Multi-scale coarse-graining for the study of assembly pathways in DNA-brick self-assembly



Pedro Fonseca

Somerville College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2020

Acknowledgements

I would like to thank my supervisor Ard Louis for giving me an interesting problem to work with, for his inspiring ideas, and for his support as I developed this work.

I am also grateful for many helpful discussions I had the pleasure to have with Flavio Romano and John Schreck.

I would also like to thank Thomas Ouldridge for his support, particularly in the early stages of this work.

Also, I would like to give a special thank you to Dr. Antonio Canas da Silva and Dr. C. Miguel for all their support and professionalism during difficult times.

Lastly, I would like to thank my parents for all their love and support.

Abstract

In this thesis, we develop a novel approach for modelling self-assembling systems that use the single-stranded-tile (SST) assembly method, an increasingly popular nanoconstruction approach in the field of DNA nanotechnology, developed primarily in the Molecular Systems Lab (Harvard University). In the SST method, a desired target structure is produced by mixing together a set of short DNA molecules (called SSTs) where each SST is provided with a judiciously chosen base sequence such that it will bind to a predetermined set of local neighbours in the target structure. Our approach for modelling these systems involves the use of simulations with a relatively fine-grained model of DNA (here we opt for oxDNA [1]) to parametrise coarser stochastic models that represent these systems at the level of the SST (the basic subunit of the method), and can probe effectively the timescales required for self assembly. As a test system for our approach we choose a rectangular-shaped SST-based object made out of 334 42-base-long SSTs, first reported in Ref. [2].

We first introduce a stochastic model that describes the growth of one target structure under a fixed background concentration of SSTs. We use the model to study nucleation pathways, finding that structures with a high number of bonds between SSTs and compact shapes are typically the most likely to form. However, we also find that for certain (recurrent) sizes, maximally-bonded structures are actually disfavoured due to statistical factors. We then devise a simple model that describes the bonding of incorrect SSTs to a (otherwise correctly formed) structure, finding that these bonds are typically very weak, and have a relatively minor impact on the overall kinetics and thermodynamics of self assembly.

We then introduce a stochastic model that describes the simultaneous growth of multiple target structures in a system where subunits are present in finite numbers (rather than present at constant supply). Under isothermal conditions the model predicts that there is only a small temperature window ($\sim$2°C) where self assembly is not hindered by excessively high nucleation barriers nor depletion traps. Finally, we simulate a cooling ramp similar to the original experiments of Refs. [2, 3]. The model predicts a peak in base-pairing formation in close agreement with Ref. [3]. However, the models predicts assembly yields that are significantly above those achieved in experiments. This differences are attributed to non-specific contacts between structures, a type of assembly that is not yet well understood, and thus not described by the model.

Contents

List of Figures **xi**

1 Introduction **1**
1.1 Deoxyribonucleic acid (DNA) 4
1.2 DNA nanotechnology 8
1.3 Theoretical approaches for modelling of DNA 12
1.3.1 Quantum and atomistic models 12
1.3.2 Continuum and statistical models 13
1.3.3 Coarse-grained models 15

2 Simulation methods (oxDNA) **21**
2.1 Monte Carlo simulation 21
2.2 Umbrella Sampling 23
2.3 Single-Histogram Reweighting 24
2.4 Weighted Histogram Analysis Method 25

3 Modelling Chemical Kinetics **27**
3.1 The Chemical System 28
3.2 Deterministic formulation of reaction kinetics 28
3.2.1 Relationship between deterministic rate constants and thermodynamic driving force 30
3.3 Stochastic formulation of reaction kinetics 31
3.3.1 Functional form of the reaction propensity 32
3.3.2 Relationship between stochastic and deterministic rate constants 33
3.3.3 Kinetic Monte Carlo (kMC) algorithm 34

4 Multiscale modelling of SST self assembly **37**
4.1 Introduction 37
4.1.1 A multiscale approach for modelling SST self assembly . . . 39
4.2 Model System 42
4.3 State definition and parametrisation of association/dissociation probabilities 44
4.3.1 Parametrisation of $c_+(\mu)$ 45

4.3.2 Parametrisation of $c_{-}(\mu)$ 45
4.4 Summary and discussion 48

5 Study of assembly pathways in single-target SST assembly 51
5.1 Simulation Details 52
5.1.1 The simulated ensemble 53
5.1.2 Enhanced kMC sampling 55
5.2 Results 56
5.2.1 Fragment growth under isothermal and thermal ramp conditions 56
5.2.2 Thermodynamic properties of SST assembly 58
5.3 Misbonds 71
5.4 The effect of misbonds on the thermodynamics of self assembly . . . 72
5.5 The effect of misbonds on the kinetics of self assembly 76
5.6 Neglected effects 78
5.6.1 Bridging events 78
5.6.2 Incomplete binding 79
5.7 Summary and discussion 79

6 Simulations in the NVT ensemble 83
6.1 Simulation Details 84
6.2 Self assembly under isothermal conditions 87
6.3 Simulation of a temperature ramp and comparison with experiments 97
6.4 Summary and discussion 101

7 Summary 103

Appendices 109

A Types of binding sites (μ) identified for R6 109

B Binding probabilities $P_{eq}^{(\mu)}(N)$ (oxDNA simulations) 113

C Standard-state binding free energies $\Delta G^{\ominus}(\mu)$ (plots) 117

D $\beta\Delta G^{\ominus}(\mu)$ (numeric values) 121

E Deviations from detailed balance 123

F Effective binding free energies for misbonds 127

G Bridging event (example) 131

H Lag time calculation 133

I $\phi_1^{\mathrm{eq}}(T)$ **135**

J t_{nuc} **and** t_{elo} **as a function of target structure size (simulations).** **137**

Bibliography **139**

List of Figures

1.1 Chemical structure of double-helical DNA. Two DNA chains run in opposite directions and form hydrogen bonds (dashed lines) according to the Watson-Crick base-pairing rules (C forms a base pair with G, and T forms a base pair with A). The location of hydrogen-bond donors is indicated with a colon $\langle:\rangle$. Adapted from Ref. [21] 5

1.2 Molecular structure of different forms of double-helical DNA. The atoms represented in the figure are coloured as follows: carbon - green, nitrogen - blue, oxygen - red, phosphorus - orange, hydrogen - white. This image was taken from Ref. [23] (Creative Commons). . 7

1.3 oxDNA representation of DNA. a) model double helix with effective potentials represented schematically. The backbone beads are represented as spheres, and the bases (containing two beads each) are shown as ellipsoids. Nucleotides also interact through excluded volume interactions (not shown in the picture). Adapted from Ref. [11]. b) Top-view cross section of the double helix showing one base pair. Each nucleotide is represented by three circles connected by a black line. The large circles represent backbone beads, whereas the small circles represent base beads. Adapted from Ref. [158]. . . 19

4.1 Coarse-grained representation of the R6 structure reported in Refs. [2, 3]. a) oxDNA and simplified representation of a 42-base-long SST. b) oxDNA and simplified representation of R6. Each SST in the figure is represented with an arrow specifying the directionality of the respective DNA chain (from the 5'-end to the 3'-end). 44

5.1 Self assembly under isothermal conditions - fragment size as a function of time. As the temperature goes up, the time to form a full fragment increases, and the fluctuations in the fragment size also increase. . . 56

5.2 Average fragment size (red curve) and fragment size fluctuations (vertical blue lines) as a function of temperature. The vertical blue lines correspond to the fragment size range $N \in [\langle N\rangle - 2\sigma, \langle N\rangle + 2\sigma]$, where σ is the standard deviation measured for fragment size. . . . 58

5.3 Fragment size as a function of temperature and under thermal ramp conditions. The curves shown in the figure correspond to averages over 10^4 independent runs. Also shown (dashed black curve) is the average fragment size, at equilibrium, as a function of temperature. Inset: derivatives with respect to temperature of the curves in the main plot. The growth peaks (during cooling) are located at 53.8°C for 3 h/°C, 53.0°C for 1 h/°C, 51.4°C for 5 min/°C, and 58.2°C at equilibrium. The melting peaks (during heating) are located at 58.7°C for 3 h/°C, 58.9°C for 1 h/°C, and 59.6°C for 5 min/°C. . . 59

5.4 Free-energy function $\beta F(N) = -\ln p_{eq}(N)$ as a function of fragment size (N). Left: $\beta F(N)$ in the size range $N \in [1, 334]$. Right: $\beta F(N)$ in the size range $N \in [1, 70]$. Also shown in the plot (black dashed lines) are CNT fits performed in the size range $N \in [1, 200]$. The critical sizes and the heights of the nucleation barriers are: $N^\ddagger = 5$ and $\beta F^\ddagger = 7.31$ for $T = 55$°C; $N^\ddagger = 7$ and $\beta F^\ddagger = 10.30$ for $T = 56$°C, $N^\ddagger = 13$ and $\beta F^\ddagger = 14.87$ for $T = 57$°C, $N^\ddagger = 40$ and $\beta F^\ddagger = 29.54$ for $T = 58$°C, $N^\ddagger = 63$ and $\beta F^\ddagger = 38.29$ for $T = T_c = 58.23$°C. . . . 60

5.5 Free-energy function $\beta F(N, E) = -\ln p_{eq}(N, E)$ as a function of fragment size (N) and connectivity (E) for the size range $N \in [1, 51]$, at $T = 58$°C. The insets show examples of bonding configurations for a given macrostate (N, E). At $T = 58$°C, the critical nucleus size is $N^\ddagger = 40$ (dashed black line). 62

5.6 Probability at equilibrium of forming a fragment with a given connectivity E, given a size N. Data points for fragment sizes satisfying $N = M_1 \times M_2$, with $|M_1 - M_2| \leq 1$, are highlighted with pink ($M_1 = M_2$) and green ($M_1 \neq M_2$) squares. Left: probability of forming a maximally-connected ($E = E_{\max}(N)$) fragment given a size N. Right: probability of forming a fragment with a less-than-maximal connectivity ($E = E_{\max}(N) - \Delta E$) given a size N. 65

5.7 Probability at equilibrium of forming a misbond relative to the reference (unbound) state, for different types of binding site (μ), as a function of temperature. In each plot we present the results obtained for a given $l(\mu)$, which corresponds to the cumulative length of the binding domains of a binding site of type μ (hence each plot covers various types of binding site μ that yield the same $l(\mu)$). Top left: $l(\mu) = 10$; top right: $l(\mu) = 11$; bottom left: $l(\mu) = 20$; bottom right: $l(\mu) = 21$. The insets show the binding probability at equilibrium for SSTs with correct sequences, relative to the reference (unbound) state. 74

5.8 Probability at equilibrium of forming a misbond relative to the reference (unbound) state, for different types of binding site (μ), as a function of temperature. In each plot we present the results obtained for a given $l(\mu)$, which corresponds to the cumulative length of the binding domains of a binding site of type μ (hence each plot covers various types of binding site μ that yield the same $l(\mu)$). Top left: $l(\mu) = 22$; top right: $l(\mu) = 31$; bottom left: $l(\mu) = 32$; bottom right: $l(\mu) = 42$. The insets show the binding probability at equilibrium for SSTs with correct sequences, relative to the reference (unbound) state. 75

5.9 Distribution of growth times with (red curves) and without (blue curves) misbonds being allowed to form. The growth time corresponds to the time needed for a fragment, initially consisting of a single SST, to reach its equilibrium size for the first time. For each temperature, we performed 2,000 independent runs. 77

6.1 Self assembly under isothermal conditions - mean concentration of fragments as a function of fragment size (N) and time (t). Each data point shown corresponds to an average over $n_{\text{sim}} = 100$ simulations. The dashed blue lines at $N = 334$ indicate the size of R6. The vertical axis denotes the monomer concentration $c(1, t)$ as a function of time. 88

6.2 Self assembly under isothermal conditions. Left, main plot - mean concentration of fragments with sizes in the target band ($N \geq 300$) as a function of time. Left, inset - lag time as function of temperature. Right - time $t_{\text{peak}}(N)$ when a concentration peak is reached for a fragment size N. Each data point shown (main figures) corresponds to an average over $n_{\text{sim}} = 100$ simulations. 89

6.3 Ratio between the rate of monomer (SST) depletion by nucleation ($N_{\text{e}}/t_{\text{nuc}}$) and the elongation rate ($1/t_{\text{elon}}$), for a SST concentration of 100 nM per SST type. We fix $N_{\text{e}} = \langle N \rangle$ as the fragment size for which the elongation phase is assumed complete. According to the criterion suggested in Refs. [205, 206], a depletion trap is expected to occur for ratios greater or equal than 1. Insets: mean nucleation time (t_{nuc}) (left), and mean elongation time (t_{elon}) (right), for a SST concentration of 100 nM per SST type. 91

6.4 Critical fragment size (left) and height of the nucleation barrier (right) at various times during the assembly reaction. For each temperature, are shown results up to the point where simulations yield $F(N) > F(1)$ for all fragment sizes (N) above one, in which case a stable nucleus can no longer form from unbound SSTs that remain in the system. Insets - results for initial reaction conditions: $t = 0$ and SST concentration of 100 nM per SST type. 93

6.5 Isothermal self assembly - concentration of nuclei (red curves) and fragments with sizes above the initial critical size (blue curves) as a function of time. 94

6.6 Simulation of a temperature ramp - mean concentration of fragments as a function of fragment size (N) and time (t). To make the figures easier to interpret, we plot the results in terms of a temperature variable $T = 65 - t/t_{\text{step}}$, where t_{step} is the cooling rate, and 65 corresponds, in Celsius, to the temperature at the start of the simulations. Note that in the simulations the temperature was decreased stepwise ($\Delta T = 1^\circ$ C). Each data point in the figure corresponds to a mean value over $n_{\text{sim}} = 100$ simulations. The dashed red lines at $N = 334$ indicate the size of R6. 95

6.7 Temperature window for productive self assembly as a function of target structure size. Left - value of $N_e t_{\text{elon}}/t_{\text{nuc}}$ as a function of temperature and target structure size. Right - value of $t_{\text{nuc}}/N_{\text{target}}$ as a function of temperature and target structure size. We use red and orange to colours regions of the parameters space for which assembly is not productive. The temperature window for productive assembly is shown in both plots (solid and dashed pink lines), and corresponds to the overlap region between both thresholds - monomer depletion and experimental accessibility. 97

6.8 Simulation of a step-wise cooling ramp ($\Delta T = 1^\circ$ C). Left, main plot - relative changes in base pairing levels with decreasing temperature, $\Delta n_b(T)/\Delta T$, where $n_b(T)$ is the number of base pairs formed (or a proxy for it - Ref. [3]) at the end of a temperature step (at temperature T). The coloured curves are the results of simulations; the dashed black curve is an experimental measurement published in Ref. [3]. A multiplicative constant was applied to each data set such that the peaks in the figure appear with the same heights. Left, inset - mean concentration of fragments with sizes in the target band ($N \geq 300$) at the end of each temperature step. Right - final assembly yield as a function of cooling rate predicted by the model (red curve), and measured experimentally in Ref. [2]. 99

6.9 Simulation of a step-wise cooling ramp ($\Delta T = 1^\circ$ C). Left - concentration of nuclei at the end of each temperature step. Right, main plot - height of the nucleation barrier at the start of each temperature step. Right, inset - size of the critical nuclei at the start of each temperature step. In both figures we use solid points if a stable nuclei can still form from unbound SSTs in the system, and open points (left figure only) if the (single-target) simulations yield $F(N) > F(1)$ for all fragment sizes above one, in which case a stable nucleus can no longer form from unbound SSTs in the system. In the latter case (open points) we estimate anyway, a concentration of nuclei $c_{\rm nuc}(t;T) = \sum_{N>N^\ddagger_{\rm (last)}} c(N,t;T)$, where $N^\ddagger_{\rm (last)}$ is the critical size at the lowest temperature where a stable nuclei could still form (for all cooling rates tested, we have $N^\ddagger_{\rm (last)} = 3$). 100

J.1 Mean nucleation (left) and elongation (right) times as a function of target structure size. The simulations were performed at T =56°C. The sizes (N) and aspect ratios (height/weight) (AR) of the target structures used in the simulations are: $N = 334$ and $AR = 23/15$; $N = 1434$ and $AR = 47/31$; $N = 3071$ and $AR = 69/45$; $N = 5324$ and $AR = 91/59$; $N = 10478$ and $AR = 127/83$; $N = 20651$ and $AR = 177/117$. The dashed black curve in the right plot is a guide to the eye. 137

1 Introduction

Deoxyribonucleic acid (DNA) has acquired iconic status for its central role in biology. Found mostly inside the cellular nucleus, DNA contains all instructions a living organism needs to assemble, maintain, and replicate itself. Owing to advances in automated DNA synthesis methods over the past decades, it has become possible, and increasingly affordable, to synthesise DNA molecules in large amounts with custom-made sequences and low error rates. These advances have enabled researchers to develop a number of applications for DNA outside a biological context. The field of DNA nanotechnology in particular, which concerns the use of DNA as a building material for static nanostructures and active nanodevices, has seen an enormous growth in the past decades. The physical and chemical properties of DNA make it optimally suited for this purpose: complementary sequences of DNA bind predictably into rigid double helices, whereas sequences without a complementary pair remain single-stranded (and floppy) and interact weakly with one another. Further, DNA is chemically stable under a wide range of conditions and can be manipulated by a large variety of DNA-acting proteins. In the field of DNA nanotechnology, nanoconstructs are created through self assembly: to create a desired target, a suitable set of DNA molecules is simply mixed together in a reaction vessel, under carefully controlled conditions, yet without the actuation of external drivers.

Since its inception in the early 80s, the field of DNA nanotechnology has developed substantially both in terms of the volume of the literature, and the size and complexity of the nanoconstructs created. Currently, DNA nanoconstructs containing thousands of different DNA molecules and virtually any shape are routinely created in experiments. We refer the reader to Ref. [4] for a thorough review of the field. However, how exactly a DNA nanoconstruct happens to self-assemble during an experiment is often not well understood. Direct observations of the assembly process are often limited to changes that occur at the level of the full sample, such as changes in (overall) base-pairing levels. To design and optimise an experiment researchers rely on empirical principles accumulated over the years, and often have to resort to a simple trial and error process. The ability to directly visualise self assembly would help identifying sources of assembly errors, direct assembly pathways through rational selection of the base sequences, and potentially increase the size and complexity of the nanoconstructs that can be produced.

Over the past decade, a few promising coarse-grained (CG) models of DNA have been proposed that offer a direct, molecular-level view of DNA systems. CG models describe the DNA molecule on the level of the nucleotide but use a reduced representation of the different groups within the nucleotide (sugar, phosphate and nucleobase) for improved computational performance. oxDNA in particular [1], by far the most widely used CG model of DNA to date, has been used extensively to study basic biophysical mechanisms that are exploited in DNA nanotechnology, such as DNA hybridisation [5], hairpin formation [6] and strand displacement [7], as well as the self assembly and operation of various small DNA nanostructures [8, 9] and nanodevices [10–14]. Nevertheless, the range of applicability of CG models is currently limited to relatively small systems and short timescales due to the large computational costs associated with simulating these models. Even if run on multi-CPU/GPU architectures, simulations with CG models can currently probe timescales up to the second range, and this for systems containing no more than a few dozen DNA molecules.

Some of the most widely used construction approaches in the field of DNA nanotechnology, such as DNA origami [15, 16] and the single-stranded-tile (SST) approach [2, 17–20] (also called "DNA brick" approach), are routinely employed to create structures on a kilo-base pair range from hundreds (and often thousands) of different DNA molecules. Furthermore, these structures often require several hours of annealing time to fully form. Describing such large and slowly-evolving systems via simulations therefore requires the development of new, more efficient models. Such models should provide a sufficiently coarse representation of the DNA molecule such that the time- and length-scales associated with self assembly can be probed effectively in simulations, while, hopefully, retaining a sufficiently realistic description of how DNA molecules interact with one another such that the basic physics is not lost.

In this thesis we develop a novel approach for modelling self-assembling systems that use the SST assembly method - an increasingly popular approach for DNA nanoconstruction developed primarily in the Molecular Systems Lab (Harvard University). In the SST method, a desired target structure is produced by mixing together a set of short, synthetic DNA molecules (called SSTs), where each SST is provided with a judiciously chosen (and unique) base sequence such that that SST will bind to a predetermined set of local neighbours in the target structure (and hence occupy a predetermined site in the target structure). Our approach for modelling these systems involves the use of simulations with a relatively fine-grained model of DNA (here we opt for oxDNA [1]) to parametrise coarser stochastic (Markovian) models that represent these systems at the level of the SST (rather than at the level of the nucleotide), where the Markov property is expected to be satisfied. While the SST assembly method is the primary focus of our thesis, our modelling approach can be applied to different assembly schemes provided subunits bind exclusively with local neighbours in the target structure. As a test system for our modelling approach we choose a rectangular-shaped SST-based object made out of 334 42-base-long SSTs first reported in Ref. [2].

This thesis is organised as follows. In chapter 1, we introduce the DNA molecule, give a brief overview of the field of DNA nanotechnology and discuss the simulation tools developed so far for modelling DNA, with a focus on the oxDNA model. In chapter 2 we describe the methods used in this thesis for simulating oxDNA. In chapter 3 we review both deterministic and stochastic formulations of chemical kinetics, with a focus on how the rate constants in both formulations relate with each other. In chapter 4, we discuss previous attempts of modelling SST-based systems, highlighting its limitations and the need for more fine-grained modelling. Also in chapter 4, we introduce and justify our modelling approach, and compute a series of stochastic rate constants via simulations with oxDNA (with the aid of a few equalities from chapter 3), for a model system chosen from Ref. [2]. In chapter 5, we introduce a stochastic model that describes the self-assembly of one target structure under a fixed background concentration of subunits (SSTs). We use the model to study dynamic and thermodynamic properties of self assembly, with a focus on the dominant nucleation pathways and the relative importance of assembly errors due to incorrectly bonded SSTs. In chapter 6, we adapt the stochastic model of chapter 5 to describe the simultaneous growth of multiple targets in a system where subunits are present in finite numbers (rather than at a constant supply). We use the (adapted) model to study self assembly under isothermal and thermal ramp conditions, with an emphasis on how saturation levels, and how these change over time, affect assembly yields and assembly times.

1.1 Deoxyribonucleic acid (DNA)

DNA is a linear polymer comprised of alternating sugar units and negatively-charged phosphate groups, with information-carrying bases attached to each sugar unit (Fig. 1.1). Sugar units (called 2-deoxyriboses) contain five carbon atoms (labeled 1' to 5'), and form covalent bonds with adjacent phosphate groups though carbons 3' and 5', and to the base through carbon 1'. DNA is often described alternatively as polymer of larger subunits called "nucleotides", each of which is comprised of a sugar unit, a phosphate group and a base. The bases are planar, aromatic moieties

containing both carbon and nitrogen atoms; they can be of one of two types: a pyrimidine, consisting of a single aromatic ring, or a purine, consisting of two aromatic rings joined together. Pyrimidines found in DNA can be either thymine (T) or cytosine (C), whereas purines can be either guanine (G) or adenine (A). The bonding pattern between sugar units and phosphate groups confers directionality to the DNA molecule: in one direction phosphate groups bind to sugar units through a carbon 3', whereas in the opposite direction phosphate groups bind to sugar units through a carbon 5'. By convention, the base sequence of a DNA molecule is spelled from the 5' end of the molecule to the 3' end of the molecule.

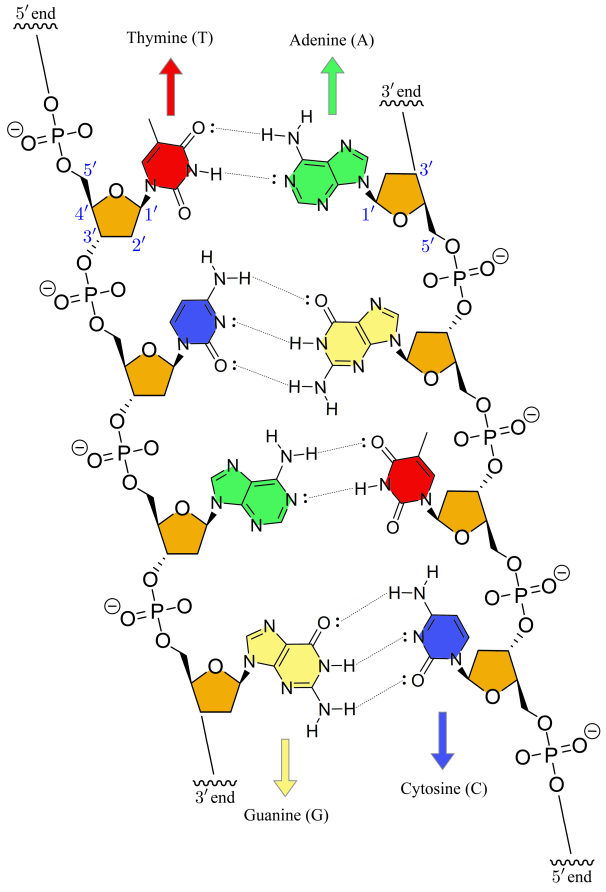


Figure 1.1: Chemical structure of double-helical DNA. Two DNA chains run in opposite directions and form hydrogen bonds (dashed lines) according to the Watson-Crick base-pairing rules (C forms a base pair with G, and T forms a base pair with A). The location of hydrogen-bond donors is indicated with a colon ⟨:⟩. Adapted from Ref. [21]

A purine and a pyrimidine can form a base pair according to the Watson-Crick (WC) base-pairing rules: A can form a base pair with T, and C can form a

base pair with G. If two bases satisfy the WC base-pairing rules they are called "complementary". Base pairing involves the formation of multiple hydrogen bonds and the alignment of both base planes. The specificity of the base-pairing rules is a consequence of the spatial arrangement of the hydrogen-bond (HB) donors and acceptors in the bases, and the requirement that a donor-acceptor pair needs to be closely aligned in order for a hydrogen bond to form. Between two complementary bases it is possible to align multiple HB donors and acceptors at a time (2 for a A-T pair and 3 for a G-C pair), whereas between non-complementary bases only one HB donor-acceptor pair can be aligned at a time (hence resulting in a less stable base pair).

In living organisms, DNA is usually found in the form of a right-handed double helix called the B-form, in which two DNA molecules with fully complementary base sequences coil around a common axis in anti-parallel orientations (see Fig. 1.2 - center). The bases of both molecules form base pairs and occupy the core of the helix, whereas the phosphate-sugar chains run on the outside and perform a structural role. The formation of a double helical structure from two unpaired DNA molecules is referred to as "DNA hybridisation". B-form DNA has a helical pitch of 3.4 nm (corresponding to approximately 10.5 base pairs) and a constant diameter of 2.0 nM, with minor and major grooves (1.2 nm and 2.2 nm wide). The molecular structure of B-form DNA was first identified by J. Watson and F. Crick in 1953 [22].

