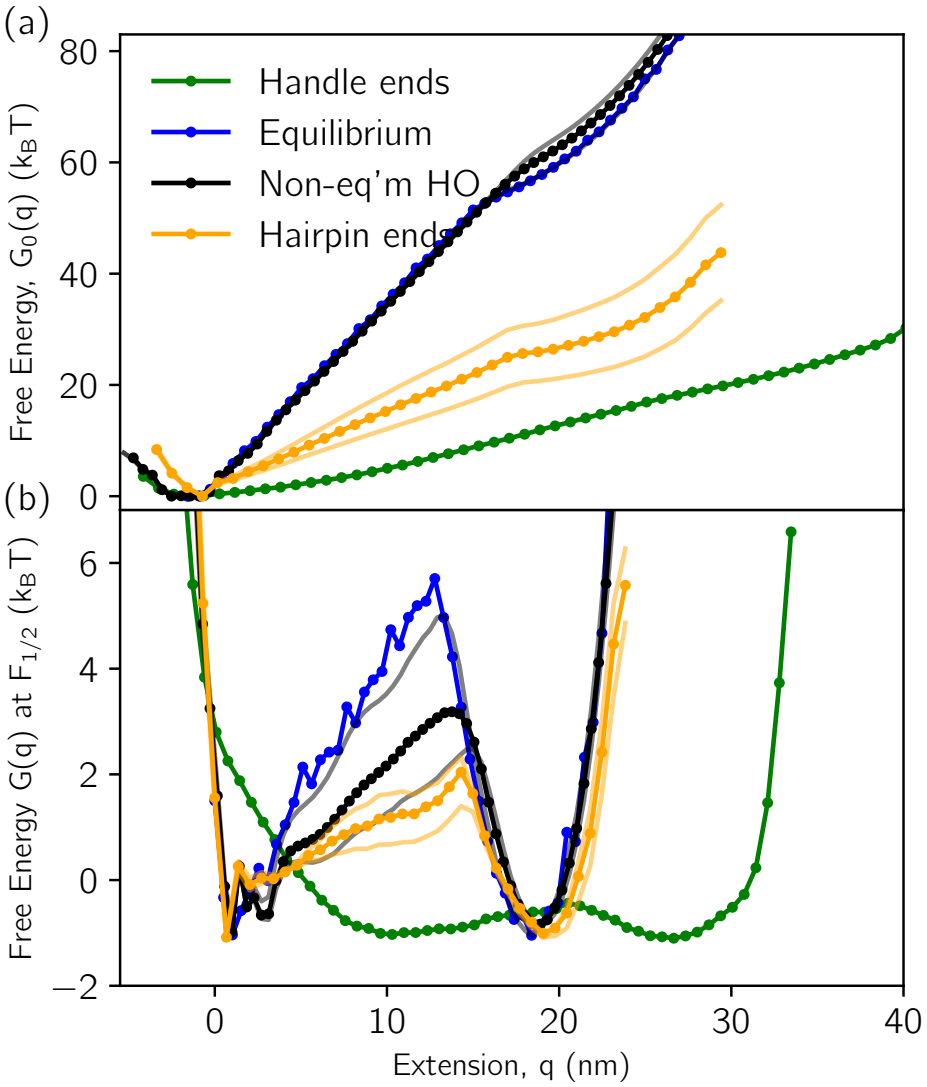


Figure 3.22: Free energy landscape reconstructions from 64 trajectories of the 20bp hairpin pulled at 5.6×10^6 nm/s. Orange landscapes are the result of using FECs collected using hairpin end-to-end extension directly; green landscapes use the end-to-end extension of the handles, and are thus 'smeared out'. For comparison, the equilibrium (blue) landscape is shown, as well as a 'hairpin-only' non-equilibrium landscape (black) collected under similar pulling conditions: 100 curves at 5.6×10^6 nm/s and 2.86 pN/nm. (a) Zero-force energy landscapes; (b) Landscapes tilted to the forces at which the folded and unfolded wells have the same free energy (4.9 pN for the orange curve, 3.0 pN for the green curve)). The grey curves bounding the black landscape are error bounds derived by taking the 'best' and 'worst' sets of 100 curves from a pool of 500. The light orange bounding curves represent landscapes using 1.58 pN/nm±25% stiffness to construct the hairpin landscapes; that they retain the discrepancy in shape with equilibrium result suggests the origami handles are intrinsically altering the landscape.

method involves using the average force at each extension to determine an extension-dependent PSF[79], and consequently, data that exhibit a broad range of forces at a given extension value will not be amenable to this approach. Since our simulations are carried out father from equilibrium than experiments, our force distributions are wider; in Fig 3.19 (a) and (b), the force varies by as much as ∼10 pN for some extensions. For this reason, de-convolution has not been attempted here.

The 20bp, 5.6×10^7 nm/s data yields an inadequately resolved barrier even when the hairpin end-to-end distance is used (Fig. 3.21, orange curve); this is expected given that even the handle-free landscape (Fig. 3.21, black curve) at this pulling rate captures the barrier poorly. We can also confirm that the landscape is inaccurate by applying the error analyses described in section 3.3.2: Figure 3.23 (b) reveals that the $g(W)$ distribution is poorly sampled, and about $\mathrm{N}_{est,WD} \approx 9 \times 10^4$ trajectories would be required, on average, to resolve its peak. However, we can glean useful insights from the hairpin landscapes in Figs. 3.20 and 3.22. In both cases, an error analyses reveals that we have sampled the $g(W)$ distribution well (Figs. 3.23 (a) and (c)) and estimating the number of trials required to resolve the peak in $g(W)$ yields $\mathrm{N}_{est,WD} \approx 300$ for the 6bp hairpin and $\mathrm{N}_{est,WD} \approx 10$ for the 20bp hairpin pulled at the slower rate. We note, however, that the small number of trajectories collected for the 20bp hairpin at the slower rate render this an extremely rough estimate. The remaining discrepancies between the hairpin + handles landscapes and the hairpin only landscapes (Figs. 3.20 and 3.22, orange and black, respectively) – namely, the lower barrier height and the increased curvature near the barrier peak visible in the orange landscapes – can be attributed to either the elastic response of the handles or an intrinsic difference between the landscapes with and without handles present. To fully remove the effect of the handles' elastic response, an extension-dependent deconvolution procedure valid across large force ranges, not currently theoretically available, would be necessary; however, we can get a rough sense of whether these differences in shape are artefacts of handles modulating the stiffness experienced by the hairpin by varying the effective stiffness used in the reconstructions by ±25%; the results are the light orange bounding curves. While

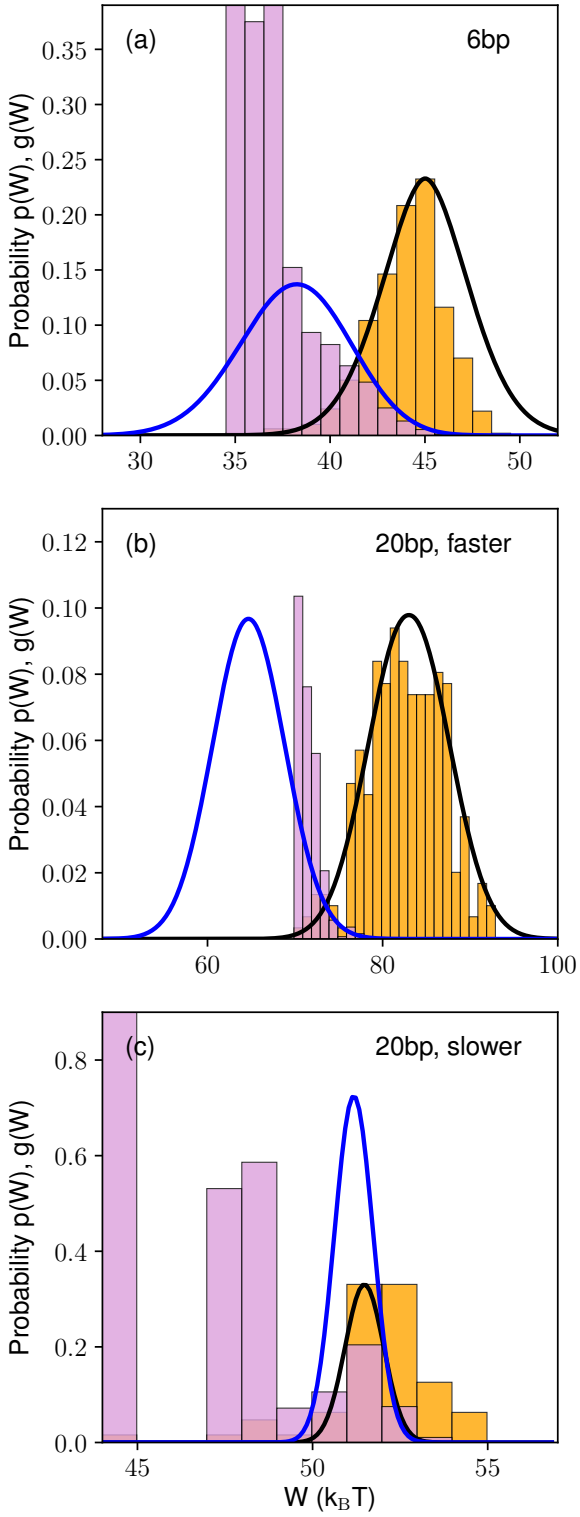


Figure 3.23: Empirical work probability distributions (orange) and the same distributions weighted by $e^{-\beta W}$ (purple) collected at the final time in pulling simulations. Black curves are $p(W)$ fits according to Eq. 3.25, and blue curves are $g(W) = e^{-\beta W} p(W)$. The extent to which the purple histograms sample the blue curves is indicative of how well-converged the free energy estimate is. Distributions shown are for (a) the 6bp hairpin (N=500 curves) and the 20bp hairpin pulled at (b) $5.6{\times}10^7$ nm/s (N=300 curves) and (c) $5.6{\times}10^6$ nm/s (N=64 curves).

the differences in barrier height may be attributable to convolution effects, the increased curvature of the barrier does not appear to be. This suggests that the steric crowding experienced by the hairpin when attached to the origami handles (see Fig. 3.17 (a), blow-up) *is* affecting the intrinsic landscape being measured. To confirm this steric effect, future work could (i) perform additional simulated pulls on the 20bp hairpin at the slower rate to increase the quality of the landscape and reliability of the $p(W)$ fit; (ii) use truncated origami handles – for which convolution effects would be lessened – that retain the features of the ends nearest the hairpin in simulations; or (iii) develop a deconvolution method valid across broad force ranges that could be used to remove handle artefacts.

3.5 Conclusions and outlook

This work has demonstrated the ability of the coarse-grained model oxDNA to successfully capture the folding free energy landscapes of DNA hairpins, both by comparing equilibrium simulation results to experiments and by validating a non-equilibrium analysis framework based on the Jarzynski equality for oxDNA. Given that the Jarzynski equation is only correct in the limit of infinite ensemble averages, an approach for assessing the error in finite-N landscape reconstructions has been outlined, building on the work of previous authors. We have illustrated the potential value of non-equilibrium landscape reconstruction methods by applying them to a novel system, DNA hairpins attached to DNA origami linker handles proposed for use in SMFS experiments, to investigate whether such handles – through steric hindrance – alter the intrinsic landscapes of molecules being studied. Future work could refine our results by collecting additional data and mitigating the convolution of the hairpin landscape with the elastic response of the handles.

Improvements to the non-equilibrium landscape analysis method presented here could include implementing an importance-sampling scheme for non-equilibrium trajectories, favouring those trajectories with lower work values that contribute more to the Jarzynski average[51], or exploring the use of alternative force protocols that studies have suggested can lead to faster convergence of the Jarzynski average[240;80].

Continued use of force protocols that mimic SMFS experiments, however, promises significant fruit: in particular, the methods we have validated can be used to explore the effect of reaction coordinate (in our case, end-to-end extension) on reconstructed landscapes, confirming whether a given experiment has chosen an appropriate degree of freedom to track the unfolding reaction[60] and suggesting more appropriate measurements if not. The effect of experimental setup on intrinsic landscapes can also be explored, as we have outlined for the DNA origami handles. Finally, structural changes underlying experimental energy landscapes can be directly visualized with the aid of coarse-grained simulations. We look forward to future use of oxDNA to complement non-equilibrium experiments.

In my own stories I have found that violence is strangely capable of returning my characters to reality and preparing them to accept their moment of grace.

— Flannery O'Connor, from *On Her Own Work* (1963).

4

Force-induced unravelling of DNA origami

Contents

4.1 Introduction 73
4.2 Methods 77
4.3 Rothemund Tile 79
4.4 10-Helix Bundle 89
4.5 7-tile Biosensor 97
4.6 Conclusions 99

This chapter is based on an article entitled *Force-induced unravelling of DNA Origami*[64] that has been submitted for publication in ACS Nano. The experiments referred to in this chapter were carried out by D. Smith, M. Jobst, M. Sajfutdinow, and T. Liedl.

4.1 Introduction

As intimated in the preceding chapters, the behavior of DNA under tension has been studied extensively both theoretically and experimentally. Single-molecule force studies have been employed to elucidate the elastic properties of double- and single- stranded DNA[259;21;175]; to characterize the mechanisms of duplex instability at higher forces[66;268;146;279;93;226]; and to examine torsional buckling of the helix under a mixture of torque and tension[71;194]. These explorations have illumined

the cellular processes that mediate structural changes in DNA, including gene regulation, transcription, and recombination[307;87;27]. Furthermore, the now well-understood behavior of duplex DNA under force has been exploited for a wide range of applications, including employing DNA in molecular force sensors[85] to characterize mechanosensitive receptors[284] and cellular traction forces[20].

Denaturing of the DNA helix is typically modelled as a thermally-activated process, in which the dissociation rate is governed by the height of the free energy barrier between the fully-zippered state – stabilized by hydrogen bonding and base stacking – and a transition state characterized by a critical number of broken base pairs[35;185]. Application of an external biasing force favors more extended configurations, lowering the effective barrier to the transition state until it becomes surmountable by thermal energy[67].

The force at which duplex DNA yields under tension depends on duplex length[100;185], base sequence and any periodicity therein[189], force application geometry, whether static or varying tension is applied and for how long[67;35;185], and solution conditions including temperature[292;185] and salt concentration[288;75]. Under shear stress, when tension is applied at either end of one strand of the duplex, DNA duplexes typically unbind at forces between 30 and 65 pN[100;93], while in an unzipping geometry, where tension is applied to both strands at the same end of the duplex, yielding occurs at forces between 10 and 20 pN[110;35;222]. The marked difference between the critical forces in shearing and unzipping geometries can be understood by considering the extension gained per base pair broken in each case. Since a force F biases the unfolding landscape by an amount $\sim F\Delta z$, with z the extension of the DNA along the force direction, the breaking of a single base pair in a shearing geometry, which allows an extension gain of $\Delta z/\mathrm{bp} \sim 0.3\,\mathrm{nm}$, will be less favorable than base pair breakage in an unzipping geometry, which allows for an extension gain of $\Delta z/\mathrm{bp} \sim 1.3\,\mathrm{nm}$[185].

The drive to produce DNA modules sufficiently rigid for use as building blocks for crystalline arrays and nanomechanical devices cultivated more complex arrangements of the well-studied DNA duplex[248;132] – namely, DNA helices linked by multiple

crossover junctions. One of the simplest and earliest of these designs is two helices bundled together by a double crossover (DX) to form a modular tile[76]. Such basic tile motifs can be expanded to construct 3D structures, including triangular tensegrity lattices[158], polyhedra[103], and nanotubes[174].

The advent of DNA origami[228] – a construction technique employing kilobase-long scaffold strands joined together by shorter, staple strands to form arbitrary shapes – allowed the construction of even larger and more intricate designs, which expanded the reach of DNA nanotechnology to nanosensing[19;141], nanomedicine[4], and nanoelectronics[210]. Finally, a similar, but scaffold-free, approach was proposed by Peng Yin and coworkers[301], in which only short single strands with four distinct binding domains are used. These versatile single-stranded tiles (SSTs) can also be fashioned into a wide array of 2D[286] and 3D[135] shapes.

As many applications depend on the increased structural rigidity of nanostructures as compared to duplex DNA, there has been a significant effort to characterize their elastic properties. Ligation closure experiments on the DX tile revealed a persistence length about twice that of duplex DNA[231]. Even larger persistence lengths, $\sim 10 - 20\times$ that of duplex DNA, have been achieved with nanotubes comprised of 6 helices[133;283]; for bundles of more helices, persistence lengths $40 - 200\times$ that of dsDNA have been observed[239]. Computational models have also contributed to the exploration of nanostructure mechanical behaviour. At the finest levels of detail, atomistic MD simulations have successfully captured the structural rigidity of simple DNA origami and SST structures[302;257]. Coarse-grained descriptions such as finite-element models[31;137] and models that coarse grain at the level of multiple base pairs[221] or single nucleotides[243] have likewise provided useful analyses of the mechanical flexibility of origami structures.

More sophisticated applications of DNA nanotechnology require an understanding of the force response beyond that of the basic elastic properties. Such understanding is crucial particularly for applications that exploit behaviour under force, such as biosensors[134;141;95], springs[121], force probes[78;193], and tensegrity structures[155;255]. Further, characterizing the long-term stability of nanostructures

under constant or intermittent stress is necessary for their use as movable joints and mechanical components like hinges[170] and as hybrid nanopores, which undergo a constant denaturing stress from ionic current flow[19;287]. The recent use of DNA nanotube hydrogels to study theoretical models of semiflexible polymers underscores the importance of knowledge of DNA nanostructural stability under force, as it hinged upon the robustness of the nanotubes under shear force[244].

Given their substantial contribution to our understanding of the DNA duplex, force spectroscopy experiments are a logical stride in the development of DNA nanotechnology. Experiments have only just begun to explore more complex force-response behaviour of DNA nanostructures, however; magnetic tweezer experiments have revealed torsional rigidities of origami 6-helix bundles in excess of $\sim 4\times$ those of dsDNA[133], and recent optical tweezer experiments on origami nanotubes found disassembly forces of 40-50 pN[253]. Here, we build a more comprehensive picture of DNA origami force response by combining coarse-grained simulations with AFM experiments to characterize the unraveling of a standard origami system: the Rothemund tile[228]. We also perform simulations of two previously-published experimental systems for which force-extension data are available: a 10-helix bundle[13] and a biosensor[141] designed such that its force-extension behaviour signals the presence of aptamers in solution.

By pursuing a combined simulation and experimental approach, we are able to correlate features in force-extension data directly with underlying structural changes, yielding insights inaccessible via experiment alone. Finite-element approaches to modelling origami and models that coarse grain on scales larger than a single nucleotide cannot capture the base-pair breakage fundamental to force spectroscopy studies[137]. On the other hand, while capable of investigating unravelling, all-atom origami simulations cannot currently probe timescales longer than $\sim$100 ns or structures larger than a few thousand base pairs[302]. To access large origami systems on timescales relevant to experiments while maintaining the ability to describe base pair breakage, we therefore use oxDNA[203;281]. Because of oxDNA's previous success in reproducing DNA nanostructures[261;243;251], capturing DNA mechanical response

to tension[226;185], and describing DNA twist[173;172], as mentioned in Chapter 1, it is well-suited to the present study.

By forcibly unravelling DNA origami, we explore departures from duplex behaviour and probe the basic physics of origami force compliance, as well as the dynamics of force propagation through the structures. We also probe the mechanical limits of DNA nanostructures under external stress, knowledge of which is crucial for applications that utilize nanostructures to apply or resist force[121;155;255]. Our results provide further validation of oxDNA's robustness in treating DNA origami. More importantly, they illustrate the possible fruits of combining experimental force spectroscopy with simulation to explore unfolding pathways and examine origami mechanical resistance to force in unprecedented detail, which can inform rational design of future nanomechanical structures.

4.2 Methods

Simulated SMFS was carried out on the three systems mentioned above, and the Rothemund tile was unravelled in complementary AFM experiments; details of these experiments are presented in Appendix A. oxDNA2 was used to simulate the Rothemund tile and the 7-tile biosensor, and the original oxDNA version was used for the 10 HB, as oxDNA2 was not released at the time these simulations were undertaken. We expect there to be very little difference between the results for the two oxDNA versions for the current systems, as both capture the thermodynamics of hybridization and the mechanical properties of double-stranded and single-stranded DNA equally well.

To mimic the AFM pulling procedure, in which one end of the enzymatically linearized scaffold is fixed to a glass coverslip and the other to the retractable AFM cantilever, simulations were performed by placing scaffold ends in 3D harmonic traps (see arrows in Figure 4.1(b)) of stiffness 5.7 pN/nm and moving one relative to the other at a constant rate. Our simulated trap stiffness is the same order of magnitude as stiffnesses typical of optical tweezer SMFS experiments, but is significantly lower than typical AFM stiffnesses. This lowers our simulated force loading rates relative

to the AFM experiments and allows us to approach experimental rates more closely. Two rates of trap movement are explored in this work: $5.6 \times 10^8\,\mathrm{nm\,s^{-1}}$ and $5.6 \times 10^7\,\mathrm{nm\,s^{-1}}$, corresponding to force loading rates of $3.2 \times 10^9\,\mathrm{pN/s}$ and $3.2\times10^8\,\mathrm{pN/s}$; experimental force loading rates were $1.4\times10^4\,\mathrm{pN/s}$ and $1.8\times10^5\,\mathrm{pN/s}$. As indicated in Chapter 3, the 'force loading rate' is the average rate of force increase as given by the product of harmonic trap stiffness and speed. This value is thus an instrumental property, and the loading rate the molecule actually 'feels' differs due to the presence of linker strands as well as the particular configuration of the molecule; estimates of the experimental force loading rate from fits to force-extension curves yield values 1-3 orders of magnitude lower than instrumental loading rates. While our simulated loading rates are many times greater than the experimental loading rates, any conversions of coarse-grained timescales to absolute time units must be interpreted with caution. The values above give a "worst-case" disparity; since coarse-graining speeds up diffusive dynamics relative to the microscopic time scales that set the coarse-grained time unit[52], the resulting telescoping of timescales[205] renders the effective force loading rates for simulations closer to experiment than they would appear from the above values.

We ran long molecular dynamics (MD) simulations on GPUs, which offered a speed-up of 25× over CPUs. The Andersen-like thermostat[230] described in Chapter 2 served to dissipate the energy imparted to the system by external force application. Temperatures relevant to experiments were used: 20 °C for the tile, 36 °C for the helix bundle[13], and 30 °C for the 7-tile assembly; and we used a high salt concentration typical of origami studies, $[\mathrm{Na^+}] = 0.5\,\mathrm{M}$.

The Rothemund tile required roughly 6×10^8 MD steps (∼9 µs) to unfold at $3.2 \times 10^9\,\mathrm{pN/s}$, or 9 days on GPU, and 6×10^9 MD steps (∼90 µs) to unfold at $3.2 \times 10^8\,\mathrm{pN/s}$, or 74 days on GPU. Unfolding the 10 HB required roughly 1.4×10^8 MD steps (∼2 µs), or 6 days on GPU, at $3.2 \times 10^9\,\mathrm{pN/s}$ and 1.3×10^9 MD steps (∼20 µs), or 60 days, at $3.2\times10^8\,\mathrm{pN/s}$. Breaking all locks in the 7-tile system required $\sim 3.3\times10^8$ MD steps (∼5 µs) at $3.2\times10^8\,\mathrm{pN/s}$, or 31 days on GPU; the simulations are significantly slower than for the tile due to the ∼ 2000 additional nucleotides.

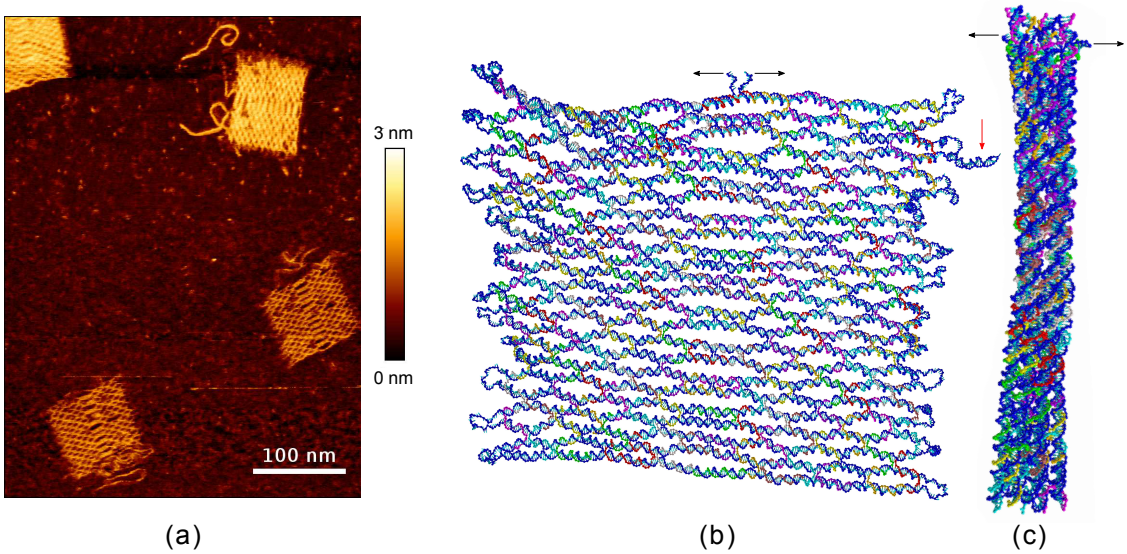


Figure 4.1: The two main systems studied in this work. (a) An AFM image of Rothemund tiles deposited on a surface. Credit: Martin Sajfutdinow. (b) An oxDNA representation of the Rothemund tile. The 20bp-stem hairpin indicated by the red arrow was excised in Rothemund's original origami design[228]. (c) An oxDNA representation of the 10-helix bundle presented in Bae et al.[13]. For both simulated systems, forces were applied to the nucleotides at the ends of the scaffold, as indicated by the black arrows.

We also incorporated sequence dependence for the Rothemund tile simulations at 3.2×10^9 pN/s[281;261]. This did not significantly alter the unfolding behaviour for the tile. Simulations of the 10 HB and 7-tile structures did not incorporate sequence dependence.

4.3 Rothemund Tile

Our system, shown in Figures 4.1(a) and (b), is very similar to Rothemund's original origami design[228], although we retain a 20-bp stem hairpin that Rothemund excised. The 6909-nucleotide scaffold is bound by 30-nucleotide (nt) and 32-nt staples, most of which have three domains, with one long (binding) and two short (linker) domains. The design has 24 rows with a central seam dividing the origami into two equivalent halves, and a very regular pattern of staples (Fig. A.1).

Figure 4.2 contains the experimental (a) and simulated ((b) and (c)) force-extension curves (FECs) for the Rothemund tile; FECs were generated according to the procedure described in section 3.3.2. The experimental curves terminate

before the full structure is unraveled – which simulations indicate would occur around 4000 nm – because the pulling handle between the molecule and the AFM tip breaks. For domains under mechanical load, there is a non-zero probability that statistically stronger ones break before weaker ones. This probability scales with the magnitude of overlap between the rupture force distributions of both domains[128;241]. In the single molecule force spectroscopy (SMFS) experiments performed here, the pulling handle is loaded throughout the whole retraction and force rarely drops to near zero between rupture events. The state of the pulling handle therefore does not get reset completely, and the probability of rupture increases with extension. It is thus unsurprising to see SMFS curves that do not unfold the whole origami structure in a single pulling cycle.

Nonetheless, several salient features are visible in the data. Immediately striking is the regular, sawtooth pattern present in all of the traces, a signature of the regularity of design in the tile. The tile unravels in units evidenced by these sawtooth rips, indicative of cooperative unfolding within each subunit. Notable, though, is that each 'rip' occurs at more or less the same rupture force: the unfolding of one unit does not render subsequent units easier to unfold.

Also evident in both experimental and simulated curves is the fundamental stochasticity of the process: even under an identical external protocol, the tile unfolds along a slightly different pathway every time. The pink and green experimental curves in (a), both collected at the same force loading rate, and the simulated curves in (c), also collected at the same rate, differ nonetheless from one another since barrier crossing is a thermally-activated diffusive process[144].

Both the experimental and simulated FECs exhibit a decrease in the slope of the force rise preceding each sawtooth 'rip' as more of the tile unravels, indicating a decrease in the apparent force loading rate[217;190]. This is a direct consequence of the effective increase in the linker length as the origami subunits closest to the handles unfold, releasing ssDNA, which behaves as an entropic spring that couples to the harmonic traps to govern the overall force response of the system. The elastic behaviour of ssDNA – typically modelled as a modified freely-jointed chain[259] or

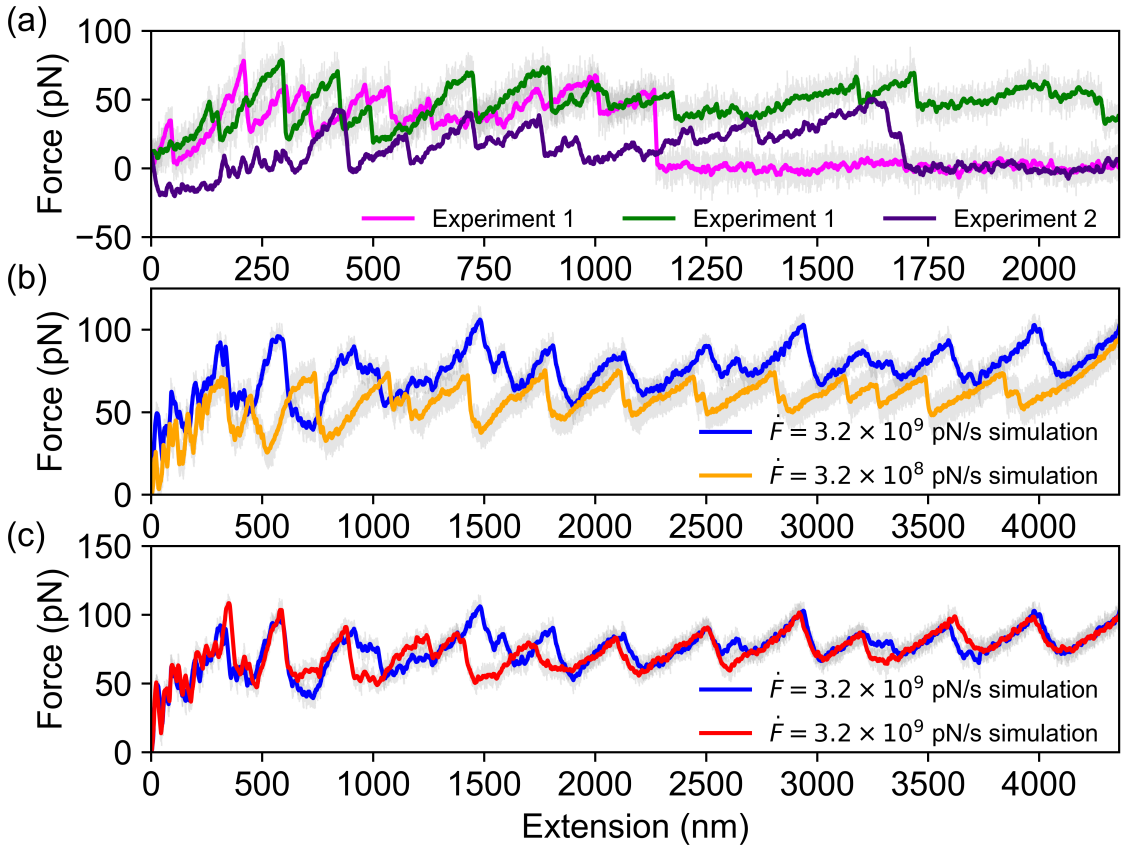


Figure 4.2: Force-extension curves for the Rothemund tile. (a) Experimental AFM curves at force loading rates of $\dot{F} = 1.8 \times 10^5$ pN/s (green and pink) and $\dot{F} = 1.4 \times 10^4$ pN/s (purple). (b) Simulated pulling curves at $\dot{F} = 3 \times 10^9$ pN/s (blue) and $\dot{F} = 3 \times 10^8$ pN/s (orange). (c) Simulated pulling curves, both collected at $\dot{F} = 3 \times 10^9$ pN/s; the blue curve is identical to the one shown in (b), reproduced here to facilitate comparison between pulling simulations carried out at the same loading rate. In all cases, colored traces are averages of gray curves and represent effective sampling rates of $250\,\mathrm{s}^{-1}$ (experiment), $1.3 \times 10^7\,\mathrm{s}^{-1}$ (simulated, slower rate), and $6.6 \times 10^7\,\mathrm{s}^{-1}$ (simulated, faster rate).

a worm-like chain[2] – is highly nonlinear, and is known to reduce the loading rate transmitted to a molecule[217]. Thus, as more ssDNA unravels, the apparent force loading rate decreases[217]. This accounts for the sharper, more jagged features visible near the beginning of the traces being softened, as sharp features are smoothed through convolution with the soft force response of the increasingly long polymeric tethers[298]. The stochasticity in the force-extension traces is more pronounced the farther from equilibrium the protocol is; in the (non-physical, $\dot{F} = 0$) limit of quasi-static pulling, the traces would appear identical for every realisation. As the apparent force loading rate decreases, therefore, so too does the variation from pull

to pull. We can rationalise this by considering that the width of the characteristic rupture force distribution for a given structural element increases with force loading rate, so a smaller spread in observed rupture forces is expected as loading rate decreases[61;26]. This explains the increased regularity and reproducibility of features as the extension grows, apparent in Fig. 4.2 (c). We note that the final segment of the simulated force-extension curves corresponds to the force response of ssDNA.