B-form DNA is primarily stabilised by intramolecular base-stacking, which consists of the layering of the base planes within a chain on top of each other. Base stacking is driven mainly by dispersion interactions (instantaneous dipole-induced dipole) between the delocalised π-electron systems of the aromatic rings, and the hydrophobic effect, which favours the sequestering of the (relatively non-polar) aromatic surfaces into the core of the double helix, where they become effectively shielded away from water molecules [24, 25]. The contribution of base pairing to the stability of B-form DNA is relatively minor in comparison. Because water molecules can form hydrogen bonds with both the HB donors and the HB acceptors in the bases, forming a B-form structure does not increase, on aggregate, the number of

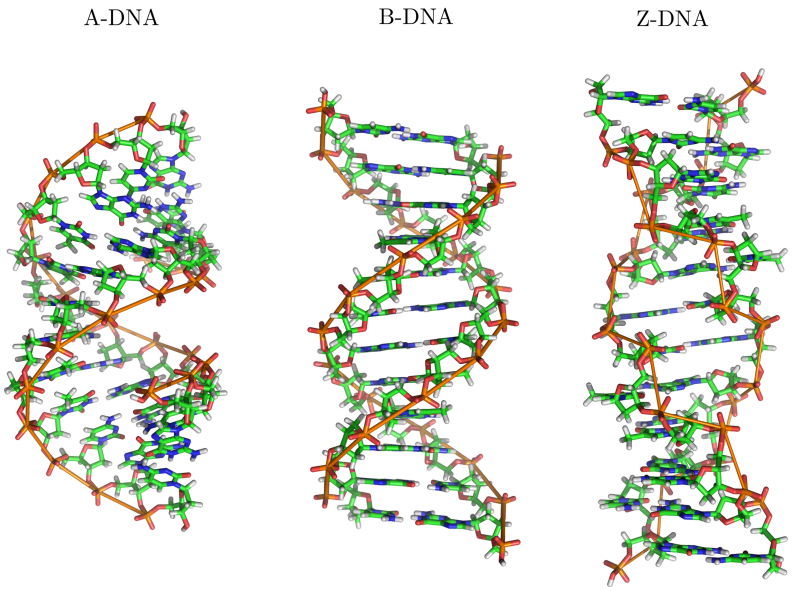


Figure 1.2: Molecular structure of different forms of double-helical DNA. The atoms represented in the figure are coloured as follows: carbon - green, nitrogen - blue, oxygen - red, phosphorus - orange, hydrogen - white. This image was taken from Ref. [23] (Creative Commons).

hydrogen bonds formed relative to the unpaired state. However, base pairing plays a crucial role with respect to the specificity of the interactions between DNA molecules. For instance, one sequence mismatch between two otherwise complementary DNA molecules leads to a net decrease in the number of hydrogen bonds formed if these molecules are made to hybridise. This is because once sequestered at the core of the helix, the mismatched pair becomes effectively inaccessible to water molecules in the environment (and therefore it is prevented from forming an optimal number of hydrogen bonds). Furthermore, a sequence mismatch also introduces significant structural distortions in the double helix. A consequence of the factors above (reduction in the number of hydrogen bonds formed and disrupted geometry) is that the thermodynamic drive for hybridisation between DNA molecules with a sequence mismatch is significantly lower than between DNA molecules that are fully complementary [26–31].

B-form DNA is made less stable due to repulsive electrostatic forces between the negatively-charged phosphate groups. The presence of cations in the medium helps improving the stability of the structure by providing screening to the electrostatic

field created by the phosphate groups. Divalent cations are known to be particularly effective at stabilising the B-form helix [32–34]. Mg^{2+}, for instance, has a stable solvation shell when in aqueous solution that can bind to the phosphate groups via hydrogen bonds and directly offset the charge of the phosphate groups [34, 35].

Under physiological conditions (low salt concentration, high hydration, and temperature within the range 20°C-40°C), DNA typically folds into the famous B-form. However, DNA can exhibit different forms if conditions in the environment are changed or if it contains certain sequences. For example, DNA in the B-form can be made to convert made into a compact right-handed helix, the A-form (Fig. 1.2 - left), if subject to low-hydration conditions and by actuation of certain DNA-protecting proteins [36–39]. Alternatively, a B-form helix containing a sequence of alternating C and G bases can be made to convert into a narrow left-handed helix, the Z-form (Fig. 1.2 - right), through actuation of DNA-binding proteins [40, 41], by increasing the salt concentration [42, 43], or by twist-straining the helix [44].

DNA has acquired iconic status for its central role in biology. Found mostly in the cellular nucleus, DNA contains genetic instructions for the assembly of proteins, which serve as building blocks for cellular structures and carry out biological functions within the cell. The genetic material of an organism (the genome) contains all information that is inherited from its progenitors and is subsequently passed on to its offspring. In conjunction with environmental factors, an organism's genome determines its observable traits, such as structure, development and behaviour.

1.2 DNA nanotechnology

The field of DNA nanotechnology traces its roots to a seminal paper published by Nadriaan Seeman in 1982 [45], in which he provided a blueprint for the first fully-synthetic DNA construct: a N-way junction. A N-way junction consists of N double helices of equal length, each emanating from a common exchange point located at the centre of the construct. In the same paper, Seeman argued that by placing unpaired DNA "overhangs" at the terminus of the helices, N-way junctions could be made to bind together to form ordered structures such as periodic lattices

and polyhedra. Seeman and coworkers successfully assembled a N-way junction (a 4-way junction) for the first time [46] in the following year (1983). 3-,5-,6-,8- and 12-way junctions have since been assembled successfully [47, 48], as well as small complexes containing N-way junctions linked together [49–51]. However, due to a high degree of flexibility of the helical angles at the exchange points, N-way junctions have proven to be too floppy to be suitable as building blocks for larger, geometrically well-defined structures.

The need for a structurally-rigid motif led to the development of the double-crossover (DX) tile [52]. A DX tile contains two parallel double helices linked together by strand exchanges at the ends points of both helices, and four unpaired overhangs (one at each helical end) that provide binding specificity to the DX tile. Using a few DX-tile types (each type provided with a specific overhang sequence), Winfree and co-workers produced a series of two-dimensional DNA crystals with programmed lattice periodicities and surface features [53]. At the time of publication (1998), these were the first geometrically well-defined, high-order structures to be created entirely from DNA. DX tiles have since been used as basic fabrication units of numerous periodic and aperiodic constructs. Examples include barcode-patterned lattices [54], nanotubes [55], crystals that can perform binary operations [56] and display fractal patterns [57] during crystal growth, to name a few. These successes have inspired the development of many other rigid motifs. N-point-stars, for instance, which resemble Seeman's N-way junctions but double the number of double helices per branch for improved rigidity, have been used to create two-dimensional lattices [58] and polyhedra [59–61]. "Tensegrity" triangles, consisting of three helical domains joined pairwise by 4-way junctions, have served as building blocks for one- and two- and three-dimensional crystals [62–64], including the first macroscopic DNA crystal ever created [63]. The triple-crossover (TX) tile - a variation of the DX tile that uses a third parallel helix per motif and facilitates the inclusion of cavities and integration of DNA with other substances - has served as a building block for patterned arrays [65], assemblies that perform basic logic operations during growth [66], and DNA-templated conductive nanowires [67], to name a few.

In parallel, a number non-hierarchical approaches have been devised where a set of DNA molecules is made to self assemble directly into a desired object without the formation of intermediate tile motifs. The first objects created this way were small polyhedra (a cube [68] and a truncated octahedron [69]), whose edges correspond to double helices and the vertices to exchange points identical to those at the centre of Seeman's N-way junctions. However, due to a high degree of flexibility of the helical angles at the exchange points, these structures lacked rigidity and did not display a well-defined geometry. Various polyhedra including a tetrahedron, an octahedron, a pyramid and a bipyramid, have since been fabricated that rely on a triangulated architecture for much improved rigidity [70–72].

The DNA origami method, introduced by Rothemund [73], uses a collection of short "staple" strands to fold a long circular scaffold strand into a desired shape. Each staple strand is provided with a judiciously chosen sequence such that it will bind to a few selected sections of the scaffold strand and bring them into close proximity. The precise shape into which the scaffold will fold depends on the sequences of the staple strands. In Rothemund's original work, several feature-rich two-dimensional shapes, including a star, a world map and a smiley face, were produced using different batches of staple strands (maintaining the same scaffold strand). DNA origami has since been extended to three dimensions through a number of approaches. For instance, in Refs. [74, 75] a few sheets of origami were linked together using single-stranded "hinges" to form hollow containers with cubic and tetrahedral shapes. By programming the double-helical domains that form between staple and scaffold strands with lengths above the helical pitch of DNA (or a multiple of it), a DNA origami can be made to fold into a curved surface, such as a sphere an ellipsoid or a bottle-like shape [76]. In a different approach, densely packed 3D objects displaying square [77], honeycomb [78] and hexagonal [79] lattices have been produced by programming sheets of DNA origami to fold onto themselves. DNA origami has also been employed to create complex wire-frame polyhedra [80–82], in which edges correspond to double helices and vertices to various types of junctions. Various hierarchical schemes have also been devised where DNA origami

is used to fabricate tile motifs which subsequently bind together to form larger structures, such as periodic lattices [83, 84] and polyhedra [78, 85].

The single-stranded-tile (SST) approach (or "DNA-brick" approach), developed in the Molecular Systems Lab (Harvard University) [86], uses only short single strands (called single-stranded tiles, or SST(s)) yet is capable of reaching structure sizes and a level of complexity comparable to that of DNA origami. In the SST approach, each SST is provided with a unique sequence containing four consecutive binding domains, each of which is complementary with one binding domain of one other SST. Once mixed together, SSTs self assemble into finite DNA lattices with coordination number equal to four. The SST approach was first introduced as fabrication tool for DNA nanotubes [87]. More recently, Wei and coworkers reported a number of intricate two-dimensional shapes [88], including smiley faces, and letters from the Latin alphabet. A central feature of Wei's approach is the use of a "molecule canvas", a large rectangular shape containing 310 different SSTs, from which a desired shape can be "carved" by one-pot annealing SSTs that define that shape, and excluding the remaining SSTs. By adjusting the lengths of the binding domains, Wei and coworkers created two-dimensional SST structures that display intrinsic curvature, twist, and a diverse range of lattice geometries [89]. The SST approach has since been extended to create solid three-dimensional objects. By "carving" SSTs from a cubic molecular canvas, Ke and coworkers created over one hundred different three-dimensional shapes that display complex surface features, internal cavities and tunnels [90]. More recently, Ong and coworkers reported the largest aperiodic DNA structure ever assembled from entirely unique components [91] - a cuboid containing over 30,000 different SST (adding up to over 1.7 million nucleotides). A finite number of SSTs can be designed to self assemble into crystals that grow indefinitely in predetermined directions, and display user-specified features such as internal cavities and tunnels [92].

A number of small active devices have been reported that rely on a mechanism called toehold-mediated strand displacement (TMSD) for its function. TMSD consists in the displacement of a short strand of DNA, initially hybridised to a

"displacement" domain, by a longer strand that is complementary to that domain plus an additional "toehold" domain. The first active device reported in the literature was a set of DNA "tweezers" [93] consisting of a pair of double helices connected by a single-stranded "hinge", which can switch between "open" and "closed" states through the addition of "fuel" and "anti-fuel" strands. Yurke et al.'s work has provided inspiration for similar multi-state devices, such as a two-state "nanoactuator" [94] and a three-state device [95], that rely on externally provided DNA for triggering structural transitions. TMSD has been used to induce structural reconfigurations in large, otherwise static nanostructures post assembly. For example, Feng and coworkers reported a two-dimensional DNA lattice, with periodically placed nanoactuators, whose unit cell dimensions can be made to change by adding a few types of "fuel" DNA [93]. Using a similar approach, Goodman et al. reported a DNA tetrahedron whose edges can switch between two different lengths through addition of specific strands of DNA [94]. Andersen et al. reported a cuboid-shaped box functionalised with a lock-key system that can be to open through addition of specific DNA "keys" [74]. In another strand of research, TSMD has been exploited to drive the movement of DNA "walkers" along a series of anchor sites of a DNA track [96–99].

1.3 Theoretical approaches for modelling of DNA

In this section we give a brief overview of the theoretical tools developed so far for modelling DNA.

1.3.1 Quantum and atomistic models

At the highest level of granularity, quantum-mechanical (QM) methods provide a first-principles framework for describing the nature and energetics of the interactions between DNA bases. QM calculations provide molecular wave functions, from which certain quantities that are difficult to access in experiments, such as charge distributions, dipole moments, and hydrogen-bonding and base-stacking energies [100, 101], can be determined. However, due to the computational costs involved,

QM calculations are only applicable to study very small systems (with no more than a few bases) in isolation. All-atom models such as AMBER [102, 103], and CHARMM [104, 105], represent the DNA molecule at the level of the atom and use analytical expressions, parametrised using experimental observables or QM calculations, to describe interactions. All-atom models have been widely used to study short-timescale phenomena (up to $\sim$50 μs) in small DNA systems. Examples include the folding of DNA "hairpins" [106–108], conformational dynamics of protein-DNA binding [109], and fast modes of motion in B-form DNA [110–112]. The use of enhanced sampling methods has made it possible to sample hybridisation thermodynamics in small double helices [113–115]. Recently, the direct simulation of a DNA-origami and a SST-object post assembly ($<$1 μs simulation time) has been carried out in Aksimentiev's group [116–118]. Despite these successes, all-atom models remain too computationally expensive to simulate large DNA systems over long periods of time. In particular, the simulation of self-assembly processes in DNA nanotechnology (even in the least complex systems) is currently well beyond the reach.

1.3.2 Continuum and statistical models

At the lowest level of granularity, continuous models describe the DNA helix as a continuous, semi-flexible rod. Continuous models are well suited for describing the behaviour of DNA at lengthscales greatly surpassing its persistence length[1]. The worm-like chain (WLC) model [120], in particular, has been widely used to model the behaviour of DNA both in isolation and under mechanical stress. In its original formulation, the WLC model describes a thin, inextensible, linearly-elastic rod with a certain bending rigidity. Despite its simplicity, the WLC model describes well the force-extension behaviour of DNA in the low force regime ($<$10 pN) [121–123]. To describe different regimes, several extensions of the WLC model have been developed over the years. For example, the extensible WLC model [124, 125] describes the elastic stretching of the rod, and reproduces well the force-extension behaviour of

[1]The persistence length of B-form DNA is about 50 nm, corresponding to roughly 150 base pairs [119].

DNA at the intermediate force regime (up to $\sim$ 35 pN), where the contour length of DNA increases slightly to resolve the stress [121, 126]. At higher forces (up to about $\sim$65 pN[2]), the DNA helix unwinds in order to resolve the strain [127]. To describe this regime, a few extensions to the WLC model have been proposed [127–130] that describe the coupling between stretching and twisting, and manage to reproduce well experimental observations, particularly the extension proposed in Ref. [130]. At very low salt concentrations ($<$ 0.1 nM [Na^+]) the screening lengths of the phosphate group's charges can approach (and even fall below) the "elastic" persistence length of B-form DNA. Marko [125] incorporated a simple description of electrostatics into the WLC model, where DNA is described as a uniformly-charged rod, that correctly reproduces the (increased) bending rigidity of DNA at low salt conditions.

Continuous models have also been used as structure prediction tools in DNA nanotechnology. Initially developed by Kim et al. [131], CanDo provides a modelling framework for predicting equilibrium shapes and mechanical properties of DNA nanostructures [132–134]. CanDo describes B-form DNA as a semi-flexible rod with a specific bending, stretching, and twisting rigidity (similar to the WLC model); and crossovers (singles-stranded segments that link double helices together in a DNA nanostructure) as rigid constraints.

Also at a low level of granularity are statistical models that describe the thermal denaturation of DNA using a small set of sequence-dependent parameters that give free-energy gains per base step[3] in a double helix, relative to the unpaired (single-stranded) state. The Poland-Sheraga (PS) model [135], for instance, describes the DNA double helix as a sequence of double-stranded segments and single-stranded loops. The PS model reproduces well the melting transition of relatively long double helices ($\gtrsim$ 20 base pairs), which can often exhibit statistically relevant, partially-open states before strands separation [136]. For shorter double helices, the widely used nearest-neighbour model (NN) by SantaLucia [31] is the model of choice.

[2]For a stretching force of about 65 pN, an "overstretching" transition takes place that involves the breaking of base pairs and the conversion of some double-helical segments into unpaired (single-stranded) segments. The WLC model, lacking any atomic detail, cannot describe this regime.

[3]a base step is defined as a pair of adjacent (and stacked) base pairs in a double helix.

The NN model describes the melting transition as a perfect two-state transition between a fully formed double helix and the unpaired single-stranded state. Despite its simplicity, the NN model is regarded in the field as a reliable predictor of melting temperatures of double helices. Since this model will be used later in this thesis, we will briefly discuss it here. In the NN model, the standard-state[4] free energy difference between double-helical and unpaired states, $\Delta G^{\ominus}_{NN}(T)$, is given as a sum of base-step terms, plus initiation, terminal and salt dependence contributions. Each term contributing to $\Delta G^{\ominus}_{NN}(T)$ is assumed decomposable into temperature-independent (and generally sequence-dependent) enthalpic and entropic parameters. For instance, the NN model has a total of 10 different enthalpic and entropic parameters for the base-step contribution that cover all possible combinations of bases: AA/TT, AT/AT, AG/CT, TA/TA, TG/CA, CG/CG, GA/TC, GT/AC, GC/GC, and GG/CC. According to the NN model, the equilibrium concentration of a double helix "AB", containing strands "A" and "B", is given by

$$\frac{[AB]}{[A][AB]} = \exp(-\Delta G^{\ominus}_{NN}(T)/RT) \times M^{-1}, \tag{1.1}$$

where $[AB]$ is the molar concentration of AB, $[A]$ and $[B]$ are the molar concentrations of strands A and B, and R is the ideal gas constant. The melting temperature (T_m) of a double-helix is defined as the temperature for which one half of the strands form double helices, and the other remain unpaired.

1.3.3 Coarse-grained models

At an intermediate level of granularity, coarse-grained (CG) models describe DNA on the level of the nucleotide but use a reduced representation of the different groups of the nucleotide (sugar, phosphate and nucleobase) for improved performance. In CG models, nucleotides are represented by a set of "beads" that interact with one another by means of effective potentials. CG models are often classified according to the approach followed to construct and parametrise the respective effective potentials. In the "bottom-up" approach, effective potentials are fitted to reproduce structural

[4] standard pressure $p^{\ominus}$ =1 bar, and strand concentration $c^{\ominus}$=1 M.

distributions (e.g. intrastrand distances, bending angles) observed in a more detailed model, typically an all-atom model. In the "top-down" approach, effective potentials are assumed to obey simple simple analytical expressions, and are fit to reproduce experimental (or experimentally-derived) observables such as melting temperatures and elastic moduli. By construction, bottom-up models provide a more faithful representation of the microscopic details of the interactions; however, these models tend to be unreliable if simulated at state points that differ (e.g. by temperature) from the state point where the effective potentials were parametrised. By contrast, top-down models are usually transferable between state points, but tend to lack the level of structural and mechanical detail that bottom-up models provide.

Examples of bottom-up DNA models include a one-bead-per-nucleotide CG model introduced by Savelyev [137, 138], which reproduces well the stiffness of B-form DNA under a wide range of salt concentrations. Savelyev's model has been used to study structural transitions in DNA nanocircles under torsional stress (caused by changes in salt concentration) [138], the unwrapping of the nucleosome [139], and the contribution of electrostatics to the bending rigidity of DNA [140]. Morriss-Andrews proposed a three-bead-per-nucleotide CG model [141] that reproduces correctly the structure and chirality of the double helix, and the stiffness of both single-stranded and double-helical DNA at different salt conditions. More recently, Naômé and coworkers introduced a one-bead-per-nucleotide CG model [142] that reproduces well cyclisation rates of small DNA helices, and geometrical and topological properties of DNA minicircles [143].

Examples of top-down models include 3SPN, a three-bead per nucleotide model developed in de Pablo's group which was initially fit to reproduce melting curves and the geometry and persistence length of B-form DNA [144]. More recently, 3SPN has been reparameterised to describe explicitly electrostatics [145], and to improve the structural and mechanical behaviour predicted for single-stranded and double-helical DNA [146]. Over the years, 3SPN has been used to study a range of biophysical processes. Examples include the packaging of DNA inside viral capsids [147], the thermodynamic and structural stability of DNA Holliday

junctions [148], and the sliding of DNA chains around histone proteins [149–151]. Dans and coworkers developed a six-bead per nucleotide model [152] which was fit to reproduce melting temperatures and geometric properties of double-helical DNA. Dans's model has been used to characterise the opening of "bubbles" (unpaired single-stranded sections) in double-helical chains [153], and the binding dynamics between DNA chains and DNA-binding proteins [154, 155].

oxDNA

Some simulations described in this thesis were performed with oxDNA [156], a top-down three-bead-per-nucleotide model developed by Ouldridge and coworkers, which was fit to reproduce melting temperatures, geometric properties, and elastic moduli for both single-stranded and double-helical DNA. In oxDNA, nucleotides are represented by three interaction beads: one bead representing the backbone units (sugar and phosphate), and two beads representing the base group (see Fig. 1.3). The effective potential of oxDNA has the form:

$$U = \sum_{\langle ij \rangle} (U_{bb} + U_{stack} + U'_{exc}) + \tag{1.2}$$

$$\sum_{i,j \notin \langle ij \rangle} (U_{HB} + U_{cr.st.} + U_{cx.st.} + U_{exc}), \tag{1.3}$$

where the first sum runs over pairs of nucleotides adjacent to each other in a DNA strand, and the second sum runs other pairs of nucleotides, including pairs of nucleotides in different DNA strands. The functional forms of the terms in Eq. 1.3 can be found in Ref. [156]. The hydrogen bonding (U_{HB}), stacking (U_{stack}), coaxial stacking ($U_{cx.st.}$) and cross stacking ($U_{cr.st.}$) terms are all anisotropic, reflecting the directional nature of these interactions, whereas the backbone potential (U_{bb}) is isotropic. All beads are provided with excluded volume terms (U'_{exc} and U_{exc}). In oxDNA, hydrogen bonding is only possible between WC-complementary bases. In its original implementation (referred to as oxDNA1), the model was parametrised to reproduce melting temperatures of helices with "average" sequences. These temperatures were inferred by taking averages over the parameters of the NN

model [31] (see Sec. 1.3.2) for all sequence combinations. An extension of oxDNA (oxDNA1.5) was later released [157] that introduces sequence dependence into the hydrogen-bonding and stacking terms. Both releases (oxDNA1 and oxDNA1.5) were fit to reproduce the behaviour of DNA at high salt conditions ($[Na^+]$=0.5 M), for which the charges of the phosphate groups are sufficiently screened that, as a good approximation, electrostatic forces can be incorporated into the excluded volume terms. In both releases, the three beads that form a nucleotide are all collinear to one another, nucleotides have the same structure independent of base identity, and the hydrogen-bonding term favours perfect alignment between the axis of two opposing nucleotides upon base pairing. As a result, B-form DNA is described as a fully-symmetrical structure, without minor or major grooves. A newer iteration of the model, oxDNA2 [158], places the backbone bead off the line of the base beads, and is capable of reproducing the minor and major grooves of B-form DNA. oxDNA2 also describes electrostatic forces between phosphate groups, thus allowing to probe low salt regimes. All three versions of oxDNA referenced above are publicly available as stand-alone programs that can be executed on a single CPU (central processing unit) or GPU (graphics processing unit). Recently, oxDNA has also been distributed alongside LAMMPS [159], a popular molecular dynamics package for mesoscopic and atomistic simulation, and can now be run in multiple CPU architectures.

Over the years, oxDNA has been used extensively to study basic biophysical mechanisms that are exploited in DNA nanotechnology, such as DNA hybridisation [5], hairpin formation [6] and strand displacement [7], as well as the self assembly and operation of various small DNA nanostructures [8, 9] and nanodevices [10–14]. OxDNA has been shown to provide surprisingly accurate predictions of experimental properties such as the effect of the toe-hold length on rates in displacement reactions, or else the effect of certain geometric constraints on the stability of more complex DNA nanostructures. Given this success, we will use oxDNA in this thesis to provide input for more coarse-grained models that can access the time and length-scales needed to understand large-scale SST assembly.

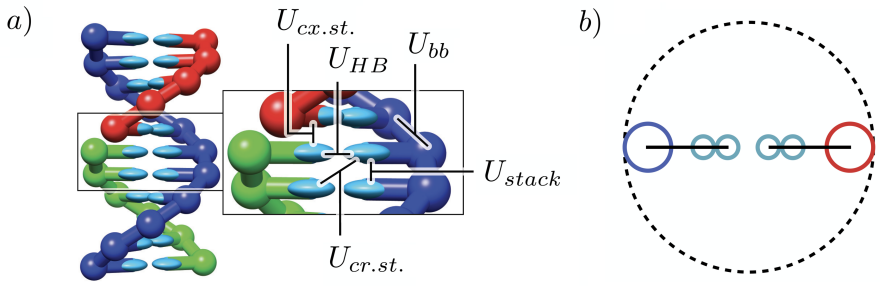


Figure 1.3: oxDNA representation of DNA. a) model double helix with effective potentials represented schematically. The backbone beads are represented as spheres, and the bases (containing two beads each) are shown as ellipsoids. Nucleotides also interact through excluded volume interactions (not shown in the picture). Adapted from Ref. [11]. b) Top-view cross section of the double helix showing one base pair. Each nucleotide is represented by three circles connected by a black line. The large circles represent backbone beads, whereas the small circles represent base beads. Adapted from Ref. [158].

2

Simulation methods (oxDNA)

Some of the results presented in this thesis are the result of simulations performed with oxDNA, a simulation package that implements a coarse-grained model of DNA introduced by Ouldridge and coworkers [156]. In this chapter, we describe the simulation and analysis techniques included in oxDNA that were used throughout this thesis.

2.1 Monte Carlo simulation

In statistical mechanics, the thermodynamic properties of a physical system are determined as ensemble averages over all points of the configurational space. For a N-body system with an Hamiltonian with the form $H(\boldsymbol{q}^N, \boldsymbol{p}^N) = T(\boldsymbol{p}^N) + U(\boldsymbol{q}^N)$, the ensemble average of an observable $A = A(\boldsymbol{q}^N)$ is given, in the canonical ensemble, as:

$$\langle A \rangle = \int d\boldsymbol{q}^N P_{eq}(\boldsymbol{q}^N) A(\boldsymbol{q}^N) = \frac{1}{\mathcal{Z}} \int d\boldsymbol{q}^N e^{-\beta U(\boldsymbol{q}^N)} A(\boldsymbol{q}^N), \tag{2.1}$$

where $\boldsymbol{q}^N$ and $\boldsymbol{p}^N$ are the generalised coordinates and momenta, U is the interaction potential, $\beta = 1/k_B T$, $\mathcal{Z} = \int d\boldsymbol{q}^N e^{-\beta U(\boldsymbol{q}^N)}$, and $P_{eq}(\boldsymbol{q}^N) = e^{-\beta U(\boldsymbol{q}^N)}/\mathcal{Z}$ is the canonical equilibrium distribution. In many systems, however, the integral in Eq. 2.1 cannot be solved analytically. Monte Carlo (MC) algorithms provide approximate, numerical solutions to the integral in Eq. 2.1. MC algorithms generate

a sequence of configurations $\{\boldsymbol{q}^N\}$ via some Markov process that has $P_{eq}(\boldsymbol{q}^N)$ as its unique steady-state distribution. A Markov process is fully defined by its transition probabilities $\pi(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$, which correspond the probability that a transition to a configuration $\boldsymbol{q}'^N$ will be chosen given $\boldsymbol{q}^N$. To ensure that a Markov process reaches $P_{eq}(\boldsymbol{q}^N)$ asymptotically, it is customary to fix its transition probabilities using the condition of detailed balance (see e.g. Ref. [160]):

$$P_{eq}(\boldsymbol{q}^N)\pi(\boldsymbol{q}^N \to \boldsymbol{q}'^N) = P_{eq}(\boldsymbol{q}'^N)\pi(\boldsymbol{q}'^N \to \boldsymbol{q}^N). \tag{2.2}$$

It is often convenient to split $\pi(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$ into two independent terms: a trial probability, $P_{\text{tr}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$, which gives the probability that a trial move $\boldsymbol{q}^N \to \boldsymbol{q}'^N$ will be proposed given $\boldsymbol{q}^N$, and an acceptance probability, $P_{\text{acc}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$, which gives the probability that, if proposed, that trial move is indeed accepted, such that $\pi(\boldsymbol{q}^N \to \boldsymbol{q}'^N) = P_{\text{tr}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)P_{\text{acc}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$. Eq. 2.2 can then be written as:

$$\frac{P_{\text{tr}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)P_{\text{acc}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)}{P_{\text{tr}}(\boldsymbol{q}'^N \to \boldsymbol{q}^N)P_{\text{acc}}(\boldsymbol{q}'^N \to \boldsymbol{q}^N)} = e^{-\beta(U(\boldsymbol{q}'^N)-U(\boldsymbol{q}^N)}. \tag{2.3}$$

The choice of the probability distributions $P_{\text{tr}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$ and $P_{\text{acc}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N)$ is arbitrary provided Eq. 2.3 is satisfied. For instance, in the widely used Metropolis algorithm [161] it is assumed that $P_{\text{tr}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N) = P_{\text{tr}}(\boldsymbol{q}'^N \to \boldsymbol{q}^N)$, and the acceptance probabilities are chosen as:

$$P_{\text{acc}}(\boldsymbol{q}^N \to \boldsymbol{q}'^N) = \min(1, e^{-\beta(U(\boldsymbol{q}'^N)-U(\boldsymbol{q}^N)}). \tag{2.4}$$

There are many approaches for generating trial moves. However, a bad choice can significantly undermine efficiency. For instance, in systems of strongly interacting particles, trial moves that involve the displacement of a single particle can result in large increases in internal energy, and consequently in high rejection rates. Also, if only single-particle moves are attempted, the diffusion of clusters of particles (which requires all particles to move in consort) tends to be suppressed. A few cluster algorithms have been developed that, according to the conditions of the system,

allow to generate trial moves where multiple particles are displaced at a time [162–164]. oxDNA uses one of such algorithms - the virtual-move Monte-Carlo (VMMC) algorithm [164], developed by Whitelam. Starting from a particle chosen at random in the system - taken as the cluster "seed" - the VMMC algorithm recursively adds new particles to the cluster depending on their interaction energies with particles already in the cluster, such that particles interacting strongly will have a high chance of being grouped together. A move is then attempted where the resulting cluster is displaced and rotated by a random (small) amount. An advantage that VMMC offers over previous cluster algorithms is that in VMMC, clusters moves are proposed based on interaction energies between particles both before, and after the trial move is attempted (rather than just before). As a consequence, VMMC tends to be more sensitive to the underlying energy landscapes, and therefore provides a better approximation to the real dynamics of the system.

2.2 Umbrella Sampling

Systems with complex free-energy landscapes can be challenging to sample using standard MC algorithms. The reason for this is that a simulation can become trapped in a local free-energy minimum for a long period of time. As a result, relevant regions of the phase space may be left improperly sampled. Umbrella sampling (US) is a simulation technique that makes transitions between local minima more likely to occur by artificially flattening out the barriers that may exist between these minima [165]. Implementing US involves specifying an (in general multi-dimensional) order parameter $Q = Q(\boldsymbol{q}^N)$, defined in terms of some microscopic property of interest. Instead of sampling from the canonical distribution $P_{eq}(\boldsymbol{q}^N)$, US simulations samples from a biased distribution $P_w(\boldsymbol{q}^N) = P_{eq}(\boldsymbol{q}^N)w(Q)$, where $w(Q)$ is a biasing factor. This is usually achieved by changing the functional of the acceptance probability ($P_{\text{acc}}(\boldsymbol{q}^N \rightarrow \boldsymbol{q}'^N)$) used by the MC algorithm. For instance, in Metropolis MC, sampling from $P_w(\boldsymbol{q}^N)$ can be achieved by introducing a multiplicative factor $w[Q(\boldsymbol{q}')]/w[Q(\boldsymbol{q})]$ into the second clause in the RHS of Eq. 2.4. Once a US simulation is finished, the unbiased distribution $P_{eq}(Q)$ can be retrieved by solving

$$p_w(Q) = f_w p_{eq}(Q) w(Q), \tag{2.5}$$

with respect to $P_{eq}(Q)$, where $p_w(Q)$ is the (biased) probability distribution obtained in the simulation, and f_w is a normalisation factor that depends on the choice of $w(Q)$. The factors $w(Q)$ can in principle be chosen such that a simulation will sample all coordinates of Q equally (hence $p_w(Q) = \text{const}$). To determine the ensemble average of an observable $A(\boldsymbol{q}^N)$, one first computes $\langle A(\boldsymbol{q}^N)/w(Q)\rangle_w$ during a run, and then unbiases the result with respect to $w(Q)$ at the end. It is easy to show that this satisfies Eq. 2.1:

$$\langle A\rangle = \frac{\int d\boldsymbol{q}^N A(\boldsymbol{q}^N) e^{-\beta U(\boldsymbol{q}^N)}}{\int d\boldsymbol{q}^N e^{-\beta U(\boldsymbol{q}^N)}} = \frac{\int d\boldsymbol{q}^N \frac{A(\boldsymbol{q}^N)}{w(Q)} w(Q) e^{-\beta U(\boldsymbol{q}^N)}}{\int d\boldsymbol{q}^N \frac{1}{w(Q)} w(Q) e^{-\beta U(\boldsymbol{q}^N)}} = \frac{\langle A/w\rangle_w}{\langle 1/w\rangle_w}. \tag{2.6}$$

2.3 Single-Histogram Reweighting

It is often relevant to know how the ensemble average of an observable $A(\boldsymbol{q}^N)$ changes with temperature. The single-histogram reweighting (SRH) method, developed by Ferrenberg [166], allows to extrapolate results from a simulation performed at a temperature T_0 to a different temperature T. Implementing SRH requires grouping configurations $\boldsymbol{q}^N$ according to their internal energies $E = U(\boldsymbol{q}^N)$, and the value of $A = A(\boldsymbol{q}^N)$, such that a simulation performed at T_0 yields a probability distribution $p(E, A; T_0)$. The density of states $\Omega(E, A)$ (assumed to be temperature independent) can be inferred from $p(E, A; T_0)$:

$$p(E, A; T_0) \propto e^{-\beta_0 E} \Omega(E, A). \tag{2.7}$$

Eq. 2.7 allows to fix ratios of $\Omega(E, A)$ for different values of E and A (the proportionality constant in Eq. 2.7 is usually not known, nor is it needed). A probability distribution $p(E, A; T)$ at a temperature T can easily be written in terms of $p(E, A; T_0)$:

$$p(E,A;T) \propto e^{-\beta E}\Omega(E,A) = e^{-\beta_0 E}e^{-(\beta-\beta_0)E}\Omega(E,A) \propto p(E,A;T_0)e^{-(\beta-\beta_0)E}, \tag{2.8}$$

where both proportionality constants in Eq. 2.8 have the same value. The ensemble average of A at a temperature T can then be determined as:

$$\langle A(T)\rangle = \frac{\int\int p(E,A;T)AdEdA}{\int\int p(E,A;T)dEdA} = \frac{\int\int p(E,A;T_0)Ae^{-(\beta-\beta_0)E}dEdA}{\int\int p(E,A;T_0)e^{-(\beta-\beta_0)E}dEdA}. \tag{2.9}$$

2.4 Weighted Histogram Analysis Method

Even given the speed-ups that umbrella sampling provides, some systems are too complex to be sampled effectively in a single simulation. In such cases, it is convenient to distribute the computational load through multiple windows, where each window covers a subset of values of an order parameter Q of interest, and combine the results at the end using some suitable method. Here, we use the weighted histogram analysis method (WHAM) by Kumar [167]. In WHAM, the equilibrium probability distribution $p_0(Q)$ that spans the whole range of Q in study is given as a weighted average of the (unbiased) probability distributions $p_0^{(i)}(Q)$ obtained in each window i:

$$p_0(Q) = \sum_i c_i(Q)p_0^{(i)}(Q), \tag{2.10}$$

where the weights $c_i(Q)$ are constrained by the normalisation condition $\sum_i c_i(Q) = 1$, and are fixed such that the expected variance in $p_0(Q)$ is minimised. This leads to (see e.g. Ref. [168]):

$$c_i(Q) = \frac{N_i w_i(Q) f_i}{\sum_j N_j w_j(Q) f_j}, \tag{2.11}$$

where N_i is the number of sampling MC moves performed in window i, $w_i(Q)$ is the biasing factor used in window i, and f_i is a normalisation constant. Each constant

f_i is determined by imposing the normalisation condition on the biased probability distribution $p_w^{(i)}$ obtained in the respective window i:

$$f_i^{-1} = \sum_Q w_i(Q) p_0(Q). \tag{2.12}$$

Eq. 2.12 can be easily verified by inserting a sum $\sum_Q$ on both side of Eq. 2.5, and equating the result to 1. In practice, Eqs. 2.10, 2.11 and 2.12 are solved self consistently with respect to $p_0(Q)$ and the normalisation constants f_i. Starting from a trial set of constants f_i (e.g. $f_1 = f_2 = \cdots = 1$), a first estimate of $p_0(Q)$ is determined by solving Eqs. 2.10 and 2.11. This first estimate of $p_0(Q)$ is then inserted into Eq. 2.12, and an improved set of constants f_i is determined. The procedure is repeated until a convergence criterion is met for the constants f_i.

3
Modelling Chemical Kinetics

Some of the key results presented in this thesis are the result of simulations of kinetic models of SST assembly. These simulations, discussed in detail in chapters 5 and 6, use stochastic algorithms based on the Gillespie algorithm [169, 170], and take as basic biophysical inputs a series of stochastic rate constants. Stochastic rate constants correspond to probabilities per unit of time that certain discrete events (e.g. the binding of a SST to a different SST) will take place in a finite system, given its current state and conditions. However, for reasons that will be made clear in chapter 4, these stochastic rate constants were parametrised using a series of deterministic rate constants, determined (mostly) from simulations with oxDNA, as a reference. Deterministic rate constants describe continuous changes in thermodynamically large systems, and are generally different from stochastic rate constants. However, provided certain assumptions apply, it can be shown that deterministic and stochastic rate constants relate to each other via statistical factors and unit conversions.

We start this chapter by discussing general properties of the chemical systems we attempt to describe (section 3.1). We then give a brief introduction to the continuous-deterministic formulation of chemical kinetics (section 3.2), with a focus on how the forward and backward rate constants of a certain reaction relate with the thermodynamic driving force of that reaction (section 3.2.1). A brief

introduction to the discrete-stochastic formulation of chemical kinetics is then provided (section 3.3), followed by a brief discussion on the functional form of the reaction propensity (section 3.3.1). We then proceed to derive a simple relationship between deterministic and stochastic rate constants (section 3.3.2). Finally, in section 3.3.3, we introduce the Gillespie algorithm [169, 170], often referred to as kinetic Monte Carlo (kMC) algorithm.

3.1 The Chemical System

In this chapter, we consider a chemical system containing molecules of m different chemical species X_i ($i = 1, \cdots, m$) in aqueous solution, where the solvent (water) is assumed unreactive and much more abundant than the solute, and acts as a thermal reservoir that maintains the reacting part of system at a fixed temperature T. The system is constrained to a (fixed) volume V and contains N_i molecules of each species X_i. These molecules can react with one another through a reaction channel of the type:

$$\sum_i \nu_{\alpha_+,i} X_i \underset{\kappa_{\alpha_-}}{\overset{\kappa_{\alpha_+}}{\rightleftharpoons}} \sum_i \nu_{\alpha_-,i} X_i, \tag{3.1}$$

where $\alpha_\pm$ specifies both the reaction channel (α) and the direction of the reaction ($\pm$), $\nu_{\alpha_\pm,i}$ are stoichiometric coefficients that give the number of molecules of X_i involved in $\alpha_\pm$, and $\kappa_{\alpha_\pm}$ are reaction rate constants. Reactive collisions are assumed to be much rarer than non-reactive ones, such that the system can be considered to be homogeneous.

3.2 Deterministic formulation of reaction kinetics

Deterministic reaction kinetics (DRK) describes the time evolution of a chemical system in the thermodynamic limit [1], where statistical fluctuations in the number of molecules of each species X_i (ΔN_i) become vanishingly small relative to the

[1] $V \to \infty$, $N_i \to \infty$, such that $N_i/V < \infty$.

average number of molecules of that species ($\overline{N}_i$). In DRK, each chemical species X_i is assumed to be sufficiently abundant such that its molar concentration $c_i = N_i/(N_A V) \approx \overline{N}_i/(N_A V)$ can be treated as continuous variable, where N_A is the Avogadro constant. The state of a system, in DRK, is specified by the concentration of each chemical species $c = (c_1, \cdots, c_m)$. In DRK, reaction rates of elementary reactions, $r_{\alpha_\pm}$, are usually assumed to obey the law of mass action (see e.g. Ref. [171]):

$$r_{\alpha_\pm} = k_{\alpha_\pm} \prod_i c_i^{\nu_{\alpha_\pm, i}}, \tag{3.2}$$

where $k_{\alpha_\pm}$ are deterministic rate constants which depend on the properties of the reaction α (e.g. activation energies, thermodynamic drive). The time evolution of the chemical system is in turn governed by a system of ordinary differential equations of the type:

$$\begin{aligned} \dot{c}_1 &= \phi_1(c_1, \cdots, c_m) \\ \dot{c}_2 &= \phi_2(c_1, \cdots, c_m) \\ &\vdots \\ \dot{c}_m &= \phi_m(c_1, \cdots, c_m) \end{aligned} \tag{3.3}$$

where the functions ϕ_i give the balance between the rate of production and consumption of X_i (in molar units, per unit of volume and per unit of time) given a system in state $c = (c_1, \cdots, c_m)$. We note that the DRK formulation does not contain in it any element of uncertainty. It describes a chemical system where concentrations change smoothly and continuously over time. Importantly, in the DRK formulation, the state of the system at any given time t can be predicted deterministically, provided the state of the system was known at a different time t_0.

3.2.1 Relationship between deterministic rate constants and thermodynamic driving force

In what follows, we will assume that at long times, the system will asymptotically reach an equilibrium state state, from now on referred to as *equilibrium*, where concentration currents $J_\alpha = r_{\alpha_+} - r_{\alpha_-}$ vanish for all reaction channels α. This leads to:

$$\frac{k_{\alpha_+}}{k_{\alpha_-}} = \frac{\prod_i c_{i,\mathrm{eq}}^{\nu_{\alpha_-,i}}}{\prod_i c_{i,\mathrm{eq}}^{\nu_{\alpha_+,i}}}, \tag{3.4}$$

where $c_{i,\mathrm{eq}}$ is the equilibrium concentration of species X_i. We now characterise the equilibrium state in terms of the thermodynamic forces that drive the system. Under the assumption of ideal dilute conditions, the chemical potential of a species X_i is given by:

$$\mu_i = \mu_i^{\ominus} + RT \ln\left(\frac{c_i}{c_{\mathrm{tot}}}\right), \tag{3.5}$$

where $\mu_i^{\ominus}$ is the chemical potential of X_i at the standard state [2], R is the ideal gas constant, $c_{\mathrm{tot}} = c_{\mathrm{sol}} + \sum_i c_i$ is the total concentration of the system, and c_{sol} is the concentration of solvent. Since the solvent is much more abundant than the solute ($c_{\mathrm{sol}} >> \sum_i c_i$), we can make the approximation $-RTc_{\mathrm{tot}} \approx -RTc_{\mathrm{sol}}$ and absorb this term, as customary, into $\mu_i^{\ominus}$. This leads to:

$$\mu_i = \mu_i^{\ominus\prime} + RT \ln\left(\frac{c_i}{c_0}\right), \tag{3.6}$$

where the term $c_0 = 1$ mol/dm^3 was inserted in the RHS of Eq. 3.6 so that the argument of the logarithm remains dimensionless. The thermodynamic driving force of a chemical reaction, also called the Gibbs free energy (ΔG), is the chemical potential difference between products and reactants. For a reaction of the type in Eq. 3.1 we have:

[2] $P^{\ominus} = 1$ bar, $c_i^{\ominus} = 1$ mol/dm^3

$$\Delta G_\alpha = \sum_i (\nu_{\alpha_+,i}\mu_i - \nu_{\alpha_-,i}\mu_i). \tag{3.7}$$

At equilibrium we must have $\Delta G_\alpha = 0$. Combining Eq. 3.7 and Eq. 3.5 leads to:

$$\Delta G_\alpha^{\ominus} = -RT \sum_i (\nu_{\alpha_+,i} - \nu_{\alpha_-,i}) \ln\left(\frac{c_{i,eq}}{c_0}\right), \tag{3.8}$$

where $\Delta G_\alpha^{\ominus} = \sum_i \nu_{\alpha_+,i}\mu_i^{\ominus'} - \sum_i \nu_{\alpha_-,i}\mu_i^{\ominus'}$ is the standard-state Gibbs free energy of the reaction α. Combining Eq. 3.4 and 3.8 leads to:

$$\frac{k_{\alpha_+}}{k_{\alpha_-}} = e^{-\frac{\Delta G_\alpha^{\ominus}}{RT}} (c_0)^{n_\alpha}, \tag{3.9}$$

where we have defined $n_\alpha = \sum_i (\nu_{\alpha_-,i} - \nu_{\alpha_+,i})$. To determine equilibrium concentrations one simply substitutes the LHS of Eq. 3.4 by the RHS of Eq. 3.9.

3.3 Stochastic formulation of reaction kinetics

Stochastic reaction kinetics (SRK) is a framework for describing chemical systems which takes explicitly into account the discrete and random nature of chemical reactions. In SRK, the state of a chemical system, such as described in section 3.1, is specified by the number of molecules of each chemical specie, $N = (N_1, \cdots, N_m)$, present. The time evolution of the system is in turn described as a continuous-time Markov jump process between these states, where each jump corresponds to a single firing in a reaction channel of the type in Eq. 3.1. Despite being more computationally intensive than DRK, SRK is necessary to describe systems where the number of molecules is low, and the continuous approximation that underlies DRK will consequently break down. In such systems, molecular populations can display large fluctuations around average values, which can lead to significant changes (both qualitative and quantitative) in the overall behaviour of the system relative to that predicted in DRK.

The fundamental hypothesis of SRK, first formulated by Gillespie [169], states that for a chemical system in a state N:

- the probability that a single firing of a reaction $\alpha_{\pm}$ will occur within a small time interval δt is given, to first order in δt, by $a_{\alpha_{\pm}}(N)\delta t$, where $a_{\alpha_{\pm}}(N)$ defines the *reaction propensity* of that reaction.
- the probability that two or more firings of a reaction $\alpha_{\pm}$ will occur within the same interval δt is given, to first order in δt, by $o(\delta t)$, where $o(\delta t)$ satisfies $o(\delta t)/\delta t \to 0$ as $\delta t \to 0$.

From this hypothesis, the chemical master equation (CME) - the fundamental equation of the SRK formulation - can be derived [172] using elements from probability theory:

$$\dot{P}(N,t) = \sum_{\alpha_{\pm}} \Big(a_{\alpha_{\pm}}(N + n_{\alpha_{\mp}}) P(N + n_{\alpha_{\mp}}, t) - a_{\alpha_{\mp}}(N) P(N,t) \Big), \tag{3.10}$$

where $P(N,t)$ is the probability of finding the system in a state N at time t, and $n_{\alpha_{\mp}} = (\Delta N_{1,\alpha_{\mp}}, \cdots, \Delta N_{m,\alpha_{\mp}})$ specifies a change in molecular population levels due to a firing of $\alpha_{\mp}$.

3.3.1 Functional form of the reaction propensity

As shown by Gillespie [169, 170], reaction propensities $a_{\alpha_{\pm}}(N)$ can be written as a product of two independent terms: a stochastic rate constant, $c_{\alpha_{\pm}}$, which depends on the physical properties of the reaction (e.g. activation energy), and a statistical factor $h_{\alpha_{\pm}}(N)$, which gives the number of different ways in which one firing of $\alpha_{\pm}$ can occur:

$$a_{\alpha_{\pm}}(N) = c_{\alpha_{\pm}} h_{\alpha_{\pm}}(N) = c_{\alpha_{\pm}} \prod_{i=1}^{m} \frac{N_i!}{(N_i - \nu_{\alpha_{\pm},i})! \nu_{\alpha_{\pm},i}!}. \tag{3.11}$$

In Eq. 3.11 we assume that there is a sufficient number of molecules in the system to produce one firing of $\alpha_{\pm}$ (otherwise we have $h_{\alpha_{\pm}}(N) = 0$). The statistical factor $h_{\alpha_{\pm}}(N)$ in Eq. 3.11 can be conveniently developed as:

$$a_{\alpha_{\pm}}(N) = c_{\alpha_{\pm}} \frac{1}{\prod_{j=1}^{m} \nu_{\alpha_{\pm},j}!} \prod_{i=1}^{m} \prod_{\nu=0}^{\nu_{\alpha_{\pm}}-1} (N_i - \nu). \tag{3.12}$$

3.3.2 Relationship between stochastic and deterministic rate constants

In this section we attempt to derive a relationship between stochastic ($c_{\alpha_\pm}$) and deterministic ($k_{\alpha_\pm}$) rate constants. In what follows, we will consider a system where the molecular populations are sufficiently large that the reaction propensities $a_{\alpha_\pm}(N)$ remain approximately constant within a small time interval τ, even if multiple firings of $\alpha_\pm$ happen to occur during τ. The probability distribution for the number of firings of $\alpha_\pm$ during τ, n_f, is therefore given by a Poisson distribution:

$$P(n_f = n) = e^{-a_{\alpha_\pm}(N)\tau}\frac{(a_{\alpha_\pm}(N)\tau)^n}{n!}, \tag{3.13}$$

with mean value $\langle n_f \rangle = a_{\alpha_\pm}(N)\tau$. We now recall that in DRK, reaction rates $r_{\alpha_\pm}$ correspond to the average number of firings of $\alpha_\pm$ in molar units, per unit of volume and per unit of time. To make a comparison with Eq. 3.2, we therefore compute the quantity $\langle n_f \rangle/(\tau N_A V)$:

$$\frac{\langle n_f \rangle}{\tau N_A V} = \frac{a_{\alpha_\pm}}{N_A V} = c_{\alpha_\pm}\frac{(N_A V)^{|\nu_{\alpha_\pm}|-1}}{\prod_{j=1}^{m}\nu_{\alpha_\pm,j}!}\prod_{i=1}^{m}\prod_{\nu=0}^{\nu_{\alpha_\pm,i}-1}\left(c_i - \frac{\nu}{N_A V}\right), \tag{3.14}$$

where we have defined $|\nu_{\alpha_\pm}| = \sum_i \nu_{\alpha_\pm,i}$, and converted numbers of molecules (N_i) to molar concentrations ($c_i = N_i/N_A V$). We next absorb unit conversions and statistical factors into the the rate constant $c_{\alpha_\pm}$:

$$c'_{\alpha_\pm} \equiv c_{\alpha_\pm}\frac{(N_A V)^{|\nu_{\alpha_\pm}|-1}}{\prod_{j=1}^{m}\nu_{\alpha_\pm,j}!}, \tag{3.15}$$

and develop Eq. 3.14 further as:

$$\begin{aligned}\frac{a_{\alpha_\pm}}{N_A V} &= c_\alpha^{\pm\prime}\prod_{i=1}^{m}\prod_{\nu=0}^{\nu_{\alpha,i}^{\pm}-1}\left(c_i - \frac{\nu}{N_A V}\right) = c'_{\alpha_\pm}\prod_{i=1}^{m}\left[(c_i)^{\nu_{\alpha_\pm,i}}\prod_{\nu=0}^{\nu_{\alpha_\pm,i}-1}\left(1 - \frac{\nu}{c_i(N_A V)}\right)\right] = \\ &= c'_{\alpha_\pm}\prod_{i=1}^{m}(c_i)^{\nu_{\alpha_\pm,i}}[1 + \mathcal{O}(V^{-1})].\end{aligned} \tag{3.16}$$

Note that the leading order term of Eq. 3.16 corresponds exactly to deterministic reaction rate $r_{\alpha_\pm}$ if $c'_{\alpha_\pm} = k_{\alpha_\pm}$. Furthermore, we have full equality between $a_{\alpha_\pm}/(N_A V)$ and $r_{\alpha_\pm}$ if a firing of $\alpha_\pm$ involves at most one molecule per chemical species, or in the large volume limit $V \to \infty$ (or both), in which case we can write:

$$c_{\alpha_\pm} = k_{\alpha_\pm} \frac{\prod_{i=1}^{m} (\nu_{\alpha_\pm,i})!}{(N_A V)^{|\nu_{\alpha_\pm}|-1}}. \tag{3.17}$$

In table 3.1 we give examples of the functional forms of $a_{\alpha_\pm}$, $c_{\alpha_\pm}$, $h_{\alpha_\pm}$, and $r_{\alpha_\pm}$ for a few elementary reactions.