The FECs also exhibit a dependence of rupture force on probe velocity, with the force required to unravel specific features increasing with pulling speed. This behaviour of rupture force under dynamic loads was first described by Evans and Ritchie[67], following Bell[18], who predicted that rupture force is an increasing function of pulling speed. Later, more refined theories have corroborated this[113;61;26], and experiments are frequently analyzed using this framework[268;82;223]. In this context, we can rationalize the maximum rupture force increasing from $\sim$50 pN for the lower experimental loading rate to $\sim$75 pN for the higher experimental loading rate, and from $\sim$65 pN for the slower simulated rate to $\sim$ 100 pN for the fastest simulated rate. We reiterate that comparing simulation and experimental timescales is a subtle task, as coarse graining speeds up intrinsic dynamics[52].

A key advantage of simulations lies in the ability to correlate force-extension data with structural changes. Figure 4.3 tracks the breakage of native base pairs as the tile unravels for two pulling speeds. For a native base pair to be classified as 'present' in generating these blue-yellow plots, the nucleotides must have a hydrogen bonding energy greater than 10% that of a fully formed hydrogen bond. At $t = 0$, the tile is fully natively bonded. Each yellow block corresponds to two half-rows: the blue 'stripes' punctuating the yellow regions are regions of single-stranded scaffold, which can be seen in Figure 4.1 (b) at the end of each row. The nucleotide labelled '0' is subjected to the moving harmonic trap, while the nucleotide labelled '6908' sits in a stationary trap. Clearly, the structure unravels from the scaffold ends inwards, corresponding to the location of applied force. While stochasticity is evident – compare the plots in Figs. 4.3(b) and 4.3(c), collected at the same pulling rate – remarkable regularity exists, suggesting a well-defined unfolding mechanism.

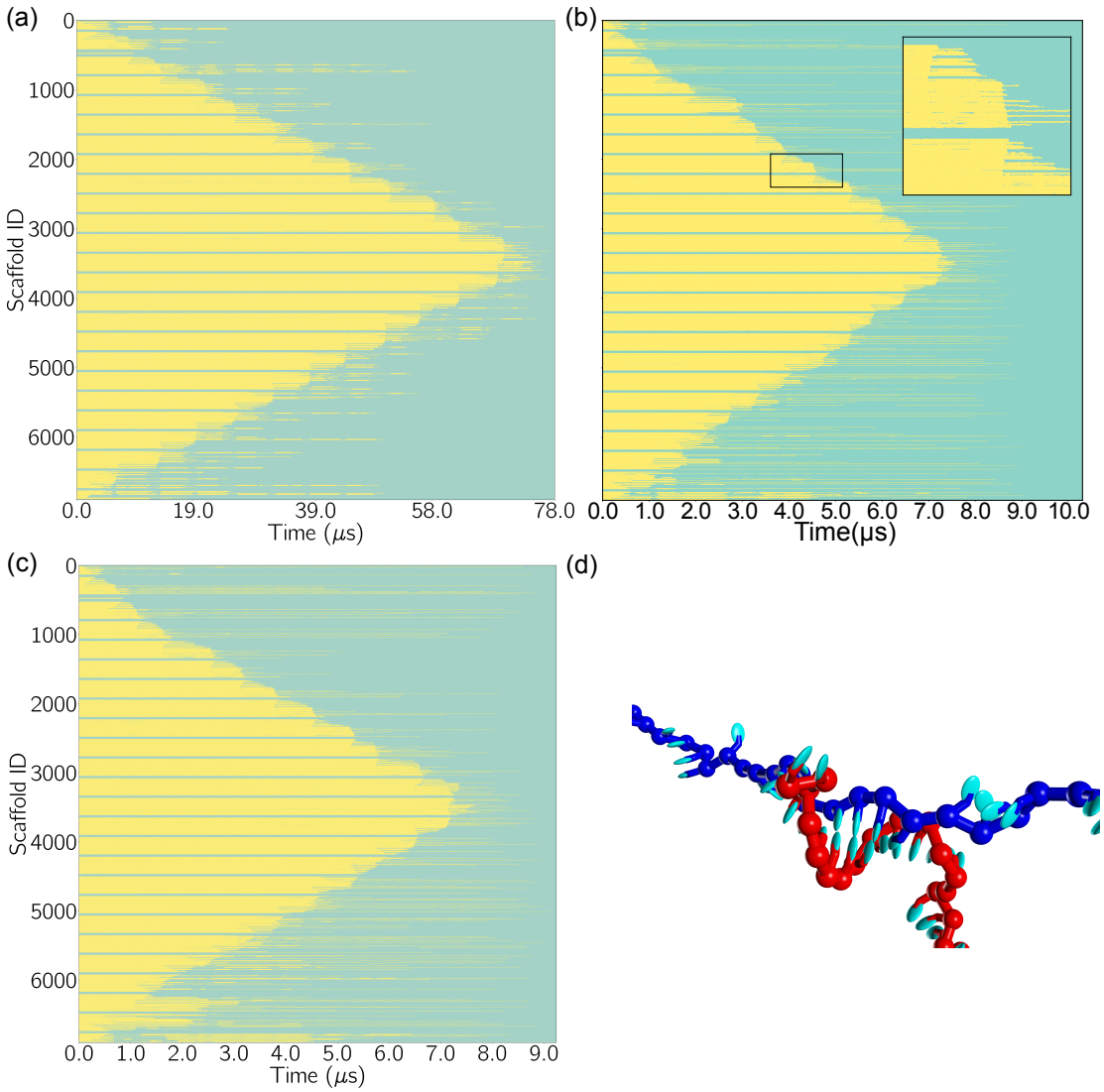


Figure 4.3: The pattern of native-bond unfolding in pulling simulations of the Rothemund tile. Yellow regions indicate bonds that are still in their native configuration; blue regions indicate an unbonded scaffold nucleotide. (a) Slower pulling rate; (b) faster pulling rate, corresponding to the blue trace in Fig. 4.2 (b) and (c); (c) second run at faster pulling rate, corresponding to the red trace in Fig. 4.2 (c). The inset in (b) shows detail of the unfolding mechanism for three half rows. Nucleotide 0 corresponds to the end associated with the moving trap. (d) An illustration of a 'persistent' staple domain (red) with three remaining base pairs connecting it to the scaffold (blue); these last base pairs do not prevent the scaffold backbone aligning with the force, which makes them more resistant to force-induced melting. By 'persistent domain', we mean a domain that remains bound after the part of the scaffold it is on has been pulled away from the main structure; such domains are visible as the thin yellow lines in (a), (b), and (c).

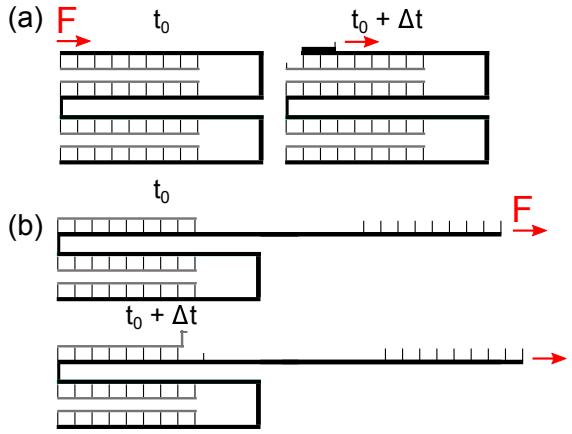


Figure 4.4: Antiparallel (a) and parallel (b) force application modalities. When force is applied antiparallel to the helix, the geometry is 'unzipping-like' and carries an extension gain per broken base pair of 9.8 Å. Applying force parallel to the helix, or in a 'shear-like' geometry, by contrast only yields an extension gain per broken base pair of ∼3 Å. This difference renders the forces observed for rows that are 'sheared' to be much higher than those for rows that are 'unzipped'.

From the plots of base pair breakage versus time, we can glean the basic mechanism by which the Rothemund tile unfolds. First, note that the routing pattern for the tile's rows means that the force is alternately antiparallel and parallel to the duplex; see Figure 4.4. The gain in free energy per broken base pair is $\sim F\Delta z/\text{bp}$, where $\Delta z/\text{bp}$ is the extension gain per base pair broken. For rows where the scaffold strand is anti-parallel to the force (Fig. 4.4(a)), each broken base pair contributes the base-base distance along the backbone plus the rise of a base pair in the duplex – in oxDNA, 9.8 Å. By contrast, when the scaffold strand is parallel to the force (Fig. 4.4(b)) $\Delta z/\text{bp}$ is a more modest ∼3 Å, the difference between the backbone length and the base-pair rise of the duplex. Consequently, the force required to break a base pair in the parallel geometry is much greater than the force required to break a base pair in the antiparallel geometry. This is analogous to the fact that the DNA duplex rupture force for a shear geometry is much higher than that for an unzipping geometry[222;267;185;226].

Consider the upper half of Figures 4.3 (a)-(c). The top half of each yellow block corresponds to a half-row with 'shear-like,' or parallel, pulling geometry, and the bottom half corresponds to a half-row with 'unzipping-like,' or antiparallel, pulling geometry. The inset of Fig. 4.3 (b) shows the unfolding of three half-rows:

in parallel pulling geometry (upper half of top yellow block), antiparallel pulling geometry (lower half of top yellow block), and again in parallel pulling geometry (upper half of bottom yellow block).

The force increases roughly linearly in the parallel pulling geometry until a maximum force is reached that allows a full half-row to be sheared off. Once the final seam staple in this half-row is removed, the subsequent half-row, now in antiparallel pulling geometry, can begin to unfold immediately because the force already exceeds the critical unzipping force. Most of the base pairs in the lower half of each yellow block – the ones that are 'unzipped' – break over a short period of time, as evident in the inset of Fig. 4.3 (b). The consequent rapid release of scaffold strand leads to the sudden drop in the force seen in the sawtooth features of Fig. 4.2. Also, for those staples whose long domain is in the half-row that unzips, yielding generally occurs by the unbinding of the shorter domain from the half-row below it, leaving the staple attached to the released scaffold. This is visible in Fig. 4.3 as a series of blue lines in the top halves of the yellow blocks (the half-rows in 'shearing' geometry), as well as the thin yellow lines associated with 'persistent domains' in the bottom halves of the yellow blocks (corresponding to 'unzipping' geometry). By 'persistent domains', we mean regions where staples remain partially bound by a single domain to the scaffold strand even after the part of the scaffold they are on has been pulled away from the main structure.

This mechanism, wherein two rows unravel in parallel, leads to the regular features in the FECs. This can be quantified by performing a Lomb-Scargle[238] periodicity analysis of one of the FECs in Figure 4.2. This analysis, shown in Fig. 4.5, reveals a strong peak that occurs very close to the length of two rows of the tile.

An additional feature evident from Fig. 4.3 is that there is a tendency for the corresponding left-hand and right-hand pairs of half-rows to yield almost simultaneously, giving these plots their rough mirror symmetry. This tendency is more evident both at the slower pulling rate, and at later times. One of the effects of this synchronization is to reduce the variability in the FECs at later times.

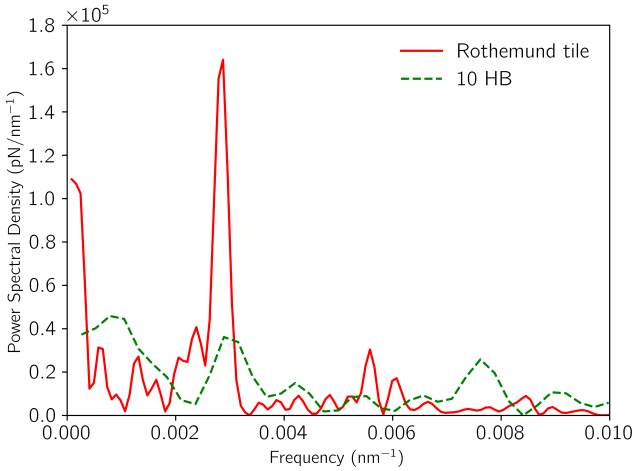


Figure 4.5: Lomb-Scargle periodogram indicating periodicity in the force-extension curves of the Rothemund tile (red) and the helix bundle (green), both pulled at $\dot{F} = 3.2 \times 10^8\,\mathrm{pN/s}$. The red peak near 0 is an artefact of the truncated curve – it would correspond to a very long timescale oscillation where the force returned to 0 at ∼8000nm. The highest peak, at $0.0029\,\mathrm{nm}^{-1}$, corresponds to a length of 345 nm, very close to the length of two rows of the tile: 2×284 nt $\times\ 0.64\,\mathrm{nm} = 363\,\mathrm{nm}$, confirming the picture of the tile unfolding two rows at a time. The helix bundle's force-extension curve lacks any clear periodicity.

Some asymmetry in the application of force is visible in the plots: persistent domains are more numerous in the lower half of the plots, corresponding to the half of the structure furthest from the moving trap. For example, in the faster simulations, at 5.0 µs in Fig. 4.3 (b) (after 15 rows have unfolded), there are 39 persistent domains in this lower half, while the upper half contains only 20 such domains. Similarly, at 8.0 µs, after all rows have unfolded, 44 persistent domains remain in the lower half and only 22 in the upper region. This asymmetry is an indication that the timescale over which force is increasing is faster than the time required for the force to propagate through the structure; that is, one half of the tile 'feels' a somewhat larger instantaneous force than the other half, making the force-induced melting of these persistent domains more likely. This effect is thus more pronounced for the faster pulling rate. Indeed, in the slower simulations, there is much less asymmetry in the persistent domains; e.g. just after ∼32.0 µs (after 11 rows have unfolded), there are 11 persistent domains in the lower half and 13 in the upper half. Such delayed propagation of stress along polymer chains has been noted elsewhere[77].

A further difference between the two rates is visible in the recurrent breakage and re-formation of base pairs in persistent staple domains, which occurs much more frequently at the slowest pulling rate. When the force approaches a maximum, the persistent domains begin to melt and the number of base pairs in these domains diminishes; but once the "rip" has occurred and the force in the system decreases, at the slowest rates there is sufficient time for the domain to reanneal. Note that the last few base pairs between staple and scaffold are significantly more resistant to force-induced melting, because the scaffold back-bone can then approximately align itself with the direction of the force and so the extensional gain from strand melting is much reduced[185]; see Figure 4.3(d).

The simulations also allow us to directly visualize the changes in configuration underlying these events. Figure 4.6 depicts the mechanisms underlying a typical "sawtooth" feature in the FEC. At point I, the top row is in a 'shear-like' pulling geometry. As force increases, the base pairs become increasingly strained, particularly on the left-hand side of the tile, corresponding to the moving harmonic trap. Structure II reveals the tile immediately after the final staple of the topmost left-hand half-row – the central seam – has been broken, but with the seam staple between the lower two half-rows still just in place. The half-row underneath then begins to unzip, and the left-hand half-row has completely unzipped by stage III. The consequent removal of the second bridging seam staple has a significant effect on the unravelling of the right-hand side. It allows the section nearest the seam to rotate about the nearest junction, dramatically changing the geometry of the pulling. At III, unzipping of the second half-row on the right-hand side has begun, even though the staples nearest the seam are still intact in a subunit that is now roughly vertical. The small peak in the FEC at III corresponds to this yielding of the right-hand side. The key role of the seam staples underlies the synchronized unravelling of the pairs of half-rows. Two rows, corresponding to the basic sawtooth unit in the force-extension plots, have been unravelled between I and IV – further confirmation of our picture of the basic mechanism of unfolding for the Rothemund tile.

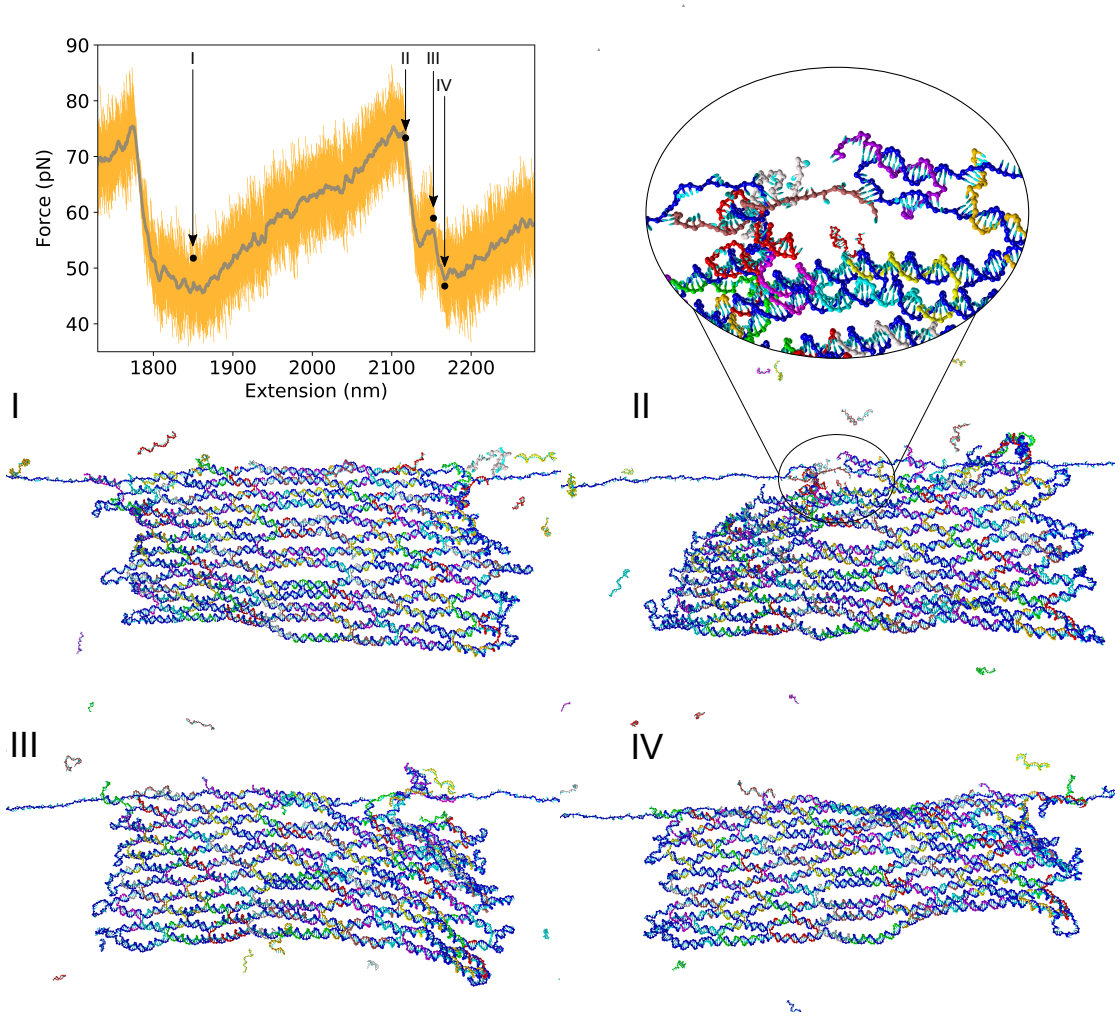


Figure 4.6: Structural illustration of the unfolding mechanism of the tile. A section of a simulated force-extension curve at $\dot{F} = 3 \times 10^8$ pN/s is shown; the gray curve is an average. I, II, III and IV are configurations at different times, as marked on the force-extension curve. I denotes a configuration at the beginning of a shearing mode. The sawtooth feature that peaks at point II corresponds to the top row being sheared open (see also blowup); subsequently, the next row unzips. The small bump at III occurs because the right half of the row unzips at a slightly later time than the left. Like I, IV is a configuration at the beginning of a shearing mode, but after a row has been pulled off.

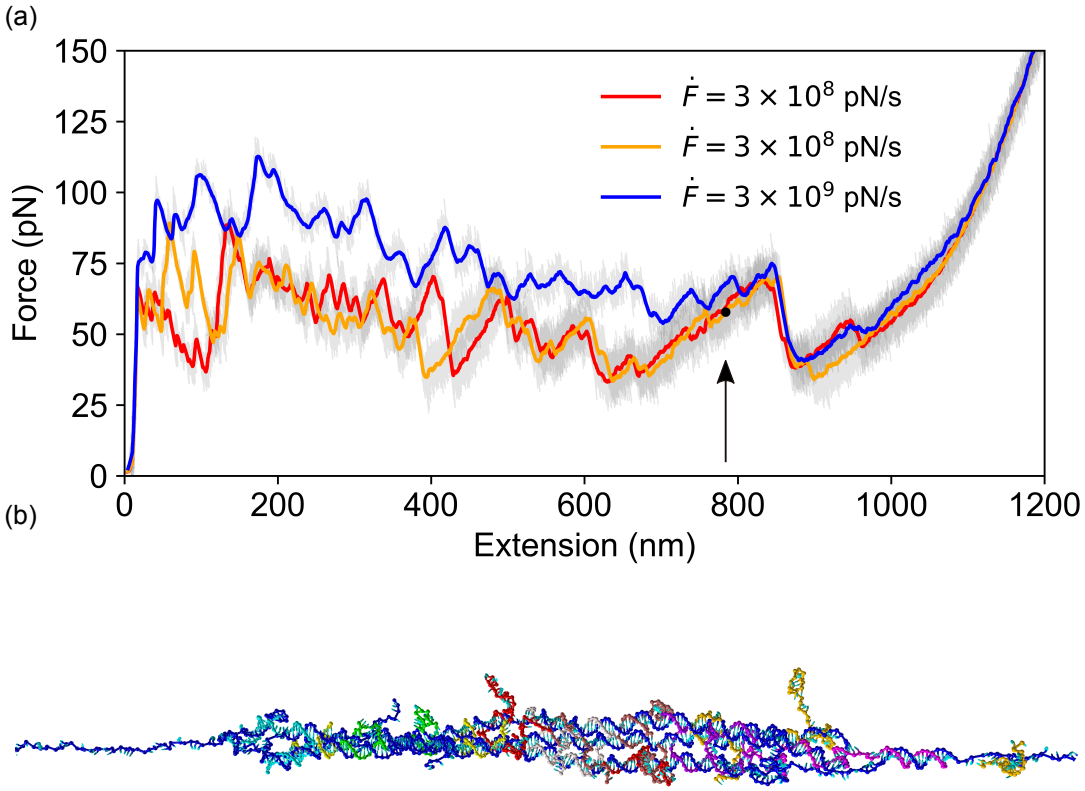


Figure 4.7: (a) Force-extension curves corresponding to unfolding the 10-helix bundle with moving harmonic traps at loading rates of $\dot{F} = 3.2 \times 10^9$ pN/s (blue) and $\dot{F} = 3.2 \times 10^8$ pN/s (red and orange). (b) Structural snapshot, corresponding to the black marker and arrow in (a), of the final rupture feature in the force-extension curves: three helices are in a shear geometry.

4.4 10-Helix Bundle

The 10-helix bundle was originally designed by Bae et al.[13] to explore the concept of mechanical origami assembly as an alternative to annealing—the scaffold was held at constant tension to eliminate transient secondary structure as folding progressed—but the force-extension curve associated with disassembly was also probed. It is both topologically and structurally more complex than the Rothemund tile. Firstly, the staple routings on the 1768-nucleotide scaffold are more non-local, with some staples spanning up to 5 domains (see Fig. A.2). Secondly, the structure is 3D rather than 2D. Despite these differences, the simulated FECs share many common features with those of the Rothemund tile: the basic sawtooth-like, 'stick-slip' dynamics, shown in Figure 4.7; an overall increase in rupture forces as pulling

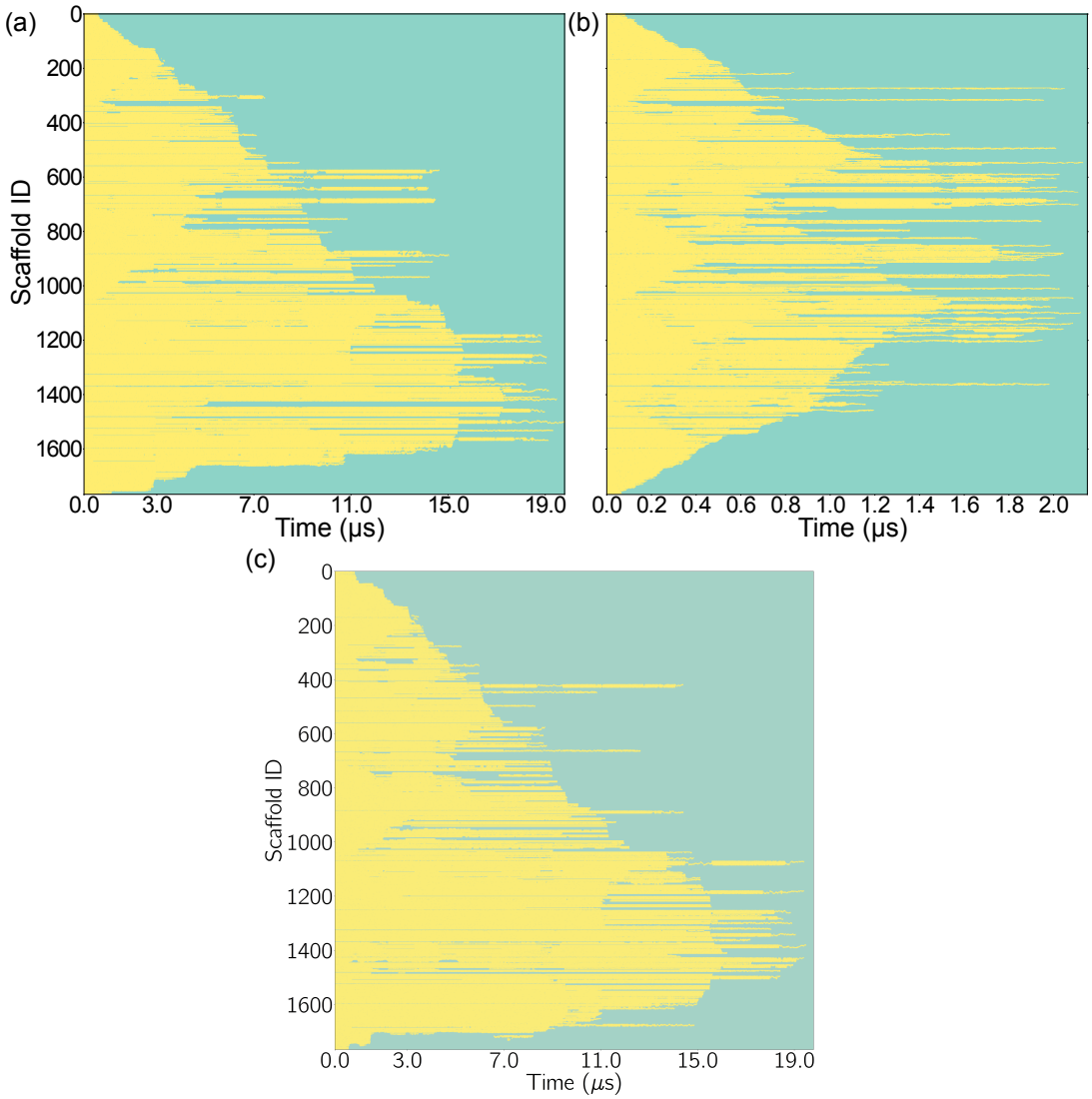


Figure 4.8: The pattern of native-bond unfolding in pulling simulations of the 10-helix bundle at (a) $\dot{F} = 3 \times 10^8$ pN/s (the dark red curve in Fig. 4.7(a)); (b) $\dot{F} = 3 \times 10^9$ pN/s (the blue curve in Fig. 4.7(a)); and (c) a second run at $\dot{F} = 3 \times 10^8$ pN/s (the orange curve in Fig. 4.7(a)). Yellow regions indicate bonds that are still in their native configuration; blue regions indicate an unbonded scaffold nucleotide. In contrast to the Rothemund tile, the plot for the faster pulling rate is more symmetric because the faster rate makes it less likely for the structure to find the more favourable asymmetric unfolding pathway.

rate increases; and increased reproducibility with increasing extension in the traces performed under the same conditions.

One consequence of the more complex structure, however, is that there is a less obvious preferred folding pathway; a multiplicity of possible unfolding events may yield the same change in extension. Figure 4.8 shows that the helix bundle

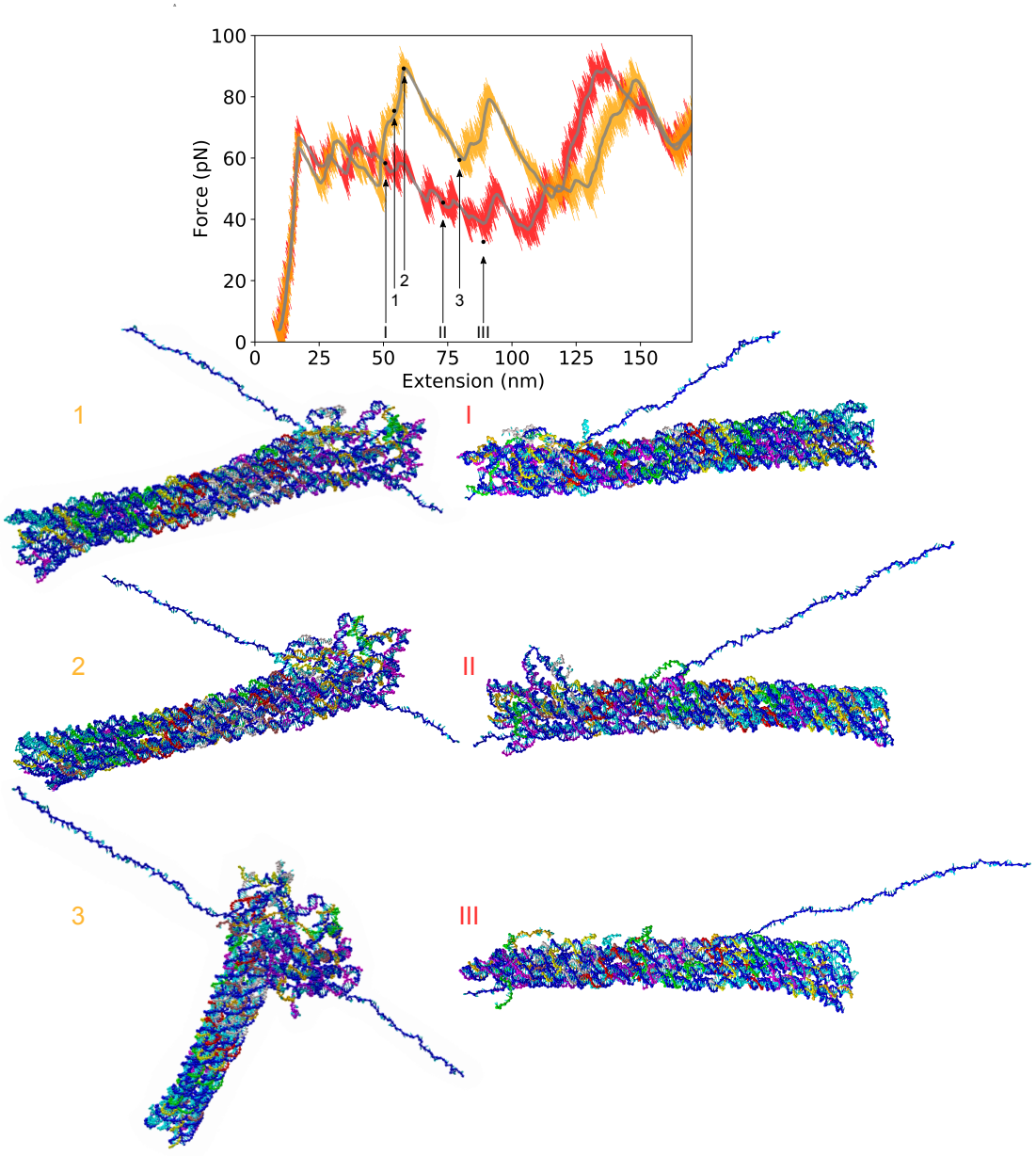


Figure 4.9: Structures exemplary of the stochasticity in the force-extension unfolding curves for the 10-helix bundle. Curves for each simulation, performed at the force loading rate $\dot{F} = 3.2 \times 10^8$ pN/s, are shown at top, and the locations of the structural snapshots are indicated. The unfolding pathway of the orange force-extension curve, shown on the left, exhibits unravelling from both ends before the first helix has completely yielded. The unfolding pathway of the red force-extension curve, however, features a complete unfolding of the first helix before others begin to yield.

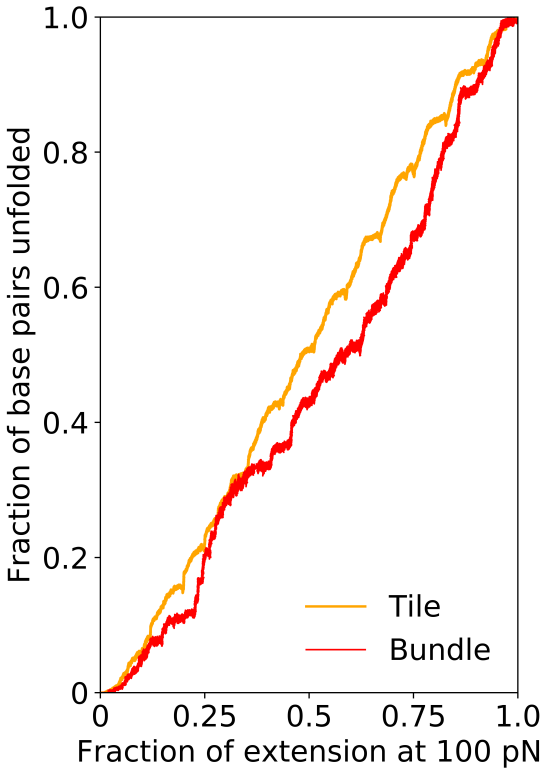


Figure 4.10: The fraction of base pairs unfolded as a function of the fractional extension (measured with respect to the extension at 100 pN) for the Rothemund tile and 10-helix bundle. The Rothemund tile exhibits a regular, linear pattern of Δz/bp as rows alternatingly shear and unzip. By contrast, the curve for the helix bundle is much more non-linear, and curvature in the plot indicates a strain-softening, or shielding, effect: once certain key basepairs unfold, subsequent basepairs are easier to break.

samples different unfolding pathways at different loading rates, and comparing Figures 4.8(a) and (c) establishes that multiple unfolding pathways are sampled even under identical simulation conditions. This is explored in structural detail in 4.9. For example, in one simulation, the bundle unfolds along a single helix before other helices unravel (Fig. 4.9, right panel). In another trace, collected under identical simulation conditions, other helices begin to yield from the opposite scaffold end before the first is finished unravelling (Fig. 4.9, left panel). Even along a single pathway, there is no clear uniformity in terms of a repeating unfolding unit; Figure 4.5 reveals an absence of periodic signatures in Fig. 4.7(a) associated with any particular length scales.

Also notable is the fact that in many regions, successive rupture events occur

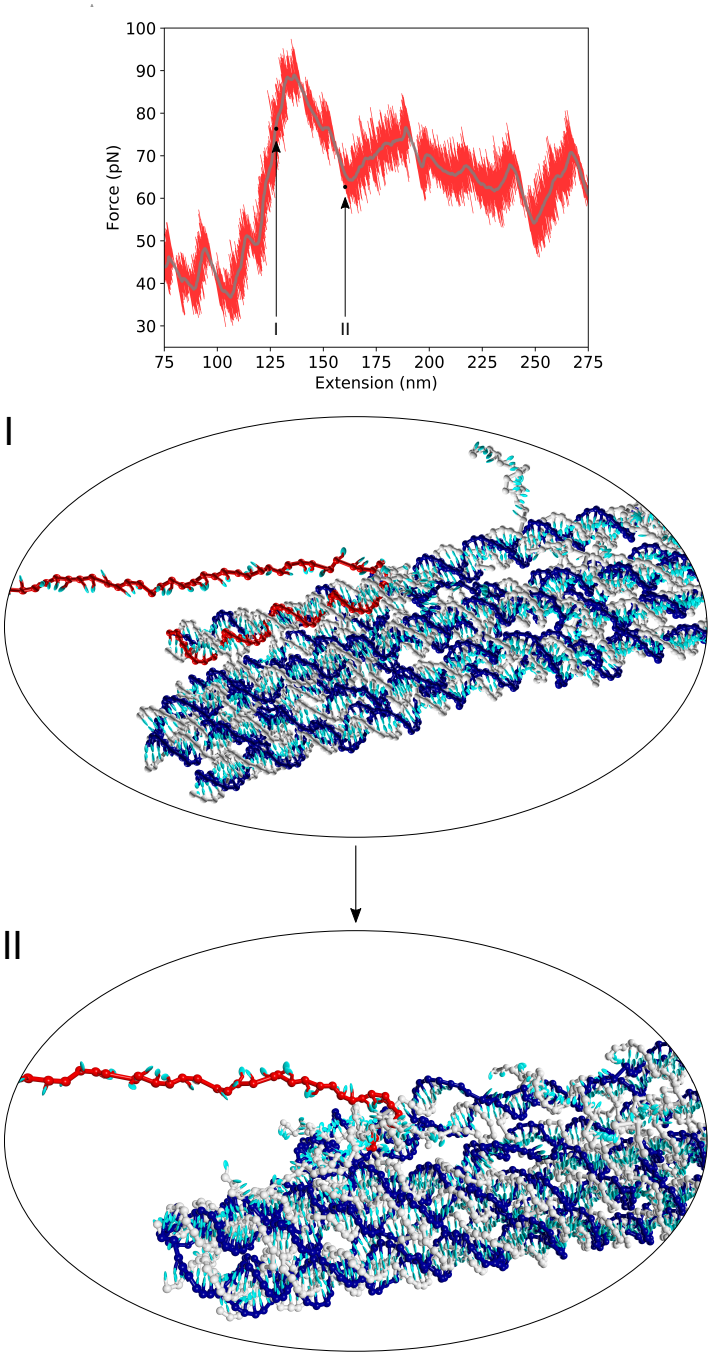


Figure 4.11: Structural illustration of the first large unfolding rip in a force-extension curve collected at $\dot{F} = 3 \times 10^8$ pN/s. The origin of the 'strain-softening effect', in which subsequent unfolding rips occur at lower forces, can be seen in the images: once the first helix (red) completely unwinds between I and II, many of the multiple-domain staples (white) attached to other helices (blue) have been compromised. Subsequent helices thus require lower forces to yield.

at progressively lower forces. For all traces, the highest forces reached occur near the start of the unravelling process. This 'strain-softening' effect arises because, unlike with the Rothemund tile, the gain in extension per base pair broken varies significantly as the helix bundle unravels. Figure 4.10 reveals the difference: whereas the behaviour of Δz/bp is essentially linear for the tile, the helix bundle Δz/bp contains regions of curvature, wherein breaking of key domains renders the structure increasingly compliant; it 'catastrophically' fails. This unfolding behaviour has been described previously as 'shielded' unfolding, in which structural topology prevents weaker barriers from being compromised before stronger ones.[214]. A specific structural example of this shielding is shown in Figure 4.11: once the first full helix has unravelled at $\sim 170\,\text{nm}$, the remainder of the bundle is comparatively easier to unfold. Given the multi-domain nature of the staples, unravelling the first helix weakens many of the staples holding subsequent helices in place; this gives rise to the strain-softening observed. The plots in Fig. 4.8 (particularly (b)) illustrate the weakening of the interior regions of the bundle. Unravelling does not simply occur from the ends inwards, as with the tile; as the final failure is approached, the structure appears to separate into multiple blocks. Strain-softening is visible up to $\sim 650\,\text{nm}$ extension. At this point, as illustrated in Figure 4.7 (b), three helices remain and force application occurs roughly in a shearing geometry, accounting for the $\sim 75\,\text{pN}$ force required to unfold this final feature – not far below the initial maximum force of $\sim 85\,\text{pN}$.

Employing a linearly increasing force protocol rather than harmonic traps makes the bundle's catastrophic failure even more evident: Figure 4.12 contains FECs for a range of force loading rates, where scaffold ends are both held at the same, linearly increasing, tension. Again, we note the increase in rupture force for increasing force loading rate. After a few initial rupture events, visible in the inset of Fig. 4.12, the bundle essentially unfolds 'all at once,' evidenced by the long linear regions in the force-extension data. Once the monotonically increasing force surpasses the critical force at which all landscape barriers are removed, unfolding enters a purely downhill regime. Thus, the force application protocol is clearly of central importance in

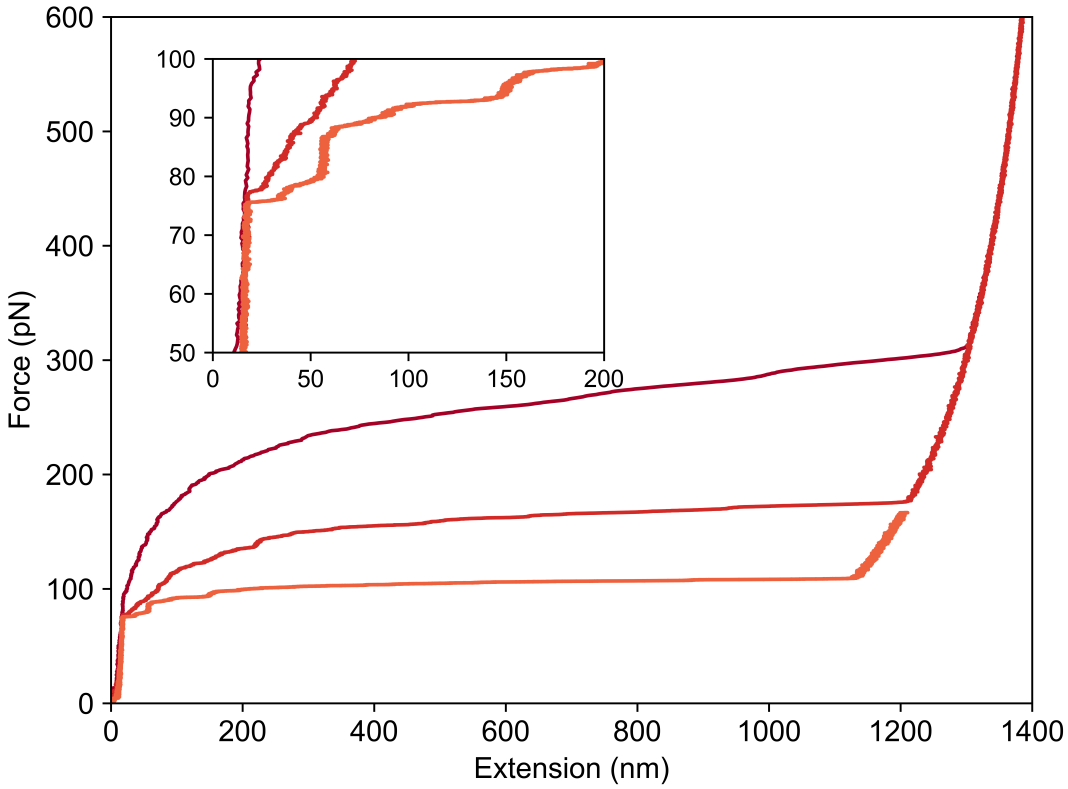


Figure 4.12: Simulated force-extension curves for the 10-helix bundle collected under a linearly increasing force protocol. Force loading rates decrease from top to bottom: $\dot{F} = 3.2 \times 10^8$ pN/s, 3.2×10^7 pN/s, 3.2×10^6 pN/s. The inset shows details of the initial rupture events preceding catastrophic unfolding.

the analysis of molecular mechanical properties; in the case of harmonic traps, the molecule has opportunities to partially relax, allowing us to distinguish more detailed features in the force-extension curves.

Bae et al.[13] also noted cooperative unfolding of the helix bundle, which in their magnetic tweezer pulling experiments began to yield at ~ 30 pN – a rupture force half as large as for our systems, which yield at ~ 60 pN under the slowest pulling rate. Their experimental force loading rate was not provided, but it was certainly smaller than our slowest rate, which probably accounts for most of the discrepancy.

It is noticeable that there is a pronounced asymmetry with respect to scaffold ends in Fig. 4.8(a) and (c), but to a much lesser extent for the faster pulling rate, Fig. 4.8(b). This pulling-rate dependence is the opposite of the behaviour observed for the tile (Fig. 4.3(a)-(c)), and is a result of an asymmetry in the helix bundle's

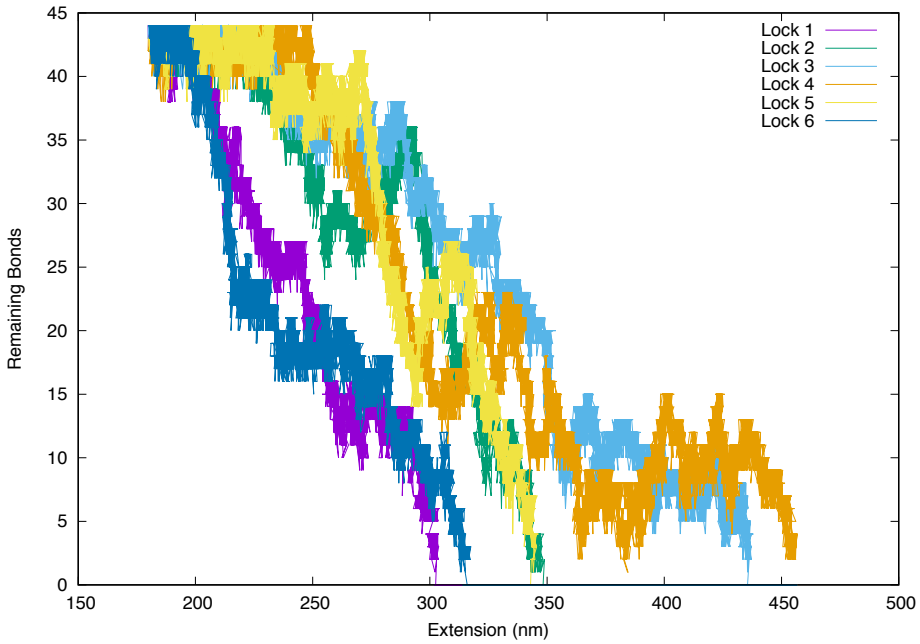


Figure 4.13: Bond breakage as a function of end-to-end extension for the six 44-base pair aptamer locks in the 7-tile assembly. Lock 1 is located nearest the moving harmonic trap (the uppermost in Fig. 4.15), lock 2 is the next furthest away, and so on to lock 6, which is closest to the stationary harmonic trap. The pulling rate of $\dot{F} = 3.2 \times 10^8$ pN/s is too rapid to differentiate the breaking of individual locks; sets of two locks yield approximately simultaneously, from the 'outside' (near the traps) to the 'inside'.

design. Specifically, there is an additional nick close to the end of helix 3 (see Fig. A.2) that will lead to a somewhat lower barrier for unravelling to be initiated from this end. If initiation occurs from just one end, the origami will then tend to rotate so that the points at which the force acts on the origami are co-linear, which in turn leads to a greater extensional release if that row continues to unravel. Furthermore, the resulting localized strain-softening will make it more likely that subsequent unraveling will continue from this end. When the pulling rate is higher, as in Figure 4.8 (b), the system has less time to find the most favorable pathway, and a more symmetric picture emerges.

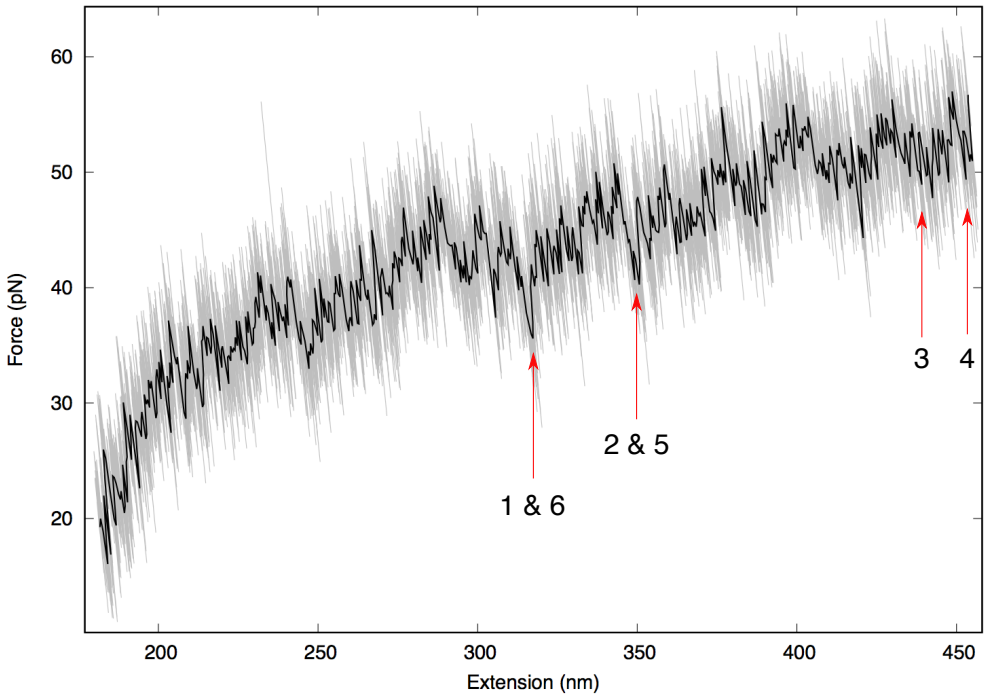


Figure 4.14: Force-extension curve for the 7-tile assembly for loading rate $\dot{F} = 3.2 \times 10^8$ pN/s. Light gray curve represents data at a sampling rate of $4 \times 10^8\,\mathrm{s}^{-1}$, and the black curve is an average, representing a sampling rate of $4 \times 10^6\,\mathrm{s}^{-1}$. Aptamer lock breakage events are indicated by the red arrows. Slower pulling rates would enable individual lock-openings to be seen more clearly. Note that as seen in Fig. 4.13 and Fig. 4.15, the last few base pairs in lock 4 do not fully unbind on the timescale of this simulation due to the tiles themselves beginning to unravel.

4.5 7-tile Biosensor

To illustrate the insights that can be obtained even for very large systems, for which simulated pulling rates are necessarily much greater than experimental rates, we also simulated the 7-tile biosensor of Koirala et al.[141]; cadnano structure is included in Appendix A. The 14,761-nucleotide origami consists of seven rectangular tiles joined by 44-base pair duplex 'locks', where the breaking of the locks can be observed as a large change in extension due to the connected tiles then being able to re-orient along the direction of applied force. In the experiments, the locks opened sequentially. Even at our slowest pulling rate of $\dot{F} = 3.2 \times 10^8$ pN/s, we observed

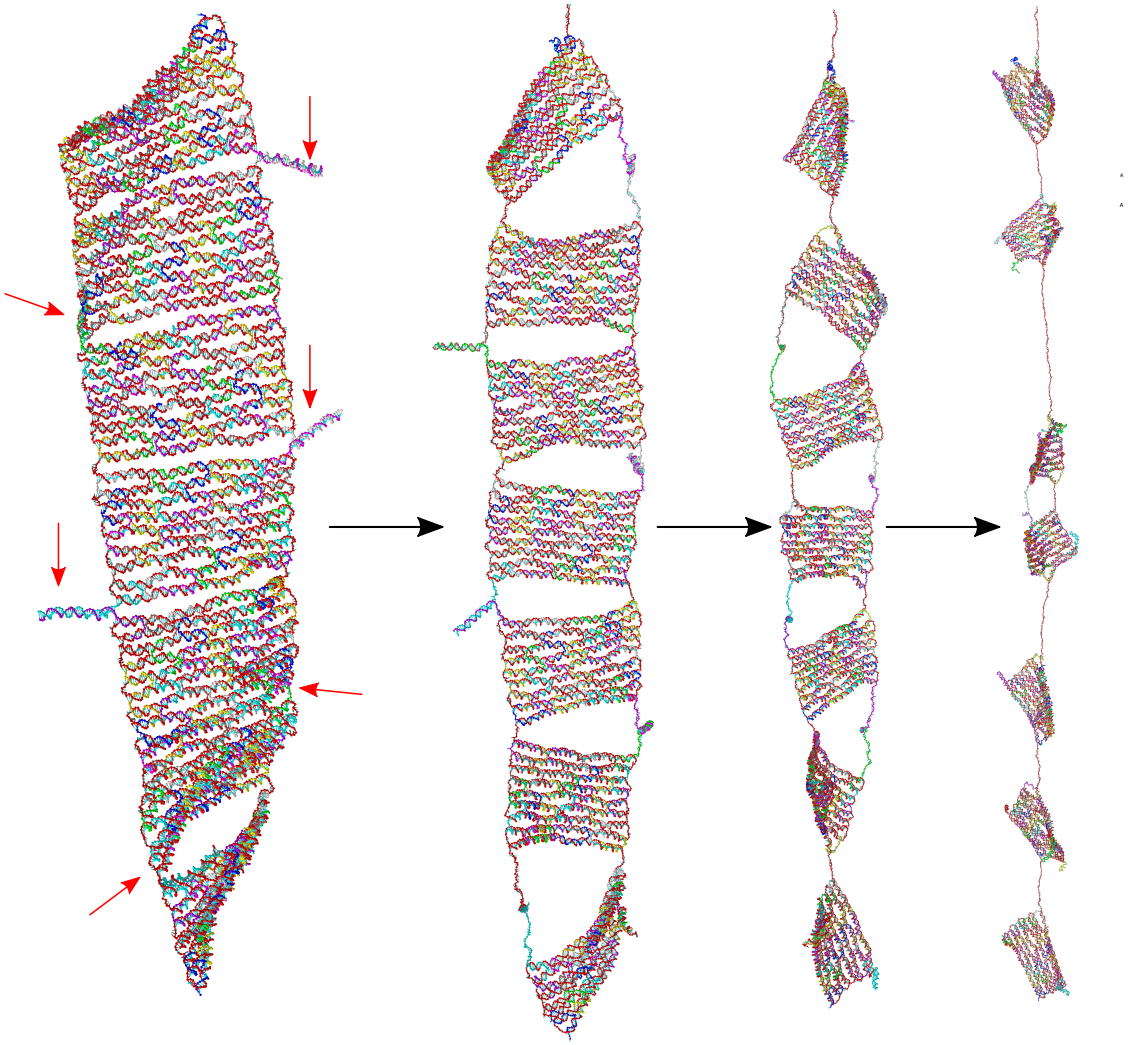


Figure 4.15: Structural snapshots illustrating the unfolding mechanism of the 7-tile biosensor of Koirala et al.[141] at a force-loading rate of $\dot{F} = 3.2 \times 10^8$ pN/s. The duplex 'locks' are labelled with red arrows in the first panel. The end-to-end length in the leftmost panel is 181 nm, and in the rightmost panel is 475 nm. As expected and experimentally observed, the locks open outermost to innermost.

the locks opening at the same time (Figs. 4.13 and 4.14). Nevertheless, the locks break in the order expected, outermost to innermost; Figure 4.15 illustrates the unlocking mechanism. Despite the fact that the resolution of individual events is not comparable to the experimental force-extension curves[141] given the relatively high force loading rate, the signature of the locks opening can still be seen in the force-extension data (Fig. 4.14). Furthermore, at this high rate, the tiles themselves begin to unravel somewhat before all of the locks have broken, a sign that our simulations are far from equilibrium. This example of a large system illustrates some

of the limits of of our simulation method; however, we are still able to capture salient insights, indicating that simulated SMFS can still be valuable for very large systems.

4.6 Conclusions

We have used molecular simulations performed using the oxDNA coarse-grained model to characterize the mechanical force response of two archetypal DNA origami systems, the Rothemund tile and a 10-helix bundle, and to explore the force response of a 7-tile biosensor. We also performed AFM pulling experiments on the Rothemund tile, enabling us to correlate features in experimental data with structural changes revealed in simulation. Both experiments and simulations exhibited regular, sawtooth unfolding behaviour in the force-extension curves for the Rothemund tile, and captured yielding at similar forces: $\sim$65 pN–75 pN.

Our force-extension data yield insight into the complex free-energy landscapes for origami unfolding, underscoring the much richer force-induced melting behaviour of origami compared to dsDNA. We observe stochasticity in the unfolding process for both systems; multiple pathways, cooperative unfolding, and strain-softening in the unfolding of the 10-helix bundle; and geometry-dependent Δz/bp. Since force biases the pathways taken through the unfolding landscape, we expect the assembly landscapes for these systems to exhibit even greater complexity.

Variations in rupture forces with pulling rates imply that any designed structural resistance to force will be loading rate dependent[185]. The effect of design on force-response behaviour has been highlighted by the disparate results obtained for the Rothemund tile, which unfolds modularly, and the 10-helix bundle, whose non-local staple routings lead to strain-softening and catastrophic failure. Differences between results obtained using linear and harmonic force pulling also underscore the importance of carefully planning force application protocol in SMFS studies.

Our results also reveal a glimpse of the complexities of force propagation through origami structures, manifesting in asymmetric unravelling behaviour. Valuable future investigations could consider in more detail how force propagation dynamics are complicated by the presence of junctions and multi-domain staples.

We have demonstrated the possibilities for improving comprehension of origami mechanical response and informing rational design of nanostructures through joint simulation and experimental studies. As it is uniquely suited to capturing the unravelling of large DNA nanostructures, our simulation strategy complements existing approaches to modelling the mechanical properties of origami, and we hope future work will continue to profit from the insights offered by DNA origami force spectroscopy.

Seldom, very seldom, does complete truth belong to any human disclosure; seldom can it happen that something is not a little disguised, or a little mistaken.

— Jane Austen, from *Emma* (1815).

5 Measuring internal forces in single-stranded DNA

Contents

5.1 Introduction 101
5.1.1 Models of ssDNA Elasticity 103
5.2 Force propagation in the oxDNA model 110
5.2.1 Theory 113
5.2.2 Secondary structure 116
5.2.3 Method for calculating internal forces in ssDNA simulations 118
5.3 Application of internal force calculation to a single-molecule force clamp 121
5.3.1 Introduction and system description 121
5.3.2 Simulations 124
5.3.3 Results 130
5.4 Conclusion and outlook 136

5.1 Introduction

Given the crucial importance of DNA in both biological and nanotechnological contexts, as outlined in Chapter 1, its elastic properties are of fundamental interest and have been the subject of extensive experimental[259;84;29], computational[308;226] and theoretical[169;194] investigation.

In both double- and single-stranded forms, DNA polymers behave as *entropic springs*, resisting any decrease in configurational entropy. Circumscribing the end-to-end distance of DNA polymers, therefore, gives rise to entropic forces. This effect is most pronounced for ssDNA, which has a bending stiffness $\sim 50\times$ lower than that of dsDNA[29] and is typically modelled as a freely-jointed random chain. Extending ssDNA by even a few nm leads to forces in the pN range[259;29]. Consequently, a growing number of nanotechnological applications that seek to measure or apply forces in the biologically relevant regime of 0-20 pN[254] make use of ssDNA.

Exploring the role of force in biological contexts is essential to the understanding of diverse processes. Cells widely employ mechanotransduction – the conversion of mechanical signals into chemical responses – through mechanosensitive ion channels that open in response to specific tension in cellular membranes[119]; modulation of gene expression through the mechanical state of the nucleus[276]; and force-induced changes to the catalytic activity of enzymes[119], among other mechanisms. The entire cellular skeleton, in fact, is a tensegrity network: its fibres balance compression and tension to maintain a prestress in the structure, allowing it to perform a central role in mechanotransduction with implications in embryogenesis and disease[119;282]. Additionally, force and torque are important to the functioning of molecular machines, from the transport molecule kinesin to ATP synthase, the energy production factory of all cells[28].

Comprehension of these mechanisms requires precise knowledge of the forces involved, which has motivated the construction of various molecular force-sensing devices[91;95;41]. Many of these capitalize on the entropic spring behaviour of ssDNA[254;177]: changes in the extension of ssDNA caused by binding events or conformational changes in a system of interest are read out as forces. Nanoscopic force-sensing devices have been applied to study, for example, proteins that connect cells to extracellular filaments[91] and DNA repair proteins[95]. Alternatively, the extension of ssDNA can be controlled to exert known forces on a molecule in force spectrometers, which have been employed to investigate nucleosome association[78] and the binding of transcription factors[193].

Beyond applications to the study of biologically-relevant systems, the entropic spring behaviour of ssDNA – and associated forces in the pN range – plays a fundamental role in manifold DNA nanotechnological devices. In so-called DNA wireframe architectures like polyhedra and buckyballs[86;103], flexible ssDNA strands act as vertices that link stiff double-stranded regions to achieve designed shapes. Some tensegrity structures – like DNA origami tensegrity prisms and kites – feature ssDNA as the mediator of constant stress in the design[155;255]. Liedl et al.[155] demonstrated the ability of DNA helix bundles in these structures to assemble against a force of ∼14 pN applied by ssDNA linkers. While tensegrity structures are as valuable as engineering elements on the nanoscale as they are on the macroscale[255], they have an added potential use as biomimetic nanostructures for the study of cellular mechanotransduction[155]. Finally, ssDNA springs have found use in nanocages designed for drug delivery[57] and nanotubes capable of the controlled release of gold nanoparticles[159].

In all of the aforementioned applications, knowledge of the precise forces involved is crucial for both the rational design and accurate interpretation of experiments. To this end, some have attempted to calibrate polymer springs using laser optical tweezer setups[254;91]. As discussed in Chapter 3, such experiments are prone to instrumental artefacts including convolution with long linker handles and the response of the dielectric beads suspended in the laser traps. Consequently, the ssDNA forces are only known to limited resolution, in one case ∼5 pN[254]. More accurate determination of forces requires an alternative strategy. One possibility is to use theoretical models of ssDNA elasticity to assign forces to particular extensions[155;193]; however, as we will see in the next section, this solution is sub-optimal because deficiencies in the accepted models lead to incorrect force estimates.

5.1.1 Models of ssDNA Elasticity

The simplest polymer description of ssDNA is the freely jointed chain (FJC) model. The FJC describes a polymer as a chain of connected rigid segments of equal length ℓ (typically called the Kuhn length) that are free to rotate about their

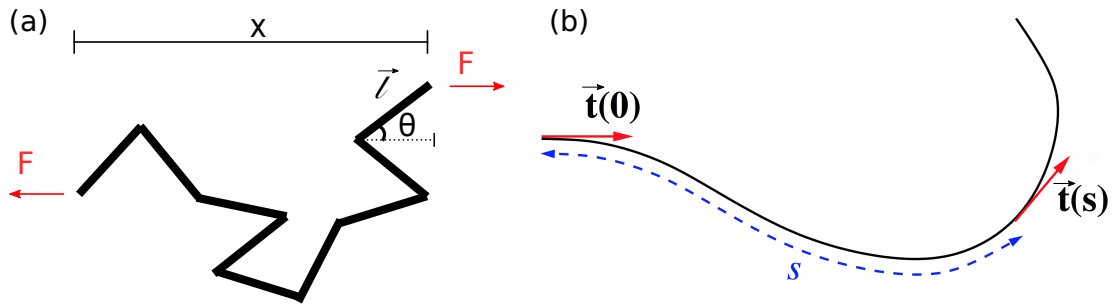


Figure 5.1: Schematics of the two polymer models most commonly used to describe DNA. (a) In the freely-jointed chain (FJC) model, DNA is modelled as a chain of rigid segments that are free to rotate about their connection points. In the presence of force, the orientations of individual segments $\vec{\ell}$ are no longer random, but are biased to point along the direction of force. Similarly, holding the chain at fixed extension x generates an entropic tension. (b) In the worm-like chain (WLC) model, DNA is treated as a continuously flexible rod characterized by a persistence length ℓ_p that quantifies the distance along the backbone over which correlations between tangent vectors $\vec{t}$ decay.

connection points; a single configuration of the chain is equivalent to a random walk. In the absence of force, the expected mean squared end-to-end extension of an N-segment FJC is the same as the mean squared displacement of an N-step random walk with step size ℓ: $\langle R_{ee}^2 \rangle = N\ell^2$. When polymer ends are subjected to a biasing force F, as in Figure 5.1(a), the orientations $\vec{\ell}$ of chain segments are no longer completely random. To find the expected end-to-end extension $\langle x \rangle$ for a given applied force, we first consider that any change in free energy $\mathrm{d}A$ will be equivalent to the work done against the force:

$$\mathrm{d}A = -\,\mathrm{d}(F\langle x \rangle) = -F\,\mathrm{d}\langle x \rangle - \langle x \rangle\,\mathrm{d}F \tag{5.1}$$

and therefore $\langle x \rangle = -\frac{\partial A}{\partial F}$. To find the Helmholtz free energy $A = -k_B T \ln Z$, with $k_B T$ the usual thermal energy, we must find the partition function, Z:

$$Z = \int \mathrm{d}\vec{\ell_1} \int \mathrm{d}\vec{\ell_2} \ldots \int \mathrm{d}\vec{\ell_N} e^{-E(\{\vec{\ell_i}\})/k_B T} \tag{5.2}$$

where $E(\{\vec{\ell_i}\})$ is the energy of a particular configuration of the FJC characterized by set of segment orientation vectors $\{\vec{\ell_i}\}$ and total end-to-end extension along the direction of force x:

$$E(\{\vec{\ell_i}\}) = -Fx = -\sum_i^N \vec{F} \cdot \vec{\ell_i} = -F\ell \sum_i^N \cos\theta_i \tag{5.3}$$

where θ_i is the angle the ith segment makes with the direction of applied force and ℓ is the magnitude of $\vec{\ell_i}$, as in Fig. 5.1(a). The energy is thus separable into N independent terms, and the overall partition function is derivable from the partition function for one segment, $Z_N = Z_1^N$, which is

$$\begin{aligned} Z_1 &= \int \mathrm{d}\vec{\ell_1} e^{-F\ell\cos\theta_1/k_BT} = \int_0^{2\pi} \mathrm{d}\phi_1 \int_0^{\pi} \mathrm{d}\theta_1 \sin\theta_1 e^{-F\ell\cos\theta_1/k_BT} \\ &= 4\pi \frac{k_BT}{F\ell} \sinh\left(\frac{F\ell}{k_BT}\right) \end{aligned} \tag{5.4}$$

where ϕ and θ have their usual definitions for a spherical coordinate system, whence we derive

$$\begin{aligned} \langle x \rangle &= -\frac{\partial A}{\partial F} = k_BT \frac{\partial \ln Z_1^N}{\partial F} \\ &= Nk_BT \frac{\partial}{\partial F} \ln\left(\frac{1}{F}\sinh\frac{F\ell}{k_BT}\right) \\ &= N\ell\left(\coth\left(\frac{F\ell}{k_BT}\right) - \frac{k_BT}{F\ell}\right). \end{aligned} \tag{5.5}$$

In the limit of low forces ($F \ll k_BT/\ell$), the force-extension behaviour of the FJC is Hookean:

$$F \approx \frac{3k_BT}{N\ell^2}\langle x \rangle \tag{5.6}$$

and in the high-force limit ($F \gg k_BT/\ell$),

$$F \approx \frac{k_BT}{\ell}\frac{1}{1-\langle x\rangle/L_c}, \tag{5.7}$$

which diverges as expected as the extension $\langle x \rangle$ approaches the total, or contour, length of the chain, L_c.