Reaction type (α_+)	c_{α_+}	h_{α_+}	a_{α_+}	R_{α_+}
$X_i \xrightarrow{\kappa_{\alpha_+}}$ reaction products	k_{α_+}	N_i	$k_{\alpha_+} N_i$	$k_{\alpha_+} c_i$
$2X_i \xrightarrow{\kappa_{\alpha_+}}$ reaction products	$\frac{k_{\alpha_+} \times 2!}{(N_A V)}$	$\frac{N_i(N_i-1)}{2!}$	$k_{\alpha_+} \frac{N_i(N_i-1)}{(N_A V)}$	$k_{\alpha_+} c_i^2$
$3X_i \xrightarrow{\kappa_{\alpha_+}}$ reaction products	$\frac{k_{\alpha_+} \times 3!}{(N_A V)^2}$	$\frac{N_i(N_i-1)(N_i-2)}{3!}$	$k_{\alpha_+} \frac{N_i(N_i-1)(N_i-2)}{(N_A V)^2}$	$k_{\alpha_+} c_i^3$
$X_i + X_j \xrightarrow{\kappa_{\alpha_+}}$ reaction products	$\frac{k_{\alpha_+}}{(N_A V)}$	$N_i N_j$	$k_{\alpha_+} \frac{N_i N_j}{(N_A V)}$	$k_{\alpha_+} c_i c_j$

Table 3.1: Examples of functional forms of reaction propensity (a_{α_+}), stochastic reaction constant (c_{α_+}), statistical factor (h_{α_+}), and deterministic reaction rate (r_{α_+}) for simple elementary reactions. k_{α_+} denotes the deterministic reaction constant for a reaction α_+. Without loss of generality, we consider here reactions that occur in the forward (+) direction only.

3.3.3 Kinetic Monte Carlo (kMC) algorithm

Given an initial condition $P(N, t_0)$, the CME provides the means of predicting $P(N, t)$ at any different time t. In many cases however, the CME is far too complex to be solved via direct methods. The main problem is that the size of the state space can grow very rapidly with the number of molecules in the system and the complexity of the reaction network. Over the years, a number of algorithms have been developed that provide approximate, numerical solutions to the CME[3]. Perhaps the most widely used of such algorithms is the Gillespie algorithm [169, 170], also

[3]We refer the reader to Refs. [173, 174] for a thorough overview of these algorithms.

known as kinetic Monte Carlo (kMC) algorithm[4]. The kMC algorithm allows to generate sampling paths following exactly the same probabilistic rules that underly the CME. Such paths consists of a sequence of discrete steps where a single random firing of a reaction $\alpha_{\pm}$ is executed by the algorithm. Each of such steps also involves a random waiting time where the system remains in the same state, before the next firing is executed by the algorithm.

In what follows, we will simplify the notation and use the variable μ to specify a certain reaction channel and direction (for which we previously used the variable $\alpha_{\pm}$). To make it easier to understand how the kMC algorithm generates a random move and a waiting time, it is helpful to introduce the reaction probability density function $P(\tau, \mu|N)d\tau$, defined as the probability that given a chemical system in a state N at time t, the the next firing will be of type μ, and it will occur during an infinitesimal time interval between $t+\tau$ and $t+\tau+d\tau$ [170]. In other words, $P(\tau, \mu|N)d\tau$ corresponds to the probability that no firings (of any type) will occur from t to $t+\tau$, $P_0(\tau)$, multiplied by the probability that one firing of type μ will occur during the infinitesimal time interval $d\tau$ following $t+\tau$, $a_\mu d\tau$:

$$P(\tau, \mu|N)d\tau = P_0(\tau|N)a_\mu(N)d\tau. \tag{3.18}$$

For an arbitrarily small value τ,[5] it can be shown that [169, 170]:

$$P_0(\tau|N) = \exp\Big(-\sum_\nu a_\nu \tau\Big). \tag{3.19}$$

Therefore, we have

[4]It is worth noting that the kMC algorithm described here is equivalent to the n-fold way algorithm (also known as BKL algorithm) developed independently (and concurrently) by Bortz et al. to simulate Ising spin systems [175].

[5]such that the probability that two or more firing of μ will occur is zero. We recall that according to fundamental hypothesis of SRK (see section 3.3), to the first order in δt, the probability of having a single firing of μ is $a_\mu(N)\delta t$, whereas the probability of having two or more firings is $o(\delta t)$ with $o(\delta t)/\delta t \to 0$ as $\delta t \to 0$. Therefore δt can be made to be arbitrarily small, such that the probability of having two or more firings is effectively zero. The probability of not having *any* firing during an arbitrarily small time interval δt is therefore $P_0(\delta t|N) = \prod_\mu (1 - a_\mu(N)\delta t) \approx 1 - \sum_\mu a_\mu(N)\delta t \approx \exp\Big(-\sum_\mu a_\mu \delta t\Big)$, as in Eq. 3.19.

$$P(\tau, \mu|N) = a_\mu(N)\exp(-a_0(N)\tau), \tag{3.20}$$

where we have defined $a_0(N) = \sum_\nu a_\nu$. The final ingredient for a working algorithm is a method for drawing a random pair (μ, τ) from the distribution $P(\tau, \mu|N)$. Since $P(\tau, \mu|N)$ can be written as product of two independent functions $P_1(\tau|N)$ and $P_2(\mu|N)$, a random value for waiting time, τ, can be drawn from $P_1(\tau|N) = \exp(-a_0(N)\tau)$, and a random value for μ can be drawn (independently) from $P_2(\mu|N) = a_\mu(N)$. Given two random numbers $r_1, r_2 \in [0, 1]$, it turns out that a random a random pair (μ, τ) can be drawn from P_2 and P_1 through the prescriptions [170]:

$$\sum_{\nu=1}^{\mu-1} a_\nu < r_2 a_0 \leq \sum_{\nu=1}^{\mu} a_\nu, \tag{3.21}$$

and

$$\tau = \frac{1}{a_0} \ln \frac{1}{r_1}. \tag{3.22}$$

To summarise, starting from a state N at time t, the kMC algorithm performs the following operations

1. compute reaction propensities a_μ for all reactions μ permitted in the system given its current state

2. generate two random numbers r_1 and r_2 and draw a random pair (τ, μ) from the distributions $P_1(\tau|N)$ and $P_2(\mu|N)$ as prescribed in Eqs. 3.22 and 3.21

3. update the state of the system according to the changes that result from a firing of μ, and increase time by an amount τ

4. repeat.

4

Multiscale modelling of SST self assembly

4.1 Introduction

Since its inception, the SST approach has been attracting a growing interest from the scientific community. At a first glance, the SST method may appear to be straightforward since each subunit is designed to occupy a predetermined position in the target structure. However, in common with other examples of self-assembly of finite objects with highly specific interactions, the relative rates of nucleation and growth need to be carefully controlled. This means that the nucleation barrier needs to be low enough that it can be crossed on experimentally accessible time scales, whilst also being sufficiently large such that the growth time for a structure to reach its equilibrium shape is significantly shorter than the typical waiting times between nucleation events. If the latter does not hold, partially formed nuclei will grow in numbers that exceed the maximum number of structures that can reach completion given the number of monomers available in the mixture (monomer starvation).

Additionally, there are hurdles to error-free assembly that are specific to addressable assembly. First, the extra entropy associated with the large number of subunit types pushes the assembly transition down to lower temperatures compared to an otherwise equivalent one-component system [176]. Second, although incorrect binding is a serious problem for any kind of self-assembly, the number of pairs

of subunits between which a misbond can form is vastly increased. Thus, correct binding must be sufficiently strong to compensate for the vastly larger number of potential incorrect bonds. Although DNA provides an optimal assembly material with the requisite specificity of interactions, it is not inherently obvious that there will be an assembly window where the conditions that there is a sufficient thermodynamic driving force and yet misbonding is not prevalent are both satisfied. Thus, the success of the first demonstrations of the single-stranded tile (SST) method was conceptually somewhat surprising.

The need to explain the success of the SST assembly method inspired a range of important theoretical work. For instance, Reinhardt and Frenkel [177], using a three-dimensional lattice model where individual SSTs are treated as patchy particles with rigid patches representing binding domains, were the first to demonstrate that SST-based structures can self-assemble successfully, albeit in a narrow temperature window. The same lattice model was later used to explore the effect of the coordination number on the nucleation barriers and hence the assembly success [178]. An off-lattice model, which provides a somewhat more robust description of the translational and rotational entropy of SSTs, was also found to reproduce the main physical phenomena observed in the simpler lattice model [179].

In parallel to these investigations, theoretical calculations by Jacobs [180, 181] helped explain why a time-dependent annealing protocol is necessary to achieve a fully assembled structure. Due to the high combinatorial entropy of these systems, as the temperature is cooled down and the critical nucleation barrier is crossed, the free-energy minimum corresponds to a partially formed structure. Further cooling is necessary for structures to grow further and eventually reach completion. More recently [182], similar calculations were used to optimise the nucleation and growth behaviour of various periodic and aperiodic SST-based structures, via changes in the ligancy of each component of the target structure. In a different strand of research [183], Murugan employed various minimal models of self-assembly to investigate strategies to minimise the effect of misbonding and monomer depletion on assembly yields, showing that the use of carefully chosen non-stoichiometric

concentrations of assembly components can improve assembly yields and help preventing monomer depletion traps.

While the current body of theoretical work on SST-based self assembly has provided the field with important insights, the theoretical tools available are minimal in detail and neglect important aspects of SST systems, such as the chain-like nature of each individual SST, and are generally unable to provide quantitative predictions that can be directly tested against experiments. More detailed modelling would be beneficial to help guiding experiments, and to work out the relative importance and interplay of the various competing mechanisms that determine the success (or failure) of an assembly reaction.

4.1.1 A multiscale approach for modelling SST self assembly

The fundamental biophysical mechanism exploited by the SST assembly method is the hybridisation of complementary domains of separate SSTs. Here, we attempt to describe these mechanisms using oxDNA [1] (see section 1.3.3), a nucleotide-level model of DNA that has been used extensively to study biophysical mechanisms that are exploited in DNA nanotechnology, such as DNA hybridisation [5], hairpin formation [6] and strand displacement [7], as well as the self assembly and operation of various small DNA nanostructures [8, 9] and nanodevices [10–14].

On the other hand, one has to keep in mind that the time- and length- scales typically associated with SST assembly makes these systems vary challenging to study via direct simulations. SST objects produced in experiments contain hundreds (and even thousands) of different SSTs ($\sim$30-50 bases long), and usually require a few hours of annealing time to successfully form. For reference, the largest DNA nanostructure whose self assembly has been simulated directly using a CG model was a small DNA origami [8], containing a 384-base-long scaffold strand and twelve 32-base-long staple strands. To make it feasible to observe the assembly process in full (from monomer phase to completion), the staple attachment rate was artificially increased by fixing the concentrations of staple strands at a much higher value

(7.7 μM) than that used in experiments (50-200 nM). Even with these speedups, assembling a single target structure required several months of computer time on a high-performance GPU running oxDNA1 [156]. According to estimates provided in Ref. [8], such computing time translated into about one second of physical time.

If SST assembly is to be studied at the level of resolution CG models provide, alternative approaches to direct simulation need to be devised. Our understanding of the thermodynamics and kinetics of DNA hybridisation can provide important guidance in this task. On a thermodynamic level, DNA hybridisation is well described as a two-state transition between an unpaired (single-stranded) state and a fully-formed double helix, provided the sequences are not too long (< 20 bases) [31]. On a kinetic level, DNA hybridisation has conventionally been viewed as a two-step process consisting of a slow, rate-limiting nucleation step in which a few base pairs form, followed the rapid "zippering" of the remaining bases. This view of DNA hybridisation was first formalised by Wetmur et al. in 1968 [184], in what become known as the "zipper" model. In the following years the zipper model was employed with some success to fit experimental hybridisation-melting curves of short DNA double helices [184–186]. In these early works it was first noted that melting rates obey simple simple first-order kinetics and are well fit by a two-state Arrhenius model. Hybridisation rates on the other hand, were found to obey simple second-order kinetics, hence confirming the nucleation step as the main rate-limiting factor of DNA hybridisation, and to display anti-Arrhenius behaviour - indicative of the existence of a metastable intermediate (negative hentalpy) at the nucleation step.

More recent research has elaborated on this initial view of DNA hybridisation. For instance, it is now understood that a fully-formed double helix can interconvert rapidly with "frayed" configurations where a few base pairs at the terminus of the helix spontaneously dissociate [187, 188]. Various short-lived intermediates have also been identified between single- and double-helical configurations [189, 190]. In addition to anti-Arrhenius behaviour, Arrhenius-like [191, 192] and non-Arrhenius behaviour [193] has also been observed in experiments for hybridisation rates. Computer simulations have provided important complements to the experimental

literature. In addition to the nucleation-plus-zippering mechanism, simulations have revealed that DNA hybridisation can proceed through "slithering" pathways [194], where two strands forming a partially formed helix slide in opposite directions, by breaking existing base pairs and forming new ones, to further extend the helix; or via pseudoknot and inchworm internal displacement pathways [5], by which two sequences initially forming non-native contacts undergo a series of complex rearrangements that eventually lead to a fully-formed helix. While much research has been carried out since the zipper model was initially proposed, the conventional view of a slow, rate-limiting nucleation step, followed by the fast propagation of the helix, still remains mostly accepted. On a thermodynamic level, this translates into free-energy landscapes that, post nucleation, are strongly funnelled down towards the native double-stranded state, where barriers that segregate local minima are not substantial and do not overtly frustrate hybridisation [5, 189, 194, 195]. Interestingly, this leads to a timescale separation between fast processes that occur at the bottom of the free-energy basin, and the much slower escaping from that basin through a sufficiently large thermal fluctuation. Conversely, the nucleation barrier imposes a separation of timescales between fast dynamics in the unpaired (single-stranded) ensemble, and nucleation events, which requires a rare reactive collision between two independently diffusing strands.

The fact that there is a timescale separation between short-timescale dynamics within single-stranded and double-helical ensembles, and transitions between both these ensembles, is the reason why studying of DNA hybridisation with CG model is often very challenging, and enhanced sampling techniques are usually required for this task [5, 146, 194]. On the other hand, the fact that both ensembles (single-stranded and double-helical) are segregated by thermodynamic barriers is also what can enable an alternative, more efficient strategy to direct simulation. Given the long lifetimes of both ensembles, it should be a good assumption that once a system finds itself in one of them, any memory of previous events will be lost before the next transition takes place. In this case, the problem of describing transitions between single- and double-stranded states simplifies into the problem of

computing transition probabilities of a two-state Markov model. These observations are particularly relevant if one's goal is to study self-assembling systems that involve thousands of consecutive hybridisation events.

In what follows, we will assume that association (dissociation) rates of a given SST to (from) a structure depends on the set of SSTs present in the structure but not on the order in which they were incorporated. From a practical viewpoint, this approximation means that instead of having to perform lengthy simulations that cover the assembly process in full, we can instead use simulations as a parametrisation for a coarser Markovian model that describes self assembly at the level of the SST. However, one should keep in mind that, given the size of the structures whose self-assembly we aim to describe (which contain hundreds (and even thousands) of different SSTs), it would be very difficult to even list all configurations that can possibly form, let alone compute transition rates (via expensive CG simulations) for each one of them. We note that in SST assembly, association and dissociation rates of SSTs are expected to depend mostly on local factors as, much like the components of a classic crystal, SSTs are designed to form bonds exclusively with nearest neighbours. To keep computational costs manageable we make a second assumption that the association (dissociation) rates of a SST to (from) a certain binding site in a structure depends on the local properties of the binding site, but is otherwise independent from the state of the structure. To parametrise transition rates, we first identify all different types of binding site to which a SST can associate to (or dissociate from). We consider two binding sites to be different if there is at least one first- or second-neighbour that is different. For each type of binding site identified, we compute association and dissociation rates with the aid of computer simulations with the oxDNA model.

4.2 Model System

As test system for our approach we consider a rectangular-shaped structure reported in Refs. [2, 3] called "R6" (see Fig. 4.1). R6 contains 334 unique SSTs, each 42 bases long, adding up to 14,028 nucleotides in total. In R6, SSTs are arranged in

23 rows; odd rows contain 15 SSTs, and even rows contain 14 SSTs. Each SST contains four concatenated binding domains of lengths 10-11-11-10 bases (odd rows) or 11-10-10-11 bases (even rows), that bind with up to four local neighbours during self-assembly. SSTs in the corner of the structure bind only with one SST, SSTs on the borders can bind with either two or four other SSTs, and internal SSTs bind with four other SSTs. In R6, each SST assumes a U-shaped conformation that branches towards two adjacent, anti-parallel helices, while also acting as a crossover that helps keeping both helices aligned. Once assembled, R6 consists of 21 double-helical chains of alternating polarity joined together by crossovers located at periodic intervals of 21 base pairs. This positioning of the crossovers (21 base pairs corresponds to roughly two full helical turns) ensures a dihedral angle between consecutive double-helical chains that is close to 180°, which ensures that the R6 structure does not overtly curve upon itself. In the appendix A we show all 32 different types of binding site, $\mu = 1, \cdots, 32$, identified for R6.

In the experiments, R6 was produced by mixing SSTs at a concentration of 10^{-7} M per SST type at high salt conditions ($[Mg^{2+}]$=25 mM). Experimental samples where R6 was produced contained about one microgram of DNA material. The samples were subject to two consecutive thermal ramps. The first ramp cooled the samples from 90° to 61°C, stepwise ($\Delta T = 1°$ C), at a rate of 5 min/°C. A second ramp cooled the samples from 60° to 25°C, also stepwise ($\Delta T = 1°$ C). Different rates were attempted for the second ramp, ranging from 5 min/°C to 120 min/°C. Native agarose gel electrophoresis [1] was applied to each sample to estimate the resulting yield. The yield was calculated as the fraction of DNA material in complexes with sizes compatible with that of R6. The assembly yield shown a mild increase with decreasing cooling rate, ranging from 6.9% for 5 min/°C, to 11.2% for 120 min/°C. For all cooling rates, the vast majority of DNA material that was not utilised to produce R6 ended up in aggregates with sizes up to to about four times the size of R6, with a small component found in unbound SST and small structures (dimers, trimers, etc.) with sizes up to about 10% of the size of R6.

[1] A method for separating a population of DNA molecules and complexes according to their molecular sizes.

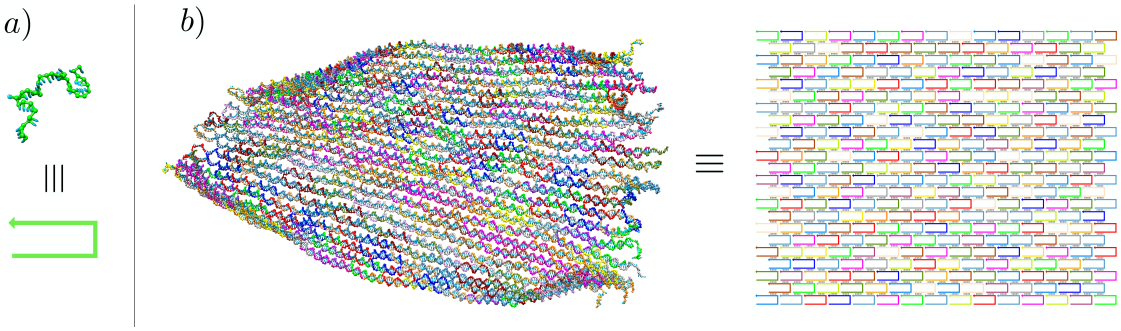


Figure 4.1: Coarse-grained representation of the R6 structure reported in Refs. [2, 3]. a) oxDNA and simplified representation of a 42-base-long SST. b) oxDNA and simplified representation of R6. Each SST in the figure is represented with an arrow specifying the directionality of the respective DNA chain (from the 5'-end to the 3'-end).

4.3 State definition and parametrisation of association/dissociation probabilities

We will neglect, for now, possible assembly errors due to incorrectly-, or incompletely-bonded SSTs. Therefore, the state of a partially formed structure, from now on also referred to as a fragment (of R6), can be specified by a vector with 334 components:

$$\vec{f} = (f_1, f_2, \cdots, f_{334}), \tag{4.1}$$

where $f_i = 1$ if an SST of a type i is present in the fragment and $f_i = 0$ otherwise, and 334 is the number of SSTs that make up R6. In what follows we will assume that a fragment can grow (and shrink) through a sequence of discrete events where a single SST either binds to (or melts from) the fragment. For consistency however, events where the removal of a SST from a fragment in a state $\vec{f}^{(1)}$ would divide it into two disconnected, smaller fragments in states $\vec{f}^{(2)}$ and $\vec{f}^{(3)}$ will be attributed, by default, a probability zero, since the reverse process ($\vec{f}^{(2)} + \vec{f}^{(3)} \rightarrow \vec{f}^{(1)}$) of fragment assembly is ignored in this formalism because it is extremely unlikely to occur. As discussed in section 4.2, a fragment can exhibit 32 different types of binding site, $\mu = 1, \cdots, 32$, to which a SST can bind to, or melt from (see appendix A). For each type μ, we aim to parametrise a pair of stochastic rate constants: a binding rate constant $c_+(\mu)$ and a melting rate constant $c_-(\mu)$. Both $c_+(\mu)$ and

$c_-(\mu)$, we recall, are related to their deterministic counterparts k_+, k_- through unit conversions and statistical factors (see section 3.3.2).

4.3.1 Parametrisation of $c_+(\mu)$

DNA hybridisation rate constants (k_+) are known to be far less dependent on temperature and conditions that melting rate constants (k_-). To keep the parameterisation of the model at a minimum, we assume that (k_+) is both temperature and binding-site independent, as a first approximation. To fix a numeric value for k_+, we consider a study by Zhang [196], where toehold-mediated strand displacement (TMSD) was characterised experimentally for various toehold lengths, and compared against predictions of a kinetic model where TMSD is treated as a simple bimolecular process. In Ref. [196], the concentration of one of the reaction products (complex Y) was measured in real time, and the resulting concentration profile was used to fit rate constants employed by the kinetic model. Of particular interest is the observation that for long toeholds ($\sim$7-8 bases and longer), the forward rate constant k_+^{km} of the kinetic model approximates the rate constant of the first step of TMSD, namely the hybridisation of a binding domain of an initially-diffusing single strand with the toehold domain. Given that the binding domains of SSTs have similar lengths (either 10 or 11 bases), we will fix k_+ at the value 6×10^6 M^{-1}, as estimated in Ref. [196] for k_+^{km} for long toeholds (>7-8 bases).

Furthermore, it will be assumed, in what follows, that the hybridisation of secondary domains during a binding event (SST-fragment) is a fast process compared to the hybridisation of the first domain, such that the binding event can be described as a bimolecular process with rate constant k_+. As shown in section 3.3.2, a rate constant c_+ can be derived from k_+ through simple unit conversions.

4.3.2 Parametrisation of $c_-(\mu)$

For a given binding site type μ, given $k_+(\mu)$, a sufficient ingredient to compute $k_-(\mu)$ (hence $c_-(\mu)$) is the standard-state free energy change $\Delta G^{\ominus}(\mu)$ (see Eq. 3.9) associated with the binding of a SST to binding site μ. We determine $\Delta G^{\ominus}(\mu)$ using

simulations with oxDNA1 [156] (see details in section 1.3.3). oxDNA1 is a version of oxDNA that allows for base pairing exclusively between WC-complementary bases (A-T, and C-G), but uses hydrogen-bonding and stacking potentials that are otherwise insensitive to base identity. The simulations were set up such that the formation of non-native hydrogen bonds is forbidden. This constraint means that we don't bias the results towards any particular sequence, and therefore don't have to repeat simulations for different combinations of sequences (one simulation per binding site suffices). For each type of binding site μ, we first determine an equilibrium probability distribution $P_{eq}^{(\mu)}(Q)$ (see annex B), where the order parameter Q specifies the number of formed base pairs between SST and binding site. The simulations were performed using the virtual-move Monte Carlo (VMMC) algorithm [164] (see details in section 2.1). To facilitate transitions between local minima we employed the umbrella sampling [165] scheme described in section 2.2. For a given type of binding site, the order parameter Q was chosen to specify the number of base pairs formed between each pair of binding domains (depending on the binding site there can be between one to four pairs of binding domains - see annex A). The order parameter Q, defined this way, was used to separate the phase space into regions biased differently by the biasing factor $W(Q)$. In the simulations, we considered a base pair to be formed if the respective interaction energy fell below the cutoff -0.596 kcal mol^{-1}, which corresponds to roughly 15% of the typical energy of a hydrogen bond. Due to the complexity of binding events where SSTs associate to binding sites through multiple binding domains, the order parameter was augmented with additional variables specifying the distance between the centre of mass of each binding domain in the SST, and the respective complementary domain in the binding site. To improve the efficiency, multiple sampling windows were used in each case, each covering a subset of values of the order parameter (and running on an independent CPU). The results were combined at the end using the weighted histogram analysis method (WHAM) [167], described in section 2.4.

All simulations were performed at T=50°C, close to the highest temperature where appreciable fragment growth was detected in experiments (T=54°C [3]).

The results were extrapolated to different temperatures using the single histogram reweighting (SHR) method [166] described in section 2.3. The simulation boxes were made sufficiently large such that periodic images of the same molecule (or complex) were unable to get in reach of each other.

To keep computational costs more manageable, we fixed the SST concentration (defined by the volume of the simulation box) at a higher value than that used in experiments (100 nM [2, 3]). Depending on the case, we used SST concentrations ranging from 4.38×10^{-6} M (for systems with the lowest number of nucleotides) and 1.72×10^{-7} M (for systems with the highest number of nucleotides). To extrapolate the results to other SST concentrations, we multiply the histogram of visits[2] for $Q = 0$ by a factor V/V_0, where V_0 is the volume of the simulation box and V is the volume that yields the desired SST concentration. For a given binding site μ, we consider that a SST is in a "bound" state if there is at least one base pair formed between the SST and a binding domain in μ. Therefore, we calculate a bound-state probability, $p^{(\mu)}_{b,eq}$, as $p^{(\mu)}_{b,eq} = 1 - p^{(\mu)}_{eq}(Q = 0)$. Note that states where all binding domains have hybridised and most base pairs possible are formed are essentially the only ones that contribute to $p^{(\mu)}_{b,eq}$ (see annex B). Finally, we compute a standard-state free-energy change (see annex C and D) as

$$\Delta G^{\ominus}(\mu) = -\beta^{-1} \ln \left\{ \frac{p^{(\mu)}_{b,eq}}{p^{(\mu)}_{eq}(Q = 0)} \frac{V_0}{V^{\ominus}} \right\}, \tag{4.2}$$

where $V^{\ominus}$ is the simulation box volume that yields a SST concentration of $c^{\ominus} = 1$ M.