In describing ssDNA, an extensible freely jointed chain (exFJC) model is usually used, in which the ability of DNA to deform under stress is incorporated into Eq. 5.5 as an additional enthalpic term[258;259]:

$$\langle x \rangle = L_c\left(\coth\left(\frac{F\ell}{k_BT}\right) - \frac{k_BT}{F\ell}\right)\left(1+\frac{F}{S}\right), \tag{5.8}$$

where S is a stretching modulus and $L_c = N\ell$ is the contour length of the chain. In the seminal work of Smith et al.[259], force-extension curves resulting from single

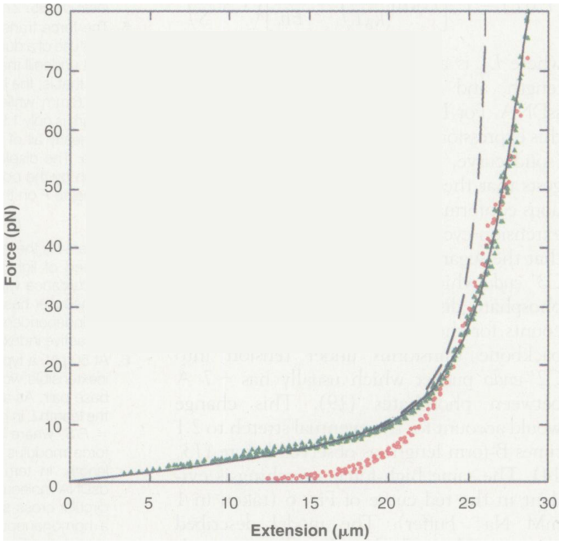


Figure 5.2: Force-extension data collected in optical tweezer experiments on 48 502-base λ phage ssDNA by Smith et al.[259]. The green curve was collected at [Na^+]=150 mM and the red curve was collected at [Na^+]=1 mM. The fit to the FJC model (Eq. 5.5) is shown as a dashed line; an extensibility correction is necessary to fit the high-force data properly (Eq. 5.8). The authors attribute the failure of the exFJC model to fit the low-force data for the red curve to electrostatic repulsion and excluded volume, more prevalent at low salt. Adapted from Smith et al.[259].

molecule force experiments on ssDNA were fit to the exFJC model; their results are reproduced in Figure 5.2. Force pulling experiments were performed on λ-DNA (48 502 bases) at 150 mM [Na^+] (green curve) and 1 mM [Na^+] (red curve). Equation 5.8 (solid line) fits the data reasonably well, and much better than does Eq. 5.5 (dashed line). The authors report fit parameters for Kuhn length ℓ=15 Å, for stretch modulus S=800 pN, and for contour length per base pair L_c/bp=0.56 nm*. While the ssDNA at low ionic strength (red curve) conforms to the exFJC at high forces, it diverges at low force; ssDNA is stiffer at low salt due to increased electrostatic repulsion, and is therefore often described in the literature by an alternate polymer elastic model: that of the worm-like chain (WLC).

First proposed for chains of discrete segments by Kratky and Porod[145], the WLC model is more suitable than the exFJC for stiff molecules because it treats polymers as continuously flexible rods; see Figure 5.1(b). The fundamental parameter in

*Subsequent literature that invokes the fit parameters of Smith et al.[259] often uses L_c/bp=0.63 nm from other sources[186;176].

the WLC model is the persistence length ℓ_p, which characterizes the distance over which correlations between tangent vectors decay. If we let $\vec{t}(0)$ and $\vec{t}(s)$ be tangent vectors separated by a distance s along the polymer backbone (as in Fig. 5.1(b)), ℓ_p is defined through

$$\langle \vec{t}(0) \cdot \vec{t}(s) \rangle = e^{-s/\ell_p}. \tag{5.9}$$

No closed-form analytical expression for the force-extension relationship of a WLC exists, but Marko & Siggia [169] derived an approximate expression:

$$F \approx \frac{k_B T}{\ell_p} \Big[\frac{1}{4(1 - \langle x \rangle / L_c)^2} - \frac{1}{4} + \frac{\langle x \rangle}{L_c} \Big], \tag{5.10}$$

where L_c is again the full extension, or contour length, of the polymer and $\langle x \rangle$ is the expected end-to-end extension along the direction of force. At low forces ($\langle x \rangle \ll L_c$), the WLC also behaves as a Hookean entropic spring:

$$F \approx \frac{3 k_B T \langle x \rangle}{2 \ell_p L_c} \tag{5.11}$$

Comparing with Eq. 5.6, we see that in this limit, the effective Kuhn length is $2\ell_p$.

While dsDNA is well-described by the WLC and variants thereof that incorporate twisting and stretching [169;168] and couplings between them [93;194], an ambiguity exists regarding which model provides the best description of ssDNA elastic behaviour, and authors have used both WLC and exFJC to fit ssDNA in different regimes, dependent on ionic conditions and force [110;21]. At low ionic concentrations, electrostatic repulsion fosters increased rigidity of ssDNA, so the WLC model may be preferable to the exFJC, which works better at higher salt concentrations [110]. Both models, however, neglect excluded volume and stacking interactions between ssDNA bases. As an illustration of how these realities can complicate the interpretation of ssDNA through the lens of polymer models, consider the results of oxDNA simulations† shown in Figure. 5.3. Equilibrium VMMC simulations at a number of different constant external tensions between 0 pN and 60 pN were performed on a 50 bp segment of poly-A ssDNA, which exhibits strong base pair stacking at the salt

†Most of these simulations were carried out by undergraduate research student Michele Fava, co-supervised by myself, but some were performed by me.

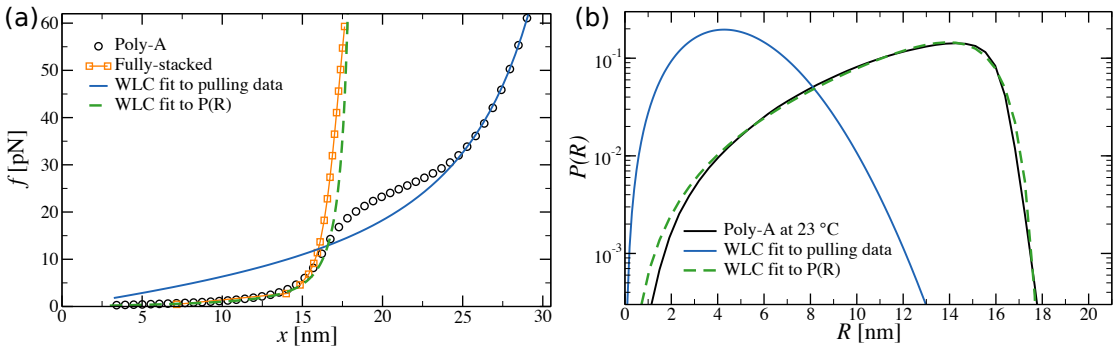


Figure 5.3: Results of oxDNA MC simulations on a 50 nt poly-A strand at $[Na^+]$=0.525 M and a variety of forces. The simulated FEC in (a) (circles) was fit to the WLC chain model (blue curve), and the resulting parameters were used to construct an expected end-to-end distance distribution, $P(R_{ee})$ (blue curve, (b)). In (b), the simulated $P(R_{ee})$ (black) was fit to the WLC model (green curve), and the resulting parameters were used to construct an expected FEC (green curve, (a)). The disagreement between fit parameters obtained under different conditions underscores the deficiencies of the WLC model, due in large part to stacking: a strand forced to retain base pair stacks is shown in orange in (a), and agrees fairly well with the parameters obtained from the zero-force data in (b). Simulations were the collaborative work of summer student Michele Fava and the author, and the figure was compiled by Dr. Lorenzo Rovigatti.

concentration used, $[Na^+] = 0.525$ M. The average extension at each force was used to construct the FEC in Fig. 5.3(a), and the $F > 20$ pN region was then fit to the WLC model (blue curve). Fig. 5.3(b) contains the probability distribution of the chain's end-to-end distance, $P(R_{ee})$, at 0 pN (black curve), which was fit to the predicted form of $P(R_{ee})$ for the WLC as approximated by Becker et al.[16] (green dashed curve). Finally, the parameters for contour length L_c and persistence length ℓ_p obtained from the FEC fit – $L_c \approx 38.8$ nm and $\ell_p \approx 0.35$ nm – were plugged into the $P(R_{ee})$ equation (Fig. 5.3(b), blue curve). Similarly, the parameters obtained from the $P(R_{ee})$ fit – $L_c \approx 18.8$ nm and $\ell_p \approx 5.7$ nm – were plugged into the WLC force-extension equation (Fig. 5.3(a), green dashed curve). What is immediately striking is the discord between the two methods of fitting ssDNA to the WLC. If zero-force data are used, the ssDNA appears much 'stiffer' and shorter than if high-force data are used. The plateau around 25 pN in the FEC corresponds to the disruption of stacking. To emphasize that the disparity between zero-force and high-force fit results is caused partly by the fact that base stacking breaks

at high forces, changing the nature of the polymer, the simulations of Fig. 5.3(a) were repeated without allowing base stacks to break (orange curve); in this case, good agreement with the zero-force WLC data (green dashed curve) is observed. Also contributing to the discrepancy are excluded volume effects, which are more pronounced at low force when the ssDNA is more compact. The disagreement between ssDNA elastic parameters obtained from experiments performed at zero-force and those obtained from FECs is widespread in the literature[259;149;21], but interestingly has not been widely acknowledged.

Clearly, polymer models which neglect excluded volume and stacking interactions are insufficient to describe ssDNA. Beyond these issues, a more fundamental problem exists: ssDNA is able to form so-called secondary structures – wherein a strand base pairs with itself to form, for example, hairpins. Like stacking interactions, secondary structures are more pronounced at low forces, as high tension causes them to break, and FEC fits will be plagued by this problem. The signature of ssDNA secondary structure has been observed in experimental FECs as a plateau at low forces[110;21], and higher ionic strengths will intensify this effect, as reduced electrostatic repulsion will facilitate hydrogen bonding. Consider again Fig. 5.2: the secondary structure plateau is much more pronounced in the curve collected at $[Na^+]$=150 mM (green) than in the curve collected at $[Na^+]$=1 mM (red). Since different ssDNA sequences will have characteristic secondary structures that alter elastic behaviour in unique ways, no general method of correcting for these issues exists.

In spite of these caveats, the original fit parameters obtained by Smith et al.[259] from the FECs in Fig. 5.2 are frequently invoked to describe the entropic spring behaviour of ssDNA[155;193;242], even for experimental conditions that differ from those under which the original data were collected, and have been used to parametrize coarse-grained models[137;221]. As discussed above and elucidated in the remainder of this chapter, for applications that require accurate estimation of forces, this practice is inadequate. As we will show, simulations can provide better estimates of the internal forces in ssDNA, not only by recreating secondary structures, salt, and temperature conditions relevant for a particular application, but by capturing

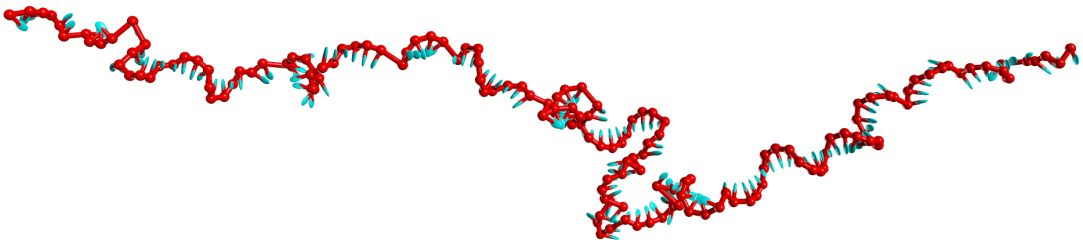


Figure 5.4: oxDNA representation of the system used in the text to illustrate the force propagation mechanism in oxDNA: a 160-nt single poly-T strand whose ends are held in harmonic traps of stiffness 285 pN/nm centred 41.9 nm apart.

more realistic ssDNA behaviours like stacking and excluded volume interactions. Motivated by the previous success of the oxDNA in describing DNA behaviour (as detailed in Chapter 2), we employ coarse-grained oxDNA simulations to calculate internal forces in ssDNA to demonstrate their suitability for this purpose. The next sections of this chapter introduce internal forces in the oxDNA model and outline the correct method for calculating tension in oxDNA structures. An application of these concepts to calibrate a recently proposed experimental force clamp[193] is then presented.

5.2 Force propagation in the oxDNA model

To illustrate the mechanism of force propagation through ssDNA in oxDNA, we consider a sample system, shown in Figure 5.4: a 160-nt single poly-T strand simulated at 21 °C and [Na^+]=10 M whose ends are held in harmonic traps of stiffness 285 pN/nm centred 41.9 nm apart[†]. Under these conditions, the exFJC model (Eq. 5.8) predicts an entropic tension in the chain of ∼4.4 pN, using the parameters reported by Smith et al.[259]: Kuhn length $\ell = 1.5$ nm, contour length L_c =160 nt × 0.56 nm, and elastic modulus $S = 800$ pN. Simulations were also performed for the strand with base stacking interactions suppressed.

The average force experienced by each nucleotide due to all other nucleotides in the chain is shown for each dimension in Figure 5.5, for both the poly-T strand and

[†]This choice of setup was inspired by the system of Nickels et al.[193], discussed in section 5.3, but differs slightly due to refinements made to more closely emulate experiments.

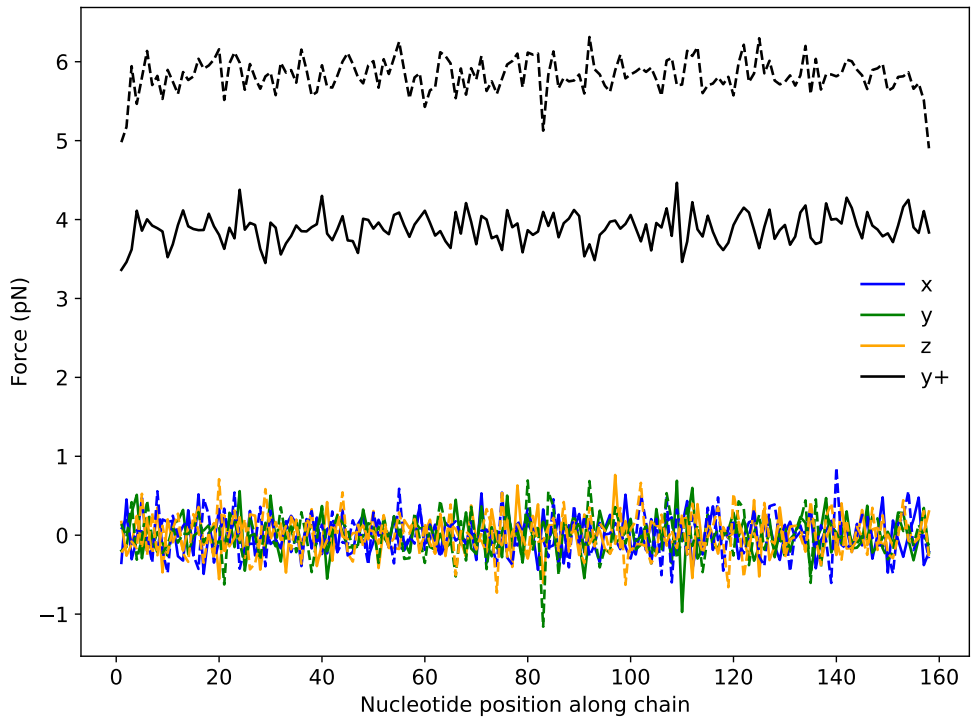


Figure 5.5: Forces experienced by nucleotides in the single strand. Coloured lines represent the total force each nucleotide experiences due to all other nucleotides in the chain, in all three directions. The black curves, labelled '$y+$', represent 'one-sided' force in the y-direction: the net force exerted on a nucleotide by all nucleotides topologically upstream of it. The dashed black curve is the result obtained when base stacking interactions in the strand are suppressed.

the strand for which stacking interactions have been suppressed; as expected, given that the chain is in equilibrium and therefore is experiencing no net acceleration, the average force for all nucleotides is ∼0 pN. We can also consider the 'one-sided' force, 'y+' in Fig. 5.5, which is the net force exerted on a nucleotide by all nucleotides topologically upstream of it. For example, the one-sided force experienced by nucleotide #50 is the sum of all forces exerted on it by nucleotides #51-160. The average one-sided force is also ∼0 pN along the axes perpendicular to the direction of constraint, but ∼3.84 pN for the stacked strand, and ∼5.79 pN for the unstacked strand, along the direction of trap separation, indicative of entropic tension. That the force is higher when stacking interactions are suppressed is physically reasonable: a stacked chain is stiffer than an unstacked one, and has

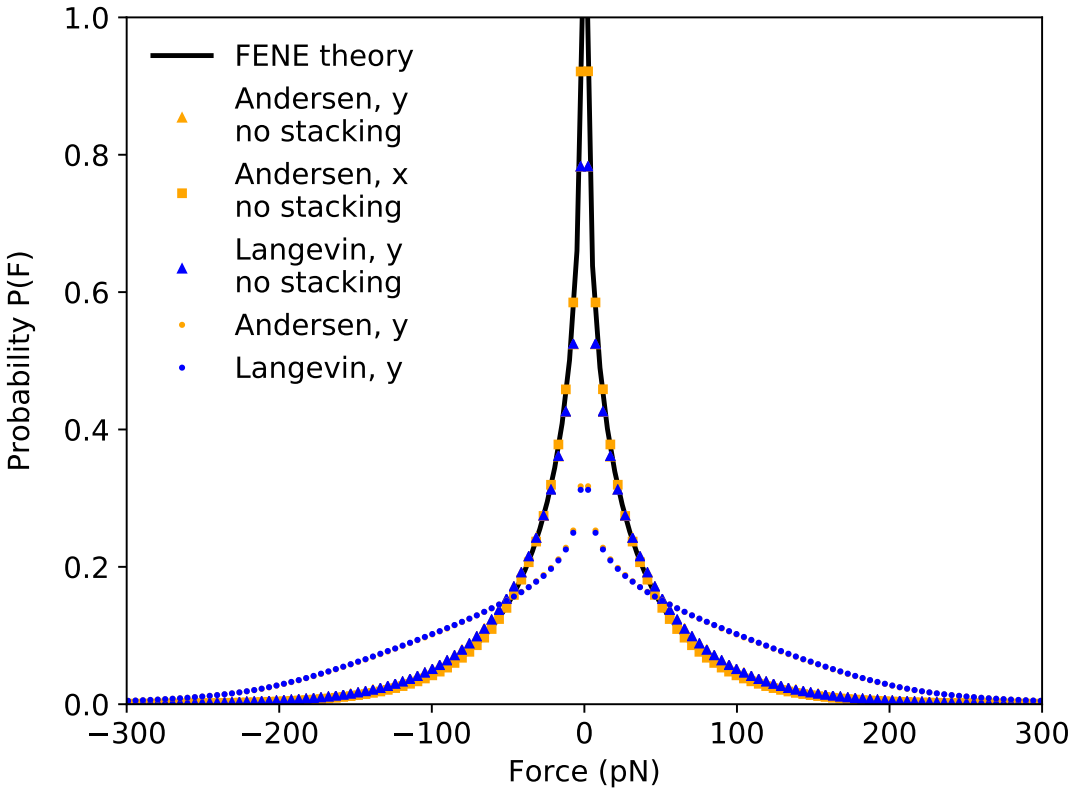


Figure 5.6: Force distribution averaged over all nucleotides in the strand of Fig. 5.5, both with the normal oxDNA model (circles) and with stacking interactions turned off (squares and triangles), using two different thermostats. Forces are net forces exerted on each nucleotide by all other nucleotides in the chain, and are shown along the direction of constraint (y) and along an orthogonal axis (x). The black line is the theoretical expectation if the FENE potential is the only mediator of force.

fewer phase space conformations available to it; thus, its entropic tension is lower.

We first note that the effective force shown in Fig. 5.5 is an average, and that instantaneous forces can be much larger. We can consider this in more detail by looking at the distribution of forces experienced by an individual nucleotide along the chain, shown in Figure. 5.6 (averaged over all nucleotides in the chain) for two directions: along the axis of constraint (y) and orthogonal to the axis of constraint (x), and for two different thermostats: Andersen-like and Langevin (described in Chapter 2). Data for the x and y axes are in good agreement, as are data for the two thermostats, which confirms that the distributions are not artefacts of a particular thermostat choice. The first thing to notice is that the distributions are very

broad. Secondly, we compare a simulation with stacking turned off to a full oxDNA simulation, and force distributions corresponding to the full simulations are broader than those with stacking turned off. In both cases, the considerable magnitude of the forces can be explained by looking in detail at the underlying oxDNA potentials.

5.2.1 Theory

To rationalise the shapes of the distributions in Fig. 5.6, we first note that for a single poly-T strand of DNA, the primary mechanisms available for internucleotide force transfer are propagation along the backbone, mediated by a finitely extensible nonlinear elastic (FENE) potential, and stacking interactions between bases.

FENE potential

It is easier to analyze the simpler system with stacking forces turned off, where the FENE potential will dominate. The FENE potential as implemented in oxDNA is given by

$$V(r) = -\frac{\epsilon}{2}\ln\left(1 - \left(\frac{r - r_0}{\Delta}\right)^2\right), \tag{5.12}$$

where $\Delta = 0.25$ (0.21 nm); $\epsilon = 2$ $k_B T^*$ with T* a reduced temperature, 3000K; and $r_0 = 0.7564$ (0.644 nm) in oxDNA 2.0[261]. The probability of finding bases separated by r is

$$P(r) = \frac{e^{-\beta V(r)}}{\int e^{-\beta V(r)} dr} = \frac{e^{-\beta V(r)}}{Z}, \tag{5.13}$$

and the free energy, A, as a function of r is

$$A(r) = -k_B T \ln(P(r)). \tag{5.14}$$

The force experienced by nucleotides separated by distance r is

$$\vec{F}(r) = -\frac{\partial A}{\partial r}\hat{\mathbf{r}} = -\frac{\partial V}{\partial r}\hat{\mathbf{r}} = \frac{\frac{\epsilon}{\Delta^2}(r - r_0)}{1 - \left(\frac{r - r_0}{\Delta}\right)^2}\hat{\mathbf{r}}, \tag{5.15}$$

and is directed along the backbone. If we let $\vec{F}(r) = F\hat{\mathbf{r}}$, the probability density function for observing a force F, $g(F)$, can be found by noting $g(F)\,dF = P(r(F))\,dr$, or $g(F) = P(r(F))\,\frac{dr}{dF}$, where $P(r)$ is given by equation 5.13 and $r(F)$ is found by inverting equation 5.15:

$$r(F) = -\frac{\epsilon}{2F} \pm \frac{1}{2F}\sqrt{\epsilon^2 + 4\Delta^2 F^2} + r_0. \tag{5.16}$$

Only the solution featuring the negative square root is bounded. Therefore,

$$g(F) = P(r(F))\frac{dr}{dF} = \frac{1}{Z}\left[1 - \frac{1}{\Delta^2}\left(- \frac{\epsilon}{2F} - \frac{1}{2F}\sqrt{\epsilon^2 + 4\Delta^2 F^2}\right)^2\right]^{\epsilon\beta/2}\frac{dr}{dF}. \tag{5.17}$$

From equation 5.16, $\frac{dr}{dF}$ is

$$\frac{dr}{dF} = \frac{\epsilon}{2F^2} + \frac{2\Delta^2}{\sqrt{\epsilon^2 + 4\Delta^2 F^2}} - \frac{\sqrt{\epsilon^2 + 4\Delta^2 F^2}}{2F^2}. \tag{5.18}$$

While keeping track of force components F_x, F_y, F_z (and thus force magnitude, $|F|$) is straightforward in oxDNA, the vector direction $\hat{\mathbf{r}}$ is not as easily accessible, and indeed will be different for all nucleotides along the chain. Therefore, for simplicity, instead of working with $g(F)$ (where F is defined by $\vec{F}(r) = F\hat{\mathbf{r}}$), we derive $P(F_z)$, the probability distribution for a single component of the force. To do so, we note that for a random variable A defined as the product of two other random variables, $A = BC$, the probability density distributions of each of the three variables are related thus[224]:

$$P_A(a) = \int_{-\infty}^{\infty} P_B(b)P_C(a/b)\frac{1}{|b|}\,db. \tag{5.19}$$

Since $F_z = \vec{F}\cdot\hat{\mathbf{z}} = F\cos\theta$, we can use equation 5.19 with $A = F_z$, $B = F$, and $C = \cos\theta$ to obtain

$$P_{F_z}(f_z) = \int_{-\infty}^{\infty} g_F(f')P_{cos\theta}(f_z/f')\frac{1}{|f'|}\,df' \tag{5.20}$$

We assume for simplicity that the instantaneous force vector is equally likely to point in all directions; namely, we assume $P(\phi, \theta) = \frac{1}{4\pi}$. In this isotropic case, $P(\theta) = \frac{1}{2}\sin\theta$, and the probability density distribution for $\cos\theta$ is given by

$$P_{cos\theta}(\cos\theta) = P(\theta)\left|\frac{d\theta}{d(\cos\theta)}\right| = \frac{1}{2}. \tag{5.21}$$

Now, we adjust the integration limits of Eq. 5.20 to reflect the fact that $F_z \leq F$ by definition:

$$P(F_z) = \int_{-\infty}^{-F_z} \frac{g_F(f')}{2|f'|}\, df' + \int_{F_z}^{\infty} \frac{g_F(f')}{2|f'|}\, df' \tag{5.22}$$

Eq. 5.22 is plotted in black in Fig. 5.6. The agreement with simulation data is excellent where stacking interactions are suppressed, as expected.

Stacking potential

To discover why the stacking interaction broadens the force distributions, we can make a simple argument based on the form of the stacking potential in oxDNA.

For simplicity, we consider only the radial part of the stacking potential:

$$V(r) = \epsilon_s\left(1 - e^{-a(r-r_{0s})}\right)^2 - \epsilon_s\left(1 - e^{-a(r_c-r_{0s})}\right)^2, \tag{5.23}$$

where $\epsilon_s = 1.3523 + 2.6717 k_B T$, $r_c = 0.9$, $r_{0s} = 0.4$, and $a = 6$ in oxDNA 2.0[261].

Figure 5.7 shows this potential alongside the radial part of the FENE potential, with the minima of the potentials shifted to the origin. The black line indicates the thermal energy the system is expected to possess in equilibrium according to equipartition, $k_B T/2$, at T=21 °C. We can evaluate the force, $F(\delta r) = -\frac{dV}{dr}$, at the intersection of each potential with the thermal energy, which occurs at radial displacements from equilibrium of $\delta r = 0.13$ nm and $\delta r = 0.07$ nm for the FENE and stacking potentials, respectively. $F_{FENE}^{thermal} \sim 89$ pN and $F_{stack}^{thermal} \sim 193$ pN for these displacements, matching well the observed widths in Fig. 5.6. We can thus confidently attribute the form of the observed force distributions to the oxDNA potentials, and the substantial force distribution width associated with stacking interactions to the steepness – and thus, large instantaneous forces – of the stacking potential.

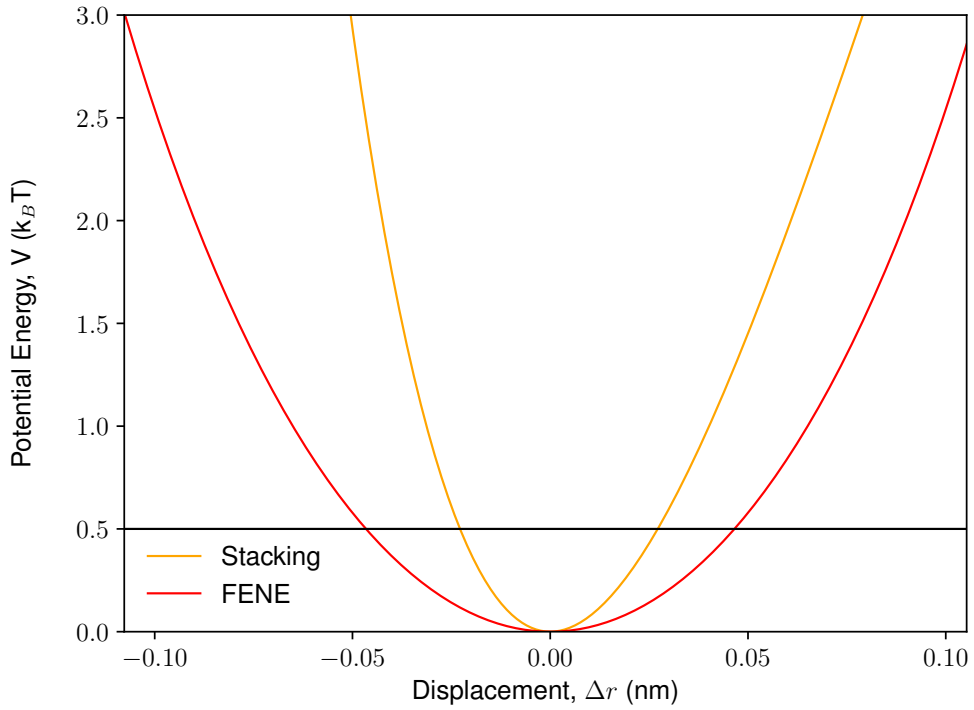


Figure 5.7: oxDNA2.0 radial FENE (red) and stacking (orange) potentials, with the minima of the potentials shifted to the origin. The stacking potential is clearly much steeper than the FENE at the thermal energy (black line), giving rise to the large force distribution width for the stacking curves in Fig. 5.6.

5.2.2 Secondary structure

The distributions in Fig. 5.6 were acquired using the forces exerted on each nucleotide by all other nucleotides in the chain. If we consider a more complex arrangement – a 160-nt strand that is poly-T save for a 10-bp stem, 4-T loop hairpin near its edge; see Figure 5.8(a) – complications arise. The average one-sided force on each nucleotide along the direction of constraint is shown in Figure 5.8(b); while one might expect it to be approximately constant and equal to the entropic tension in the chain, this is not the case in the vicinity of the hairpin. Clearly, when analyzing tension in a DNA chain, it is no longer sufficient to simply calculate the topologically 'upstream' and 'downstream' forces on each nucleotide where secondary structures are involved, as force propagation in that case is mediated by more complex interactions, including hydrogen bonding and cross-stacking, that

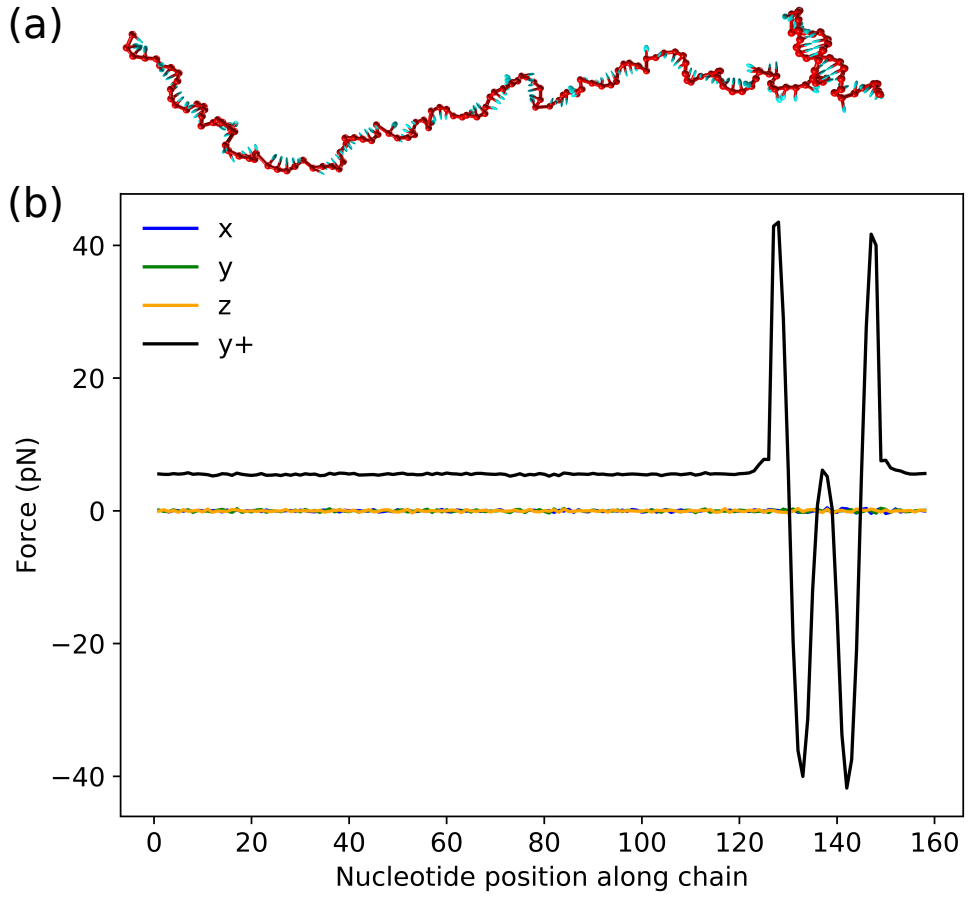


Figure 5.8: (a) oxDNA representation of a strand identical to that of Fig. 5.5, but with a 10 bp-stem, 4 base loop hairpin near its edge. (b) The net forces exerted on each nucleotide by all other nucleotides in three dimensions (blue, green yellow), as well as the one-sided force – the net force exerted on each nucleotide by all nucleotides topologically upstream of it – along the axis of constraint (black, $y+$). The presence of secondary structure complicates force propagation in the strand.

communicate entropic tension through the chain in non-obvious ways. The bases in the hairpin stem cannot simultaneously minimize hydrogen-bonding, stacking, FENE, and cross-stacking potentials, regardless of whether there is a net tension in the strand; this is evidenced by the fact that, for example, a stacked single strand has a tighter helix than dsDNA[203]. Because they are displaced from the minima of one of more of these potentials, the nucleotides in the vicinity of the hairpin are consequently subjected to large forces, so calculating a per-nucleotide force is insufficient to extract entropic tension in the chain.

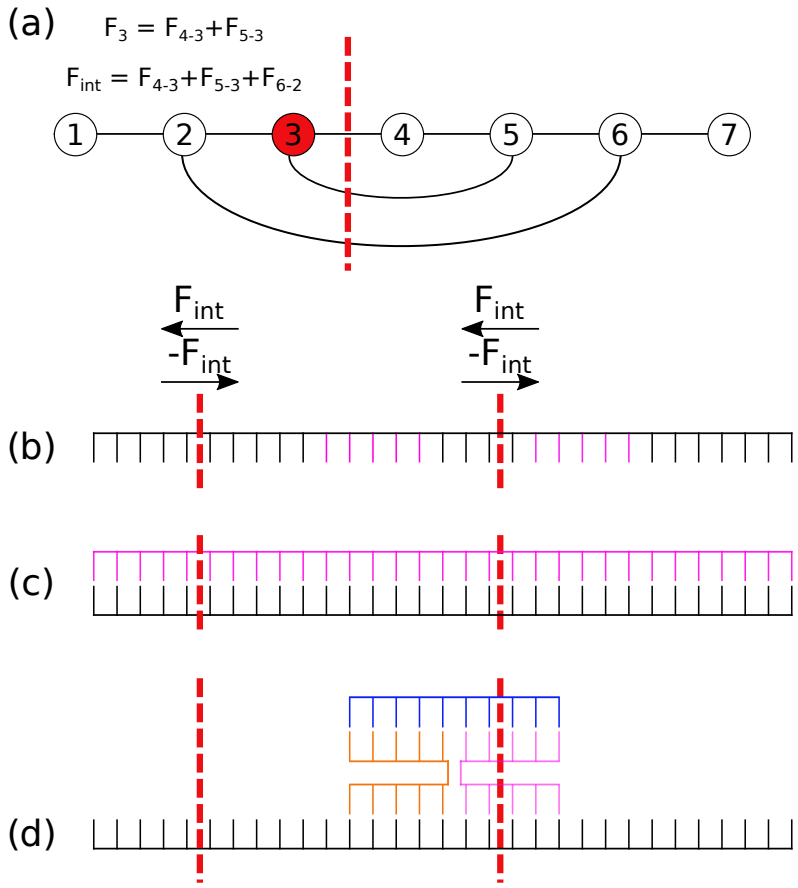


Figure 5.9: (a) Schematic of the method for determining internal tension in DNA structures. For simplicity, consider nucleotides linked only through a FENE backbone (straight lines) and hydrogen bonds (curved lines). Calculating the one-sided force on nucleotide 3 from topologically upstream yields $F_3 = F_{4-3} + F_{5-3}$. The net force thus computed neglects an important counterbalancing term and yields too-large forces (Fig. 5.8(b)) in the vicinity of secondary structures. Instead, the total force exerted on *all* nucleotides to the left of a virtual interface (red dashed line) by *all* nucleotides to the right of that interface should be computed: $F_3 = F_{4-3} + F_{5-3} + F_{6-2}$. The method is illustrated for three representative DNA structures: (b) single strand with a hairpin (pink); (c) double strand; (d) single strand containing a Holliday junction. Red dashed lines are interfaces across which net force is calculated.

5.2.3 Method for calculating internal forces in ssDNA simulations

To correctly determine the average tension in DNA structures, we propose a method using 'topological interfaces'. A virtual interface is drawn between two neighbouring nucleotides in a DNA structure, and the forces exerted by *all* nucleotides topologically upstream of this interface on *all* nucleotides topologically downstream

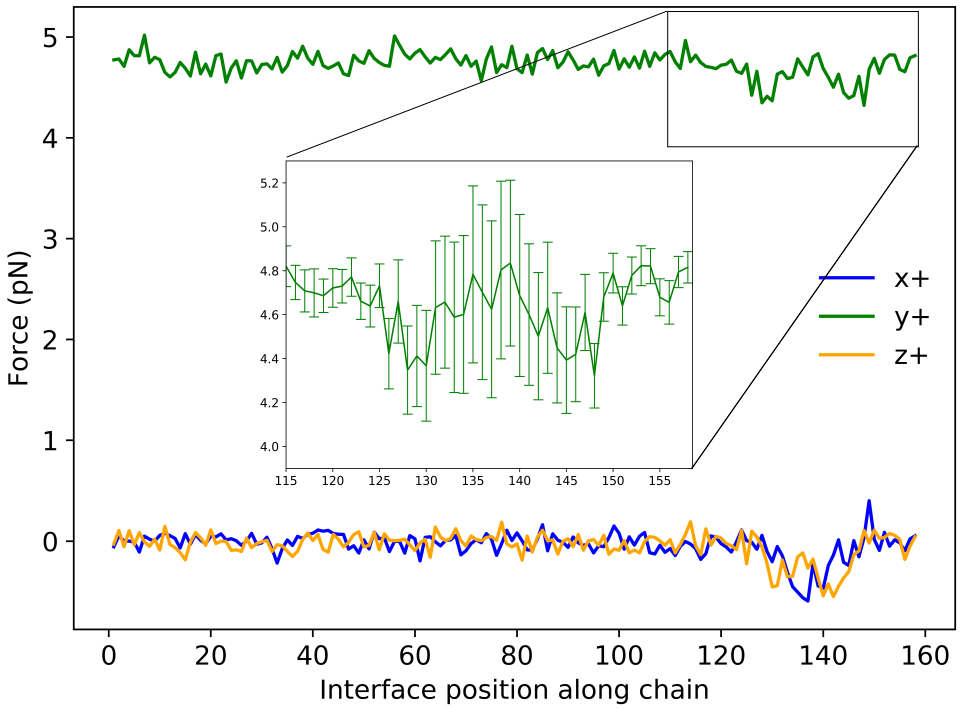


Figure 5.10: Net forces exerted by all nucleotides topologically upstream of a given interface on all nucleotides topologically downstream of the interface, for interfaces positioned throughout the chain (just upstream of the nucleotide number indicated in the abscissa) and for all three directions. Inset: zoom in of the region containing a 10-base pair hairpin. Errors are standard errors of the mean of 64 identical parallel simulations, and tend to be larger in the vicinity of the hairpin because of the larger magnitude of forces in that region. As expected, a non-zero entropic tension exists along the direction of constraint (y).

of this interface (or *vice versa*) is calculated. On average and at equilibrium, the net force directed topologically up $\rightarrow$ down will be equal in magnitude to the net force directed topologically down $\rightarrow$ up, and will equal the net tension in the chain. To see this more clearly, consider the ssDNA schematic in Figure 5.9(a). For simplicity, consider only FENE backbone interactions –indicated by straight lines – and hydrogen bonding interactions – indicated by the curved lines. The one-sided force on nucleotide 3 from upstream, F_3, is the sum of the forces exerted on it by nucleotides 4 and 5. As we saw in Fig. 5.8(b), this leads to forces much larger than the expected entropic tension in the chain. Using the topological interface method, the force exerted by all nucleotides to the right of the red barrier on all

nucleotides to the left of the red barrier, F_{int}, is calculated instead; this amounts to adding the force exerted by nucleotide 6 on nucleotide 2 to F_3. This method correctly yields a constant tension throughout the chain, as shown in Figure 5.10 for the hairpin system in Fig. 5.8(a). The inset shows the region containing the hairpin in more detail, and contains the standard errors of the mean from 64 parallel simulations. Because of the magnitude of forces involved in the vicinity of the hairpin, errors tend to be larger there.

The procedure is further illustrated by the simple examples in Figure 5.9(b)-(d), wherein two possible choices of virtual interface are drawn as dashed red lines intersecting three sample systems. Fig. 5.9(b) is ssDNA containing a 5 bp stem, 4 base loop hairpin, indicated by the pink bases. For the first interface, the net force exerted on the first 5 nucleotides (nts) in the strand by the remaining 25 nts is calculated. The second interface intersects the hairpin, but the procedure remains unchanged: forces exerted on the first 18 nts in the strand by the remaining 12 nts is calculated. Fig. 5.9(c) is dsDNA, and thus slightly more complex. In the case of the first interface here, the net force exerted on the first 5 nts in each strand (total of 10 nucleotides) by the remaining sum total of 50 nts is calculated, and similarly for the second interface. Finally, a strand containing a Holliday junction[108] is shown in Fig. 5.9(d). The net force acting through the first interface is given by the sum of forces exerted on the first 5 nts in the strand by the remaining 25 nts in the long strand and the 30 nts comprising the other three strands. The net force acting through the second interface is the sum of all forces exerted on the first 18 nts in the black strand, all 10 nts in the orange strand, the first 7 nts in the blue strand, and 4 nts in the yellow strand as indicated by all nts to the right of the interface.

While in principle, we can get the tension in the DNA structure using the net force across a single interface, in practice it makes sense to perform the force calculation for multiple interfaces throughout the structure to improve statistics. We apply our method for calculating internal forces in the next section.

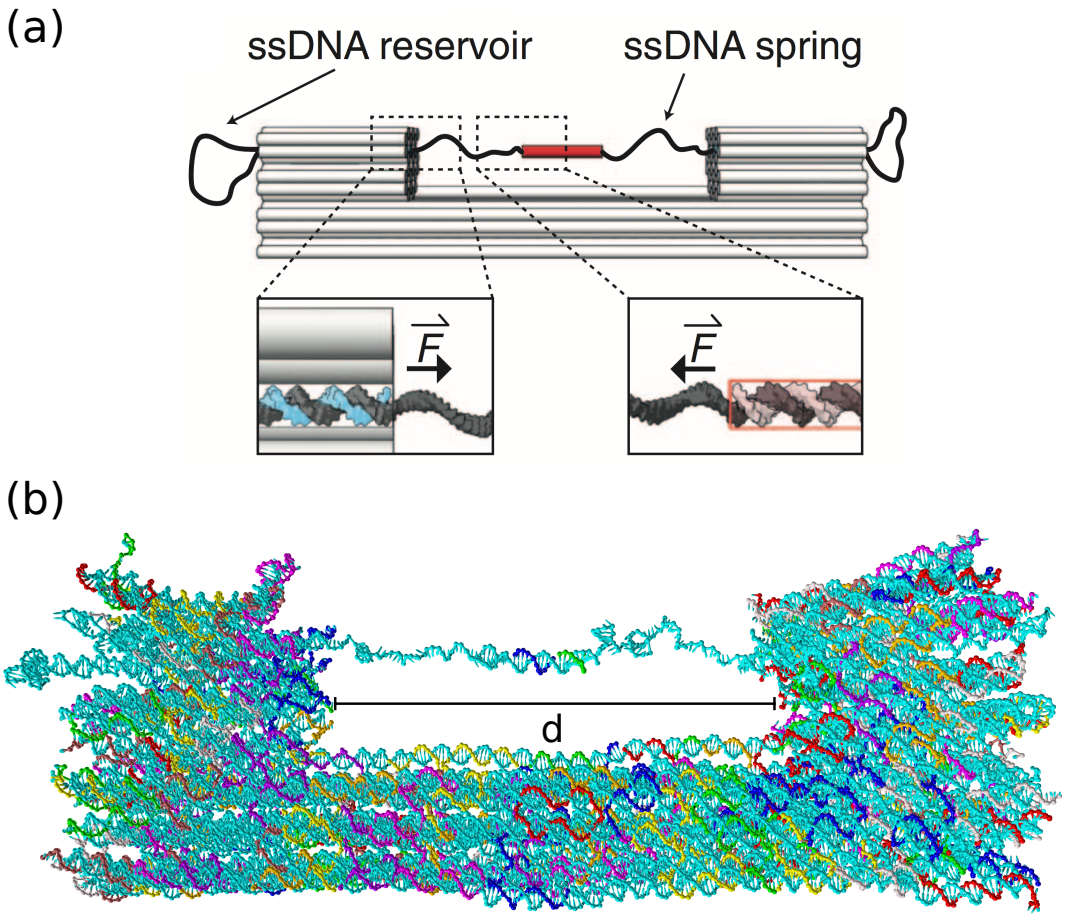


Figure 5.11: Illustration of the force clamp design of Nickels et al. [193]. (a) Schematic of the force clamp: a portion of ssDNA to which a system of interest is bound (red) is suspended in a DNA origami frame. Its length can be adjusted by altering staple sequences such that material in the ssDNA 'reservoirs' shifts in or out of the central gap. Adapted from Nickels et al. [193]. (b) oxDNA representation of the structure. The designed gap distance d is 42.84 nm.

5.3 Application of internal force calculation to a single-molecule force clamp

5.3.1 Introduction and system description

A recent proposal for a DNA-origami force clamp by Nickels et al. [193] solves many of the issues with traditional force spectroscopy apparatus like AFM and LOT, and is a

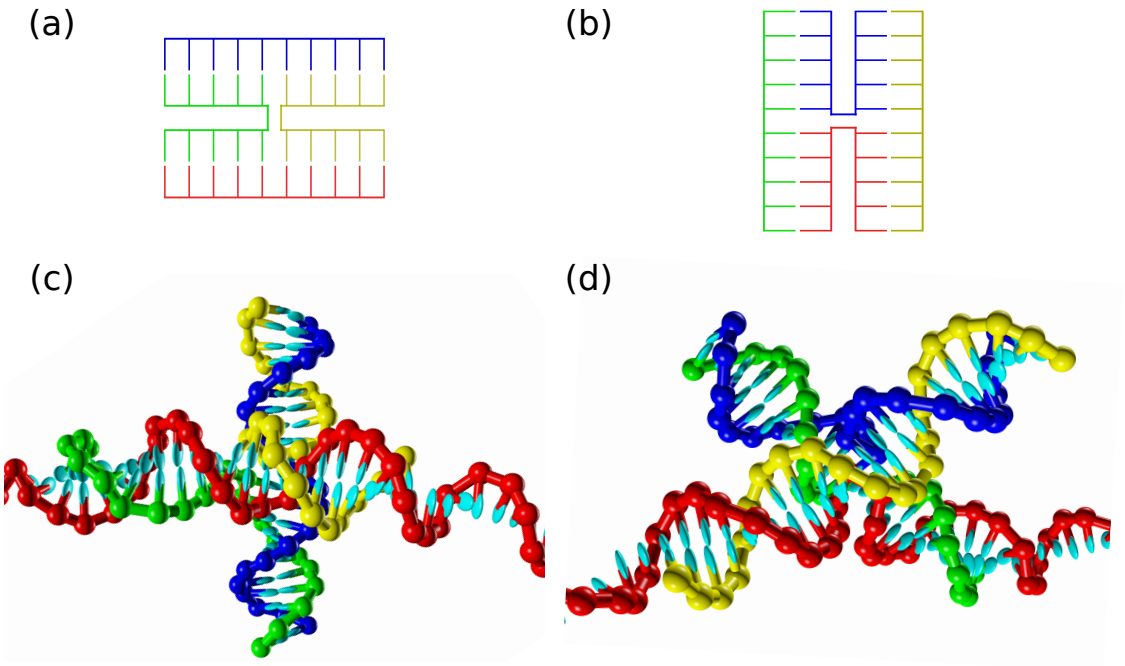


Figure 5.12: Schematics and oxDNA representations of the two possible coaxially-stacked isomers of the Holliday junction. (a), (c) IsoII, in which coaxial stacking occurs between yellow and green strands. (b), (d) IsoI, in which coaxial stacking occurs between red and blue strands. Here, the red strand is part of a longer ssDNA segment (see Fig. 5.14 for zoom-out) along which force is applied. This force adds a thermodynamic bias favouring IsoII.

powerful addition to the toolbox used to investigate mechanical processes in the cell. In particular, because it is instrument-free, many of the artefacts that plague SMFS experiments – for example, convolution with long linker handles, as discussed in Chapter 3 – are not a concern. Additionally, the nanostructure is highly scalable, so single-molecule experiments can be carried out in parallel in large numbers; contrast this with the fundamentally serial nature – and consequently poorer statistics – of other experimental SMFS strategies. Finally, the need for attachment to microscopes prevents traditional SMFS experiments from being performed *in vivo*, while no such limitation exists in theory for nanoscopic force clamps. Figure 5.11 contains the design of Nickels et al.[193] as well as its oxDNA representation. The usual M13mp18 origami scaffold, containing 7253 nt, was used. Additional base pairs from ssDNA 'reservoirs' on either side of the structure can be threaded into and out of the central gap region by appropriately altering staple sequences in the body of the origami;

this changes the number of nts, and consequently alters the tension, in the ssDNA spring. A molecule of interest can be attached at the location marked in red in Fig. 5.11(a) and indicated by blue and green staples in Fig. 5.11(b), on which a particular force is then exerted based on the length of the ssDNA spring.

Nickels et al.[193] demonstrate the operation of their nanoscopic force clamp by studying the four-way Holliday junction[108] (HJ), a fundamental component of DNA origami designs. HJs are comprised of 4 strands hydrogen-bonded to each other and arranged in an 'X' shape. In the presence of salt, the strands coaxially stack to adopt one of two possible isomeric conformations, illustrated in Figure 5.12. In Fig. 5.12(a) and (c), the green and yellow strands are coaxially stacked, and the blue and red strands are regularly stacked such that they are straight; the reverse is true in Fig. 5.12(b) and (d). The population ratio between the isomeric states can be altered via the application of external force; pulling on either end of the red strand in Fig. 5.12, for example, will favour the isomer in which the red strand is straight. Nickels et al.[193] refer to this more favourable isomeric state as 'IsoII', and the less favourable isomer, in which the tension-bearing strand is bent, as 'IsoI'; we adopt the same convention.

To illustrate the force-dependence on the population ratios of IsoI and IsoII, Nickels et al.[193] construct three versions of the origami force clamp: 4.0 pN, 2.5 pN, and 1.2 pN, named according to the forces the authors calculate the ssDNA springs to exert on the molecule of interest. Values for these designed forces were derived from the exFJC model, Eq. 5.8, using the Kuhn length and stretch modulus parameters reported by Smith et al.[259], the number of nts spanning the central gap in each design (160, 4.0 pN; 226, 2.5 pN; 422, 1.2 pN) minus the 22 nts corresponding to the HJ region, a contour length per base pair of 0.63 nm[176], and the designed central gap width of 42.84 nm – labelled d in Fig. 5.11 – minus the average end-to-end extension of the HJ as the polymer end-to-end distance, $\langle x \rangle$. Notably, the magnesium ion concentration at which the force clamp experiments were performed – 100 mM [Mg^{2+}] – represents much higher electrostatic screening than the salt concentration used in the experiments of Smith et al.[259], 150 mM [Na^{+}], because Mg^{2+} has been

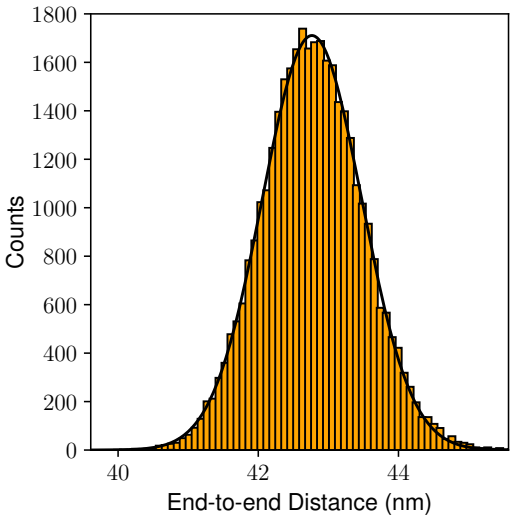


Figure 5.13: End-to-end distance distribution for two scaffold nucleotides on either end of the gap in Fig. 5.11. The width of this distribution can be used to determine an effective stiffness of the origami base.

estimated to have an effect that is 20-100× stronger than Na^+ [171;21], depending on the property that is being studied. High ionic strengths are often used in DNA origami experiments to facilitate origami assembly.

5.3.2 Simulations

To highlight the desirability of performing simulations to estimate ssDNA internal forces for applications like nanoscopic force clamps, we simulated the system of Nickels et al. [193] to determine the forces being exerted on the HJ. For our purposes, it was unnecessary to simulate the full origami structure repeatedly. Instead, we determined the effective stiffness of equivalent harmonic traps located in the body of the origami clamp and simulated only the ssDNA strand in the central gap, vastly improving the statistics we were able to collect in our available simulation time. To do this, an MD simulation of the origami force clamp with no connecting single strand was performed, and the end-to-end distance r between scaffold nucleotides on either end of the clamp was monitored; the results are shown in Figure 5.13. The resulting distribution was fit to a Gaussian form:

$$P(r) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5.24}$$

to extract the mean $\langle r \rangle = 42.77$ nm and variance $\sigma_r^2 = 0.50$ nm^2. Note that $\langle r \rangle$ is approximately equal to the value used by Nickels et al.[193] based on simple geometrical arguments: 42.84 nm. Consider two nucleotides separated along the z axis – the long axis of the force clamp – by d. We assume each nucleotide moves in an isotropic 3D harmonic potential for simplicity, such that the Boltzmann probability of observing position r_1 for nucleotide 1, positioned at the origin, is

$$\begin{aligned} P(\mathbf{r_1}) = P(x_1, y_1, z_1) &= \frac{e^{-\beta k(x_1^2+y_1^2+z_1^2)/2}}{\int_{-\infty}^{\infty} e^{-\beta k(x_1^2+y_1^2+z_1^2)/2}\,\mathrm{d}x_1\,\mathrm{d}y_1\,\mathrm{d}z_1} \\ &= \Big(\frac{\beta k}{2\pi}\Big)^{3/2} e^{-\beta k(x_1^2+y_1^2+z_1^2)/2} \end{aligned} \tag{5.25}$$

and for nucleotide 2 is

$$P(\mathbf{r_2}) = P(x_2, y_2, z_2) = \Big(\frac{\beta k}{2\pi}\Big)^{3/2} e^{-\beta k(x_2^2+y_2^2+(z_2-d)^2)/2}, \tag{5.26}$$

where β is the inverse thermal energy and k is the stiffness of each trap. The distance between the two nucleotides is $r = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$, so the mean r and r^2 values are given by:

$$\langle r \rangle = \int_{-\infty}^{\infty} \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2} P(x_1, y_1, z_1) P(x_2, y_2, z_2)\,\mathrm{d}^2x_{1,2}\,\mathrm{d}^2y_{1,2}\,\mathrm{d}^z_{1,2} \tag{5.27}$$

$$\langle r^2 \rangle = \int_{-\infty}^{\infty} \Big((x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2\Big) P(x_1, y_1, z_1) P(x_2, y_2, z_2)\,\mathrm{d}^2x_{1,2}\,\mathrm{d}^2y_{1,2}\,\mathrm{d}^2z_{1,2} \tag{5.28}$$

where the joint probability distribution $P(x_1, y_1, z_1, x_2, y_2, z_2)$ is simply the product $P(x_1, y_1, z_1)P(x_2, y_2, z_2)$ since the variables are independent. Equations 5.27 and 5.28 give the mean and variance ($\langle r^2 \rangle - \langle r \rangle^2$) of the $P(r)$ distribution, and can be numerically integrated with different values of k until the observed parameters are obtained. Doing so yields an effective trap stiffness of $k = 16.32$ pN/nm and separation of $d = 42.76$ nm. These values are only slightly different from the results obtained by neglecting position fluctuations in x and y altogether, assuming $r = \Delta z = z_2 - z_1$, and taking the variance of $P(r)$ to be the sum $\sigma_r^2 = \sigma_{z_1}^2 + \sigma_{z_2}^2 = \frac{1}{\beta k} + \frac{1}{\beta k}$,

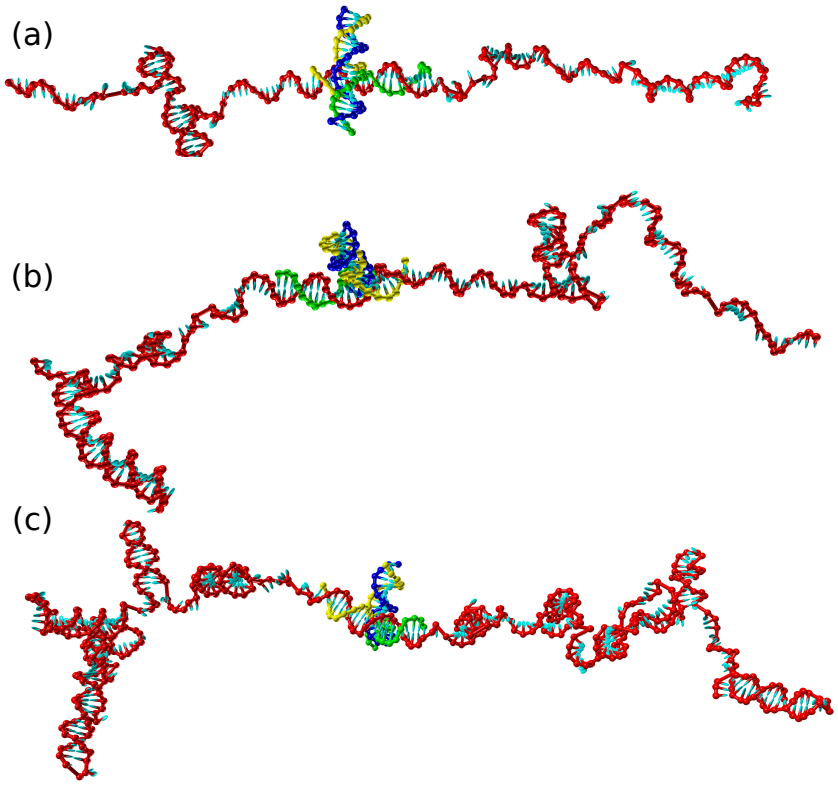


Figure 5.14: oxDNA representations of the gap-spanning ssDNA segment of the force clamp, with Holliday junction attached, for the (a) 4.0 pN, (b) 2.5 pN, and (c) 1.2 pN designs. The red strand contains a total of 162 nt, 228 nt, and 424 nt in each of these three cases, respectively. Clearly, this sequence is prone to secondary structure formation.

with the individual variances $\sigma_{z_i}^2$ equal and determined using the equipartition theorem. Under these assumptions, $k = 16.23$ pN/nm and $d = 42.77$ nm.

We performed MD simulations on the ssDNA + HJ system, with end nucleotides harmonically trapped, for the three different force designs of Nickels et al.[193]; see Figure 5.14. Including the two trapped endpoint nucleotides, the main ssDNA strands comprised 162 nt, 228 nt, and 424 nt for the 4.0 pN, 2.5 pN, and 1.2 pN designs, respectively. Clearly, secondary structures are present for this ssDNA sequence, shortening the effective contour length of the ssDNA. We used the same temperature as Nickels et al.[193] and a monovalent salt concentration of $[Na^+]$=5 M in order to mimic the experimental $[Mg^{2+}]$ concentration. To explore the effect of origami compliance on the force in the structure (the exFJC model used by

Nickels et al.[193] assumes fixed endpoints), we also performed simulations on the 4.0 pN system using harmonic trap stiffnesses an order of magnitude larger and smaller than $k = 16.32$ pN/nm. Finally, to explore the effect of salt concentration on the results, we performed additional simulations on the 4.0 pN system at $[Na^+]$=0.15 M, 0.5 M, and 1.0 M.

In all instances, we simulated the strands both with secondary structure and without allowing secondary structures to form. In the case of the 4.0 pN design, we performed an additional simulation with secondary structures suppressed *and* stacking interactions between bases (with the exception of those comprising the HJ) prohibited, to illustrate the effect base stacking has on ssDNA entropic force. Transitions between HJ IsoI and IsoII conformations are very rare on the timescale of our simulations†, so we were unable to probe the effect of force on transition frequency, but we simulated each isomer separately and report results for both.

Simulations on these systems are still ongoing, and results for some of the slowest simulations (larger systems, low salt concentrations) are based on fewer steps, as reported in Table 5.1. For each set of simulation parameters, multiple parallel replicas were run; in most cases, 24, but in a few instances some replicas were discarded due to a transition between isomers occurring (see Table 5.1).

It is important to ensure that our simulations are sampling all relevant secondary structures, a notoriously difficult task given the possibility of the system becoming trapped in local minima for long stretches of simulation time. To gauge whether we had sampled secondary structures sufficiently, we split our simulation output for each system in half and computed histograms of which bases formed hydrogen bonds throughout the simulation, shown in Figures 5.15-5.17 for all three designed forces. Only IsoII results at the highest salt concentration simulated, 5 M $[Na^+]$, are pictured, as secondary structures become less favourable at lower ionic strengths due to electrostatic repulsion and secondary structure sampling is therefore most difficult at high salt. Though well over 10 000 possible pairs of bases exist for these strands, the figures show only those hydrogen bonds which were actually observed

†Over $\sim 4 \times 10^{11}$ cumulative MD simulation steps (∼6 ms), only 9 transitions were observed, and in each case, the HJ never reverted.

Table 5.1: Summary of the simulations performed on the system of Nickels et al.[193], including the average ssDNA entropic force in each case. Errors in force are the standard errors of a weighted mean across a number of different nucleotides at which forces were measured, as shown in Fig. 5.18. For almost all sets of parameters, 24 identical replica simulations were run; where fewer replicas are reported below, some runs were discarded due to the (rare) occurrence of a transition between isomer states.

Design	**Isomer**	**Secondary structures**	**k (pN/nm)**	**Salt (M)**	**# MD steps (×24)**	**CPU days**	**Force (pN)**
4.0 pN	IsoI	On	16.32	5	1×10^9 (×23)	85	6.54±0.05
			16.32	1	9×10^8	113	6.46±0.06
			16.32	0.5	5.8×10^8	96	5.71±0.07
			16.32	0.15	4.8×10^8 (×19)	102	3.6±0.1
		Off	16.32	5	1.3×10^9	104	2.91±0.03
	IsoII	On	16.32	5	1.3×10^9	113	6.35±0.04
			16.32	1	9×10^8	113	6.52±0.05
			16.32	0.5	5.8×10^8	96	5.43±0.06
			16.32	0.15	4.8×10^8	130	3.4±0.1
		Off	1.632	5	7.8×10^8	64	2.23±0.05
			16.32	5	7.8×10^8	64	2.61±0.05
			163.2	5	8.8×10^8	64	2.61±0.05
		Off, no stacking	16.32	5	7.8×10^8	59	3.26±0.04
2.5 pN	IsoI	On	16.32	5	9×10^8	110	4.64±0.05
		Off	16.32	5	1.4×10^9 (×23)	131	1.38±0.03
	IsoII	On	16.32	5	1.1×10^9	134	4.66±0.06
		Off	16.32	5	1.4×10^9 (×23)	131	1.19±0.03
1.2 pN	IsoI	On	16.32	5	6×10^8	134	3.52±0.08
		Off	16.32	5	1×10^9 (×23)	155	0.43±0.03
	IsoII	On	16.32	5	6×10^8	134	3.30±0.08
		Off	16.32	5	7.5×10^8	40	0.36±0.03

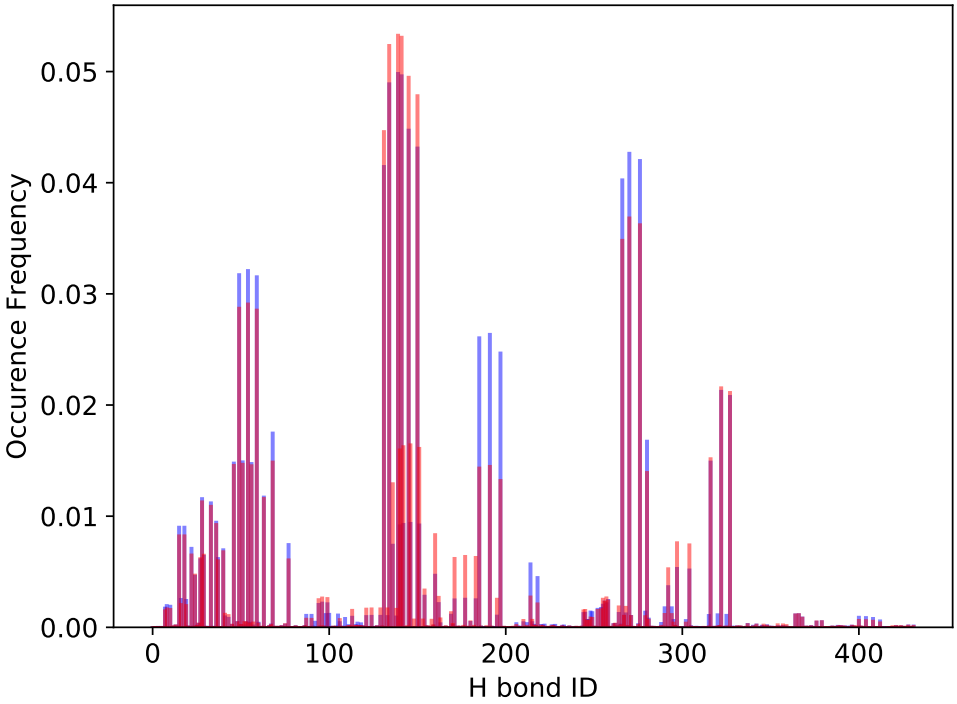


Figure 5.15: Map of all unique base pairs observed in oxDNA simulations of the 4.0 pN IsoII structure at $[Na^+]$=5 M. The data were split in half (12 identical replicas of 1×10^9 MD steps each), indicated by red and blue histograms. The amount of overlap gives a rough indication of whether the important secondary structures are being sampled sufficiently.

and assigns a numeric ID to each unique bond. The amount of overlap between blue and red histograms is taken as a proxy of how well secondary structures were sampled. For the 4.0 pN design, the main patterns of secondary structures visited are the same for both histograms, and we conclude that 1×10^9 MD steps ($\times$ 24 replicas) is an appropriate simulation length for this system. We expect equilibration over secondary structures to be easier at higher entropic forces, since force helps to destabilize secondary structures. Indeed, the histogram overlap is less for the larger 2.5 pN and 1.2 pN designs, and more different pairs of bases are seen to interact via hydrogen bonding as the designed force decreases. The results reported for these low-force systems should be treated correspondingly more cautiously. While the accuracy of the results can be improved with longer simulations, as discussed in the next section, we observe the same trends for all three force designs, and thus

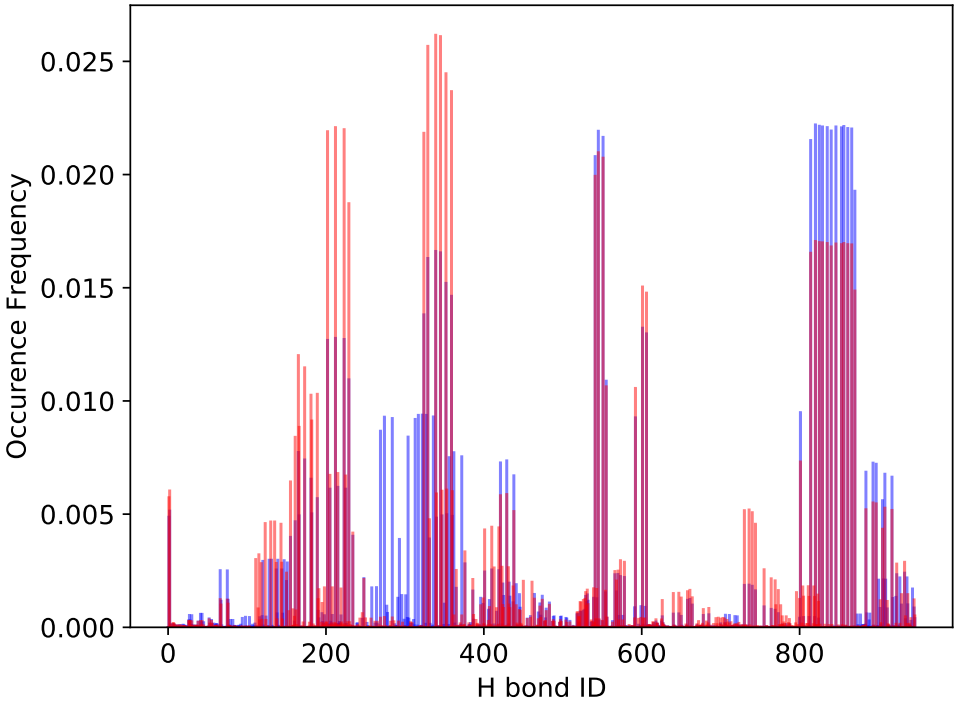


Figure 5.16: Map of all unique base pairs observed in oxDNA simulations of the 2.5 pN IsoII structure at $[Na^+]$=5 M. The data were split in half (12 identical replicas of 1.1×10^9 MD steps each), indicated by red and blue histograms. The amount of overlap gives a rough indication of whether the important secondary structures are being sampled sufficiently.

have confidence in the results even for the 2.5 pN and 1.2 pN systems. Finally, the simulations at $[Na^+]$=0.15 M and $[Na^+]$=0.5 M and the 1.2 pN simulations with secondary structure on are only ∼1/2 as long as the others, and the corresponding results should therefore be regarded as more preliminary.