One might ask whether a less computationally intensive approach could be used to estimate the free-energy changes $\Delta G^{\ominus}(\mu)$. For example, the nearest-neighbour (NN) model of SantaLucia [31] (the predictions of which were used to fit oxDNA) is regarded in the field as a good estimator of free-energy changes $\Delta G^{\ominus}$ associated with the hybridisation of small (<20 bases) DNA sequences. The NN model could in principle be used to estimate $\Delta G^{\ominus}(\mu)$ for binding sites with just one binding

[2] in which we record the number of times configurations with a given Q are visited during a run.

domain ($\mu = 1, \cdots, 6$). However, the NN model does not provide estimates for the free-energy barriers associated with the initiation of a second and subsequent domains (see annex B). The capturing of these barriers, which arise due to sudden losses of conformational entropy upon the initiation of secondary domains, in a physically motivated way, is the key extra ingredient that simulations with oxDNA provide. Similarly, we argue, this may be a key limitation of patchy-particle (pp) models of SST assembly [177, 178], which use the NN model to fix the strength of patch-patch interactions (in pp models, each patch represents a binding domain), but rely on simple geometric arguments to describe the initiation of secondary domains. These models do exhibit a free-energy barrier associated with forming secondary domains, but this barrier arises from the loss of rotational freedom of the patchy particle (e.g., a patchy particle with one bond formed can rotate freely around the axis defined by the bond, but becomes constrained once a second bond forms) and not from a realistic description of DNA's chain-like nature. In chapter 5, we will highlight some of the implications that a more structurally-rich description of the initiation of secondary domains (such as provided by oxDNA) has on the assembly pathways in SST assembly.

4.4 Summary and discussion

In this chapter we devise a multiscale modelling approach for SST assembly. As a test system we chose the rectangular-shaped R6 structure produced in Refs. [2, 3]. We specify the state of R6 (or a fragment of it) by the types of SSTs that are present in it, but not on the order in which they were incorporated. The growth of a fragment is described as a Markov jump process between fragment states, where in each "jump" a single SST can either associate to (or dissociate from) a binding site in a fragment. We make the simplification that association/dissociation rates depend only on the local properties of the binding site (we identify 32 different types of binding site for R6). This simplification drastically reduces the number parameters needed for a working model. For each type of binding site identified, we determine a free-energy change associated with the binding of a SST via simulations with oxDNA.

To keep computational costs manageable, we simulate binding events involving fragments containing a small number of SSTs. Implicit in our modelling approach is the assumption that the chemical activity of a binding site remains unchanged if additional SSTs happen to bind to the fragment at distant sites (further away than second neighbours of the binding site). A degree of correlation between the chemical activity of a binding site and the composition of the fragment is, of course, expected. We expect such correlations to be small, and arise mostly due to the volume excluded by distant SSTs. We performed several tests where, for a few types of binding site, the fragment used in the simulations was augmented with additional SSTs (in a way that kept the type of the binding site unchanged). Changing the composition of the fragments led to small changes (up to a few tenths of one k_BT) in the binding free energies predicted by oxDNA (which, depending on the type of binding site, range from about $-100\ k_BT$ to about $-10\ k_BT$ (see appendix D)).

We make the approximation that the SST association rate is independent from temperature and type of binding site, and fix the respective rate constant at 6×10^6 M^{-1}, a value estimated in Ref. [196] for the hybridisation of DNA sequences whose lengths are similar to that of a binding domain of a SSTs. Given an association rate constant and a binding free-energy, we derive a dissociation rate constant using elements from statistical thermodynamics.

In this chapter and throughout the rest of the thesis, we describe SST assembly as a non-hierarchical process driven by single SST-association. We expect hierarchical modes of assembly involving the fusion of two fragments into a larger, correctly-formed one to be relatively unimportant. This is because each subcomponent in the SST method is unique, and thus for two fragments to fuse correctly they would need to share fully-complementary facets and collide in specific orientations. We assume (productive) fragment-fragment collisions to be rare, and growth modes involving single-SST association to be much more rapid in comparison.

It should be noted that non-specific bonds (which can be SST-fragment or fragment-fragment) are well known to be a major source of assembly errors in SST assembly. As noted in Refs. [2, 3] (where R6 was assembled in experiments) as well

as in more recent realisations of SST assembly (see, e.g. Refs. [17, 18]), most SSTs in the samples typically end up in malformed aggregates with sizes above the size of the target structure. In section 5.3, a simple model will be introduced that describes non-specific interactions of the type SST-fragment. Our main finding is that these interactions are quite weak over a wide range of conditions. Non-specific fragment-fragment interactions are not well understood and present a serious challenge from a modelling perspective, which we leave to further work.

5

Study of assembly pathways in single-target SST assembly

In this chapter, we use the rate constants $c_+(\mu)$ and $c_-(\mu)$ determined in chapter 4 (see sections 4.3.1 and 4.3.2), to parametrise a stochastic model that describes the time evolution of one fragment of R6 under a fixed background concentration of SSTs. We use the model to study kinetic and thermodynamic properties of self assembly, with a focus on the dominant nucleation pathways and the relative importance of assembly errors due to incorrectly bonded SSTs. This chapter is organised as follows. In sections 5.1, 5.1.1 and 5.1.2 we introduce the model, and describe the methods used to simulate it and analyse the results of the simulations. In section 5.2.1 we use the model to simulate fragment growth under isothermal and thermal ramp conditions. We next focus on the thermodynamics of self assembly (section 5.2.2). In section 5.3 we examine the impact of incorrectly bonded SSTs on the assembly kinetics and thermodynamics. Finally, in section 5.6.1, we discuss the relative importance of two types of binding events that are not described in the model: binding events involving "bridging SSTs" (SSTs whose association involves bringing together two distant sections of the fragment), and binding events where a SST binds incompletely with the binding site (such that some binding domains are left unhybridised).

5.1 Simulation Details

In this chapter we describe the time evolution of one fragment in a thermal bath at temperature T, where unbound SSTs are maintained at a fixed concentration of $c = 100$ nM per SST type. As discussed in section 4.3, if assembly errors due to incorrectly- or incompletely-bonded SSTs are neglected, the state of a fragment is specified by a vector $\vec{f} = (f_1, \cdots, f_{334})$, where $f_i = 1$ if a SST of type i is present in the fragment and $f_i = 0$ otherwise. For generating state-to-state dynamics we use a custom algorithm based on the kinetic Monte-Carlo (kMC) scheme described in section 3.3.3. The algorithm stores information about the fragment using an integer-valued "structural" matrix $\sigma_{i,j}$ with $i = 1, \cdots, 334$ and $j = 1, 2$; plus a real-valued "propensities" vector α_i with $i = 1, \cdots, 334$:

- The first column of $\boldsymbol{\sigma}$ specifies if a SST of type i is present ($\sigma_{i,1} = 1$) or absent ($\sigma_{i,1} = 0$) from the fragment. Therefore, we have $\sigma_{i,1} \equiv f_i$.
- If a SST of type i is present in the fragment (hence $\sigma_{i,1} = 1$), the second column of $\boldsymbol{\sigma}$ stores a value between 1 and 32 that specifies the type of binding site to which that SST is bonded to. If a SST of type i is not present in the fragment (hence $\sigma_{i,1} = 0$) the second column of $\boldsymbol{\sigma}$ stores either the value 0 - if that SST can bind to an empty binding site - or -1 (otherwise).
- If a SST of type i is present in the fragment, α_i stores a dissociation propensity, computed as $k_+ e^{\beta \Delta G^{\ominus}(\sigma_{i,2})}$, according to the type of binding site ($\sigma_{i,2}$) to which that SST is bonded to. For a SST of type i that is absent from fragment but can bind to it (hence $\sigma_{i,1} = 0$, $\sigma_{i,2} = 0$), α_i stores an association propensity $k_+ c$ (with $c = 100$ nM). If a SST of type i is not present in the fragment and cannot bind to it (hence $\sigma_{i,1} = 0$, $\sigma_{i,2} = -1$), α_i stores the value 0.

During a simulation the algorithm keeps track of the size of the fragment, $N = \sum_i \sigma_{i,1}$, defined by the number of SSTs present in it; the cumulative propensity function $A(l) = \sum_{k=1}^{l} \alpha_k$, and the cumulative reaction propensity $A_0 = A(334)$. To propagate the system in time, the algorithm cycles through the following steps:

1 Compute the cumulative propensity function[1] $A(l)$ and the cumulative reaction propensity A_0.

2 Generate two random numbers $r_1, r_2 \in [0, 1]$. Advance time by an amount $\delta t = (1/A_0)\ln(1/r_1)$. Select the type of SST involved in the next event by solving $A(i-1) < r_2 A_0 \leq A(i)$ with respect to i.

3 If a SST of type i is present in the fragment then $\sigma_{i,1} = 1 \to 0$, otherwise $\sigma_{i,1} = 0 \to 1$. Reevaluate $\sigma_{i,2}$ for binding sites neighbouring the site where the transition took place (first- and second-neighbours).

4 Run any relevant subroutines (e.g. evaluation of some observable, temperature change, check for stopping conditions).

5.1.1 The simulated ensemble

If a simulation is run for an arbitrarily long time under constant conditions, it will reach a steady state. Here we discuss the nature of the steady state and its relation with the equilibrium yield of fragments. In an ideal (dilute) solution, in which fragments evolve independently in a large bath of SSTs at a fixed concentration $c = 100$ nM per SST type, the steady-state concentration of fragments in state $\vec{f}$, $c_{\vec{f}}$ is given by (see, e.g., Ref. [197]):

$$c_{\vec{f}} = c\left(\frac{c}{1\mathrm{M}}\right)^{N-1} \exp\left(-\beta G^{\ominus}(\vec{f})\right), \tag{5.1}$$

where $G^{\ominus}(\vec{f})$ is the standard-state free-energy of formation of a fragment in state $\vec{f}$ from its constituent SSTs. Perfectly accurate values of $\Delta G^{\ominus}(\vec{f}, \vec{f}')$ would satisfy

$$\Delta G^{\ominus}(\vec{f}, \vec{f}') = G^{\ominus}(\vec{f}') - G^{\ominus}(\vec{f}). \tag{5.2}$$

In principle, one could study the assembly process, given by a well-defined $\Delta G^{\ominus}(\vec{f}, \vec{f}')$, k_+ and c in a fixed volume, and infer the equilibrium concentration of fragments

[1] for consistency, if the fragment contains only two SSTs, we only count one dissociation propensity in $A(l)$.

from the frequency with which they are observed in the simulation volume [198]. Instead of simulating a fixed volume, however, we pursue an "assembly-eye view" of the same process in which we simply track the growth and shrinkage of one fragment. In this case, the steady-state probability of finding the system in a given state $\vec{f}$, in equilibrium, is given as:

$$p_{eq}(\vec{f}) = \begin{cases} \frac{1}{2Z}, & \text{if } N = 1 \\ \frac{1}{Z}\left(\frac{c}{1\text{M}}\right)^{N-1} \exp\left(-\beta G^{\ominus}(\vec{f})\right), & \text{if } N > 1, \end{cases} \tag{5.3}$$

where Z is a normalisation factor. Eq. 5.3 can be easily verified by observing that given these probabilities, the net probability flux between states is zero under the simulation procedure described in section 5.1 (yielding a detailed-balanced steady state). We note that the ratio $p_{eq}(\vec{f})/p_{eq}(\vec{f'})$ is consistent with the equilibrium concentrations in Eq. 5.1 unless either $N(\vec{f}) = 1$ or $N(\vec{f'}) = 1$, in which case the ratio differs by a factor 2. In order for the equilibrium probabilities in the "assembly-eye view" to map directly into equilibrium concentrations, $p_{eq}(\vec{f})$, for $N(\vec{f}) = 1$, will be corrected, for the remaining of this thesis, by a multiplicative factor 2.

In fact, due to the local approximations in $\Delta G^{\ominus}(\vec{f}, \vec{f'})$ (see section 4.1.1) and the finite sampling accuracy of the simulations with oxDNA, Eq. 5.2 is only approximately satisfied. Detailed balance is therefore marginally violated in simulations. We perform the following test to gauge whether deviations from detailed balance are significant. Let $\vec{f_i}$ and $\vec{f_j}$ be two fragment states, and $\wp_{ij} \equiv \vec{f_i} \rightarrow \vec{f}_{i+1} \rightarrow \vec{f}_{i+2} \rightarrow \cdots \rightarrow \vec{f}_{j-1} \rightarrow \vec{f_j}$ a reaction pathway linking states $\vec{f_i}$ and $\vec{f_j}$. If detailed balance is satisfied exactly, the free-energy change

$$G^{\ominus}(\vec{f_j}) - G^{\ominus}(\vec{f_i}) = \sum_{\vec{f} \rightarrow \vec{f'} \in \wp_{ij}} \Delta G^{\ominus}(\vec{f}, \vec{f'}) \tag{5.4}$$

should be pathway-independent. In Annex E we consider a few examples of fragment states $\vec{f_i}$ and $\vec{f_j}$ and compute the factor on the RHS of Eq. 5.4 for a different paths $\wp_{ij}$. We find path-dependence to be relatively small. While a non-zero difference between paths means that detailed balance is not strictly satisfied, we assume that these

differences are sufficiently small that they can be neglected as a first approximation. A similar approach was taken recently in a model of DNA origami folding [199, 200].

5.1.2 Enhanced kMC sampling

Computing ensemble averages using the simulation scheme in section 5.1 can be very inefficient if the fragment state space is segregated by free-energy barriers. If that is the case, a simulation can easily become trapped in a local free-energy minimum and leave other regions of the fragment state space unsampled. Here, we address this issue by using a biasing scheme, similar to Umbrella sampling (see section 2.2), that artificially flattens free-energy barriers and facilitates transitions between local minima. The biasing scheme devised here involves specifying an (in general multi-dimensional) order parameter $Q = Q(\vec{f})$, defined in terms of some property of interest (e.g. fragment size, number of bonded domains in a fragment). Instead of sampling from the equilibrium distribution $p_{eq}(\vec{f})$, we set our simulations to sample from a biased distribution $p_w(\vec{f}) = w(Q)^2 p_{eq}(\vec{f})$. This is achieved by changing the rules followed by the algorithm for choosing transitions to carry out. In what follows, we will assume that the condition of detailed balance

$$p_{eq}(\vec{f_i})a(\vec{f_i} \to \vec{f_j}) = p_{eq}(\vec{f_j})a(\vec{f_j} \to \vec{f_i}), \tag{5.5}$$

is satisfied exactly between all fragment states $\vec{f_i}$ and $\vec{f_j}$. In Eq. 5.5, $a(\vec{f_i} \to \vec{f_j})$ is the propensity of a reaction that leads the system from $\vec{f_i}$ to $\vec{f_j}$. By making a few substitutions, we can rewrite Eq. 5.5 as

$$p_w(\vec{f_i})a(\vec{f_i} \to \vec{f_j})\frac{w[Q(\vec{f_j})]}{w[Q(\vec{f_i})]} = p_w(\vec{f_j})a(\vec{f_j} \to \vec{f_i})\frac{w[Q(\vec{f_i})]}{w[Q(\vec{f_j})]}. \tag{5.6}$$

Eq. 5.6 means that sampling from the (biased) probability distribution $p_w(\vec{f})$ can be achieved by making the following change to the reaction propensities:

$$a(\vec{f_i} \to \vec{f_j}) \to \frac{w[Q(\vec{f_j})]}{w[Q(\vec{f_i})]}a(\vec{f_i} \to \vec{f_j}). \tag{5.7}$$

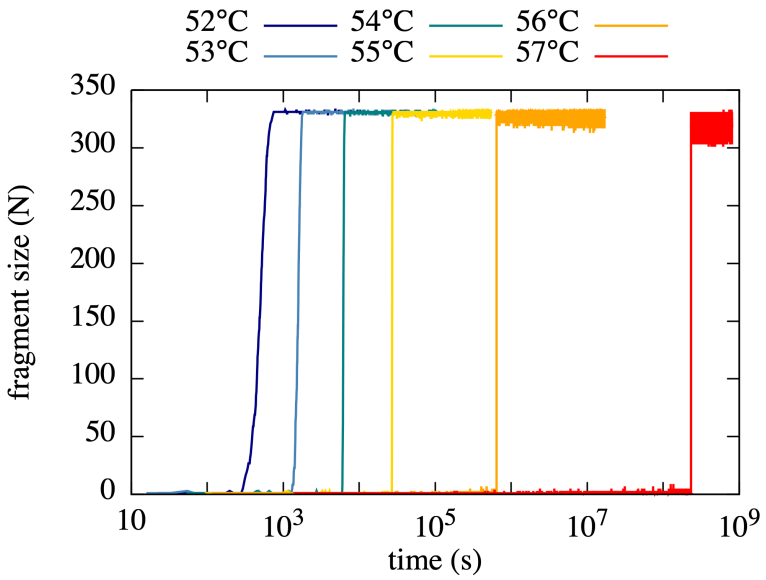


Figure 5.1: Self assembly under isothermal conditions - fragment size as a function of time. As the temperature goes up, the time to form a full fragment increases, and the fluctuations in the fragment size also increase.

In general, the biasing factor $w(Q)$ can be chosen such that a simulation will sample all coordinates Q equally (hence $p_w(Q) = \text{const.}$). To compute the ensemble average of an observable $A = A(\vec{f})$, $\langle A \rangle$, we simply compute $\langle A(\vec{f})/w(Q)^2 \rangle_w$ during a (biased) run. It is easy to show that this yields the correct result:

$$\langle A \rangle = \sum_{\vec{f}} p_{eq}(\vec{f}) A(\vec{f}) = \sum_{\vec{f}} w^2[Q(\vec{f})] p_{eq}(\vec{f}) \frac{A(\vec{f})}{w^2[Q(\vec{f})]} = \langle A/w(Q)^2 \rangle_w. \tag{5.8}$$

5.2 Results

5.2.1 Fragment growth under isothermal and thermal ramp conditions

We start our analysis by simulating fragment growth under isothermal conditions. The simulations where initiated from a fragment state $f_i = \delta_{ij}$, where $j \in [1, 334]$ is a random (integer) number, corresponding to a single SST. The time evolution of fragment size (Fig. 5.1) shows a few interesting features. At lower temperatures, the fragment grows quickly after the simulation is initiated. As the temperature is increased, however, we see increasingly long lag periods, during which the fragment

does not grow past a few SSTs, followed by a rapid increase towards a large stable intermediate (yet, on average, below completion). The increase of the lag period with temperature is significant. At $T = 57°$C, it typically takes around a few years (physical time) before appreciable fragment growth is observed in the simulations. A small drop in temperature, to $T = 56.00°$C, results in a large drop in lag period, down to about a few hours. At $T = 54.00°$C, the lag period is reduced to a few minutes. This behaviour suggests the existence of a nucleation barrier that decreases rapidly in height with decreasing temperature. We perform additional simulations starting from the state $f_i = 1 \; \forall \; i$, corresponding to a fully-formed R6 object. After a period of equilibration, we measure the fragment size over a long simulation time (10^8 kMC steps). At the critical temperature[2] $T_c = 58.23°$C, the average fragment size is 260 SSTs, with large fluctuations around this average (Fig. 5.2). As the temperature decreases, the average fragment size increases and the fluctuations in size become smaller. A fully-formed R6 object is observed with a frequency of 50% and above for temperatures of 41°C and below.

It should be emphasised that the concentration of SSTs was kept fixed during the simulations discussed above, whereas in the experiments [2, 3] where R6 was produced it was the total number of SSTs present inside the reaction vessel that is kept fixed. This means that in experiments, (SST) binding rates are expected to decrease with time as SSTs are gradually consumed. In any case, the fact that a clear timescale separation between lag period and growth (a necessary condition for preventing monomer depletion traps from occurring) only exists in a narrow temperature window ($\sim 54° - 58°$C) below T_c, suggests that below this range, kinetic traps due to monomer depletion are likely to dominate kinetics.

It is instructive to study the effect a thermal ramp on the growth and shrinkage of a fragment. Starting from a fragment state $f_i = \delta_{ij}$, where $j \in [1, 334]$ is a random (integer) number, at $T = 65°$C, we decreased the temperature in discrete steps (assumed instantaneous) of $\delta T = 0.1°$C until the temperature fell to $T = 35°$C.

[2]Here, we define critical temperature as the highest temperature for which, for a relatively large fragment size ($N \gtrsim 50$), we have $F(N) - F(1) \leq 0$, where $F(N) = -\beta^{-1} \ln p_{eq}(N)$. In section 5.2.2 we will detail how $F(N)$ is determined.

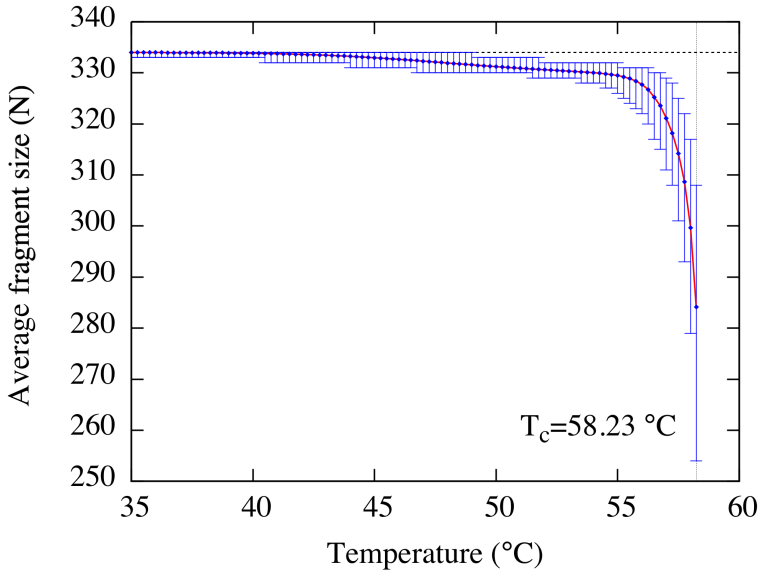


Figure 5.2: Average fragment size (red curve) and fragment size fluctuations (vertical blue lines) as a function of temperature. The vertical blue lines correspond to the fragment size range $N \in [\langle N\rangle - 2\sigma, \langle N\rangle + 2\sigma]$, where σ is the standard deviation measured for fragment size.

Next, we increased the temperature (also step wise ($\delta T = 0.1$°C)) back to $T = 65$°C. We used different cooling-heating rates, ranging from 5 min/°C to 3h/°C, such as in Refs. [2, 3]. During both cooling and heating stages, the fragment, on average, never reached its equilibrium size for temperatures close to T_c (Fig. 5.3). The hysteresis width on fragment size decreased markedly with decreasing cooling rate, from about 4.5°C for 5 min/°C to about 0.8°C for 3h/°C. However, the fragment followed more closely its equilibrium size during the heating stage than during the cooling stage, suggesting that fragment melting is a more cooperative process than fragment growth.

5.2.2 Thermodynamic properties of SST assembly

To better interpret the behaviour described above, we determined the free-energy function $F(N) = -\beta^{-1} \ln p_{eq}(N)$ as a function of fragment size (N), where $p_{eq}(N) = \sum_i p_{eq}(\vec{f}_i)\delta_{N,N_i}$, and $N_i = N(\vec{f}_i)$ (see Fig. 5.4). To determine the probability distribution $p_{eq}(N)$ we used the biasing scheme described in section 5.1.2, where fragment size (N) was used as the order parameter. To make the calculations more

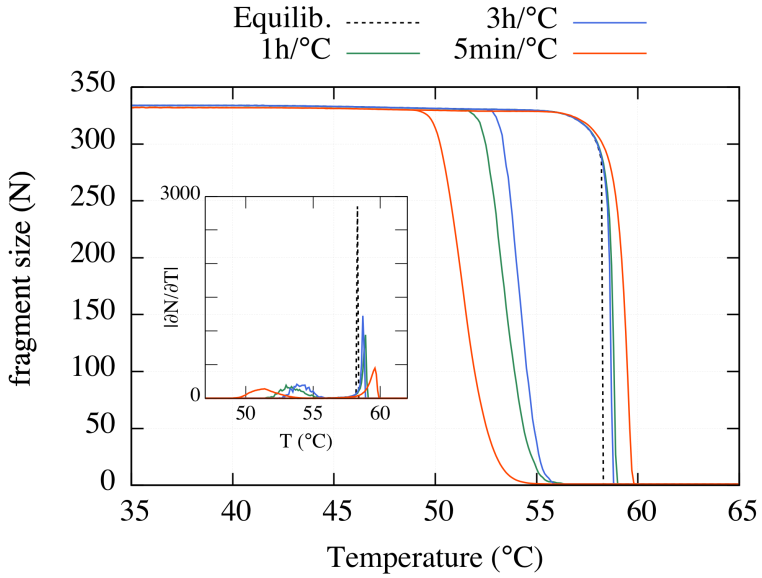


Figure 5.3: Fragment size as a function of temperature and under thermal ramp conditions. The curves shown in the figure correspond to averages over 10^4 independent runs. Also shown (dashed black curve) is the average fragment size, at equilibrium, as a function of temperature. Inset: derivatives with respect to temperature of the curves in the main plot. The growth peaks (during cooling) are located at 53.8°C for 3 h/°C, 53.0°C for 1 h/°C, 51.4°C for 5 min/°C, and 58.2°C at equilibrium. The melting peaks (during heating) are located at 58.7°C for 3 h/°C, 58.9°C for 1 h/°C, and 59.6°C for 5 min/°C.

efficient we divided the range of values of N to sample ($N = 1, \cdots, 334$) into multiple windows $s_i : N \in [N_{\min(s_i)}, N_{\max(s_i)}]$ that overlap with each other as $N_{\max(s_i)} - 2 = N_{\min(s_i+1)}$, such that there are 3 overlapping points between any two consecutive windows. Each sampling window yielded a partial $F_{s_i}(N) = -\beta^{-1} \ln p_{eq}^{(s_i)}(N) + c_{s_i}$, defined up to an additive constant c_{s_i}. To reconstruct the function $F(N)$, spanning the whole range $N = 1, \cdots, 334$, we first fixed c_{s_1} such that $F_{s_1}(0) = 0$ and equaled the function $F_{s_1}(N)$ to $F(N)$ in the range $N \in [1, N_{\max(s_1)}]$. The second profile $F_{s_2}(N)$ was then set to be equal to $F(N)$ in the range $N \in [N_{\max(s_1)} + 1, N_{\max(s_2)}]$, where the constant c_{s_2} was chosen such that the sum of squared distances between $F_{s_1}(N)$ and $F_{s_2}(N)$ was minimised at the overlap region between s_1 and s_2. The third and remaining functions $F_{s_i}(N)$ were used sequentially to reconstruct $F(N)$ in the rest of the range, in a similar way as $F_{s_2}(N)$.