5.3.3 Results

Figure 5.18 shows net forces crossing several interfaces along the ssDNA chains; the interfaces are located immediately topologically upstream of the nucleotides indicated in the abscissa. Error bars represent the standard deviation of the results of 24† identical simulations run in parallel. In general, because of the more complex structure in the region of the HJ (centre of the strands), and hence the larger the

†Exceptions noted in Table 5.1.

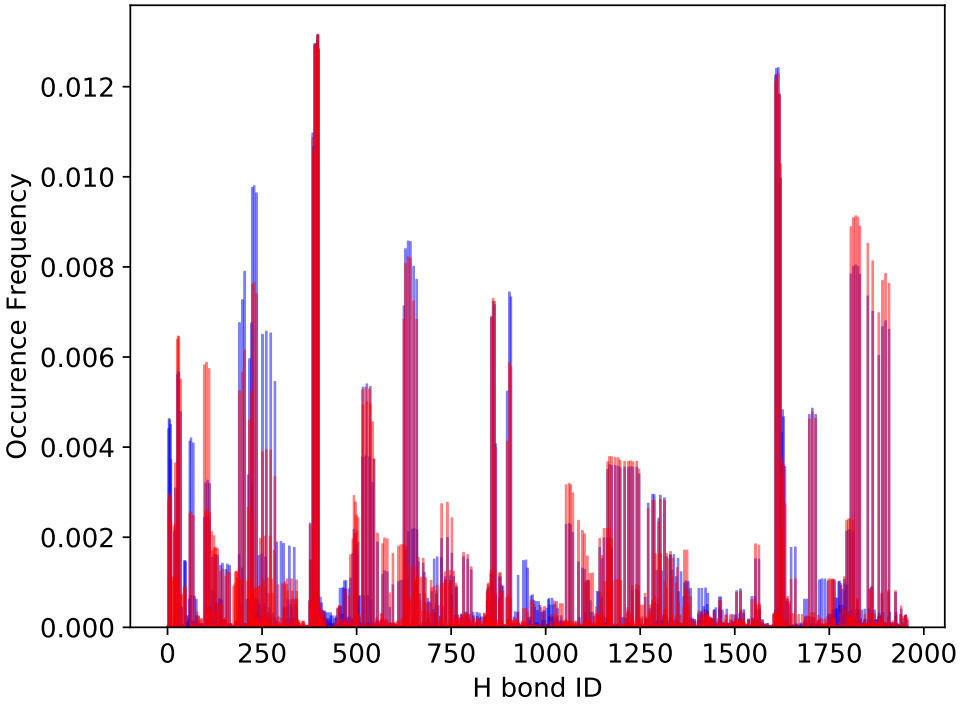


Figure 5.17: Map of all unique base pairs observed in oxDNA simulations of the 1.2 pN IsoII structure at $[Na^+]$=5 M. The data were split in half (12 identical replicas of 6×10^8 MD steps each), indicated by red and blue histograms. The amount of overlap gives a rough indication of whether the important secondary structures are being sampled sufficiently.

magnitude of the (positive and negative) forces across the interfaces, errors tend to be larger there; the same effect was observed in Fig. 5.10. As aforementioned, results at $[Na^+]$=0.5 M and $[Na^+]$=0.15 M and for the 1.2 pN design with secondary structures turned on are more preliminary than those at higher concentrations. Only the projection of force onto the direction of trap separation is shown; forces along the orthogonal directions average to 0, as expected. The average forces over all interfaces in Fig. 5.18 are reported in Table 5.1 and illustrated in Figures 5.19-5.21, which summarize our main results. Error bars represent the standard error of a weighted mean across the interfaces shown in Fig. 5.18.

Immediately evident in Fig. 5.19 is the substantial discrepancy between the forces as measured in simulation (blue and yellow curves) and both the Smith et al.[259] exFJC force predictions (black dashed line; ∼2.3 pN average discrepancy)

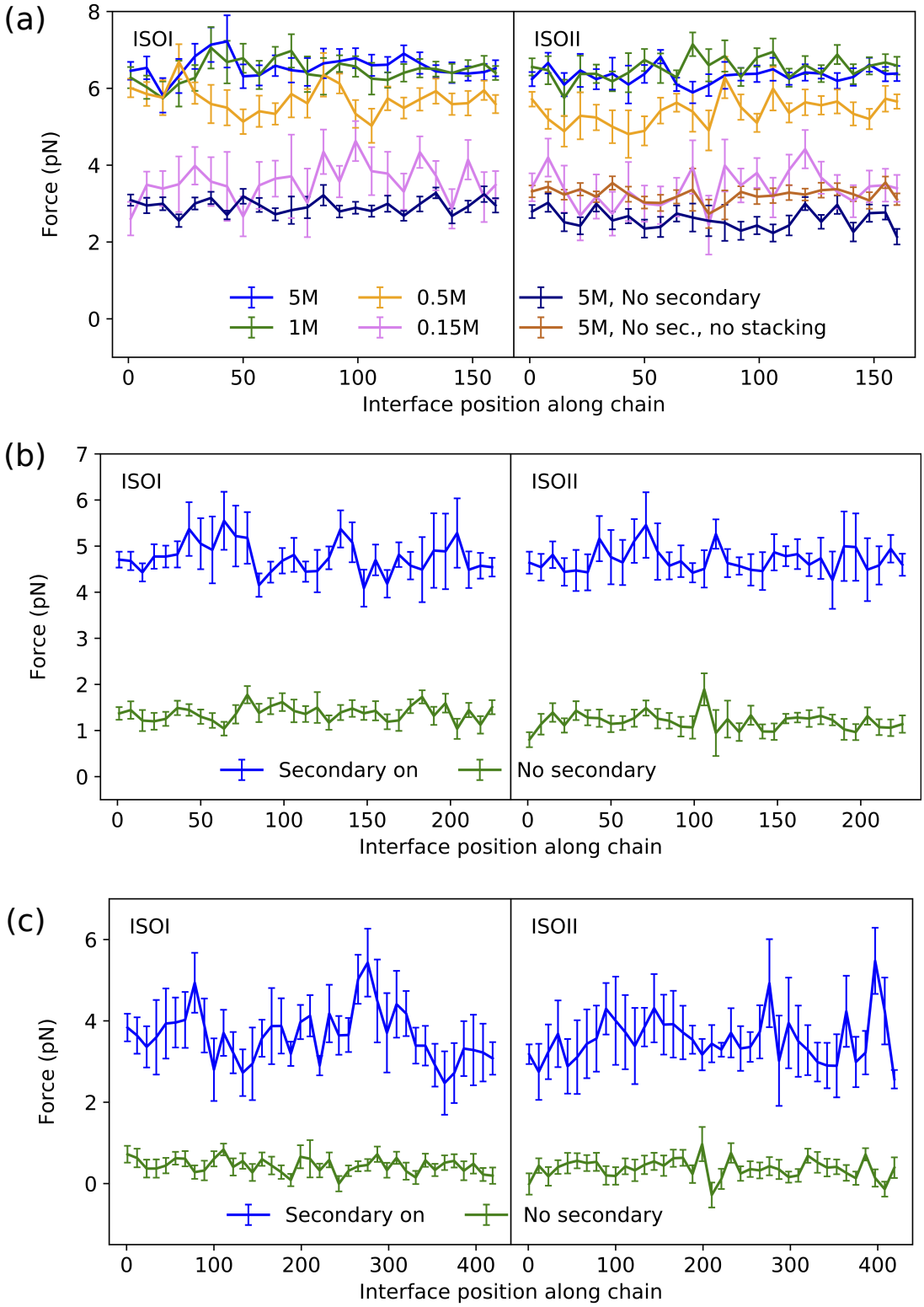


Figure 5.18: Net forces crossing interfaces at several locations along the ssDNA strand for the (a) 4.0 pN, (b) 2.5 pN, and (c) 1.2 pN designs. Errors are standard deviations across multiple identical replica simulations.

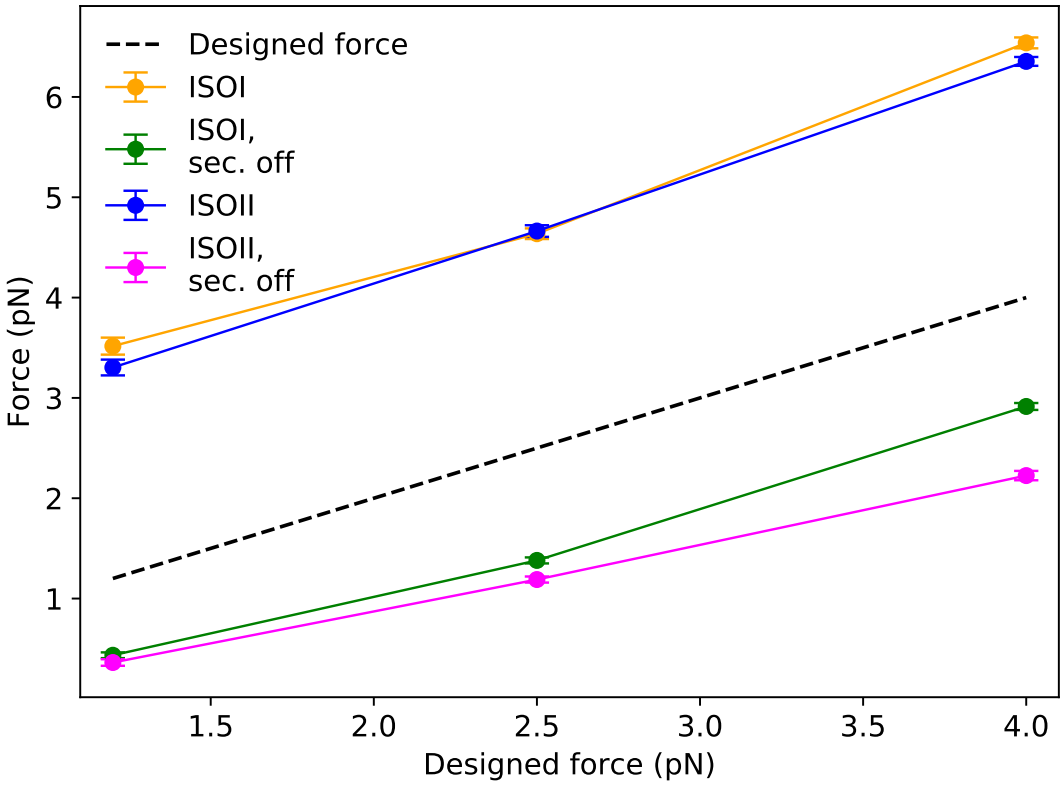


Figure 5.19: Force measured in simulations versus 'designed' force, based on the exFJC parameters of Smith et al.[259] 1:1 equivalence between the two is indicated by the dashed black line. In all cases, the measured force is larger than predicted (by around ∼2.3 pN) and larger than the result when secondary structures are suppressed (by around ∼3.4 pN). In all cases, the force difference between IsoI and IsoII variations is small.

and the forces in the absence of secondary structures (∼3.4 pN average discrepancy). The fact that the secondary-structure-free forces are lower than the exFJC estimates can be understood by considering that the λ-phage ssDNA used by Smith et al.[259] likely contained some inherent secondary structure, which their fits averaged over.

While it is clear that secondary structures have a significant effect on entropic force, another key driver of the disparity between measured forces and exFJC predictions in this case may simply be the different salt concentrations under which the experiments of Smith et al.[259] and Nickels et al.[193] were performed. Figure 5.20 shows the average forces for the 4.0 pN design at four different salt concentrations; at the lowest of these, $[Na^+]$=0.15M – equivalent to the concentration at which Smith et al.[259] performed their fits – the measured force is ∼3 pN lower than at

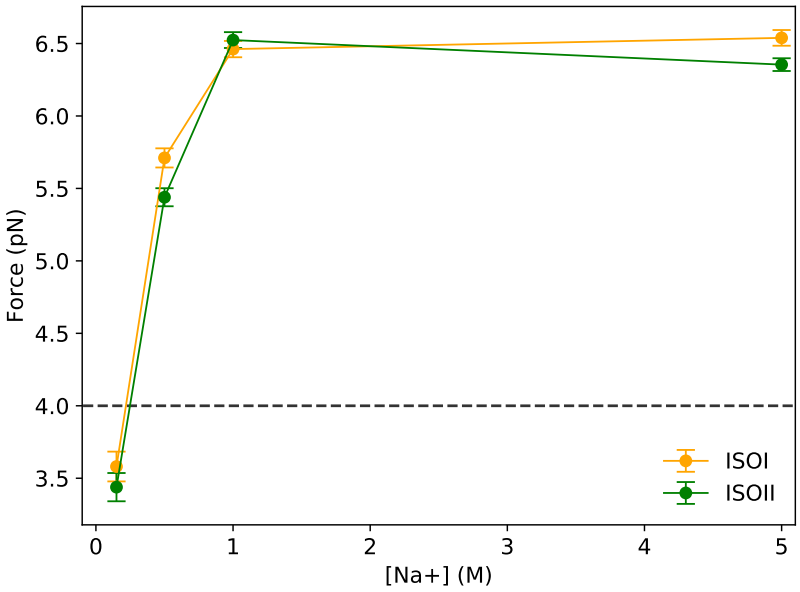


Figure 5.20: Effect of salt concentration on entropic force in ssDNA, for the 4.0 pN design of Nickels et al.[193] (the designed force is indicated by the dashed black line). At low salt, secondary structures are partially suppressed by electrostatic repulsion, leading to lower entropic forces.

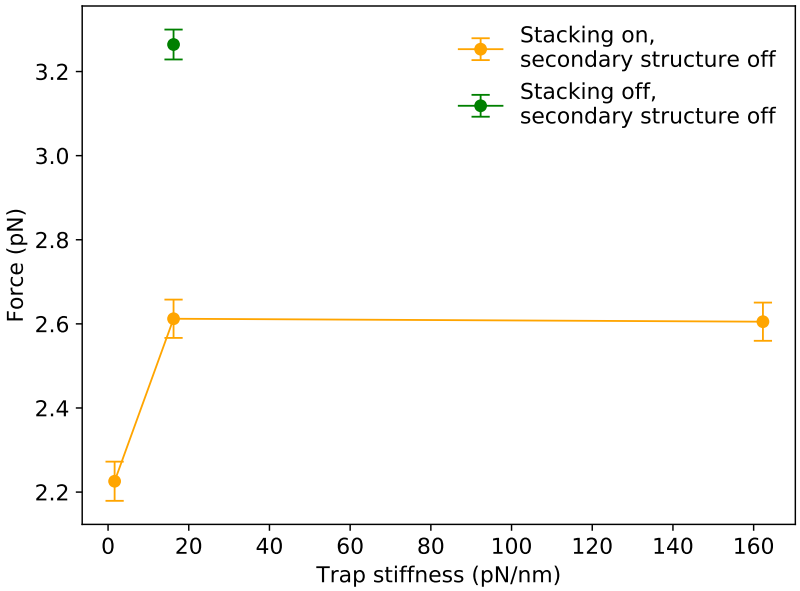


Figure 5.21: Effect of stiffness of the harmonic traps applied to end nucleotides of the ssDNA. The effect of suppressing stacking interactions is also shown for trap stiffness 16.32 pN/nm.

the higher salt concentrations used by Nickels et al.[193]. Caution is clearly needed when translating ssDNA elastic fit results performed under different conditions.

Interestingly, the oxDNA force at [Na^+]=0.15M is close (just ∼0.6 pN below) the Smith et al.[259] prediction‡. The discrepancy can be attributed to a combination of experimental error, sequence differences resulting in different secondary structures and stacking propensities, and the fact that although oxDNA captures the salt dependence of hairpin thermodynamics well, it doesn't do so perfectly[261]. This low-salt agreement between the oxDNA result and the Smith et al.[259] exFJC prediction underscores again the fact that the fits of Smith et al.[259] average over secondary structures. Their exFJC fits therefore attempt to simultaneously describe multiple different polymer regimes. At low forces, secondary structures and stacking reduce the effective contour length of a ssDNA molecule. As force increases, secondary structures and base stacks are disrupted, altering this contour length. While the exFJC form apparently has enough parameters to fit the FECs, one ought to be cautious in physically interpreting these parameters.

Fig. 5.21 explores the effect of compliant attachment points on internal ssDNA forces. While decreasing the trap stiffness an order of magnitude decreases the force noticeably, ∼20 pN/nm appears to be a threshold beyond which increasing the stiffness does not appreciably increase the entropic force. As oxDNA has been elsewhere shown to reproduce the elastic properties of DNA origamis well[251;180], we can conclude that for these designs the assumption of Nickels et al.[193] that the origami is rigid does not affect the force estimate.

As two final notes, we point out firstly the effect that base stacking interactions have on the elastic behaviour of ssDNA: the entropic force jumps 0.65 pN when stacking interactions are suppressed. This effect has already been discussed in connection with Fig. 5.5. Secondly, the difference in force between the two isomers is in all cases rather small: ∼0.1-0.3 pN, with the IsoI conformation leading to higher (or the same within error) forces than IsoII, as expected. We don't expect

‡This is when a contour length per base pair of 0.63 nm is used; with Smith et al.[259]'s original parameters, the predicted force is 4.6 pN, 1.2 pN higher than the oxDNA force.

to perfectly reproduce this difference, however, as the oxDNA model cannot fully capture the magnitude of the twist angle between the HJ arms[263].

5.4 Conclusion and outlook

We have demonstrated a method for calculating the internal entropic forces in ssDNA to high (∼0.05 pN) precision using the coarse-grained simulator oxDNA, performing much better than the ∼5 pN precision of experimental calibration methods[254]. In addition to its usefulness for computing forces in ssDNA-based force sensors, force clamp devices, and tensegrity nanostructures, the method can be expanded to calculating forces in more complex structures as outlined in section 5.2.3.

A few caveats should be noted. Firstly, the current work has assumed that origami structures were perfectly assembled – with all staples present and no topologically incorrect strand routings – despite the possibility of defects occurring experimentally which could compromise the effective stiffness of the origami or alter the length of the ssDNA spring. Furthermore, oxDNA has been parametrized to capture the $[Na^+]$ dependence of secondary structure thermodynamics. To describe experiments featuring $[Mg^{2+}]$, we relied on a mapping to an equivalent $[Na^+]$. In the current work, this was not problematic, as experiments were carried out in a high-salt regime where the predicted force had reached its limiting value, but for lower $[Mg^{2+}]$, comparison of simulation and experiment may be less straightforward. Lastly, it is possible that for high entropic forces or low ionic strengths, the ssDNA tension may lead to non-elastic yielding events within origami designs. Future origami systems for which these effects are relevant can be simulated in their entirety.

Simulating precise experimental sequences under relevant conditions is crucial for obtaining accurate estimates of internal force, as using previously-obtained fit parameters with WLC and exFJC theoretical models can lead to incorrect estimates due to differences in secondary structure, stacking propensities, salt concentration, and stiffness of the endpoint traps, as well as fundamental deficiencies in the WLC and exFJC models, such as their neglect of excluded volume. Suggested strategies to mitigate secondary structure-related force estimate errors include

constructing "G-free" origami structures (using T, C, and A nucleotides only), which reduces the accessible number of secondary structures[193], or using poly-T sequences instead – though this is more difficult as the standard DNA origami scaffold sequence lacks poly-T regions. Both of these suggestions are being pursued currently by our experimental collaborators, and ongoing oxDNA simulations will confirm whether such strategies allow for reasonable theoretical exFJC-based force estimates. Ultimately, however, we recommend that experimental groups pursue simulations to obtain a clearer idea of the internal forces in their structures. Future work furthering this end could translate the techniques we have described into an open-source, accessible user interface that takes sequence input, performs oxDNA simulations, and provides internal force estimates. In any case, the balance between realistic detail and computational speed offered by oxDNA render it a valuable tool in the ongoing study of DNA internal forces in cellular mechanotransduction and artifical DNA nanostructures.

The Road goes ever on and on
Down from the door where it began
Now far ahead the Road has gone
Let others follow it who can!

— Bilbo Baggins, in J.R.R. Tolkien's *The Return of the King* (1955).

6 Conclusions

The preceding chapters have surveyed the gamut of force-DNA interactions, from small systems to complex ones, from violent external pulling to subtle internal entropic forces. In Chapter 3, we explored the forcible unravelling of simple DNA hairpin systems. Through equilibrium simulations, we verified oxDNA's ability to capture DNA free energy landscapes by means of comparison to diverse experimental results. Next, using non-equilibrium single molecule force spectroscopy (SMFS) simulations, we validated a free energy landscape analysis framework based on the Jarzynski equality from non-equilibrium physics for use with oxDNA. Excellent agreement with equilibrium hairpin landscapes was achieved in reasonable simulation time.

Treating the errors in non-equilibrium analyses is a subtle task, and, drawing from the literature, we have outlined a practical procedure for doing so. We also illustrated the potential usefulness of non-equilibrium simulations by reconstructing hairpin unfolding landscapes in the presence of long, DNA origami linker handles; our results suggest that such handles may introduce undesirable changes to molecular landscapes via steric interactions. To improve the convergence of results in the future, the force application protocol could be optimized, likely through the use of discrete force jumps[80]; however, this would render the simulations less relevant to experiments.

A major advantage of performing SMFS simulations *in silico* is the absence of experimental apparatus: experimental pulling landscapes are convolved with the elastic responses of long, polymeric tethers, dielectric beads, and AFM cantilevers. Deconvolution of non-equilibrium experiments is difficult, and a framework for doing so for systems whose unfolding curves span a broad range of forces – RNA pseudoknots, for example[63] – has not been developed. Using simulations in conjunction with pulling experiments can thus provide a refined picture of underlying free energy landscapes. Coarse-grained simulations could also play a role in developing and testing suitable far-from-equilibrium deconvolution methods for experimentalists. A logical next step would be the validation of the methods described here with oxRNA[281], which would open a broader range of nucleic acid systems for study.

Another exciting application of future non-equilibrium simulated landscape reconstructions is the exploration of the effect of different choices of reaction coordinate on measured landscapes. In oxDNA, multiple reaction coordinates can be monitored simultaneously, granting access to multi-dimensional non-equilibrium energy landscapes. As the choice of experimentally measured reaction coordinate can mean the difference between a reliable landscape and a poorly-projected one[60], using simulations to inform which coordinate to track is of great potential benefit for experimentalists. Furthermore, simulations enable correlation of structural dynamics with energy landscape features.

Our second foray into simulated SMFS involved systems hundreds of times larger than the DNA hairpins of Chapter 3. Chapter 4 featured the unravelling of two archetypal DNA origami systems: the Rothemund tile[228] and a 10-helix bundle (10 HB)[13] – the largest systems to be studied with *in silico* SMFS to date. Simulations of the Rothemund tile were supplemented with experiments, which agreed well with simulation results. While both origami systems exhibit sawtooth-like force-extension curves (FECs), the dissimilarities in folding response we observed underscore the differences in nanostructure topology and design. The 2D Rothemund tile exhibited regular, modular unfolding of discrete subunits, corresponding to repeated rows in its

design. By exploring detailed structural snapshots and base-pair breakage data, we isolated the unfolding mechanism for the tile, revealing how rows alternately shear and unzip. The unfolding of the 3D 10 HB, by contrast, featured strain-softening – wherein the unfolding of certain key features lowered the barriers for subsequent unfolding events – as well as a richer diversity of unfolding pathways; this can be attributed to its increased topological complexity as compared to the Rothemund tile. To test the limits of our simulation capabilities, we also performed SMFS on a 7-tile biosensor, and though our pulling speeds were too fast to reproduce experimental resolution, we observed aptamer locks breaking in the order expected.

Our studies highlight the potential value of using simulated SMFS – preferably in conjunction with experiments – to inform the rational design of DNA nanostructures. This is particularly relevant for applications in which mechanical stability is important, such as in tensegrity structures[155], which would make excellent targets for future work building on this thesis. The ability to correlate unfolding behaviour with nanostructure design suggests another interesting research direction: the exploration of whether FECs can be used as a unique 'fingerprint' that can be tuned using, for instance, staple routing and sequence. Our collaborators are currently working on developing smaller, $\sim$1000 nucleotide origami systems – easier to simulate and to unfold experimentally – to explore this idea, which could potentially have applications in information storage.

The final stop in our expedition examined internal, rather than external, forces. In Chapter 5, first, we outlined deficiencies with widely-applied DNA polymer models, aided by oxDNA simulations of ssDNA. Next, we characterized the propagation of forces via oxDNA potentials, rationalizing the observed very large instantaneous force distribution widths. A method for computing internal forces in DNA structures was presented, and applied to a recently-proposed origami force clamp[193] that employs ssDNA entropic forces to study single molecules. Our simulations reveal that the polymer model employed by Nickels et al.[193] leads to incorrect estimates of exerted forces due to its failure to capture the salt concentration, secondary structures, and stacking interactions relevant to their experiments. To capture these

effects, we recommend coarse-grained simulations be used to supply more realistic entropic force estimates, relevant to a multitude of DNA nanostructures in which entropic forces play a role. Future work building on this thesis could implement, using oxDNA, a user-friendly online platform which experimentalists could use to generate entropic force estimates for their particular systems. As internal forces in oxDNA can be calculated for more complex systems as well as single strands, future studies could also explore forces in more complex, highly stressed structures, such as DNA nanocircles[271]. For example, a recent study attributed the dependence on DNA circularization efficiency with helix length to the amount of strain in the circles; oxDNA force measurements could confirm this and related effects.[130]

The diversity of DNA-force interactions featured in this study serves to underscore the versatility of oxDNA, which allowed us to access the elastic behaviour of small single strands and hairpins as well as the unfolding of immense nanostructures. No code is an island, however, and an underlying theme of this thesis – and the strength of coarse-grained modelling more generally – is that the greatest gains are possible when modelling is used to complement experimental studies. This complementarity can be expanded to other modelling strategies: for example, oxDNA estimates of internal forces could be augmented by force calculations in more realistic, atomistic simulations. We look forward eagerly to the continuing, fruitful alliance of oxDNA DNA-force simulations with experiments and other modelling modalities.

I cannot remember the books I have read any more than the meals I have eaten; even so, they have made me.

— Ralph Waldo Emerson

Bibliography

[1] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P., 2008. *Molecular Biology of the Cell*. Garland Science, 5th ed. ISBN 0815341067.

[2] Alemany, A. & Ritort, F., 2014. Determination of the elastic properties of short ssDNA molecules by mechanically folding and unfolding DNA hairpins. *Biopolymers*, **101**, 1193–1199.

[3] Allen, M. P. & Tildesley, D. J., 1991. *Computer simulation of liquids*. Oxford University Press, 2nd ed. ISBN 9780191841439.

[4] Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L. P., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V. & Kjems, J., 2009. Self-assembly of a nanoscale DNA box with a controllable lid. *Nature*, **459**, 73–6.

[5] Andersen, H. C., 1980. Molecular dynamics simulations at constant pressure and/or temperature. *J. Comput. Phys.*, **72**, 2384–2393.

[6] Andrec, M., 2010. *The Weighted Histogram Analysis Method (WHAM)*.

[7] Anfinsen, C. B., 1973. Principles that Govern the Folding of Protein Chains. *Science*, **181**, 223–230.

[8] Anfinsen, C. B., Haber, E., Sela, M. & White, F. H., 1961. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Nat. Aca. Sci. of the U.S.A.*, **47**, 1309–1314.

[9] Araque, J. C., Panagiotopoulos, A. Z. & Robert, M. A., 2011. Lattice model of oligonucleotide hybridization in solution. I. Model and thermodynamics. *The Journal of Chemical Physics*, **134**(16), 165103.

[10] Araque, J. C. & Robert, M. A., 2016. Lattice model of oligonucleotide hybridization in solution. II. Specificity and cooperativity. *The Journal of Chemical Physics*, **144**(12), 125101.

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

[12] Athènes, M. & Marinica, M.-C., 2010. Free energy reconstruction from steered dynamics without post-processing. *J. Comput. Phys.*, **229**, 7129 – 7146.

[13] Bae, W., Kim, K., Min, D., Ryu, J.-K., Hyeon, C. & Yoon, T.-Y., 2014. Programmed folding of DNA origami structures through single-molecule force control. *Nat. Commun.*, **5**, 5654.

[14] Bauer, M. S., Milles, L. F., Sedlak, S. M. & Gaub, H. E., 2018. Monomeric streptavidin: a versatile regenerative handle for force spectroscopy. *bioRxiv*.

[15] Baumann, F., Bauer, M. S., Milles, L. F., Alexandrovich, A., Gaub, H. E. & Pippig, D. A., 2015. Monovalent Strep-Tactin for strong and site-specific tethering in nanospectroscopy. *Nat. Nanotechnol.*, **11**, 89.

[16] Becker, N. B., Rosa, A. & Everaers, R., 2010. The radial distribution function of worm-like chains. *Eur. Phys. J. E*, **32**, 53–69.

[17] Belkin, M., Chao, S.-H., Jonsson, M. P., Dekker, C. & Aksimentiev, A., 2015. Plasmonic Nanopores for Trapping, Controlling Displacement, and Sequencing of DNA. *ACS Nano*, **9**(11), 10598–10611.

[18] Bell, G., 1978. Models for the specific adhesion of cells to cells. *Science*, **200**, 618–627.

[19] Bell, N. A. W., Engst, C. R., Ablay, M., Divitini, G., Ducati, C., Liedl, T. & Keyser, U. F., 2012. DNA Origami Nanopores. *Nano lett.*, **12**, 512–517.

[20] Blakely, B. L., Dumelin, C. E., Trappmann, B., McGregor, L. M., Choi, C. K., Anthony, P. C., Duesterberg, V. K., Baker, B. M., Block, S. M., Liu, D. R. & Chen, C. S., 2014. A DNA-based molecular probe for optically reporting cellular traction forces. *Nat. Methods*, **11**, 1229–1232.

[21] Bosco, A., Camunas-Soler, J. & Ritort, F., 2014. Elastic properties and secondary structure formation of single-stranded DNA at monovalent and divalent salt conditions. *Nucleic Acids Res.*, **42**, 2064–2074.

[22] Brandner, A., Schüller, A., Melo, F. & Pantano, S., 2018. Exploring DNA dynamics within oligonucleosomes with coarse-grained simulations: SIRAH force field extension for protein-DNA complexes. *Biochemical and Biophysical Research Communications*, **498**(2), 319–326.

[23] Braun, O., Hanke, A. & Seifert, U., 2004. Probing Molecular Free Energy Landscapes by Periodic Loading. *Phys. Rev. Lett.*, **93**, 158105.

[24] Brooks, B., Brooks, C., MacKerell, A., Nilsson, L., Petrella, R. et al., 2009. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, **30**, 1545–1614.

[25] Bryngelson, J. D. & Wolynes, P. G., 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, **84**, 7524–7528.

[26] Bullerjahn, J. T., Sturm, S. & Kroy, K., 2014. Theory of rapid force spectroscopy. *Nat. Commun.*, **5**, 4463.

[27] Bustamante, C., Bryant, Z. & Smith, S. B., 2003. Ten years of tension: single-molecule DNA mechanics. *Nature*, **421**, 423–427.

[28] Bustamante, C., Chemla, Y. R., Forde, N. R. & Izhaky, D., 2004. Mechanical Processes in Biochemistry. *Annu. Rev. Biochem.*, **73**, 705–748. PMID: 15189157.

[29] Camunas-Soler, J., Ribezzi-Crivellari, M. & Ritort, F., 2016. Elastic Properties of Nucleic Acids by Single-Molecule Force Spectroscopy. *Annu. Rev. Biochem.*, **45**, 65–84. PMID: 27145878.

[30] Cao, Q., Zuo, C., Ma, Y., Li, L. & Zhang, Z., 2011. Interaction of double-stranded DNA with a nanosphere: a coarse-grained molecular dynamics simulation study. *Soft Matter*, **7**(2), 506–514.

[31] Castro, C. E., Kilchherr, F., Kim, D.-N., Shiao, E. L., Wauer, T., Wortmann, P., Bathe, M. & Dietz, H., 2011. A primer to scaffolded DNA origami. *Nat. Methods*, **8**, 221.

[32] Chakraborty, D., Hori, N. & Thirumalai, D., 2018. Sequence-dependent Three Interaction Site (TIS) Model for Single and Double-stranded DNA. *ArXiv e-prints*.

[33] Chelli, R., 2012. Local Sampling in Steered Monte Carlo Simulations Decreases Dissipation and Enhances Free Energy Estimates via Nonequilibrium Work Theorems. *J. Chem. Theory Comput.*, **8**, 4040–4052.

[34] Chen, A. A. & García, A. E., 2012. Mechanism of enhanced mechanical stability of a minimal RNA kissing complex elucidated by nonequilibrium molecular dynamics simulations. *Proc. Nat. Aca. Sci.*, **109**, E1530–E1539.

[35] Cocco, S., Monasson, R. & Marko, J. F., 2001. Force and kinetic barriers to unzipping of the DNA double helix. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 8608–8613.

[36] Colizzi, F. & Bussi, G., 2012. RNA Unwinding from Reweighted Pulling Simulations. *J. Am. Chem. Soc.*, **134**, 5173–5179. PMID: 22313354.

[37] Collin, D., Ritort, F., Jarzynski, C., Smith, S. B., Tinoco Jr, I. & Bustamante, C., 2005. Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, **437**, 231.

[38] Commons, W., 2015. File:A-DNA, B-DNA and Z-DNA.png — Wikimedia Commons, the free media repository. [Online; accessed 18-April-2018].

[39] Cook, S. M., Scäffer, T. E., Chynoweth, K., Wigton, M., Simmonds, R. W. & Lang, K. M., 2006. Practical implementation of dynamic methods for measuring atomic force microscope cantilever spring constants. *Nanotech.*, **17**, 2135.

[40] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A., 1995. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

[41] Cost, A.-L., Ringer, P., Chrostek-Grashoff, A. & Grashoff, C., 2015. How to Measure Molecular Forces in Cells: A Guide to Evaluating Genetically-Encoded FRET-Based Tension Sensors. *Cell. Mol. Bioeng.*, **8**, 96–105.

[42] Cragnolini, T., Chakraborty, D., Šponer, J., Derreumaux, P., Pasquali, S. & Wales, D. J., 2017. Multifunctional energy landscape for a DNA G-quadruplex: An evolved molecular switch. *The Journal of Chemical Physics*, **147**(15), 152715.

[43] Cragnolini, T., Derreumaux, P. & Pasquali, S., 2013. Coarse-Grained Simulations of RNA and DNA Duplexes. *The Journal of Physical Chemistry B*, **117**(27), 8047–8060.

[44] Crick, F., 1970. Central Dogma of Molecular Biology. *Nature*, **227**, 561.

[45] Crooks, G. E., 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, **60**, 2721–2726.

[46] Cuendet, M. A., 2006. The Jarzynski identity derived from general Hamiltonian or non-Hamiltonian dynamics reproducing NVT or NPT ensembles. *J. Chem. Phys.*, **125**, 144109.

[47] Dannenberg, F., Dunn, K. E., Bath, J., Kwiatkowska, M., Turberfield, A. J. & Ouldridge, T. E., 2015. Modelling DNA origami self-assembly at the domain level. *J. Chem. Phys.*, **143**, 165102.

[48] Dans, P. D., Darré, L., Machado, M. R., Zeida, A., Brandner, A. F. & Pantano, S., 2013. Assessing the Accuracy of the SIRAH Force Field to Model DNA at Coarse Grain Level. In J. C. Setubal & N. F. Almeida, eds., *Advances in Bioinformatics and Computational Biology*, 71–81. Springer International Publishing, Cham. ISBN 978-3-319-02624-4.

[49] Dans, P. D., Walther, J., Gómez, H. & Orozco, M., 2016. Multiscale simulation of DNA. *Current Opinion in Structural Biology*, **37**, 29–45.

[50] Dans, P. D., Zeida, A., Machado, M. R. & Pantano, S., 2010. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *Journal of Chemical Theory and Computation*, **6**(5), 1711–1725.

[51] Dellago, C. & Hummer, G., 2014. Computing Equilibrium Free Energies Using Non-Equilibrium Molecular Dynamics. *Entropy*, **16**, 41–61.

[52] Depa, P., Chen, C. & Maranas, J. K., 2011. Why are coarse-grained force fields too fast? A look at dynamics of four coarse-grained polymers. *J. Chem. Phys.*, **134**, 014903.

[53] Dey, P. & Bhattacherjee, A., 2018. Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Scientific Reports*, **8**(1), 844.

[54] Dietz, H., Berkemeier, F., Bertz, M. & Rief, M., 2006. Anisotropic deformation response of single protein molecules. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 12724–12728.

[55] Dill, K. A. & MacCallum, J. L., 2012. The Protein-Folding Problem, 50 Years On. *Science*, **338**, 1042–1046.

[56] Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T. et al., 2012. Landscape of transcription in human cells. *Nature*, **489**, 101.

[57] Douglas, S. M., Bachelet, I. & Church, G. M., 2012. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science*, **335**, 831–834.

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

[59] Doye, J. P. K., Ouldridge, T. E., Louis, A. A., Romano, F., Šulc, P., Matek, C., Snodin, B. E. K., Rovigatti, L., Schreck, J. S., Harrison, R. M. & Smith, W. P. J., 2013. Coarse-graining DNA for simulations of DNA nanotechnology. *Phys. Chem. Chem. Phys.*, **15**, 20395–20414.

[60] Dudko, O. K., Graham, T. G. W. & Best, R. B., 2011. Locating the Barrier for Folding of Single Molecules under an External Force. *Phys. Rev. Lett.*, **107**, 208301.

[61] Dudko, O. K., Hummer, G. & Szabo, A., 2006. Intrinsic Rates and Activation Free Energies from Single-Molecule Pulling Experiments. *Phys. Rev. Lett.*, **96**, 108101.

[62] Dunn, K. E., Dannenberg, F., Ouldridge, T. E., Kwiatkowska, M., Turberfield, A. J. & Bath, J., 2015. Guiding the folding pathway of DNA origami. *Nature*, **525**, 82.

[63] Engel, M. C., Ritchie, D. B., Foster, D. A. N., Beach, K. S. D. & Woodside, M. T., 2014. Reconstructing Folding Energy Landscape Profiles from Nonequilibrium Pulling Curves with an Inverse Weierstrass Integral Transform. *Phys. Rev. Lett.*, **113**, 238104.

[64] Engel, M. C., Smith, D. M., Jobst, M. A., Sajfutdinow, M., Liedl, T., Romano, F., Rovigatti, L., Louis, A. A. & Doye, J. P. K., 2018. Force-induced unravelling of DNA Origami. *ACS Nano*, Submitted.

[65] Erlich, K. R., Baumann, F., Pippig, D. A. & Gaub, H. E., 2017. Strep-Tag II and Monovalent Strep-Tactin as Novel Handles in Single-Molecule Cut-and-Paste. *Small Methods*, **1**, 1700169. 1700169.

[66] Essevaz-Roulet, B., Bockelmann, U. & Heslot, F., 1997. Mechanical separation of the complementary strands of DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 11935–11940.

[67] Evans, E. & Ritchie, K., 1997. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, **72**, 1541–1555.

[68] Fichthorn, K. A. & Weinberg, W. H., 1991. Theoretical foundations of dynamical Monte Carlo simulations. *J. Chem. Phys.*, **95**, 1090–1096.

[69] Fonseca, P., Romano, F., Schreck, J. S., Ouldridge, T. E., Doye, J. P. K. & Louis, A. A., 2018. Multi-scale coarse-graining for the study of assembly pathways in DNA-brick self-assembly. *The Journal of Chemical Physics*, **148**(13), 134910.

[70] Forloni, G., Terreni, L., Bertani, I., Fogliarino, S., Invernizzi, R., Assini, A., Ribizzi, G., Negro, A., Calabrese, E., Volonté, M. A., Mariani, C., Franceschi, M., Tabaton, M. & Bertoli, A., 2002. Protein misfolding in Alzheimer's and Parkinson's disease: genetics and molecular mechanisms. *Neurobiology of Aging*, **23**, 957 – 976. Brain Aging: Identifying the Brakes and Accelerators.

[71] Forth, S., Deufel, C., Sheinin, M. Y., Daniels, B., Sethna, J. P. & Wang, M. D., 2008. Abrupt Buckling Transition Observed during the Plectoneme Formation of Individual DNA Molecules. *Phys. Rev. Lett.*, **100**, 148301.

[72] Freeman, G. S., Hinckley, D. M., Lequieu, J. P., Whitmer, J. K. & de Pablo, J. J., 2014. Coarse-grained modeling of DNA curvature. *The Journal of Chemical Physics*, **141**(16), 165103.

[73] Freeman, G. S., Lequieu, J. P., Hinckley, D. M., Whitmer, J. K. & de Pablo, J. J., 2014. DNA Shape Dominates Sequence Affinity in Nucleosome Formation. *Phys. Rev. Lett.*, **113**, 168101.

[74] Frenkel, D. & Smit, B., 2002. *Understanding Molecular Simulation: From Algorithms to Applications*. Elsevier Academic Press, 2nd ed. ISBN 0-12-267351-4.

[75] Fu, H., Chen, H., Koh, C. G. & Lim, C. T., 2009. Effects of magnesium salt concentrations on B-DNA overstretching transition. *Eur. Phys. J. E*, **29**, 45–49.

[76] Fu, T. J. & Seeman, N. C., 1993. DNA double-crossover molecules. *Biochemistry*, **32**, 3211–3220.

[77] Fugmann, S. & Sokolov, I. M., 2009. Scaling of the rupture dynamics of polymer chains pulled at one end at a constant rate. *Phys. Rev. E*, **79**, 021803.

[78] Funke, J. J., Ketterer, P., Lieleg, C., Schunter, S., Korber, P. & Dietz, H., 2016. Uncovering the forces between nucleosomes using DNA origami. *Sci. Adv.*, **2**, 1600974.

[79] Gebhardt, J. C. M., Bornschlögl, T. & Rief, M., 2010. Full distance-resolved folding energy landscape of one single protein molecule. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 2013–2018.

[80] Geiger, P. & Dellago, C., 2010. Optimum protocol for fast-switching free-energy calculations. *Phys. Rev. E*, **81**, 021127.

[81] Genot, A. J., Bath, J. & Turberfield, A. J., 2011. Reversible Logic Circuits Made of DNA. *J. Am. Chem. Soc.*, **133**, 20080–20083.

[82] Getfert, S. & Reimann, P., 2012. Hidden Multiple Bond Effects in Dynamic Force Spectroscopy. *Biophys. J.*, **102**, 1184–1193.

[83] Gkionis, K., Kruse, H., Platts, J. A., Mládek, A., Koča, J. & Šponer, J., 2014. Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *Journal of Chemical Theory and Computation*, **10**(3), 1326–1340.

[84] Goddard, N. L., Bonnet, G., Krichevsky, O. & Libchaber, A., 2000. Sequence Dependent Rigidity of Single Stranded DNA. *Phys. Rev. Lett.*, **85**, 2400–2403.

[85] Goktas, M. & Blank, K. G., 2017. Molecular Force Sensors: From Fundamental Concepts toward Applications in Cell Biology. *Adv. Mater. Interfaces*, **4**, 1600441.

[86] Goodman, R. P., Schaap, I. A. T., Tardin, C. F., Erben, C. M., Berry, R. M., Schmidt, C. F. & Turberfield, A. J., 2005. Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science*, **310**, 1661–1665.

[87] Goodman, S. D. & Nash, H. A., 1989. Functional replacement of a protein-induced bend in a DNA recombination site. *Nature*, **341**, 251–254.

[88] Göpfrich, K., Li, C.-Y., Mames, I., Bhamidimarri, S. P., Ricci, M., Yoo, J., Mames, A., Ohmann, A., Winterhalter, M., Stulz, E., Aksimentiev, A. & Keyser, U. F., 2016. Ion Channels Made from a Single Membrane-Spanning DNA Duplex. *Nano Letters*, **16**(7), 4665–4669.

[89] Göpfrich, K., Li, C.-Y., Ricci, M., Bhamidimarri, S. P., Yoo, J., Gyenes, B., Ohmann, A., Winterhalter, M., Aksimentiev, A. & Keyser, U. F., 2016. Large-Conductance Transmembrane Porin Made from DNA Origami. *ACS Nano*, **10**(9), 8207–8214.

[90] Gore, J., Ritort, F. & Bustamante, C., 2003. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 12564–12569.

[91] Grashoff, C., Hoffman, B. D., Brenner, M. D., Zhou, R., Parsons, M., Yang, M. T., McLean, M. A., Sligar, S. G., Chen, C. S., Ha, T. & Schwartz, M. A., 2010. Measuring mechanical tension across vinculin reveals regulation of focal adhesion dynamics. *Nature*, **466**, 263.

[92] Greenleaf, W. J., Woodside, M. T. & Block, S. M., 2007. High-Resolution, Single-Molecule Measurements of Biomolecular Motion. *Annu. Rev. Biophys. Biomol. Struct.*, **36**, 171–190.

[93] Gross, P., Laurens, N., Oddershede, L. B., Bockelmann, U., Peterman, E. J. G. & Wuite, G. J. L., 2011. Quantifying how DNA stretches, melts and changes twist under tension. *Nat. Phys.*, **7**, 731–736.

[94] Grubmüller, H., Heymann, B. & Tavan, P., 1996. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science*, **271**, 997–999.

[95] Gu, H., Yang, W. & Seeman, N. C., 2010. DNA Scissors Device Used to Measure MutS Binding to DNA Mis-pairs. *J. Am. Chem. Soc.*, **132**, 4352–4357. PMID: 20205420.

[96] Gupta, A. N., Vincent, A., Neupane, K., Yu, H., Wang, F. & Woodside, M. T., 2011. Experimental validation of free-energy-landscape reconstruction from non-equilibrium single-molecule force spectroscopy measurements. *Nat. Phys.*, **7**, 631.

[97] Harris, N. C. & Kiang, C.-H., 2009. Velocity convergence of free energy surfaces from single-molecule measurements using Jarzynski's equality. *Phys. Rev. E*, **79**, 041912.

[98] Harris, N. C., Song, Y. & Kiang, C.-H., 2007. Experimental Free Energy Surface Reconstruction from Single-Molecule Force Spectroscopy using Jarzynski's Equality. *Phys. Rev. Lett.*, **99**, 068101.

[99] Harris, S. A., Sands, Z. A. & Laughton, C. A., 2005. Molecular Dynamics Simulations of Duplex Stretching Reveal the Importance of Entropy in Determining the Biomechanical Properties of DNA. *Biophys. J.*, **88**, 1684–1691.

[100] Hatch, K., Danilowicz, C., Coljee, V. & Prentiss, M., 2008. Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs. *Phys. Rev. E*, **78**, 011920.

[101] He, C., Genchev, G. Z., Lu, H. & Li, H., 2012. Mechanically Untying a Protein Slipknot: Multiple Pathways Revealed by Force Spectroscopy and Steered Molecular Dynamics Simulations. *J. Am. Chem. Soc.*, **134**, 10428–10435.

[102] He, Y., Maciejczyk, M., Ołdziej, S., Scheraga, H. A. & Liwo, A., 2013. Mean-Field Interactions between Nucleic-Acid-Base Dipoles can Drive the Formation of a Double Helix. *Phys. Rev. Lett.*, **110**, 098101.

[103] He, Y., Ye, T., Su, M., Zhang, C., Ribbe, A. E., Jiang, W. & Mao, C., 2008. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature*, **452**, 198.

[104] Henrich, O., Gutiérrez Fosado, Y. A., Curk, T. & Ouldridge, T. E., 2018. Coarse-grained simulation of DNA using LAMMPS. *The European Physical Journal E*, **41**(5), 57.

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

[106] Hinckley, D. M., Lequieu, J. P. & de Pablo, J. J., 2014. Coarse-grained modeling of DNA oligomer hybridization: Length, sequence, and salt effects. *The Journal of Chemical Physics*, **141**(3), 035102.

[107] Hinczewski, M., Gebhardt, J. C. M., Rief, M. & Thirumalai, D., 2013. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 4500–4505.

[108] Hohng, S., Zhou, R., Nahas, M. K., Yu, J., Schulten, K., Lilley, D. M. J. & Ha, T., 2007. Fluorescence-Force Spectroscopy Maps Two-Dimensional Reaction Landscape of the Holliday Junction. *Science*, **318**, 279–283.

[109] Huang, P.-S., Boyken, S. E. & Baker, D., 2016. The coming of age of de novo protein design. *Nature*, **537**, 320.

[110] Huguet, J. M., Bizarro, C. V., Forns, N., Smith, S. B., Bustamante, C. & Ritort, F., 2010. Single-molecule derivation of salt dependent base-pair free energies in DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 15431–15436.

[111] Hummer, G., 2001. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.*, **114**, 7330–7337.

[112] Hummer, G. & Szabo, A., 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 3658–3661.

[113] Hummer, G. & Szabo, A., 2003. Kinetics from Nonequilibrium Single-Molecule Pulling Experiments. *Biophys. J.*, **85**, 5–15.

[114] Hummer, G. & Szabo, A., 2005. Free Energy Surfaces from Single-Molecule Force Spectroscopy. *Acc. Chem. Res.*, **38**, 504–513.

[115] Hummer, G. & Szabo, A., 2010. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 21441–21446.

[116] Hutter, J. L. & Bechhoefer, J., 1993. Calibration of atomic force microscope tips. *Rev. Sci. Instrum.*, **64**, 1868–1873.

[117] Imparato, A., Luccioli, S. & Torcini, A., 2007. Reconstructing the Free-Energy Landscape of a Mechanically Unfolded Model Protein. *Phys. Rev. Lett.*, **99**, 168101.

[118] Imparato, A., Pelizzola, A. & Zamparo, M., 2007. Ising-Like Model for Protein Mechanical Unfolding. *Phys. Rev. Lett.*, **98**, 148102.

[119] Ingber, D. E., 2006. Cellular mechanotransduction: putting all the pieces together again. *The FASEB Journal*, **20**, 811–827.

[120] Isralewitz, B., Gao, M. & Schulten, K., 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, **11**, 224 – 230.

[121] Iwaki, M., Wickham, S. F., Ikezaki, K., Yanagida, T. & Shih, W. M., 2016. A programmable DNA origami nanospring that reveals force-induced adjacent binding of myosin VI heads. *Nat. Commun.*, **7**, 13715.

[122] Izrailev, S., Stepaniants, S., Balsera, M., Oono, Y. & Schulten, K., 1997. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, **72**, 1568–1581.

[123] Jarzynski, C., 1997. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, **78**, 2690–2693.

[124] Jarzynski, C., 2006. Rare events and the convergence of exponentially averaged work values. *Phys. Rev. E*, **73**, 046105.

[125] Jensen, M. Ø., Park, S., Tajkhorshid, E. & Schulten, K., 2002. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 6731–6736.

[126] Jiří, Š., E., Š. J., Arnošt, M., Pavel, B., Petr, J. & Otyepka, M., 2013. How to understand quantum chemical computations on DNA and RNA systems? A practical guide for non-specialists. *Methods*, **64**(1), 3 – 11. Nucleic Acid Structure.

[127] Jiří, Š., E., Š. J., Arnošt, M., Petr, J., Pavel, B. & Michal, O., 2013. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers*, **99**(12), 978–988.

[128] Jobst, M. A., Milles, L. F., Schoeler, C., Ott, W., Fried, D. B., Bayer, E. A., Gaub, H. E. & Nash, M. A., 2015. Resolving dual binding conformations of cellulosome cohesin-dockerin complexes using single-molecule force spectroscopy. *eLIFE*, **4**, 180–183.

[129] Jobst, M. A., Schoeler, C., Malinowska, K. & Nash, M. A., 2013. Investigating Receptor-ligand Systems of the Cellulosome with AFM-based Single-molecule Force Spectroscopy. *J. Visualized Exp.*, 50950.

[130] Joffroy, B., Uca, Y. O., Prešern, D., Doye, J. P. K. & Schmidt, T. L., 2018. Rolling circle amplification shows a sinusoidal template length-dependent amplification bias. *Nucleic Acids Res.*, **46**, 538–545.

[131] Jonathan, B., J., G. S. & J., T. A., 2005. A Free-Running DNA Motor Powered by a Nicking Enzyme. *Angew. Chem., Int. Ed.*, **44**, 4358–4361.

[132] Jones, M. R., Seeman, N. C. & Mirkin, C. A., 2015. Programmable materials and the nature of the DNA bond. *Science*, **347**, 1260901.

[133] Kauert, D. J., Kurth, T., Liedl, T. & Seidel, R., 2011. Direct Mechanical Measurements Reveal the Material Properties of Three-Dimensional DNA Origami. *Nano lett.*, **11**, 5558–5563.

[134] Ke, Y., Lindsay, S., Chang, Y., Liu, Y. & Yan, H., 2008. Self-Assembled Water-Soluble Nucleic Acid Probe Tiles for Label-Free RNA Hybridization Assays. *Science*, **319**, 180–183.

[135] Ke, Y., Ong, L. L., Shih, W. M. & Yin, P., 2012. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science*, **338**, 1177–1183.

[136] Khara, D. C., Schreck, J. S., Tomov, T. E., Berger, Y., Ouldridge, T. E., Doye, J. P. K. & Nir, E., 2018. DNA bipedal motor walking dynamics: an experimental and theoretical study of the dependency on step size. *Nucleic Acids Res.*, **46**, 1553–1561.

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

[138] Klaus, M., Prokoph, N., Girbig, M., Wang, X., Huang, Y.-H., Srivastava, Y., Hou, L., Narasimhan, K., Kolatkar, P. R., Francois, M. & Jauch, R., 2016. Structure and decoy-mediated inhibition of the SOX18/Prox1-DNA interaction. *Nucleic Acids Research*, **44**(8), 3922–3935.

[139] Knotts, T. A., Rathore, N., Schwartz, D. C. & de Pablo, J. J., 2007. A coarse grain model for DNA. *The Journal of Chemical Physics*, **126**(8), 084901.

[140] Kočar, V., Schreck, J. S., Čeru, S., Gradišar, H., Bašić, N., Pisanski, T., Doye, J. P. K. & Jerala, R., 2016. Design principles for rapid folding of knotted DNA nanostructures. *Nature Communications*, **7**, 10803 EP –.

[141] Koirala, D., Shrestha, P., Emura, T., Hidaka, K., Mandal, S., Endo, M., Sugiyama, H. & Mao, H., 2014. Single-Molecule Mechanochemical Sensing Using DNA Origami Nanostructures. *Angew. Chem., Int. Ed.*, **126**, 8275–8279.

[142] Korolev, N., Luo, D., Lyubartsev, A. P. & Nordenskiöld, L., 2014. A Coarse-Grained DNA Model Parameterized from Atomistic Simulations by Inverse Monte Carlo. *Polymers*, **6**, 1655–1675.

[143] Korolev, N., Nordenskiöld, L. & Lyubartsev, A. P., 2016. Multiscale coarse-grained modelling of chromatin components: DNA and the nucleosome. *Advances in Colloid and Interface Science*, **232**, 36–48.

[144] Kramers, H., 1940. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, **7**, 284 – 304.

[145] Kratky, O. & Porod, G., 1949. Röntgenuntersuchung gelöster Fadenmoleküle. *Rec. Trav. Chim. Pays-Bas.*, **68**, 1106–1123.

[146] Kühner, F., Morfill, J., Neher, R. A., Blank, K. & Gaub, H. E., 2007. Force-Induced DNA Slippage. *Biophys. J.*, **92**, 2491 – 2497.

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

[148] Laughton, C. A. & Harris, S. A., 2011. The atomistic simulation of DNA. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**, 590–600.

[149] Laurence, T. A., Kong, X., Jäger, M. & Weiss, S., 2005. Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 17348–17353.

[150] Lechner, W. & Dellago, C., 2007. On the efficiency of path sampling methods for the calculation of free energies from non-equilibrium simulations. *J. Stat. Mech. - T. and E.*, **2007**, P04001.

[151] Lequieu, J., Córdoba, A., Schwartz, D. C. & de Pablo, J. J., 2016. Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Central Science*, **2**(9), 660–666.

[152] Lequieu, J. P., Hinckley, D. M. & de Pablo, J. J., 2015. A molecular view of DNA-conjugated nanoparticle association energies. *Soft Matter*, **11**(10), 1919–1929.

[153] Levinthal, C., 1969. How to Fold Graciously. In J. T. P. DeBrunner & E. Munck, eds., *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois.*, 22–24. University of Illinois Press.

[154] Li, C.-Y., Hemmig, E. A., Kong, J., Yoo, J., Hernández-Ainsa, S., Keyser, U. F. & Aksimentiev, A., 2015. Ionic Conductivity, Structural Deformation, and Programmable Anisotropy of DNA Origami in Electric Field. *ACS Nano*, **9**(2), 1420–1433.

[155] Liedl, T., Hogberg, B., Tytell, J., Ingber, D. E. & Shih, W. M., 2010. Self-assembly of three-dimensional prestressed tensegrity structures from DNA. *Nat. Nanotechnol.*, **5**, 520–524.

[156] Linak, M. C., Tourdot, R. & Dorfman, K. D., 2011. Moving beyond Watson–Crick models of coarse grained DNA dynamics. *The Journal of Chemical Physics*, **135**(20), 205102.

[157] Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. & Bustamante, C., 2002. Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski's Equality. *Science*, **296**, 1832–1835.

[158] Liu, D., Wang, M., Deng, Z., Walulu, R. & Mao, C., 2004. Tensegrity: Construction of Rigid DNA Triangles with Flexible Four-Arm DNA Junctions. *J. Am. Chem. Soc.*, **126**, 2324–2325.

[159] Lo, P. K., Karam, P., Aldaye, F. A., McLaughlin, C. K., Hamblin, G. D., Cosa, G. & Sleiman, H. F., 2010. Loading and selective release of cargo in DNA nanotubes with longitudinal variation. *Nat. Chem.*, **2**, 319.

[160] Louis, A. A., 2002. Beware of density dependent pair potentials. *J. Phys.: Condens. Matter*, **14**, 9187.

[161] Machado, M. R. & Pantano, S., 2015. Exploring LacI–DNA Dynamics by Multiscale Simulations Using the SIRAH Force Field. *Journal of Chemical Theory and Computation*, **11**(10), 5012–5023.

[162] Maffeo, C. & Aksimentiev, A., 2017. Molecular mechanism of DNA association with single-stranded DNA binding protein. *Nucleic Acids Res.*, **45**, 12125–12139.

[163] Maffeo, C., Ngo, T. T. M., Ha, T. & Aksimentiev, A., 2014. A Coarse-Grained Model of Unstructured Single-Stranded DNA Derived from Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.*, **10**, 2891–2896.

[164] Maffeo, C., Yoo, J., Comer, J., Wells, D. B., Luan, B. & Aksimentiev, A., 2014. Close encounters with DNA. *J. Phys.: Condens. Matter*, **26**, 413101.

[165] Manuel, A. P., Lambert, J. & Woodside, M. T., 2015. Reconstructing folding energy landscapes from splitting probability analysis of single-molecule trajectories. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 7183–7188.

[166] Markegard, C. B., Fu, I. W., Reddy, K. A. & Nguyen, H. D., 2015. Coarse-Grained Simulation Study of Sequence Effects on DNA Hybridization in a Concentrated Environment. *The Journal of Physical Chemistry B*, **119**(5), 1823–1834.

[167] Markegard, C. B., Gallivan, C. P., Cheng, D. D. & Nguyen, H. D., 2016. Effects of Concentration and Temperature on DNA Hybridization by Two Closely Related Sequences via Large-Scale Coarse-Grained Simulations. *The Journal of Physical Chemistry B*, **120**(32), 7795–7806.

[168] Marko, J. F., 1997. Stretching must twist DNA. *Europhys. Lett.*, **38**, 183.

[169] Marko, J. F. & Siggia, E. D., 1995. Stretching DNA. *Macromolecules*, **28**, 8759–8770.

[170] Marras, A. E., Zhou, L., Su, H.-J. & Castro, C. E., 2015. Programmable motion of DNA origami mechanisms. *Proc. Natl. Acad. Sci. U. S. A.*, **112**, 713–718.

[171] Martin, T. G. & Dietz, H., 2012. Magnesium-free self-assembly of multilayer DNA objects. *Nat. Commun.*, **3**, 1103.

[172] Matek, C., Ouldridge, T., Levy, A., Doye, J. & Louis, A., 2012. DNA cruciform arms nucleate through a correlated but asynchronous cooperative mechanism. *J. Phys. Chem. B*, **116**, 11616–11625.

[173] Matek, C., Ouldridge, T. E., Doye, J. P. K. & Louis, A. A., 2015. Plectoneme tip bubbles: Coupled denaturation and writhing in supercoiled DNA. *Sci. Rep.*, **5**, 7655.

[174] Mathieu, F., Liao, S., Kopatsch, J., Wang, T., Mao, C. & Seeman, N. C., 2005. Six-Helix Bundles Designed from DNA. *Nano lett.*, **5**, 661–665.

[175] McIntosh, D. B., Ribeck, N. & Saleh, O. A., 2009. Detailed scaling analysis of low-force polyelectrolyte elasticity. *Phys. Rev. E*, **80**, 041803.

[176] McKinney, S. A., Joo, C. & Ha, T., 2006. Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys. J.*, **91**, 1941–1951.

[177] Meng, F. & Sachs, F., 2011. Visualizing dynamic cytoplasmic forces with a compliance-matched FRET sensor. *J. Cell Sci.*, **124**, 261–269.

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

[179] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E., 1953. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, **21**, 1087–1092.

[180] Mishra, G., Prešern, D., Tortora, M., Cao, Y. & Doye, J. P. K., In preparation. Elastic mechanical properties of rod-like DNA mechanical properties.

[181] Mitchell, J. S., Laughton, C. A. & Harris, S. A., 2011. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.*, **39**, 3928–3938.

[182] Mitternacht, S., Luccioli, S., Torcini, A., Imparato, A. & Irbäck, A., 2009. Changing the Mechanical Unfolding Pathway of FnIII10 by Tuning the Pulling Strength. *Biophys. J.*, **96**, 429 – 441.

[183] Mládek, A., Krepl, M., Svozil, D., Čech, P., Otyepka, M., Banáš, P., Zgarbová, M., Jurecká, P. & Šponer, J., 2013. 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.

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

[185] Mosayebi, M., Louis, A. A., Doye, J. P. K. & Ouldridge, T. E., 2015. Force-Induced Rupture of a DNA Duplex: From Fundamentals to Force Sensors. *ACS Nano*, **9**, 11993–12003.

[186] Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M. & Ha, T., 2004. Probing Single-Stranded DNA Conformational Flexibility Using Fluorescence Spectroscopy. *Biophys. J.*, **86**, 2530–2537.

[187] Naômé, A., Laaksonen, A. & Vercauteren, D. P., 2014. A Solvent-Mediated Coarse-Grained Model of DNA Derived with the Systematic Newton Inversion Method. *Journal of Chemical Theory and Computation*, **10**(8), 3541–3549.

[188] Naômé, A., Laaksonen, A. & Vercauteren, D. P., 2015. A Coarse-Grained Simulation Study of the Structures, Energetics, and Dynamics of Linear and Circular DNA with Its Ions. *Journal of Chemical Theory and Computation*, **11**(6), 2813–2826.

[189] Neher, R. A. & Gerland, U., 2005. DNA as a Programmable Viscoelastic Nanoelement. *Biophys. J.*, **89**, 3846–3855.

[190] Neuert, G., Albrecht, C., Pamir, E. & Gaub, H., 2006. Dynamic force spectroscopy of the digoxigenin antibody complex. *FEBS Lett.*, **580**, 505–509.

[191] Neukirch, S. & Marko, J. F., 2011. Analytical Description of Extension, Torque, and Supercoiling Radius of a Stretched Twisted DNA. *Phys. Rev. Lett.*, **106**, 138104.

[192] Neupane, K., Wang, F. & Woodside, M. T., 2017. Direct measurement of sequence-dependent transition path times and conformational diffusion in DNA duplex formation. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 1329–1334.

[193] Nickels, P. C., Wünsch, B., Holzmeister, P., Bae, W., Kneer, L. M., Grohmann, D., Tinnefeld, P. & Liedl, T., 2016. Molecular force spectroscopy with a DNA origami–based nanoscopic force clamp. *Science*, **354**, 305–307.