As discussed in section 5.2.1, we defined the critical temperature T_c as the highest temperature for which $F(N) - F(1) \leq 0$ for a relatively large fragment

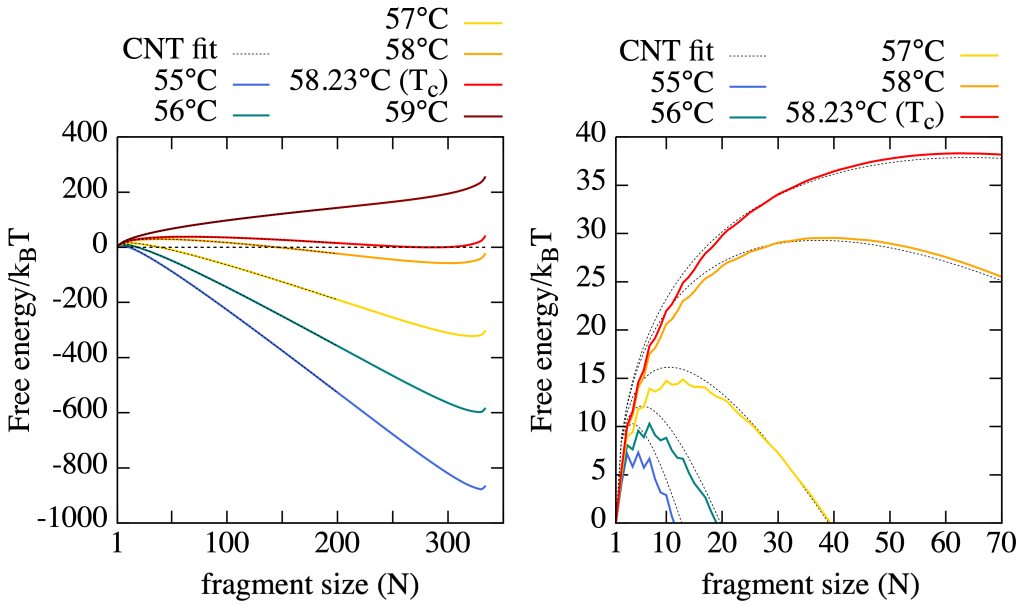


Figure 5.4: Free-energy function $\beta F(N) = -\ln p_{eq}(N)$ as a function of fragment size (N). Left: $\beta F(N)$ in the size range $N \in [1, 334]$. Right: $\beta F(N)$ in the size range $N \in [1, 70]$. Also shown in the plot (black dashed lines) are CNT fits performed in the size range $N \in [1, 200]$. The critical sizes and the heights of the nucleation barriers are: $N^{\ddagger} = 5$ and $\beta F^{\ddagger} = 7.31$ for $T = 55°\text{C}$; $N^{\ddagger} = 7$ and $\beta F^{\ddagger} = 10.30$ for $T = 56°\text{C}$, $N^{\ddagger} = 13$ and $\beta F^{\ddagger} = 14.87$ for $T = 57°\text{C}$, $N^{\ddagger} = 40$ and $\beta F^{\ddagger} = 29.54$ for $T = 58°\text{C}$, $N^{\ddagger} = 63$ and $\beta F^{\ddagger} = 38.29$ for $T = T_c = 58.23°\text{C}$.

size $N \gtrsim 50$. The critical temperature predicted by the model ($T_c = 58.23°\text{C}$) is consistent with experimental observations in Refs. [2, 3]. In Ref. [3], a fluorometric assay is provided (which measures real-time changes in (target-structure related) base-pairing levels) performed during the production of R6 for a cooling rate of 3h/°C. The fluorometric assay shows a large increase in (target-related) base-pairing levels at 54°C (peak at 53°C). T_c is expected to be close to the onset temperature below which there is a sufficient thermodynamic drive for fragments with appreciable sizes to form under experimental conditions, where the concentration of SSTs is not fixed[3]. This onset temperature is expected to be slightly above 54°C given the length of the temperature steps (1°C) used in Ref. [3] and relatively short incubation times per temperature step (3h), therefore within a $\sim 1.5\%$ margin of the critical temperature T_c predicted by the model, in Kelvin [4].

[3] as as it will be shown in chapter 6, above this onset temperature almost all SSTs in the assembly mixture are unbound, at a concentration close to 100nM, with a small component $< 5\%$ used to form small fragments (dimers, trimers, etc.).

[4] The appropriate temperature scale for statistical mechanical models.

The results (Fig. 5.4) confirm the existence of a nucleation barrier that decreases quickly in height as the temperature drops below T_c. To confirm that the critical sizes have been correctly identified, we performed unbiased simulations starting from fragment states visited during the sampling runs, and verified that, on average, a fragment with a critical size has a roughly $50\% - 50\%$ chance of either growing further or melting away. Both the critical size and the height of the nucleation barrier show sharp decreases with decreasing temperature. At $T_c = 58.23°\text{C}$, the critical size and the height of the nucleation barrier are $N^\ddagger = 63$ and $\beta F^\ddagger = 38.29$, respectively, whereas at $T = 57°\text{C}$ these values have already dropped to $N^\ddagger = 13$ and $\beta F^\ddagger = 14.87$. Just two degrees lower ($T = 55°\text{C}$) the critical fragment size has dropped to $N^\ddagger = 5$ and the nucleation barrier height to $\beta F^\ddagger = 7.30$. Below T_c, the functions $F(N)$ exhibit a global minimum at a fragment size below 334 (the size of R6). The location of the global minimum gradually approaches 334 as the temperature is decreased. This is a statistical effect: as the size approaches 334, the number of fragment states for a given size decreases rapidly, reaching 1 at $N = 334$ (hence, zero configurational entropy). As the temperature is decreased, bonds between SSTs become less likely to dissociate and the statistical tendency for the system to maximise its entropy decreases as well. This entropic effect is typical of finite-sized assemblies and is absent, for instance, in conventional crystals, where the number configurations in which a given set of subunits (e.g. atoms) can be arranged will, in principle, always grow with the number of subunits in the crystal.

The features of the free-energy function $F(N)$ discussed above explain why a thermal ramp is necessary for targets (R6) to form successfully in experiments. At temperatures just below T_c, the nucleation barrier is sufficiently large to set a timescale separation between nucleation (slow) and growth (fast), and prevent, in principle, monomer depletion traps; however, fragments reach a ceiling at sizes below completion due to a combination of insufficiently stable SST-SST bonds and statistical effects. At low temperatures, fragments are able to reach completion in equilibrium, however, depletion traps are likely to dominate kinetics since the nucleation barrier is far too small to control nucleation in a meaningful way. A

thermal ramp allows to exploit advantages from both regimes. As the temperature is dropped below T_c, fragments start nucleating and grow in a controlled fashion. The decrease in thermodynamic drive that results from SSTs being consumed is compensated by gradual decreases in temperature. These decreases should be performed sufficiently slowly, such that nucleation remains rate-controlling. Ideally, the concentration of fragments above the critical size (supercritical) should never exceed the threshold 100nM (above which some supercritical fragments will need to denucleate spontaneously for a maximal yield to be reached). If the thermal ramp is set up perfectly, the concentration of supercritical fragments should approach 100 nM at an intermediate temperature, at which point only a small fraction of SSTs should remain unbound. Further cooling should not lead to additional nucleation events, but rather to increase the size of the fragments already supercritical towards completion.

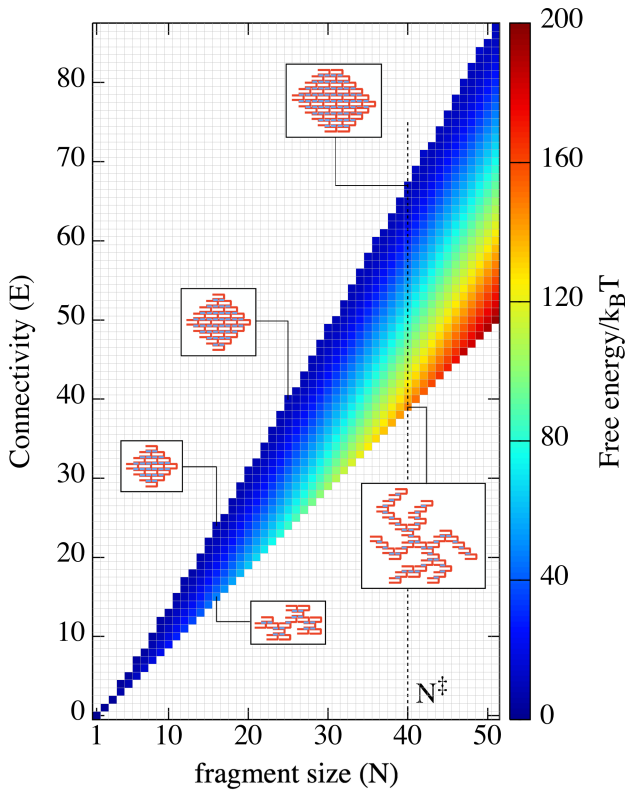


Figure 5.5: Free-energy function $\beta F(N, E) = -\ln p_{eq}(N, E)$ as a function of fragment size (N) and connectivity (E) for the size range $N \in [1, 51]$, at $T = 58°\text{C}$. The insets show examples of bonding configurations for a given macrostate (N, E). At $T = 58°\text{C}$, the critical nucleus size is $N^{\ddagger} = 40$ (dashed black line).

The number of bonded domains between SSTs gives a proxy for the shape of a

fragment: for a given size, a fragment with a higher number of bonded domains is typically more compact, with less protuberances, than a fragment with a lower number of bonded domains. We investigate which shapes are the most likely to form in equilibrium by computing the free-energy function $F(N,E) = -\beta^{-1} \ln p_{eq}(N,E)$, where is E is the number of bonded domains (from now on referred to as the "connectivity"), $p_{eq}(N,E) = \sum_i p_{eq}(\vec{f}_i)\delta_{N,N_i}\delta_{E,E_i}$, $N_i = N(\vec{f}_i)$, and $E_i = E(\vec{f}_i)$. To determine the probability distribution $p_{eq}(N,E)$ we perform simulations using the biasing scheme introduced in section 5.1.2, where both fragment size (N) and connectivity (E) are used as order parameters. We choose a biasing factor with the form $w(N,E) = w(N)w_N(E)$, where $w(N)$ is fixed as $w(N)^2 p_{eq}(N) = \text{const.}$[5] For convenience, we split the range of values of N to sample, $N \in [1, N_{\max}]$, into multiple windows s_i: $N_{(s_i)} \in [N_{\min(s_i)}, N_{\max(s_i)}]$, where each window s_i overlaps with the next, s_{i+1}, as $N_{\max(s_i)} - 2 = N_{\min(s_{i+1})}$. For each window s_i, we perform a series of simulations where the factor $w_N^{(s_i)}(E)$ is gradually improved, up to the point where the (biased) probability distribution $p_w^{(s_i)}(N,E)$ meets a flatness criterion[6]. First, we adjust $w_N^{(s_i)}(E)$ manually until all macrostates (N,E) can be observed with some frequency in one simulation. Afterwards, we improve $w_N^{(s_i)}(E)$ by fixing:

$$w(N)^2 w_N^{(s_i)}(E)^2 p_{eq}^{(s_i)}(N,E) = c_{s_i}, \tag{5.9}$$

where $p_{eq}^{(s_i)}(N,E)$ is the unbiased probability distribution obtained in s_i, and c_{s_i} is an arbitrary constant. Once all factors $w_N^{(s_i)}(E)$ have been improved sufficiently, we merge them into a single $w_N(E)$ that spans the whole range $N \in [1, N_{\max}]$:

$$w_N(E) = \begin{cases} \frac{w_N^{(s_1)}(E)}{\alpha_{s_1}}, & \text{if } \; N \in [1, N_{\max(s_1)} - 1] \\ \frac{w_N^{(s_i)}(E)}{\alpha_{s_i}}, & \text{if } \; N \in [N_{\min(s_i)} + 1, N_{\max(s_i)} - 1], \end{cases} \tag{5.10}$$

[5] the probability distribution $p_{eq}(N)$ is, at this point, already known (see Fig. 5.4)

[6] In this chapter, we consider a probability distribution $p(N,E)$ to be sufficiently flat if $p(N_1,E_1)$ and $p(N_2,E_2)$ do not differ by more than 10% for any two macrostates (N_1,E_1) and (N_2,E_2).

where the multiplicative constants α_{s_i} were introduced to ensure that macrostates (N, E) from different windows are sampled equally. A good choice for α_{s_i} is given by solving:

$$\left(\frac{1}{\alpha_{s_i}^2}\right) \sum_{\substack{N=N_{\min(s_i)}+1,\\ N_{\max(s_i)}-1}} \sum_{\substack{E=E_{\min}(N),\\ E_{\max}(N)}} \left\{ w_N^{(s_i)}(E)^2 w(N)^2 p_{eq}^{(s_i)}(N,E) \right\} = \Gamma N_{\text{states}}(s_i), \quad (5.11)$$

with respect to α_{s_i}, where $N_{\text{states}}(s_i)$ is the number of macrostates (N, E) in the range $N \in [N_{\min(s_i)} + 1, N_{\max(s_i)} - 1]$, $E_{\min}(N)$ and $E_{\max}(N)$ are the minimal and maximal connectivities for a fragment of size N, and the constant Γ is the same for all windows s_i. Eq. 5.11 simply enforces that a simulation that uses the biasing factor $w_N(E)$ in Eq. 5.10 will visit macrostates (N, E) in any given range $N \in [N_{\min(s_i)} + 1, N_{\max(s_i)} - 1]$ with a frequency that is proportional to the number of macrostates (N, E) in that range. Finally, to determine $p_{eq}(N, E)$, we perform a long simulation ($> 10^{12}$ kMC steps) using the bias $w(N)w_N(E)$ as described above, over the whole range $N \in [1, N_{\max}]$.

As can be seen in Fig. 5.6, fragments with a high number of bonded domains, and compact shapes, are the most likely to form in equilibrium. However, a competition is visible between the tendency for maximising the number of bonded domains, and the tendency for maximising entropy, by forming fragments that while less connected, can be realised in a larger number of ways (Fig. 5.6). For small sizes ($N \lesssim 15$), maximally-connected fragments are the most likely to form. However, as the size increases, entropic factors become increasingly dominant. Interestingly, the probability of finding a fragment in a maximally-connected state, $p_{eq}(E_{\max}(N)|N)$, does not decrease in a monotonic way with fragment size. Instead, $p_{eq}(E_{\max}(N)|N)$ shows a pattern of recurrence, with local minima at $N_{\min} = M_1 \times M_2$ with $|M_1 - M_2| \leq 1$ followed by local maxima at a size just above $(N_{\min} + 1)$. This is a statistical effect: the number of ways in which a maximally-connected fragment can form shows a pattern of recurrence with fragment size analogous

to that of $p_{eq}(E_{\max}(N)|N)$ (see tables 5.1 and 5.2). For $N = N_{\min}$, for which maximally-connected fragments display compact, rhomboid-like shapes, the number of realisations has a minimum (which is steepest if N happens to be a perfect square). For $N = N_{\min} + 1$, maximally-connected fragments are the most anisotropic and can be realised in a large number of ways. As the size increases past $N = N_{\min} + 1$, maximally-connected fragments become increasingly more compact, and the number of possible realisations decreases until the next local minimum is reached.

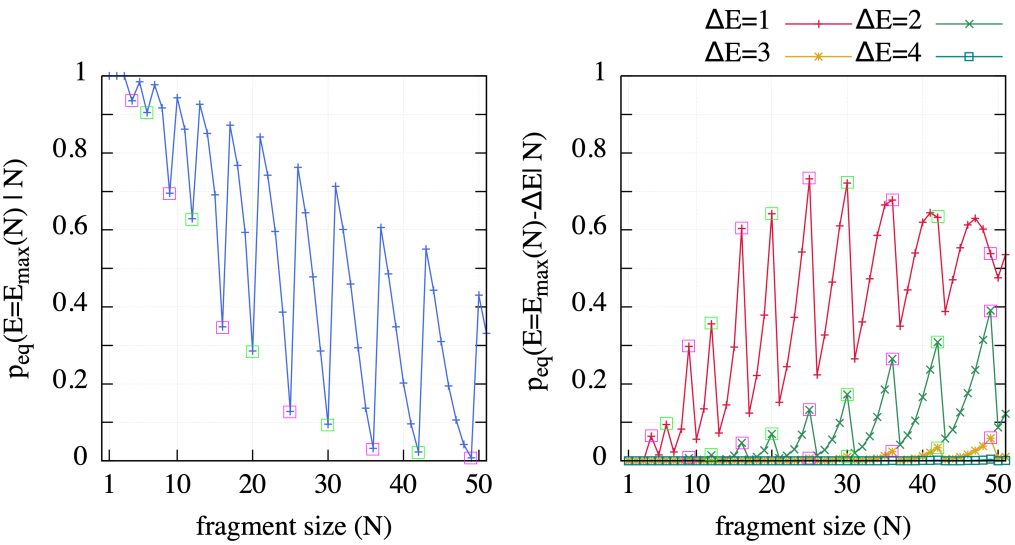


Figure 5.6: Probability at equilibrium of forming a fragment with a given connectivity E, given a size N. Data points for fragment sizes satisfying $N = M_1 \times M_2$, with $|M_1 - M_2| \leq 1$, are highlighted with pink ($M_1 = M_2$) and green ($M_1 \neq M_2$) squares. Left: probability of forming a maximally-connected ($E = E_{\max}(N)$) fragment given a size N. Right: probability of forming a fragment with a less-than-maximal connectivity ($E = E_{\max}(N) - \Delta E$) given a size N.

The fact that for small sizes ($N \lesssim 15$), maximally-connected fragments are the most likely to form, helps to explain the small-scale features observed in $F(N)$ (Fig. 5.4) in this size range. Looking closely at Fig. 5.4 we can see, for $T = 55°\text{C}$ and $T = 56°\text{C}$, clear decreases in $F(N)$ for $N = 4, 6, 8$ and 9, and clear increases for $N = 2, 3, 5, 7$ and 10 (for $N = 10$ there is a small decrease in $F(N)$ for $T = 55°\text{C}$). An analogous behaviour is observed at higher temperatures, where, for sizes $N = 4, 6, 8$ and 9, $F(N)$ shows small increases, and for sizes $N = 2, 3, 5, 7$ and 10, $F(N)$ shows relatively larger increases. These features are explained by the

size-dependence of the connectivity, in maximally-connected fragments. For a size increase by 1, an increase in connectivity by 2 is expected to be more stabilising than an increase in connectivity by 1. Above $N = 3$, an increase in size by 1 results in an increase in connectivity by 2 unless $N = N_{\min}$, in which case the connectivity increases by 1 (see tables 5.1 and 5.2). Fragments with $N = 1, 2$ and 3 have always connectivities $E = 0, 1$ and 2, respectively (hence a size increase by 1 results in a connectivity increase by 1). Therefore, there is a connectivity change of 2 if $N = 3 \to 4$, $N = 5 \to 6$, $N = 7 \to 8$, $N = 8 \to 9$, and of 1 in the remaining cases. These differences are expressed in $F(N)$ as local decreases at $T = 55°\mathrm{C}$ and $T = 56°\mathrm{C}$ (or, analogously, relatively small increases at higher temperatures) for a connectivity change of 2, and as local increases at $T = 55°\mathrm{C}$ and $T = 56°\mathrm{C}$ (or, analogously, relatively larger increases at higher temperatures) for a connectivity change of 1. For sizes above $N \sim 15$, fragments that are not maximally connected become increasingly more likely to form. This explains the rapid smoothing out of the curves $F(N)$ for sizes above $N \sim 15$, as discrete changes in connectivity become replaced by changes of average connectivity values (over fragments states with a fixed size and different connectivity values).

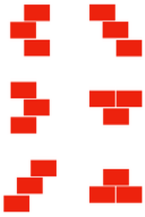
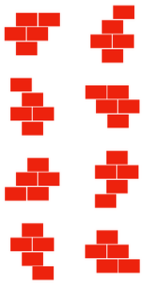
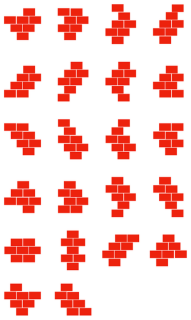
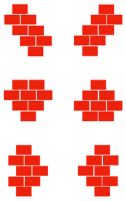
$N = 1$ ($E_{\max} = 0$)	$N = 2$ ($E_{\max} = 1$)	$N = 3$ ($E_{\max} = 2$)	$N = 4$ ($E_{\max} = 4$)
			
$N = 5$ ($E_{\max} = 5$)	$N = 6$ ($E_{\max} = 7$)	$N = 7$ ($E_{\max} = 8$)	$N = 8$ ($E_{\max} = 10$)
			

Table 5.1: Bonding configurations possible for maximally-connected fragments ($N \leq 8$).

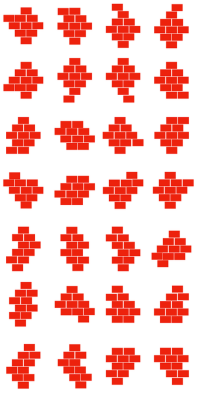
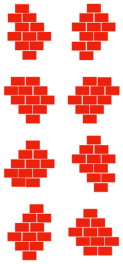
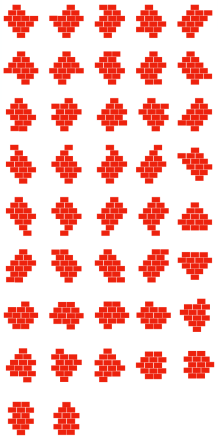
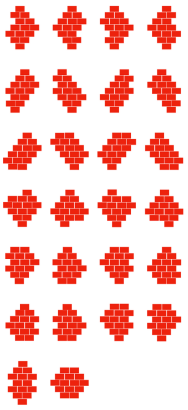
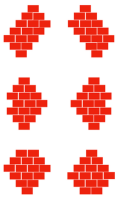
$N = 9$ ($E_{\max} = 12$)	$N = 10$ ($E_{\max} = 13$)	$N = 11$ ($E_{\max} = 15$)	$N = 12$ ($E_{\max} = 17$)
			
$N = 13$ ($E_{\max} = 18$)	$N = 14$ ($E_{\max} = 20$)	$N = 15$ ($E_{\max} = 22$)	$N = 16$ ($E_{\max} = 24$)
			

Table 5.2: Bonding configurations possible for maximally-connected fragments ($8 < N \leq 16$).

The behaviour predicted here has qualitative differences with that of Refs. [180, 181], where a maximally-connected fragment is always the most likely to form regardless of size, and consequently the function $F(N)$ shows much stronger features associated with discrete changes in connectivity. The source for this discrepancy is likely to be the way in which the entropic barriers associated with forming secondary domains (upon SST association) are determined. In Refs. [180, 181] forming a second domain has an entropic cost of $\Delta S/k_B = \ln(4) \approx 1.39$, whereas forming a third and fourth domains has an entropic cost of zero. As can be seen in annex B, oxDNA predicts a substantially higher entropic cost for forming a second ($\Delta S/k_B$ ~6-8), third ($\Delta S/k_B$ ~3-6) and fourth ($\Delta S/k_B$ ~2-5) domains. As a result, for a given fragment size, a decrease in connectivity by 1 has a more destabilising

effect in the model of Refs. [180, 181] than in oxDNA, which explains the stronger bias towards maximally-connected fragments in Refs. [180, 181].

We next investigate whether classical nucleation theory (CNT) is able to provide a good fit to the free-energy function $F(N)$ in Fig. 5.4. CNT is a model that describes the formation of a nucleus of a stable, ordered phase from a metastable mother phase. In CNT, nucleation is described under a few assumptions that, in conjunction, are referred to as the capillarity approximation: i) a nucleus of the new (stable) phase has, regardless of its size, the same properties (e.g. density, structure) as a bulk system of the same phase, ii) a nucleus is perfectly spherical (or circular, in 2D) regardless of its size, and iii) the interface between the nucleus and the metastable (mother) phase is a sharp boundary. In CNT, the formation free-energy of a nucleus is the sum of a bulk term that gives the free-energy gain of forming a new phases and scales with the volume of the nucleus (or area, in 2D), with a second term that gives the free-energy cost of forming an interface between both phases and scales with the surface area (or perimeter, in 2D) of the nucleus. Since the R6 structure is two-dimensional, we consider here the 2D variant of CNT. According to 2D-CNT, the formation free-energy of a circular nucleus of radius r is given by (see eg. Ref. [201]):

$$\Delta F(r) = -\pi r^2 \rho_s \Delta\mu + 2\pi r\gamma, \tag{5.12}$$

where ρ_s is the bulk (surface) density of monomers in the ordered (stable) phase, $\Delta\mu$ is the bulk free-energy difference per monomer between the mother phase and the new (stable) phase, and γ is the free-energy cost per unit of length of forming an interface between the two phases (also called the line tension). For convenience, we rewrite Eq. 5.12 in terms of the number of monomers $N = \pi r^2 \rho_s$ contained in the nucleus:

$$\Delta F(N) = -\Delta\mu N + 2\pi^{1/2} \rho_s^{-1/2} \gamma N^{1/2}. \tag{5.13}$$

To determine the surface density of monomers (ρ_s), we simulated a fully-formed R6 structure with oxDNA2 (see section 1.3.3), a version of oxDNA that is optimised for

describing structural properties of large DNA nanostructures, using the molecular dynamics package. The simulation was performed at $T = 25°C$, for which the structure remained stable (without any SST falling off) during the simulation. After a period of equilibration, we determined the average distance between the centres of mass of adjacent (first and second neighbours) SSTs in R6. Apart from SSTs close to the outer border of the structure, these distances were found to be largely independent of location. The area occupied by each SST (the inverse of ρ_s) was then estimated using simple geometric arguments. We obtained the value $\rho_s = 5.82 \times 10^{16}$ m^{-2} for the surface density of monomers, in reasonable agreement with the value $\rho_s = 4.80 \times 10^{16}$ m^{-2} estimated (without direct measurement) in Ref. [2]. We use Eq. 5.13 to fit the free-energy function $F(N)$ predicted by our model for sizes up to $N = 200$, excluding larger sizes to prevent finite size effects from overtly affecting the fit (conventional CNT does not describe such effects). In order for Eq. 5.13 to satisfy $F(1) = 0$, we use $N' \equiv N + 1$ as the fitting variable. The fits were performed using the curve-fitting app of MATLAB 9.6.

As can be seen in Fig. 5.4, CNT provides a fairly good fit to the function $F(N)$ predicted by our model. The parameters $\Delta\mu$ and λ that yielded best fits are given in table 5.3. Not surprisingly, the fit is worst in the small-size range ($N \approx 15$ and below), where the curves $F(N)$ display small-scale features that are system-specific, and naturally cannot be captured by a model as generic as CNT. As observed in many other systems, the line tension does not change much with temperature. Instead, the rapid collapse of the nucleation barrier below T_c is caused primarily by a rapid decrease in $\Delta\mu$. To the best of our knowledge, no measurements of line tension has been performed so far for a SST structure (nor for any other DNA nanoconstruct). Nevertheless, it is perhaps interesting to observe that the line tension reported here is fairly close to previous experimental and simulational measurements in lipid bilayers, e.g.: $\lambda = 1$ pN in Ref. [202], $\lambda = 4$ pN in Ref. [203], and $\lambda = 4 \pm 2$ pN in Ref. [204].

T(°C)	$\beta\Delta\mu$	λ (pN)
55	3.497	3.698
56	2.595	3.471
57	1.697	3.246
58	0.806	3.021
58.23	0.601	2.969
59	-0.082	2.794

Table 5.3: CNT best fit parameters.