[194] Nomidis, S. K., Kriegel, F., Vanderlinden, W., Lipfert, J. & Carlon, E., 2017. Twist-Bend Coupling and the Torsional Response of Double-Stranded DNA. *Phys. Rev. Lett.*, **118**, 217801.

[195] Oberhofer, H. & Dellago, C., 2008. Optimum bias for fast-switching free energy calculations. *Comput. Phys. Commun.*, **179**, 41 – 45. Special issue based on the Conference on Computational Physics 2007.

[196] Oberhofer, H., Dellago, C. & Geissler, P. L., 2005. Biased Sampling of Nonequilibrium Trajectories: Can Fast Switching Simulations Outperform Conventional Free Energy Calculation Methods? *J. Phys. Chem. B*, **109**, 6902–6915. PMID: 16851777.

[197] Odijk, T., 1995. Stiff Chains and Filaments under Tension. *Macromolecules*, **28**(20), 7016–7018.

[198] Onuchic, J. N., Luthey-Schulten, Z., & Wolynes, P. G., 1997. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.*, **48**, 545–600. PMID: 9348663.

[199] Ott, W., Jobst, M. A., Bauer, M. S., Durner, E., Milles, L. F., Nash, M. A. & Gaub, H. E., 2017. Elastin-like Polypeptide Linkers for Single-Molecule Force Spectroscopy. *ACS Nano*, **11**, 6346–6354.

[200] Otten, M., Ott, W., Jobst, M. A., Milles, L. F., Verdorfer, T., Pippig, D. A., Nash, M. A. & Gaub, H. E., 2014. From genes to protein mechanics on a chip. *Nat. Methods*, **11**, 1127.

[201] Ouldridge, T. E., 2011. *Coarse-grained modelling of DNA and DNA self-assembly*. Ph.D. thesis, University of Oxford, UK.

[202] Ouldridge, T. E., Hoare, R. L., Louis, A. A., Doye, J. P. K., Bath, J. & Turberfield, A. J., 2013. Optimizing DNA Nanotechnology through Coarse-Grained Modeling: A Two-Footed DNA Walker. *ACS Nano*, **7**, 2479–2490.

[203] Ouldridge, T. E., Louis, A. A. & Doye, J. P. K., 2011. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Comput. Phys.*, **134**, 085101.

[204] Paci, E. & Karplus, M., 1999. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J. Mol. Bio.*, **288**, 441 – 459.

[205] Padding, J. T. & Louis, A. A., 2006. Hydrodynamic interactions and Brownian forces in colloidal suspensions: Coarse-graining over time and length scales. *Phys. Rev. E*, **74**, 031402.

[206] Palassini, M. & Ritort, F., 2011. Improving Free-Energy Estimates from Unidirectional Work Measurements: Theory and Experiment. *Phys. Rev. Lett.*, **107**, 060601.

[207] Pan, K., Bricker, W. P., Ratanalert, S. & Bathe, M., 2017. Structure and conformational dynamics of scaffolded DNA origami nanoparticles. *Nucleic Acids Research*, **45**(11), 6284–6298.

[208] Pan, K., Kim, D.-N., Zhang, F., Adendorff, M. R., Yan, H. & Bathe, M., 2014. Lattice-free prediction of three-dimensional structure of programmed DNA assemblies. *Nature Communications*, **5**, 5578.

[209] Pfitzner, E., Wachauf, C., Kilchherr, F., Pelz, B., Shih, W. M., Rief, M. & Dietz, H., 2013. Rigid DNA Beams for

High-Resolution Single-Molecule Mechanics. *Angew. Chem., Int. Ed.*, **52**, 7766–7771.

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

[211] Pippig, D. A., Baumann, F., Strackharn, M., Aschenbrenner, D. & Gaub, H. E., 2014. Protein-DNA Chimeras for Nano Assembly. *ACS Nano*, **8**, 6551–6555.

[212] Plotkin, S. S. & Onuchic, J. N., 2002. Understanding protein folding with energy landscape theory Part I: Basic concepts. *Q. Rev. Biophys.*, **35**, 111–167.

[213] Powner, M. W., Gerland, B. & Sutherland, J. D., 2009. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, 239.

[214] Puchner, E. M. & Gaub, H. E., 2012. Single-Molecule Mechanoenzymatics. *Annu. Rev. Biophys.*, **41**, 497–518.

[215] Pud, S., Chao, S.-H., Belkin, M., Verschueren, D., Huijben, T., van Engelenburg, C., Dekker, C. & Aksimentiev, A., 2016. Mechanical Trapping of DNA in a Double-Nanopore System. *Nano Letters*, **16**(12), 8021–8028. PMID: 27960493.

[216] Randall, G. L., Zechiedrich, L. & Pettitt, B. M., 2009. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res.*, **37**, 5568–5577.

[217] Ray, C., Brown, J. R. & Akhremitchev, B. B., 2007. Correction of Systematic Errors in Single-Molecule Force Spectroscopy with Polymeric Tethers by Atomic Force Microscopy. *J. Phys. Chem. B*, **111**, 1963–1974.

[218] Reich, D., 2018. *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past*. Oxford University Press, 1st ed. ISBN 9780198821250.

[219] Reinhardt, A. & Frenkel, D., 2014. Numerical Evidence for Nucleated Self-Assembly of DNA Brick Structures. *Phys. Rev. Lett.*, **112**, 238103.

[220] Reinhardt, A. & Frenkel, D., 2016. DNA brick self-assembly with an off-lattice potential. *Soft Matter*, **12**(29), 6253–6260.

[221] Reshetnikov, R. V., Stolyarova, A. V., Zalevsky, A. O., Panteleev, D. Y., Pavlova, G. V., Klinov, D. V., Golovin, A. V. & Protopopova, A. D., 2018. A coarse-grained model for DNA origami. *Nucleic Acids Res.*, **46**, 1102–1112.

[222] Rief, M., Clausen-Schaumann, H. & Gaub, H. E., 1999. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Mol. Biol.*, **6**, 346–349.

[223] Ritchie, D. B., Foster, D. A. N. & Woodside, M. T., 2012. Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 16167–16172.

[224] Rohatgi, V. K. & Ehsanes Saleh, A. K. M., 2001. *An Introduction to Probability and Statistics*. John Wiley & Sons, Inc, 2nd ed. ISBN 978-0-471-34846-7.

[225] Rohwer, C. M., Angeletti, F. & Touchette, H., 2015. Convergence of large-deviation estimators. *Phys. Rev. E*, **92**, 052104.

[226] Romano, F., Chakraborty, D., Doye, J. P. K., Ouldridge, T. E. & Louis, A. A., 2013. Coarse-grained simulations of DNA overstretching. *J. Chem. Phys.*, **138**, 085101.

[227] Romeo, M., Da Costa, V. & Bardou, F., 2003. Broad distribution effects in sums of lognormal random variables. *Euro. Phys. J. B*, **32**, 513–525.

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

[229] Rubinstein, M. & Colby, R., 2003. *Polymer Physics*. OUP Oxford. ISBN 9780198520597.

[230] Russo, J., Tartaglia, P. & Sciortino, F., 2009. Reversible gels of patchy particles: Role of the valence. *J. Chem. Phys.*, **131**, 014504.

[231] Sa-Ardyen, P., Vologodskii, A. V. & Seeman, N. C., 2003. The Flexibility of DNA Double Crossover Molecules. *Biophys. J.*, **84**, 3829–3837.

[232] Sajfutdinow, M., Uhlig, K., Prager, A., Schneider, C., Abel, B. & Smith, D. M., 2017. Nanoscale patterning of self-assembled monolayer (SAM)-functionalised substrates with single molecule contact printing. *Nanoscale*, **9**, 15098–15106.

[233] Sambriski, E., Schwartz, D. & de Pablo, J., 2009. A Mesoscale Model of DNA and Its Renaturation. *Biophys. J.*, **96**, 1675–1690.

[234] Sancho, D. D. & Best, R. B., 2017. Complex Dynamics in Single Molecule Force Spectroscopy from Simple Simulation Models. *Biophys. J.*, **112**, 196a.

[235] SantaLucia, J. & Hicks, D., 2004. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 415–440.

[236] Savelyev, A., 2012. Do monovalent mobile ions affect DNA's flexibility at high salt content? *Physical Chemistry Chemical Physics*, **14**(7), 2250–2254.

[237] Savelyev, A. & Papoian, G. A., 2010. Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 20340–20345.

[238] Scargle, J. D., 1982. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.*, **263**, 835–853.

[239] Schiffels, D., Liedl, T. & Fygenson, D. K., 2013. Nanoscale Structure and Microscale Stiffness of DNA Nanotubes. *ACS Nano*, **7**, 6700–6710.

[240] Schmiedl, T. & Seifert, U., 2007. Optimal Finite-Time Processes In Stochastic Thermodynamics. *Phys. Rev. Lett.*, **98**, 108301.

[241] Schoeler, C., Verdorfer, T., Gaub, H. E. & Nash, M. A., 2016. Biasing effects of receptor-ligand complexes on protein-unfolding statistics. *Phys. Rev. E*, **94**, 042412.

[242] Schöneweiß, E.-C. & Sacca, B., 2017. The collective behaviour of spring-like motifs tethered to a DNA origami nanostructure. *Nanoscale*, **9**, 4486–4496.

[243] Schreck, J. S., Romano, F., Zimmer, M. H., Louis, A. A. & Doye, J. P. K., 2016. Characterizing DNA Star-Tile-Based Nanostructures Using a Coarse-Grained Model. *ACS Nano*, **10**, 4236–4247.

[244] Schuldt, C., Schnauß, J., Händler, T., Glaser, M., Lorenz, J., Golde, T., Käs, J. A. & Smith, D. M., 2016. Tuning Synthetic Semiflexible Networks by Bending Stiffness. *Phys. Rev. Lett.*, **117**, 197801.

[245] Schuss, Z., 1980. *Theory and Applications of Stochastic Processes*, vol. 170. Springer-Verlag New York, 1st ed. ISBN 978-1-4419-1604-4.

[246] Sedeh, R. S., Pan, K., Adendorff, M. R., Hallatschek, O., Bathe, K.-J. & Bathe, M., 2016. Computing Nonequilibrium Conformational Dynamics of Structured Nucleic Acid Assemblies. *Journal of Chemical Theory and Computation*, **12**(1), 261–273.

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

[248] Seeman, N. C., 2010. Nanomaterials Based on DNA. *Annu. Rev. Biochem.*, **79**, 65–87.

[249] Seeman, N. C. & Sleiman, H. F., 2017. DNA nanotechnology. *Nature Rev. Mater.*, **3**, 17068.

[250] Severin, P. M. D., Zou, X., Gaub, H. E. & Schulten, K., 2011. Cytosine methylation alters DNA mechanical properties. *Nucleic Acids Res.*, **39**, 8740–8751.

[251] Sharma, R., Schreck, J. S., Romano, F., Louis, A. A. & Doye, J. P. K., 2017. Characterizing the Motion of Jointed DNA Nanostructures Using a Coarse-Grained Model. *ACS Nano*, **11**, 12426–12435.

[252] Shi, Z., Castro, C. E. & Arya, G., 2017. Conformational Dynamics of Mechanically Compliant DNA Nanostructures from Coarse-Grained Molecular Dynamics Simulations. *ACS Nano*, **11**(5), 4617–4630.

[253] Shrestha, P., Emura, T., Koirala, D., Cui, Y., Hidaka, K., Maximuck, W. J., Endo, M., Sugiyama, H. & Mao, H., 2016. Mechanical properties of DNA origami nanoassemblies are determined by Holliday junction mechanophores. *Nucleic Acids Res.*, **44**, 6574–6582.

[254] Shroff, H., Reinhard, B. M., Siu, M., Agarwal, H., Spakowitz, A. & Liphardt, J., 2005. Biocompatible Force Sensor with Optical Readout and Dimensions of 6 nm^3. *Nano lett.*, **5**, 1509–1514. PMID: 16178266.

[255] Simmel, S. S., Nickels, P. C. & Liedl, T., 2014. Wireframe and Tensegrity DNA Nanostructures. *Acc. Chem. Res.*, **47**, 1691–1699.

[256] Slone, S. M., Li, C.-Y., Yoo, J. & Aksimentiev, A., 2016. Molecular mechanics of DNA bricks: in situ structure, mechanical properties and ionic conductivity. *New Journal of Physics*, **18**(5), 055012.

[257] Slone, S. M., Li, C.-Y., Yoo, J. & Aksimentiev, A., 2016. Molecular mechanics of DNA bricks: in situ structure, mechanical properties and ionic conductivity. *New J. Phys.*, **18**, 055012.

[258] Smith, S., Finzi, L. & Bustamante, C., 1992. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, **258**, 1122–1126.

[259] Smith, S. B., Cui, Y. & Bustamante, C., 1996. Overstretching B-DNA: the elastic response of individual double-stranded and sinle-stranded DNA molecules. *Science*, **271**, 795–799.

[260] Snodin, B. E. K., 2016. *Simulating large DNA nanostructures with a coarse-grained model*. Ph.D. thesis, University of Oxford, UK.

[261] Snodin, B. E. K., Randisi, F., Mosayebi, M., Šulc, P., Schreck, J. S., Romano, F., Ouldridge, T. E., Tsukanov, R., Nir, E., Louis, A. A. & Doye, J. P. K., 2015. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.*, **142**, 234901.

[262] Snodin, B. E. K., Romano, F., Rovigatti, L., Ouldridge, T. E., Louis, A. A. & Doye, J. P. K., 2016. Direct Simulation of the Self-Assembly of a Small DNA Origami. *ACS Nano*, **10**, 1724–1737.

[263] Snodin, B. E. K., Schreck, J., Romano, F., Louis, A. A. & Doye, J. P. K., In preparation. Coarse-grained modelling of DNA origami.

[264] Sotomayor, M. & Schulten, K., 2007. Single-Molecule Experiments in Vitro and in Silico. *Science*, **316**, 1144–1148.

[265] Srinivas, N., Ouldridge, T. E., Šulc, P., Schaeffer, J. M., Yurke, B., Louis, A. A., Doye, J. P. K. & Winfree, E., 2013. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.*, **41**, 10641–10658.

[266] Stahl, E., Martin, T. G., Praetorius, F. & Dietz, H., 2014. Facile and Scalable Preparation of Pure and Dense DNA Origami Solutions. *Angew. Chem., Int. Ed.*, **53**, 12735–12740.

[267] Strackharn, M., Stahl, S. W., Puchner, E. M. & Gaub, H. E., 2012. Functional Assembly of Aptamer Binding Sites by Single-Molecule Cut-and-Paste. *Nano lett.*, **12**, 2425–2428.

[268] Strunz, T., Oroszlan, K., Schäfer, R. & Güntherodt, H.-J., 1999. Dynamic force spectroscopy of single DNA molecules. *Proc. Natl. Acad. Sci. U. S. A.*, **96**, 11277–11282.

[269] Šulc, P., Ouldridge, T. E., Romano, F., Doye, J. P. K. & Louis, A. A., 2014. Simulating a burnt-bridges DNA motor with a coarse-grained DNA model. *Natural Computing*, **13**(4), 535–547.

[270] Sun, W., Boulais, E., Hakobyan, Y., Wang, W. L., Guan, A., Bathe, M. & Yin, P., 2014. Casting inorganic structures with DNA molds. *Science*, **346**(6210), 1258361.

[271] Sutthibutpong, T., Matek, C., Benham, C., Slade, G. G., Noy, A., Laughton, C., K. Doye, J. P., Louis, A. A. & Harris, S. A., 2016. Long-range correlations in the mechanics of small DNA circles under topological stress revealed by multi-scale

simulation. *Nucleic Acids Res.*, **44**, 9121–9130.

[272] Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R., 1982. 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.

[273] Tay, C. Y., Yuan, L. & Leong, D. T., 2015. Nature-Inspired DNA Nanosensor for Real-Time in Situ Detection of mRNA in Living Cells. *ACS Nano*, **9**, 5609–5617.

[274] Then, H. & Engel, A., 2008. Computing the optimal protocol for finite-time processes in stochastic thermodynamics. *Phys. Rev. E*, **77**, 041105.

[275] Torrie, G. & Valleau, J., 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, **23**, 187 – 199.

[276] Uhler, C. & Shivashankar, G. V., 2017. Regulation of genome organization and gene expression by nuclear mechanotransduction. *Nat. Rev. Mol. Cell Biol.*, **18**, 717.

[277] Uusitalo, J. J., Ingólfsson, H. I., Akhshi, P., Tieleman, D. P. & Marrink, S. J., 2015. Martini Coarse-Grained Force Field: Extension to DNA. *Journal of Chemical Theory and Computation*, **11**(8), 3932–3945.

[278] Vaikuntanathan, S. & Jarzynski, C., 2011. Escorted free energy simulations. *J. Chem. Phys.*, **134**, 054107.

[279] van Mameren, J., Gross, P., Farge, G., Hooijman, P., Modesti, M., Falkenberg, M., Wuite, G. J. L. & Peterman, E. J. G., 2009. Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 18231–18236.

[280] Šulc, P., 2014. *Coarse-grained modelling of nucleic acids*. Ph.D. thesis, University of Oxford, UK.

[281] Šulc, P., Romano, F., Ouldridge, T. E., Rovigatti, L., Doye, J. P. K. & Louis, A. A., 2012. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Comput. Phys.*, **137**, 135101.

[282] Wang, N., 2017. Review of cellular mechanotransduction. *J. Phys. D: Appl. Phys.*, **50**, 233002.

[283] Wang, T., Schiffels, D., Martinez Cuesta, S., Kuchnir Fygenson, D. & Seeman, N. C., 2012. Design and Characterization of 1D Nanotubes and 2D Periodic Arrays Self-Assembled from DNA Multi-Helix Bundles. *J. Am. Chem. Soc.*, **134**, 1606–1616.

[284] Wang, X. & Ha, T., 2013. Defining Single Molecular Forces Required to Activate Integrin and Notch Signaling. *Science*, **340**, 991–994.

[285] Watson, J. D. & Crick, F. H. C., 1953. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature*, **171**, 737.

[286] Wei, B., Dai, M. & Yin, P., 2012. Complex shapes self-assembled from single-stranded DNA tiles. *Nature*, **485**, 623.

[287] Wei, R., Martin, T. G., Rant, U. & Dietz, H., 2012. DNA Origami Gatekeepers for Solid-State Nanopores. *Angew. Chem., Int. Ed.*, **51**, 4864–4867.

[288] Wenner, J. R., Williams, M. C., Rouzina, I. & Bloomfield, V. A., 2002. Salt Dependence of the Elasticity and Overstretching Transition of Single DNA Molecules. *Biophys. J.*, **82**, 3160 – 3169.

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

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

[291] Wickham, S. F. J., Bath, J., Katsuda, Y., Endo, M., Hidaka, K., Sugiyama, H. & Turberfield, A. J., 2012. A DNA-based molecular motor that can navigate a network of tracks. *Nat. Nanotechnol.*, **7**, 169.

[292] Williams, M. C., Wenner, J. R., Rouzina, I. & Bloomfield, V. A., 2001. Entropy and Heat Capacity of DNA Melting from Temperature Dependence of Single Molecule Stretching. *Biophys. J.*, **80**, 1932 – 1939.

[293] Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C., 1998. Design and self-assembly of two-dimensional DNA crystals. *Nature*, **394**, 539.

[294] Wolde, P. R. t. & Frenkel, D., 1997. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science*, **277**, 1975–1978.

[295] Wood, R. H., Muhlbauer, W. C. F. & Thompson, P. T., 1991. Systematic errors in free energy perturbation calculations due to a finite sample of configuration space: sample-size hysteresis. *J. Phys. Chem.*, **95**, 6670–6675.

[296] Woodside, M. T., Anthony, P. C., Behnke-Parks, W. M., Larizadeh, K., Herschlag, D. & Block, S. M., 2006. Direct Measurement of the Full, Sequence-Dependent Folding Landscape of a Nucleic Acid. *Science*, **314**, 1001–1004.

[297] Woodside, M. T., Behnke-Parks, W. M., Larizadeh, K., Travers, K., Herschlag, D. & Block, S. M., 2006. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 6190–6195.

[298] Woodside, M. T. & Block, S. M., 2014. Reconstructing Folding Energy Landscapes by Single-Molecule Force Spectroscopy. *Annu. Rev. Biophys.*, **43**, 19–39.

[299] Woodside, M. T., García-García, C. & Block, S. M., 2008. Folding and unfolding single RNA molecules under tension. *Curr. Opin. Chem. Biol.*, **12**, 640 – 646.

[300] Yagyu, H., Lee, J.-Y., Kim, D.-N. & Tabata, O., 2017. Coarse-Grained Molecular Dynamics Model of Double-Stranded DNA for DNA Nanostructure Design. *The Journal of Physical Chemistry B*, **121**(19), 5033–5039.

[301] Yin, P., Hariadi, R. F., Sahu, S., Choi, H. M. T., Park, S. H., LaBean, T. H. & Reif, J. H., 2008. Programming DNA Tube Circumferences. *Science*, **321**, 824–826.

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

[303] Yu, H., Gupta, A. N., Liu, X., Neupane, K., Brigley, A. M., Sosova, I. & Woodside, M. T., 2012. Energy landscape analysis of native folding of the prion protein yields the diffusion constant, transition path time, and rates. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 14452–14457.

[304] Yunger Halpern, N. & Jarzynski, C., 2016. Number of trials required to estimate a free-energy difference, using fluctuation relations. *Phys. Rev. E*, **93**, 052144.

[305] Zhang, D. Y., Turberfield, A. J., Yurke, B. & Winfree, E., 2007. Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science*, **318**, 1121–1125.

[306] Zhang, Q., Brujić, J. & Vanden-Eijnden, E., 2011. Reconstructing Free Energy Profiles from Nonequilibrium Relaxation Trajectories. *Curr. Opin. Chem. Biol.*, **144**, 344–366.

[307] Zhang, Y., McEwan, A., Crothers, D. & Levene, S., 2006. Analysis of In-Vivo LacR-Mediated Gene Repression Based on the Mechanics of DNA Looping. *PLoS ONE*, **1**, e136.

[308] Zhu, Y.-L., Lu, Z.-Y. & Sun, Z.-Y., 2016. The mechanism of the emergence of distinct overstretched DNA states. *J. Chem. Phys.*, **144**, 024901.

[309] Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D. & Chu, S., 2000. A Single-Molecule Study of RNA Catalysis and Folding. *Science*, **288**, 2048–2051.

[310] Zuckerman, D. M. & Woolf, T. B., 2002. Theory of a Systematic Computational Error in Free Energy Differences. *Phys. Rev. Lett.*, **89**, 180602.

Appendices

A

Experimental methods and additional details supplementary to Chapter 4

Additional Structural Details

Rothemund tile

In his seminal DNA origami paper, Rothemund presented a rectangular tile folded from M13mp18 DNA[228]. Here, we use a nearly identical structure: a 284-nt wide (Rothemund's original system was 288-nt wide), 24-helix long rectangle with two extended ends, corresponding to elongating sequences inserted into the multiple cloning site (MCS), protruding at a central seam. We also retain a 73-nt region Rothemund excised due to its containing a 20-bp stem hairpin. In the AFM experiments, the MCS sites were 433-nt and 217-nt long and were incorporated into double-stranded handles for attachment to the AFM tip and cover slide; the total handle lengths were 433-nt and 295-nt. For the simulations, since the unfolding of the origami rectangle itself was of interest, these MCS regions were replaced by 10-nucleotide poly-A segments. The folded origami rectangle in our case comprises a 6909-nucleotide scaffold bound by 192 staple strands: 48 30-nt

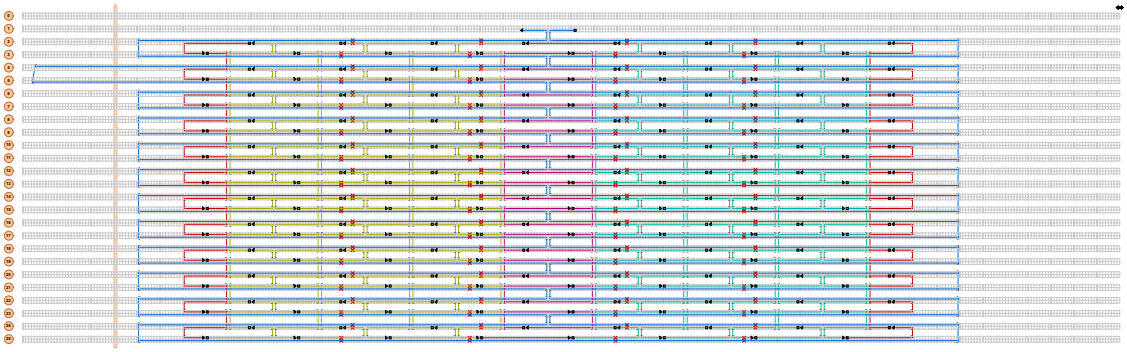


Figure A.1: Cadnano structure for the Rothemund tile.

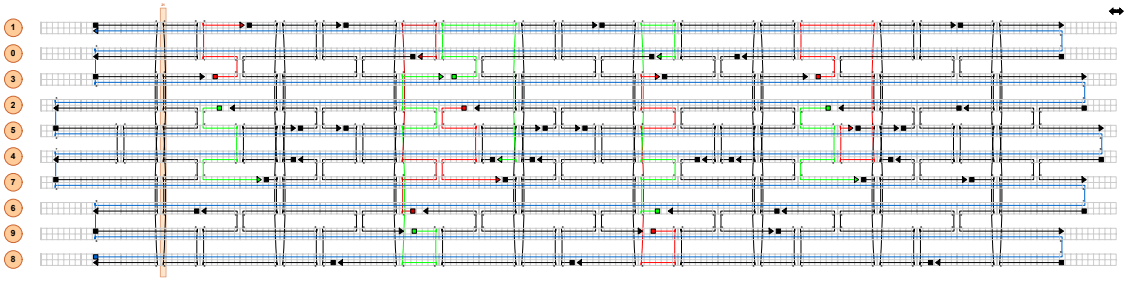


Figure A.2: Cadnano structure for the 10-helix bundle of Bae et al. [13].

staples and 144 32-nt staples. Most of the staples bind to 3 domains and a few are 1- or 2-domain staples; see Figure A.1.

10-helix bundle

In contrast to the Rothemund tile, our second system is based on a hexagonal lattice and is comprised of 10 helices: four 171-nt long, two 175-nt long, two 181-nt long, and two 185-nt long. The 1768-nt scaffold has 51 staples, many of which have 4 or 5 domains (see Figure A.2). A key difference between the helix bundle (HB) and the Rothemund tile lies in the relative non-locality of the bundle's staple domains.

7-tile biosensor

The 7-tile biosensor of Koirala et al. [141] comprises 14 761 nucleotides arranged as rectangular tiles joined by 44-base pair duplex 'locks'. In Koirala's work, the ends of the system are held in harmonic traps. The sequences of the 'locks' can be specifically designed to bind with aptamers of interest such that their presence in

solution is revealed by the opening of one or more locks, manifesting as a sudden increase in extension and decrease in force.

Experimental Methods*

Scaffold '7560'[58] from bacteriophage M13mp18[228] was double functionalized by ligation with biotin and thiol groups to facilitate attachment to either the streptavidin-coated AFM tips or maleimide-functionalized substrate, respectively. The circular, single stranded scaffold was digested by BamHI-HF and EcoRI-HF after partial pre-hybridization at restriction sites. The linearized scaffold was purified by solid phase extraction. Afterwards, partially double-stranded scaffold was ligated with 10× excess of double-stranded inserts carrying either biotin or thiol (5'-BIO; 3'-THIOL + 5'-Phosphate) by T4 DNA ligase. This ligated scaffold was either used in its unpurified form (Fig. 4.2(a), green and pink curves) to fold DNA origami rectangles[232], or was purified from excess functionalization strands by size exclusion (Fig. 4.2(a), purple curve) in spin filtration columns and subsequently folded. The folding procedure has been reported elsewhere[232]. The freshly folded DNA origami was purified by PEG precipitation[266]. Prior to 'unzipping' experiments, the origami pellet was resolved in 1× TE buffer (pH 8) supplemented with 10 mM MgCl2 and 2 mM TCEP. TCEP is required to reduce oxidized thiol-groups and enable maleimide conjugation reaction on surfaces. Afterwards, the resolved origami samples were coupled to AFM sample slides.

Correct DNA origami folding from digested scaffold was confirmed by AFM imaging; see Fig 4.1. Freshly cleaved mica surfaces were coated by 0.01% poly-L-ornithine to facilitate origami immobilization. Imaging was carried out in intermittent contact mode.

*The experimental description was drafted primarily by D. Smith, M. Jobst, and M. Sajfutdinow, though during a visit to their AFM lab in Munich I observed the procedure.

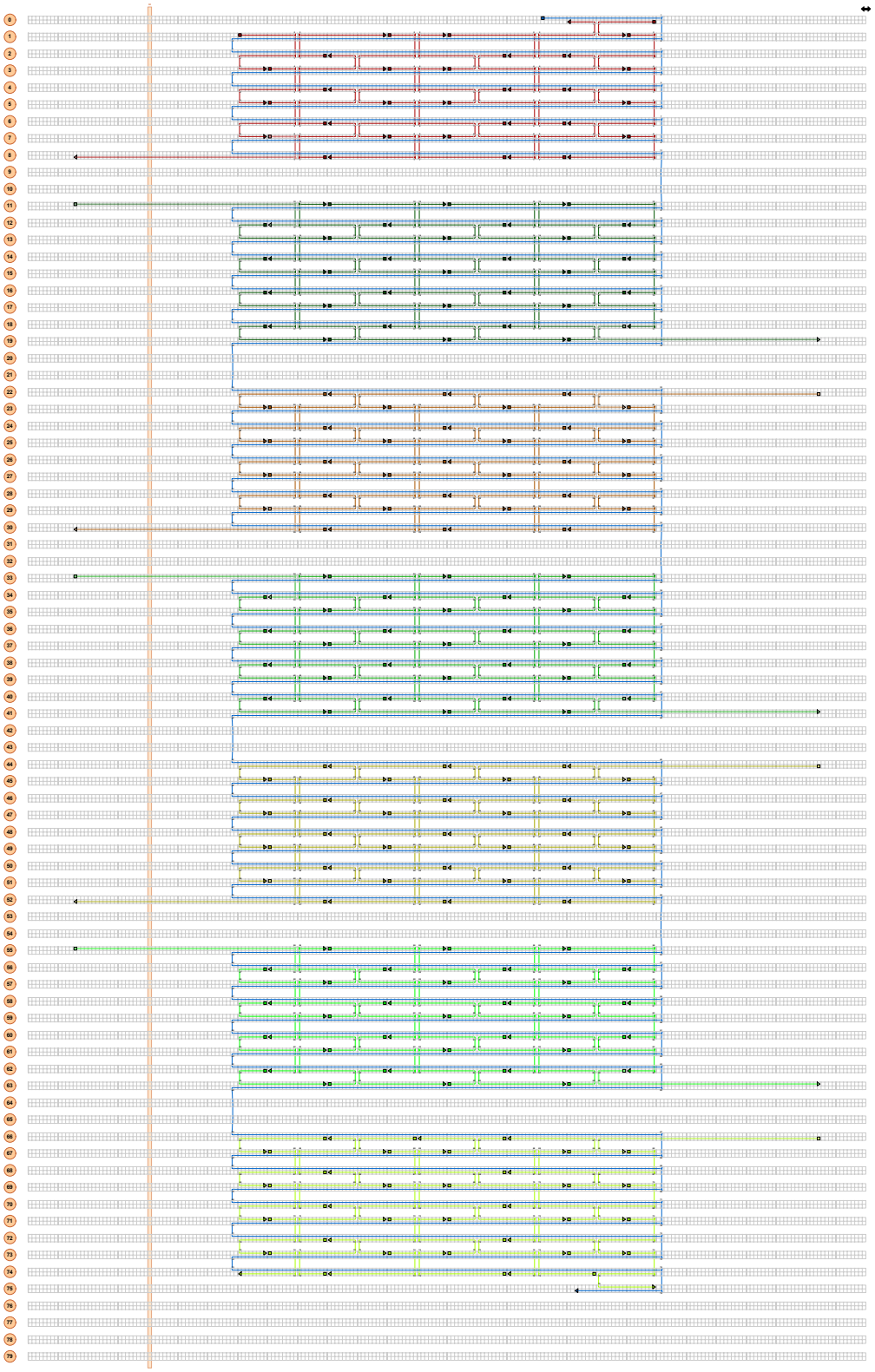


Figure A.3: Cadnano structure for the 7-tile assembly of Koirala et al. [141].

AFM Sample Preparation

Clean glass cover slides and UV-ozone cleaned silicon nitride AFM cantilevers were covalently functionalized first with amino-silane and subsequently with 5kDa NHS-PEG-maleimide polymer linkers. Reduced sulfhydryl groups of DNA origami were coupled to the free maleimide groups on the cover slides by forming permanent thiol-ether bonds[129;128]. Maleimide groups on the cantilevers were bound to coenzyme A, and subsequently to a monovalent variant of StrepTactin (monoST)[15;65] or a monomeric variant of streptavidin[14] via Sfp phosphopantetheinyl transferase (SFP) with ybbR tags (amino acid sequence DSLEFIASKLA) for use as specific AFM pulling handles[211;200].

SMFS Measurement Protocol

SMFS experiments were performed in TE buffer supplemented with magnesium at room temperature with custom-built instruments. Pulling velocities were set to $1\,\mu m\,s^{-1}$ or $1.6\,\mu m\,s^{-1}$. Cantilever and glass slide positioning was controlled by PID feedback loops during measurement. For analysis, curves with peaks at more than 150 nm extension and more than 30 pN force were sorted automatically, after thermal cantilever calibration, interferometric piezo calibration, and data conversion, similar to previous work[200;199;116;39].