5.3 Misbonds

So far, we have neglected the possibility SSTs of incorrect types may attach themselves to the binding sites of a fragment. In this section we devise a simple model that describes the association (and dissociation) of incorrect SSTs, and use this model study the impact of incorrect bonds (or "misbonds") on the assembly kinetics and thermodynamics.

A problem with describing misbonds is that the stability of a misbond is expected to be very sensitive to the length (and number) of complementary sections between the sequence of the SST, and the sequence of the binding domains at the binding site. Given the very large number of different combinations of sequences for SST and binding domains, this problem alone makes its unworkable to parametrise reaction propensities for misbonds via simulations with oxDNA. To make progress, we make a first simplification that a binding site can be represented as a continuous single-stranded segment, with a length equal to the cumulative length of the binding domains, and a sequence given by the concatenation of sequences of the binding domains. As can be seen in the appendix A, the cumulative length of the binding domains can be either 10 or 11 bases for binding sites with one binding domain ($\mu = 1, \cdots, 6$); 20, 21 or 22 bases for binding sites with two binding domains ($\mu = 7, \cdots, 18$); 31 or 32 bases for binding sites with three binding domains ($\mu = 19, \cdots, 26$), and 42 bases for binding sites with four binding domains ($\mu = 27, \cdots, 32$). We note that this simplification overestimates the stability of a misbond in cases where the pattern of complementarity between a SST and

the binding domains (at the binding site) is such that a continuous double helix could form between the SST and two binding domains simultaneously. As can be seen in appendix B, initiating secondary domains involve substantial entropic penalties that are naturally not described under the simplification above. An advantage of describing binding sites in such a way is that we can rely on the NN model [31] (see section 1.3.2) for computationally inexpensive estimates of the formation free energies of misbonds.

5.4 The effect of misbonds on the thermodynamics of self assembly

We start by focusing on the thermodynamic properties of misbonds. For each type of binding site μ, we first list down all combinations ν of base sequences possible for the binding domains, taking the sequences used in Ref. [2] as a reference. As discussed in section 5.3, we make the simplification that a binding site of type μ with sequences ν can be represented as a single (continuous) sequence $S_{\mu\nu}$. Given $S_{\mu\nu}$, there are 333 types of SST, s_ξ, that can form misbonds with $S_{\mu\nu}$, and only one type of SST that can bind correctly with $S_{\mu\nu}$. For each pair $S_{\mu\nu}, s_\xi$ we use a custom algorithm that identifies complementary sections η between the sequences of $S_{\mu\nu}$ and s_ξ that are at least two-bases long. A complementary section is rejected if it overlaps fully with a longer section already identified.[7] For each complementary section η identified, we compute a binding free energy $\Delta G_\eta^{\ominus}(S_{\mu\nu}, s_\xi)$ using the average-base parametrisation of the NN model [31] (see section 1.3.2). The probability at equilibrium that s_ξ will form bonds with $S_{\mu\nu}$, $p(S_{\mu\nu}, s_\xi)$, is given as:

$$p(S_{\mu\nu}, s_\xi) = \sum_\eta p_\eta(S_{\mu\nu}, s_\xi) = p_\varnothing \frac{c_{s_\xi}}{c_0} \sum_\eta e^{-\beta \Delta G_\eta^{\ominus}(S_{\mu\nu}, s_\xi)}, \tag{5.14}$$

[7] For example, if there is a stretch of 4 consecutive complementary bases between $S_{\mu\nu}$ and s_ξ, that (4-base-long) complementary domain will be accepted, but smaller domains from subsets of these 4 bases (namely two 3-bases-long complementary domains and three 2-bases-long complementary domains.) will be rejected, as they overlap fully with the longer (4-base-long) domain.

where $p_\eta(S_{\mu\nu}, s_\xi)$ is the probability at equilibrium that s_ξ will bind with $S_{\mu\nu}$ through a complementary section η, $c_{s_\xi} \equiv c$ =100 nM is the concentration of a SST of type ξ (assumed the same for all SST types), $c_0 = 1$ M, and $p_\varnothing$ is the probability at equilibrium of finding the pair $S_{\mu\nu}$, s_ξ in the unbound state $\varnothing$. For a pair $S_{\mu\nu}, s_\xi$, we define an effective binding free energy, $\Delta G^{\ominus}_{\text{eff}}(S_{\mu\nu}, s_\xi)$, as:

$$p(S_{\mu\nu}, s_\xi) = p_\varnothing \frac{c_{s_\xi}}{c_0} e^{-\beta \Delta G^{\ominus}_{\text{eff}}(S_{\mu\nu}, s_\xi)}. \tag{5.15}$$

$\Delta G^{\ominus}_{\text{eff}}(S_{\mu\nu}, s_\xi)$ can be determined by comparing Eq. 5.14 with Eq. 5.15. As can be seen in annex F, a misbond is typically far less stable than a correctly bonded SST. However, for a given $S_{\mu\nu}$, there are 333 different SSTs that can bind incorrectly to $S_{\mu\nu}$. The probability that a misbond (of any kind) will form with $S_{\mu\nu}$ is given as:

$$p_{\text{misbond}}(S_{\mu\nu}) = \sum_\xi p(S_{\mu\nu}, s_\xi) = p_\varnothing \frac{c_s}{c_0} \sum_\xi \sum_{\eta_{(\xi)}} e^{-\beta \Delta G^{\circ}_{\eta_{(\xi)}}(S_{\mu\nu}, s_\xi)}. \tag{5.16}$$

As can be seen in Figs. 5.7 and 5.8, for all types of binding site, a misbond is far less likely to be observed in equilibrium than a correctly bonded SST. Furthermore, for a relatively wide range of temperatures (down to ~ 40°C), we find $p_{\text{misbond}}(S_{\mu\nu}) < p_\varnothing$, implying that the melting propensities of misbonds are sufficiently high to compensate for their formation propensities (however high) in this temperature range. The observations above imply that misbonding, such as described here, has at most a minor effect on the thermodynamics of self assembly provided the temperature is not decreased too much below T_c (at least not below ~ 40°C).

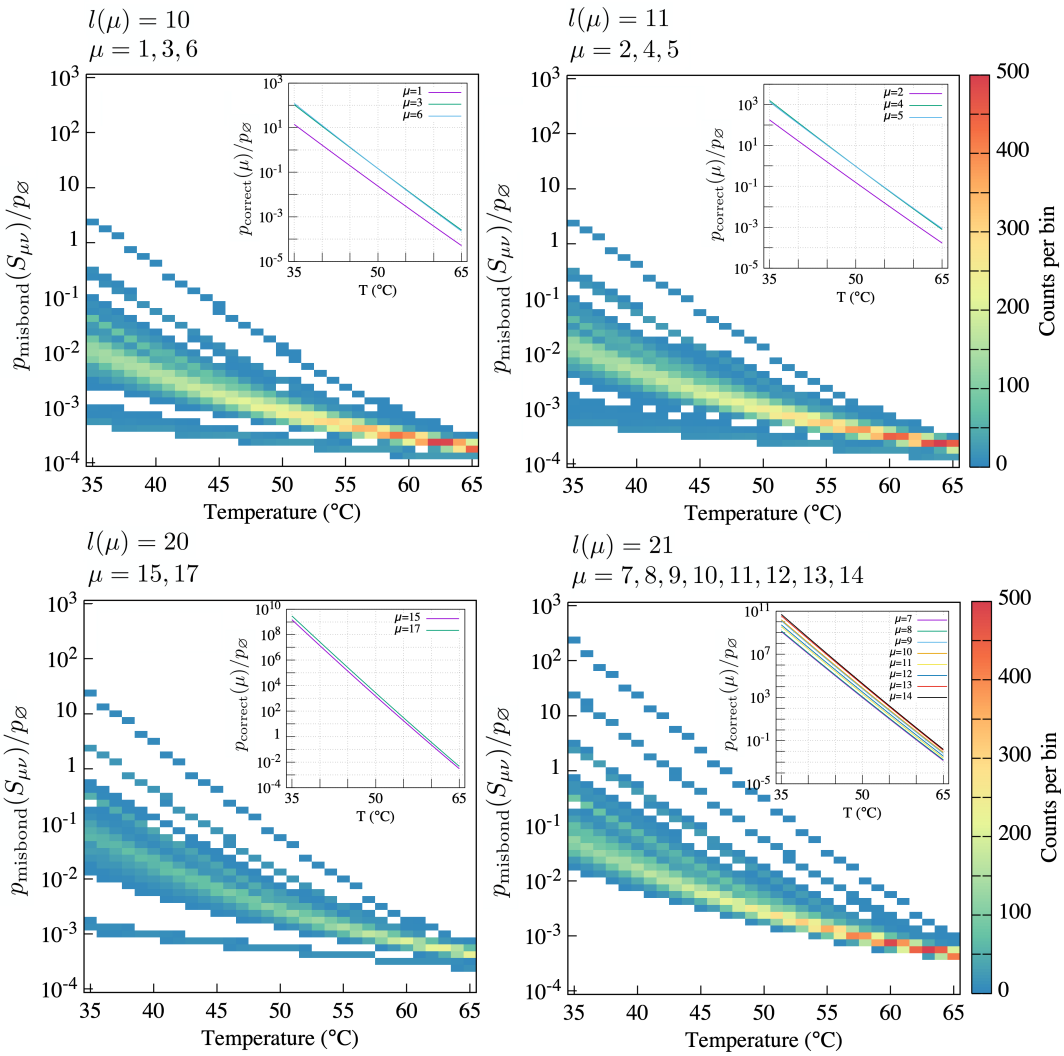


Figure 5.7: Probability at equilibrium of forming a misbond relative to the reference (unbound) state, for different types of binding site (μ), as a function of temperature. In each plot we present the results obtained for a given $l(\mu)$, which corresponds to the cumulative length of the binding domains of a binding site of type μ (hence each plot covers various types of binding site μ that yield the same $l(\mu)$). Top left: $l(\mu) = 10$; top right: $l(\mu) = 11$; bottom left: $l(\mu) = 20$; bottom right: $l(\mu) = 21$. The insets show the binding probability at equilibrium for SSTs with correct sequences, relative to the reference (unbound) state.

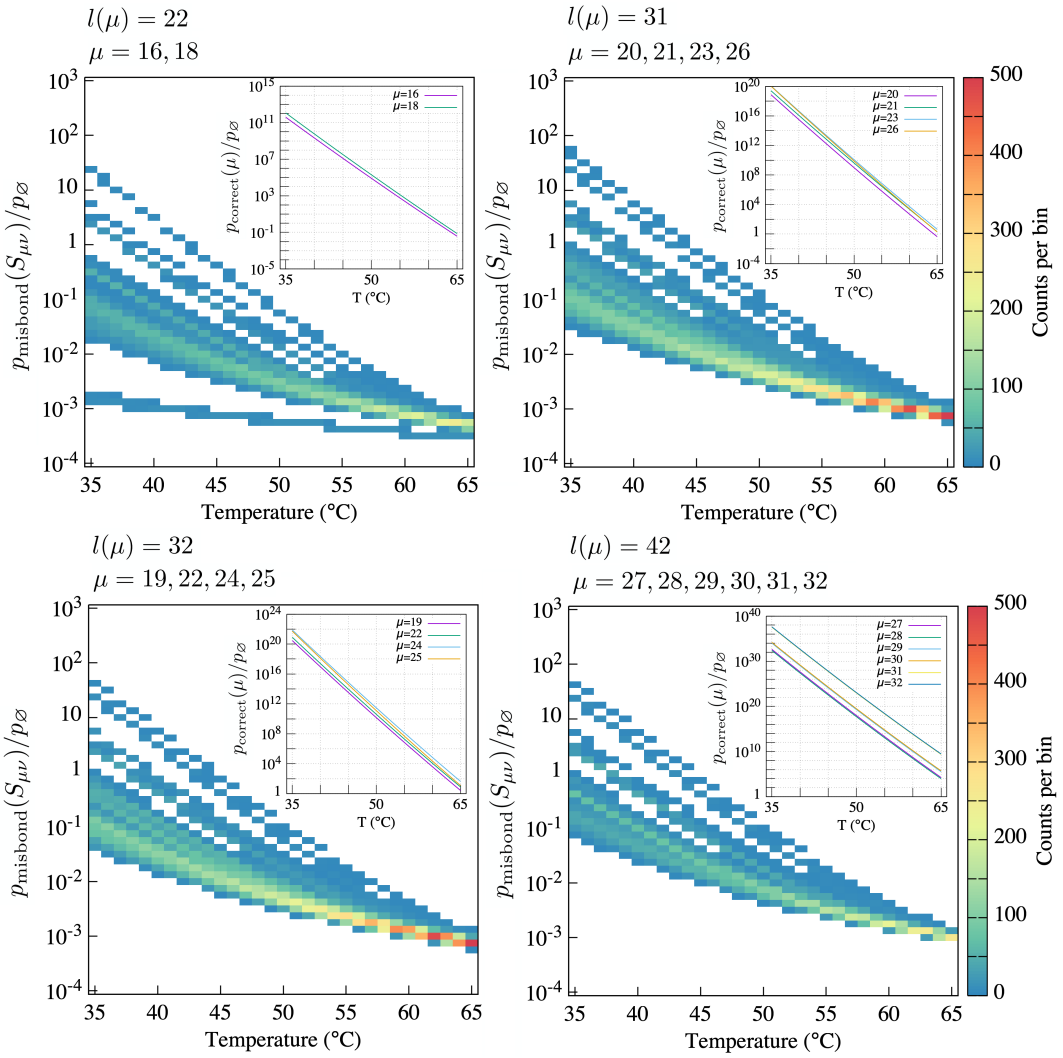


Figure 5.8: Probability at equilibrium of forming a misbond relative to the reference (unbound) state, for different types of binding site (μ), as a function of temperature. In each plot we present the results obtained for a given $l(\mu)$, which corresponds to the cumulative length of the binding domains of a binding site of type μ (hence each plot covers various types of binding site μ that yield the same $l(\mu)$). Top left: $l(\mu) = 22$; top right: $l(\mu) = 31$; bottom left: $l(\mu) = 32$; bottom right: $l(\mu) = 42$. The insets show the binding probability at equilibrium for SSTs with correct sequences, relative to the reference (unbound) state.

5.5 The effect of misbonds on the kinetics of self assembly

Next, we examine the impact of misbonds on the kinetics of self assembly. To parametrise the binding rate constant c_+ for SSTs with incorrect sequences we use the same value $k_+ = 6 \times 10^{-6}\mathrm{M}^{-1}$ as for SSTs with correct sequences. Therefore, the binding propensity for each type of SST to a binding site is given (regardless if that SST is of the correct type or not) as k_+c, with $c = 100$ nM. Dissociation propensities for misbonds are determined as $k_+e^{-\beta\Delta G^\circ_{\mathrm{eff}}(S_{\mu\nu},s_\xi)}$, where the effective free energies $\Delta G^\circ_{\mathrm{eff}}(S_{\mu\nu}, s_\xi)$ are calculated on-the-fly as described in section 5.4.

We perform simulations using a variant of the algorithm described in section 5.1. Here, the structural matrix $\boldsymbol{\sigma}$ is augmented with a sufficient number of extra rows such that fragment growth is not limited in an artificial manner[8]. Additionally, we add a third column to $\boldsymbol{\sigma}$ that specifies, for a SST present in the fragment in a coordinate i (hence $\sigma_{i,1} = 1$), if that SST is correctly bonded ($\sigma_{i,3} = 0$) or not ($\sigma_{i,3} = 1$); and a fourth column that specifies, for a SST present in the fragment in a coordinate i, the respective SST type (hence $\sigma_{i,4} = 1, \cdots, 334$). SSTs that are considered correctly bonded are always found in a coordinate i of $\sigma_{i,j}$ that corresponds to the respective SST type, similar to the algorithm described in section 5.1. For example, a fully-formed R6 structure (without misbonds) is represented as $\sigma_{i,1} = 1$, $\sigma_{i,3} = 0$ and $\sigma_{i,4} = i$ for $i \in [1, 334]$; and $\sigma_{i,1} = 0$ for $i > 334$. If a SST binds to one (or various) correctly bonded SST(s) in the fragment via native bonds, that SST is considered correctly bonded as well. If a SST present in the fragment is correctly bonded, it is assumed that, if there are any other SSTs incorrectly bonded to it, these bonds are assumed to be too weak to change meaningfully the dissociation propensity. In other words, we determine dissociation propensities for correctly bonded SSTs as if misbonds where not present. For misbonds on the other hand, we determine dissociation propensities taking into account all SSTs present in their local environments (regardless if they are misbonds as well, or not).

[8] We recall that the number of rows of $\boldsymbol{\sigma}$ corresponds to the maximum size a fragment can reach.

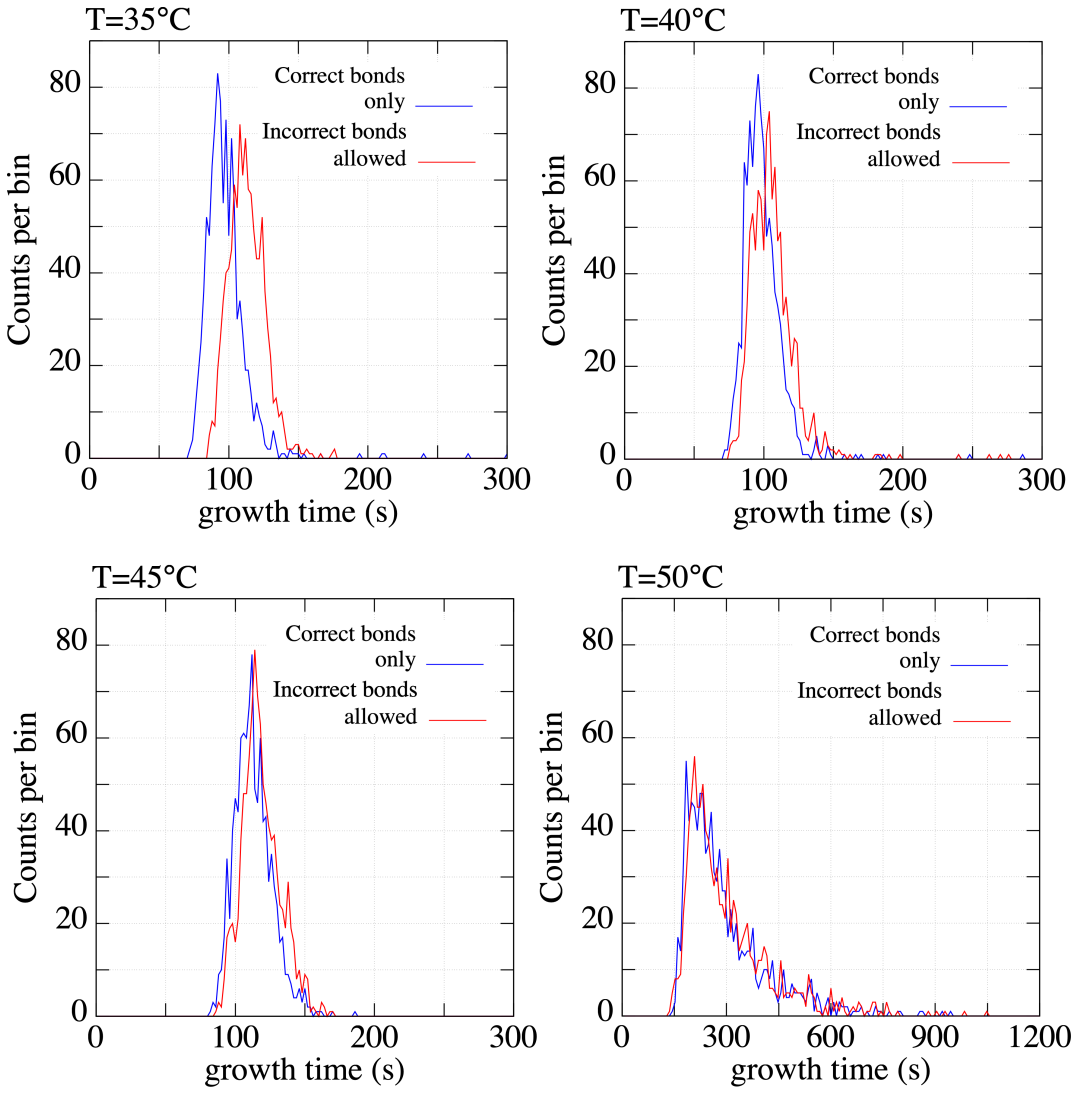


Figure 5.9: Distribution of growth times with (red curves) and without (blue curves) misbonds being allowed to form. The growth time corresponds to the time needed for a fragment, initially consisting of a single SST, to reach its equilibrium size for the first time. For each temperature, we performed 2,000 independent runs.

We simulated fragment growth under isothermal conditions at several temperatures in the range $35° - 60°$C. The simulations were initiated from a fragment state $f_i = \delta_{ij}$, where $j \in [1, 334]$ is a random (integer) number, corresponding to a single SST (deemed "correct"). For each temperature value, we recorded the time required for the fragment to reach for the first time the equilibrium size for that temperature, as determined in section 5.2.1. We find growth times to be largely insensitive to misbonds in the temperature range considered. For instance, in Fig. 5.9 we show a distribution of fragment growth times at the lower end of

the temperature range ($35° - 50°$C), where misbonds have a somewhat greater impact on kinetics. While misbonds form much more often than correct bonds, they typically have short lifetimes and melt away very quickly once formed.

5.6 Neglected effects

In this section we discuss the relative importance of two types of binding events that are not described in our model: binding events that involve bringing into close proximity two distant parts of a fragment (bridging events); and binding events where a SST binds with a subset of the binding domains in the binding site, and some binding domains are left unhybridised (incomplete binding).

5.6.1 Bridging events

The binding of a SST is expected to involve additional entropic costs if it involves constraining parts of the fragment that, previous to the binding event, were distant, and could move relative to each other. We gauge the relative importance of such "bridging" events by performing simulations with two variants of the algorithm described in section 5.1. In the first variant, the association and dissociation propensities for bridging SSTs are assumed to be the same as for an otherwise equivalent binding site, where both sections of the fragment are constrained, and the local approximation (see section 4.1.1) applies. As an example, the association and dissociation propensities for the binding site in the appendix G were assumed to be the same as for a binding site of type $\mu = 7$ (see appendix A). In this variant, the entropic penalty associated with forming a bridge is neglected, hence $\Delta S_{\text{bridge}} = 0$. In the second variant, the formation of bridges of is not allowed. Therefore, we effectively set the limit $\Delta S_{\text{bridge}} \to \infty$. Of course, the entropic cost of forming a bridge will lie between both limits (0 and ∞). We made repeats of the simulations described in sections 5.2.1 (fragment growth under isothermal and thermal ramp conditions) and 5.2.2 (free-energy functions $F(N)$), finding no perceptible differences in the results. In all cases, we find bridge formation to be exceedingly rare. Therefore, we assume that bridging events are sufficiently rare

that not modelling them properly does not change the results in a meaningful way. The results presented throughout this thesis were obtained using the second variant of the algorithm described above (where association (dissociation) events that involve forming (severing) bridges are not allowed to occur).

5.6.2 Incomplete binding

Our model does not describe events where a SST binds with a subset of the binding domains in a binding site, leaving some binding domains unhybridised. Our rational for not describing such events is that we expect the hybridisation of secondary domains to be a fast process compared to the hybridisation of the first domain, particularly given the low concentrations of SSTs (10^{-7} M per type). This is because the hybridisation of a secondary domain is a first order process that depends only on local factors, namely a reactive collision between single strands that are already in close proximity (hence a small entropic barrier - see appendix B). Conversely, the hybridisation of the first domain is a second order process involving dilute reactants (hence a high entropic barrier - see appendix B). Therefore, once a first binding domain is initiated successfully it should be a good assumption that there is sufficient time for the system to equilibrate locally and reach its most stable configuration, before the next binding event, involving a nearby site, starts. For all types of binding site, the most stable configurations (where base pairs are formed) involve a large amount of base pairing between all complementary domains involved (see appendix B). It should be noted that at sufficiently high SST concentrations, the assumption of a timescale separation between first- and second-order events will, of course, break down; but we expect this to occur at much higher SST concentration than that considered here.

5.7 Summary and discussion

In this chapter, we introduce a stochastic model that describes the growth of one fragment of R6 under a fixed background concentration of (unbound) SSTs. The model predicts that there is only a narrow temperature window ($\sim 4 - 5°$C) where

the nucleation barrier can be crossed in experimentally accessible timescales, while also being sufficiently high that growth times, post nucleation, are much shorter than the typical lag period before a nucleation event takes place. Such a separation of timescales is necessary for kinetic traps due to monomer depletion to be prevented in systems where monomers are present in finite supply (such as in the experiments of Refs. [2, 3]). However, in this temperature window we find the size of the fragment at equilibrium to fall below that of R6. Further cooling (down to about 15°C below the lower limit of the window) was found necessary for a completed structure to be observed in the simulations with high probability.

We find that classical nucleation theory (CNT) fits quite well the free-energy functions $F(N)$ predicted by the model close to the thermodynamic T_c, where the nucleation barrier is the highest, and the critical fragment size is relatively large. The fit becomes worse for lower temperatures where the nucleation barrier quickly becomes very small (both in width and height). This is because for small sizes, $F(N)$ displays strong features associated with discrete changes in connectivity, which naturally cannot be captured in a model as coarse as CNT.

We find that fragments with a high connectivity to be the the ones that are the most likely to form in equilibrium. However, maximally-connected fragments form with a probability that decreases as the fragment size increases. This decrease in probability is found to be not non-monotonic. It shows instead a pattern of recurrence with local minima at sizes $N_{\min} = M_1 \times M_2$ with $|M_1 - M_2| \leq 1$,[9] followed by local maxima at sizes just above $(N_{\min} + 1)$. This is found to be a statistical effect: for sizes $N = N_{\min}$ the number of different bonding configurations for which the connectivity is maximal has a sharp minimum, whereas for $N = N_{\min} + 1$ it has a local maximum.

We find that undesirable bonds between SSTs with incorrect sequences and a (otherwise correctly formed) fragment - a type of assembly error that was thought in the past to render the SST method unviable - has a relatively minor impact

[9] e.g. $N_{\min} = 2 \times 2$, 2×3, 3×3, 4×3, 4×4, 4×5, etc.

on both the kinetics and thermodynamics of self assembly, for a fairly wide range of temperatures below T_c.

Finally, we note that in Refs. [2, 3] (where the production of R6 was reported), the total number of SSTs in the reaction vessels was kept fixed during the experiments. In such conditions, binding rates are expected to decrease with time as SSTs are gradually consumed to form fragments. In the next chapter, we will describe reacting systems where SST are present in limited supply, and monomer depletion effects can be described explicitly.

6

Simulations in the NVT ensemble

In chapter 5 we have worked in a grand canonical ensemble with a constant chemical potential for the SSTs. In experiments, however, the number of SSTs are typically fixed, and so an NVT ensemble is more appropriate. As the target shapes grow, the local concentration (and effectively the chemical potential) of the SSTs goes down. This monomer-depletion effect can have important implications for the assembly dynamics[205–207]. The basic prediction is that monomer depletion effects become important when nucleation is fast compared to (further) growth. In that case, many more fragments will nucleate and grow further than needed for full assembly, up to the point monomers in the system become depleted. Full assembly can only occur via exchanges of SSTs among fragments, which is typically a very slow process, often referred to as Oswald ripening [208]. In the other limit, where nucleation is too slow compared to growth, assembly will occur sequentially, but the process may also be very slow. The picture above suggests that for SST assembly there should be an optimal temperature between both limits where assembly is fastest. In this chapter, we study assembly in the NVT ensemble, with a focus on the basic physics of nucleation and elongation. One advantage of our coarse-grained kinetic model is that it allows us to reach the time-scales of extremely slow processes such as nucleation and Oswald ripening, thus providing access to fine details of competing processes in assembly.

In this chapter, we use the rate constants $c_+(\mu)$ and $c_-(\mu)$ determined in chapter 4 (sections 4.3.1 and 4.3.2), to parametrise a stochastic model that describes the simultaneous growth of multiple fragments in a finite system, where the total number of SSTs is fixed. We use the model to simulate an assembly reaction under isothermal and thermal ramp conditions, with a focus on how the level of undercooling applied to the system determines the reaction timescales and the occurrence or not of depletion traps. This chapter is organised as follows. In section 6.1 we introduce the model, and describe the methods used to simulate it. In section 6.2 we use the model to simulate an assembly reaction under isothermal conditions. We next use the model to simulate an assembly reaction under a temperature ramp (section 6.3), and compare the results against experimental measurements of Refs. [2] and [3].

6.1 Simulation Details

In this chapter we describe the time evolution of a finite system with volume V containing M fragments and unbound SSTs, in contact with a thermal bath at temperature T. The populations of unbound SSTs are specified by a vector $n = (n_1, \cdots, n_{334})$ where n_i is the number of unbound SSTs of type i present in the system. M and n can change with time but the total number of SSTs in the system of each SST type remains fixed. As discussed in section 4.3, if assembly errors due to incorrectly- or incompletely-bonded SSTs are neglected, the state of a fragment k ($k = 1, \cdots, M$) is specified by a vector with 334 components $\vec{f}^{(k)} = (f_1^{(k)}, \cdots, f_{334}^{(k)})$, where $f_i^{(k)} = 1$ if a SST of type i is present in the fragment and $f_i^{(k)} = 0$ otherwise.

For generating state-to-state dynamics we use a custom algorithm based on the kinetic Monte-Carlo (kMC) scheme described in section 3.3.3. A fragment k ($k = 1, \cdots, M$) is represented in the algorithm by an integer-valued "structural" matrix $\sigma_{i,j}^{(k)}$, with $i \in [1, 334]$ and $j \in [1, 2]$, and a real-valued "propensity" vector $\alpha_i^{(k)}$ with $i \in [1, 334]$:

- The first column of $\boldsymbol{\sigma}^{(k)}$ specifies if a SST of type i is present ($\sigma_{i,1}^{(k)} = 1$) or absent ($\sigma_{i,1}^{(k)} = 0$) from fragment k. Hence, we have $\sigma_{i,1}^{(k)} \equiv f_i^{(k)}$.

- If fragment k contains a SST of type i (hence $\sigma_{i,1}^{(k)} = 1$), the second column of $\boldsymbol{\sigma}^{(k)}$ stores an integer between 1 and 32 that specifies the type of binding site to which that SST is bonded to. If a SST of type i is absent from fragment k (hence $\sigma_{i,1}^{(k)} = 0$) the second column of $\boldsymbol{\sigma}^{(k)}$ stores the value 0 if that SST type is compatible with an empty binding site in the fragment, or -1 (otherwise).

- If a SST of type i is present in fragment k (hence $\sigma_{i,1}^{(k)} = 1$), $\alpha_i^{(k)}$ stores a dissociation propensity $k_+ e^{\beta \Delta G^{\ominus}(\sigma_{i,2}^{(k)})}$ according to the type of binding site ($\sigma_{i,2}^{(k)}$) to which that SST is bonded to. If a SST of type i is absent from fragment k but can bind to it (hence $\sigma_{i,1}^{(k)} = 0$ and $\sigma_{i,2}^{(k)} = 0$), $\alpha_i^{(k)}$ stores a binding propensity $k_+ n_i/(N_A V)$ where N_A is the Avogadro constant. If a SST of type i is not present in fragment k and cannot bind to it (hence $\sigma_{i,1}^{(k)} = 0$ and $\sigma_{i,2}^{(k)} = -1$), $\alpha_i^{(k)}$ stores the value 0.

Additionally, the algorithm keeps track of the fragment size, $N^{(k)} = \sum_i \sigma_{i,1}^{(k)}$, the cumulative propensity function $A(l)^{(k)} = \sum_{o=1}^{l} \alpha_o^{(k)}$, and the overall propensity for an event that involves fragment k, $A_0^{(k)} \equiv A(334)^{(k)}$. Also, the algorithm keeps track of the cumulative propensity function $A_{\text{frag}}(l) = \sum_{o=1}^{l} A_0^{(o)}$, where $A_{\text{frag}}(M)$ is the overall propensity for an event that involves any of the (M) fragments in the system.

Given the complementarity pattern between the different types of SST, it turns out that there are 616 different combinations of two SST types that share complementary domains, and from which a dimer can form. A dimmer of type m ($m \in [1, 616]$) containing SST types i and j has a formation propensity $a_d^{(m)} = k_+ n_i n_j/(N_A V)$ (see table 3.3.2). The algorithm keeps track of the formation propensities $a_d^{(m)}$ as well as the cumulative propensity function $A_d(l) = \sum_{o=1}^{l} a_d^{(o)}$, where $A_d(616)$ is the overall propensity for forming a dimer (of any type). The overall propensity for any type of event in the system, A_0, can be written as $A_0 = A_{\text{frag}}(M) + A_d(616)$.

To propagate the system in time, the algorithm cycles through the following steps:

1 Generate two random numbers $r_1, r_2 \in [0,1]$. Advance time by an amount $\delta t = (1/A_0)\ln(1/r_1)$. If $r_2 A_0 \leq A_d(616)$ go to step 2.1 (and form one dimer); otherwise skip to step 2.2 (and execute one move on a fragment already formed).

2.1 Find the type of dimer to form by solving $A_d(m-1) < r_2 A_0 \leq A_d(m)$ with respect to m. Allocate and determine the entries of a new matrix $\boldsymbol{\sigma}^{(M+1)}$ and a vector $\boldsymbol{\alpha}^{(M+1)}$ used to represent the newly formed dimer. Update the populations SST types i and j contained in the dimer as $n_i \to n_i - 1$ and $n_j \to n_j - 1$, and recalculate the formation propensity of all dimers that contain these SST types. For fragments k' to which these SST types (i and j) can bind to, recalculate the binding propensities $\alpha_i^{(k')}$ and $\alpha_j^{(k')}$, and update $A(l)^{(k')}$ accordingly. Finally, update $A_d(l)$ and $A_{\text{frag}}(l)$, compute A_0 for the new state, and skip to step 3.

2.2 Find the fragment in which the move will be executed by solving $A_{\text{frag.}}(k-1) < r_2 A_0 - A_d(616) \leq A_{\text{frag.}}(k)$ with respect to k. Find the coordinate in fragment k where the move will be executed by solving $\sum_{i'=1}^{i-1} \alpha_{i'}^{(k)} < r_2 A_0 - A_d(616) - A_{\text{frag.}}(k-1) \leq \sum_{i'=1}^{i} \alpha_{i'}^{(k)}$ with respect to i. If fragment k is a dimer and coordinate i contains a SST, go to step 2.2.A (and erase the dimer); otherwise skip to step 2.2.B.

2.2.A Deallocate $\boldsymbol{\sigma}^{(k)}$ and $\boldsymbol{\alpha}^{(k)}$. Update the populations of SSTs of types i and j as $n_i \to n_i + 1$ and $n_j \to n_j + 1$, where j was the type of the second SST in fragment k, and recalculate the formation propensity of all dimers that use these SST types. For fragments k' to which these SST types (i and j) can bind to, recalculate the binding propensities $\alpha_i^{(k')}$ and $\alpha_j^{(k')}$, and update $A(l)^{(k')}$ accordingly. Finally, update $A_d(l)$ and $A_{\text{frag.}}(l)$, compute A_0 for the new state, and skip to step 3.

2.2.B If coordinate i contains a SST, remove that SST ($\sigma_{i,1}^{(k)} = 1 \to 0$) and update n_i as $n_i \to n_i + 1$; otherwise, add a SST ($\sigma_{i,1}^{(k)} = 0 \to 1$) and

update n_i as $n_i \rightarrow n_i - 1$. Recalculate $\sigma_{i,2}^{(k)}$ and $\alpha_i^{(k)}$ for first- and second-neighbouring sites of coordinate i, and update $A(l)^{(k)}$ accordingly. Recalculate the formation propensity of all dimers that contain a SST of type i. For fragments k' to which a SST of type i can bind to, recalculate the binding propensity $\alpha_i^{(k')}$, and update $A(l)^{(k')}$ accordingly. Finally, update $A_d(l)$ and $A_{\text{frag.}}(l)$, compute A_0 for the new state, and go to step 3.

3 Run any relevant subroutines (e.g. measurements, temperature changes, check for stopping conditions), and go to step 1.

6.2 Self assembly under isothermal conditions

We begin by making an estimate of the thermodynamical critical temperature of the NVT system. Assuming equal consumptions among SST types, the total free energy F_{TOT} of the system is given by:

$$\beta F_{\text{TOT}}/\nu_0 = \sum_{N=1}^{334} \beta^{-1}\rho_N[\log(\rho_N\nu_0) - 1] + \rho_N G(N), \tag{6.1}$$

where ρ_N is the density of fragments of size N, ν_0 is a standard-state volume, and $G(N)$ is a one-dimensional projection of the free-energy onto the number of SSTs in a fragment (see section 5.2.2). To determine equilibrium concentrations one minimises Eq. 6.2 subject to the constraint that the total concentration of SSTs in the system, ρ_{TOT}, is conserved:

$$\sum_{N=1}^{334} N\rho_N = \rho_{\text{TOT}}. \tag{6.2}$$

This leads to the law of mass action (see e.g. Ref. [209]):

$$\begin{aligned} \rho_N\nu_0 &= \exp[-\beta(G(N) - N\mu)] \\ \beta\mu &= \log(\nu_0\rho_1) \end{aligned} \tag{6.3}$$

In practice, Eqs. 6.2 and 6.3 are solved self consistently with respect to ρ_N. We determine a thermodynamical critical temperature $T_c = 57.43°\text{C}$ for a total SST

concentration of 100 nM per SST type. Here, we define thermodynamical critical temperature as the temperature for which half of the SSTs in the system remain unbound at the other half is used in fragments. The fraction of SSTs that remain unbound, $\phi_1^{\mathrm{eq}}(T)$, as a function of temperature is shown in Appendix I.

We start our analysis by simulating an assembly reaction under isothermal conditions. The simulations were initiated from a configuration with 100 unbound SSTs per SST type (33400 SSTs total) and no fragments, such that 100 R6 structures can in principle form in the system. The volume (V) was fixed at 3.321×10^{-15} dm^3 such that each SST type is initially present at a concentration 100 nM as done in the experiments of Refs. [2, 3]. The simulations were performed at different temperatures in the range 50°C-60°C; for each temperature, $n_{\mathrm{sim}} = 100$ independent simulations were carried out, each of which for $2.5{\times}10^8$ kMC steps.

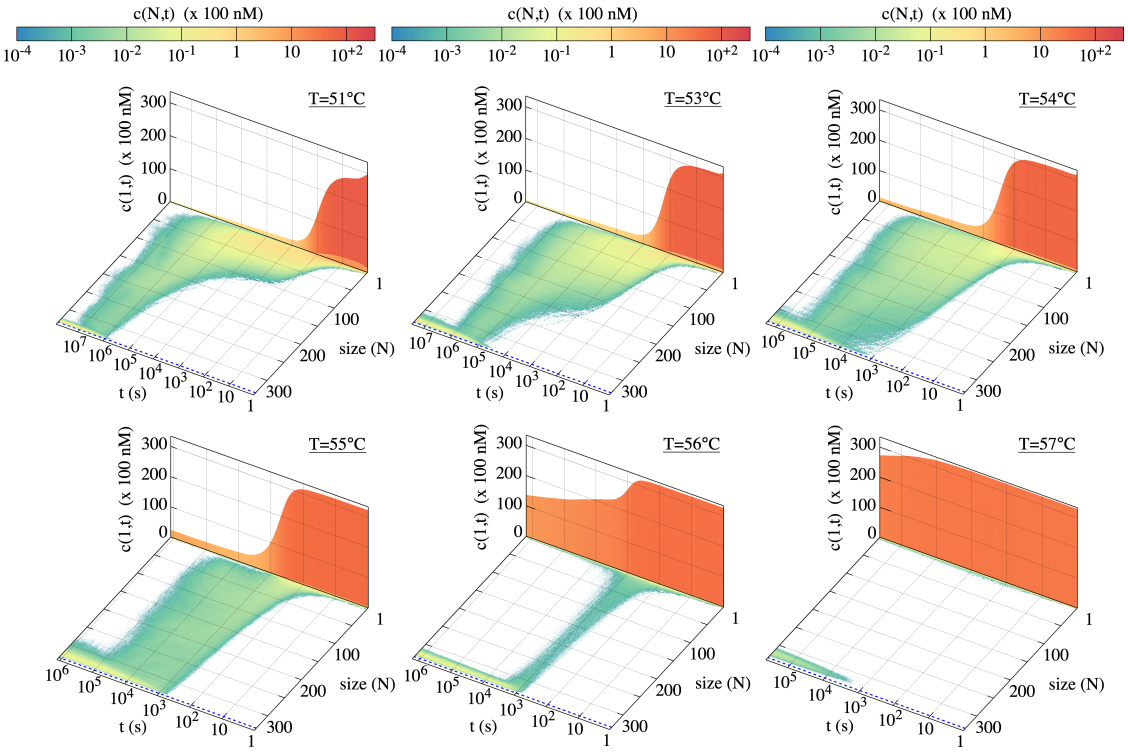


Figure 6.1: Self assembly under isothermal conditions - mean concentration of fragments as a function of fragment size (N) and time (t). Each data point shown corresponds to an average over $n_{\mathrm{sim}} = 100$ simulations. The dashed blue lines at $N = 334$ indicate the size of R6. The vertical axis denotes the monomer concentration $c(1, t)$ as a function of time.

Fig. 6.1 presents mean concentration values, $\langle c(N, t)\rangle_{n_{\mathrm{sim}}}$, where N is a fragment

size, t is time, and the angular brackets $\langle\,\rangle_{n_{\rm sim}}$ denote an average over $n_{\rm sim} = 100$ simulations. To keep the notation simple, we will omit the angular brackets $\langle\,\rangle_{n_{\rm sim}}$ when referring to averages between simulations. A few things to note from this figure. For temperatures 57°C and below (Fig. 6.1), a steady state is approached at long times where almost all SSTs are found in fragments with sizes close to R6, or unbound, with a small component in small fragments (dimers, trimers, etc.). For temperatures 58°C and above (not shown) no appreciable fragment growth was observed in the simulations. Note that as the temperature goes down, the growth modes change from a simple nucleated growth from monomers to fully formed structures, to something more akin to Oswald ripening, because there are many partially formed fragments that rapidly form and deplete the available monomers.

To make it easier to interpret the simulations we introduce a minimum size $N_{\rm tb} = 300$ (∼90% of the size of R6) above which a fragment is assumed to have finished its growth.

The size range $N \geq N_{\rm tb}$ will be referred to as the "target band", and we define $c_{\rm tb}(t) \equiv \sum_{N \geq N_{\rm tb}} c(N,t)$ (see Fig. 6.2 - left).

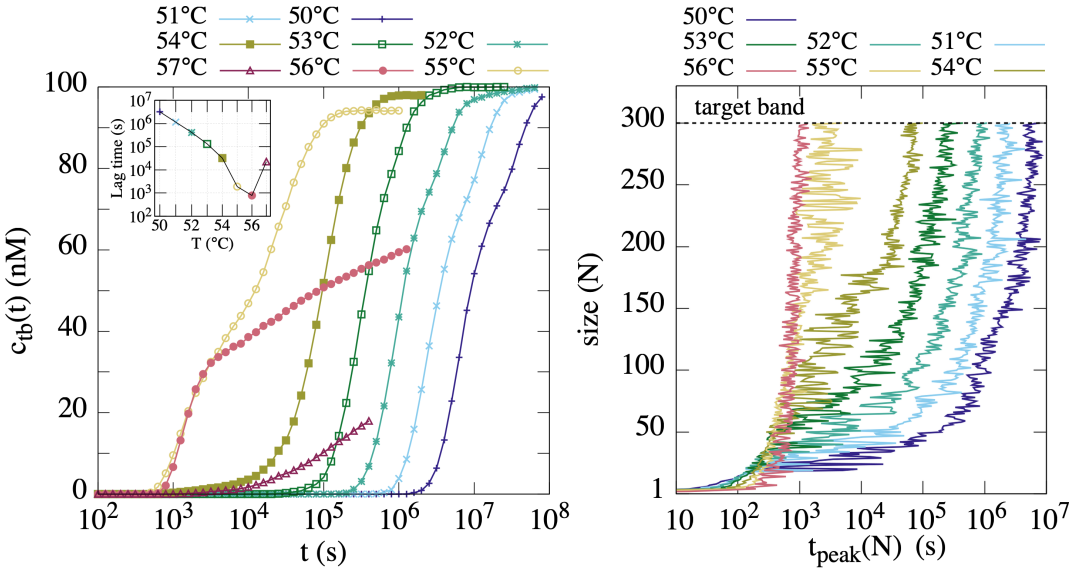


Figure 6.2: Self assembly under isothermal conditions. Left, main plot - mean concentration of fragments with sizes in the target band ($N \geq 300$) as a function of time. Left, inset - lag time as function of temperature. Right - time $t_{\rm peak}(N)$ when a concentration peak is reached for a fragment size N. Each data point shown (main figures) corresponds to an average over $n_{\rm sim} = 100$ simulations.

Starting from a configuration containing only unbound SSTs, the assembly reaction is characterised by a sequential increase in fragment concentrations with increasing fragment size. For a fragment size N, a peak concentration is first reached at a time $t_{\mathrm{peak}}(N)$ (Fig. 6.2 - right), to which follows a slow decrease towards a steady-state value. The production of fragments with sizes in the target band is negligible until fragments with sizes just below start to form in appreciable numbers. The lag time before the reaction chain reaches the target band first decreases with decreasing temperature, from 57°C to 56°C, and then increases with decreasing temperature for temperatures below 56°C (Fig. 6.2 - left, inset). For a given temperature, the lag time was determined by making a linear fit to $c_{\mathrm{tb}}(\log t)$ at the point where the difference $\Delta_{\{\log t\}} c_{\mathrm{tb}}(\log t)$ has a maximum (the fits are plotted against the main data in annex H). Lag time was defined as the point where the linear fit crosses the (log-)time axis.

Broadly, the changes observed in lag time were expected given the existence of a nucleation barrier that collapses quickly with decreasing temperature (see section 5.2.2). At higher temperatures, while nucleation is slow compared to growth, lowering the temperature decreases the lag time as the nucleation rate increases; at lower temperatures however, where nucleation is fast compared to growth, lowering the temperature increases the lag time as depletion traps become more severe. Nonetheless, the onset of monomer depletion at $T = 55$°C asks for further explanation.

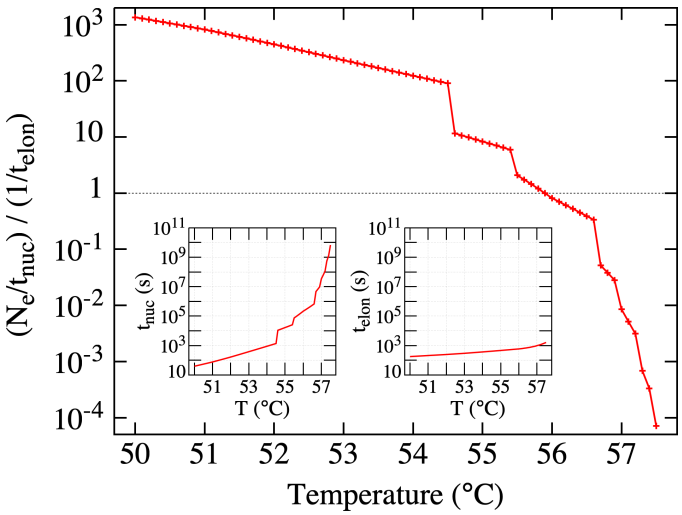


Figure 6.3: Ratio between the rate of monomer (SST) depletion by nucleation (N_e/t_{nuc}) and the elongation rate ($1/t_{elon}$), for a SST concentration of 100 nM per SST type. We fix $N_e = \langle N \rangle$ as the fragment size for which the elongation phase is assumed complete. According to the criterion suggested in Refs. [205, 206], a depletion trap is expected to occur for ratios greater or equal than 1. Insets: mean nucleation time (t_{nuc}) (left), and mean elongation time (t_{elon}) (right), for a SST concentration of 100 nM per SST type.

The mean nucleation (t_{nuc}) and elongation (t_{elon}) timescales of individual fragments under initial reaction conditions (where SST concentrations are at their peaks) can be used to asses if a depletion trap will occur in the course of the reaction. Here, we refer to elongation as the phase following nucleation where a fragment grows towards a stable size. It has been suggested [205, 206] that for finite-sized assemblies, depletion traps are expected to occur if the initial rate of monomer depletion by nucleation N_e/t_{nuc} is at least equal to the elongation rate $1/t_{elon}$, where N_e is the size of the assembly at the end of the elongation phase. To test the criterion suggested in Refs. [205, 206], we could in principle track down individual fragments that form early in the reaction, while SST concentrations remain close to their peak values, and measure directly t_{nuc} and t_{elon}. However, given the finite size of the system and the computational costs involved, it turns out to be very challenging to obtain statistically meaningful averages via direct measurements (nucleation times in particular show large fluctuations around mean values).

To determine t_{nuc} and t_{elon}, we simulate independently the growth of one fragment under a background concentration of SSTs fixed at 100 nM per SST type, using the

model described in section 5.1. For a given temperature, we perform a first simulation to determine the critical fragment size. Using the biasing scheme described in section 5.1.2, we determine the free-energy function $F(N) = -\beta^{-1} \ln p_{\rm eq}(N)$, where $p_{\rm eq}(N)$ is the probability at equilibrium of finding the fragment with size N,[1] from which we deduce the critical size $N^{\ddagger}$ for which $F(N)$ has a maximum.[2] We then perform 10^6 (unbiased) simulations, each initiated from a single SST of a type chosen at random[3], and determine $t_{\rm nuc}$ as the mean time that it takes for the fragment to grow past $N^{\ddagger}$ for the first time in a simulation. The last states visited during the nucleation simulations (where the simulations were terminated once a fragment reached size $N^{\ddagger}+1$ for the first time) were used as seed states for elongation simulations. In the latter, the critical size $N^{\ddagger}$ was set to act as a reflecting boundary, and elongation was assumed to be complete once a fragment reached its average (steady-state) size $\langle N \rangle$, determined previously in section 5.2.1 for the same conditions (see Fig. 5.2). $t_{\rm elon}$ was determined as the mean time that it takes for a fragment, initially with size $N^{\ddagger}+1$, to reach size $\langle N \rangle$ for the first time in a simulation.

The criterion suggested in Refs. [205, 206] correctly predicts the onset of monomer depletion at $T = 55$°C (Fig. 6.3). Nucleation times show, evidently, much larger changes with temperature (∼5 orders of magnitude between 50°C-56°C) than elongation times (which remain in the same order of magnitude between 50°C-56°C). The rapid increase in nucleation times for $T > 54$°C results from a rapid rise in both the height of the nucleation barrier and the critical fragment size (Fig. 6.4 - insets).

[1] The procedure for determining $p_{\rm eq}(N)$ is explained in the first paragraph of section 5.2.2

[2] Here, we determine $F(N)$ for temperatures 50°C-54°C; for temperatures 55°C and above, $F(N)$ was determined previously (see Fig. 5.4) for the same conditions.

[3] Or, in chapter's 5 notation: fragment state $f_i = \delta_{ij}$ where $j \in [1, 334]$ is a random integer.

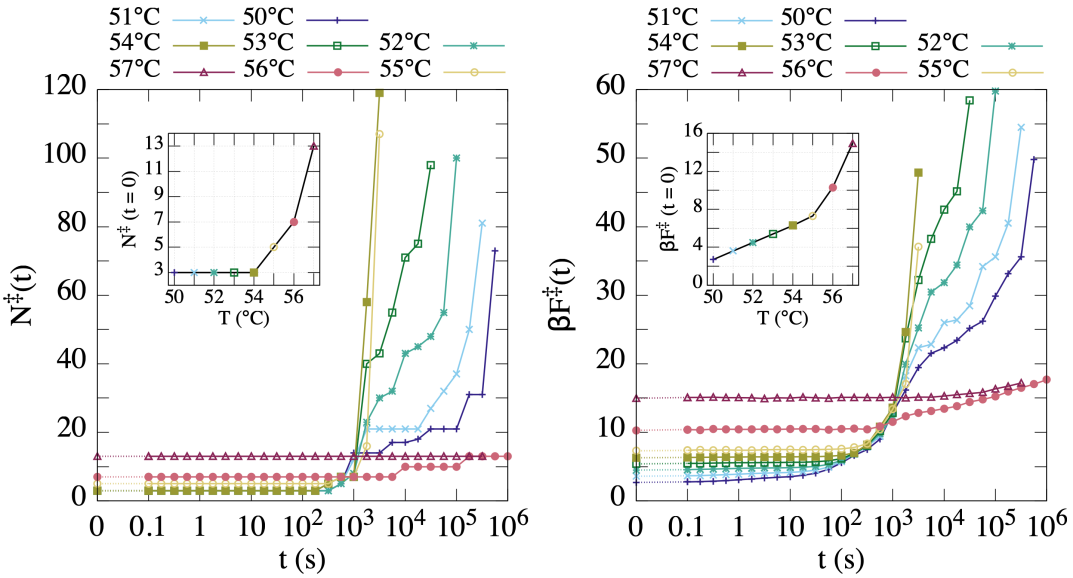


Figure 6.4: Critical fragment size (left) and height of the nucleation barrier (right) at various times during the assembly reaction. For each temperature, are shown results up to the point where simulations yield $F(N) > F(1)$ for all fragment sizes (N) above one, in which case a stable nucleus can no longer form from unbound SSTs that remain in the system. Insets - results for initial reaction conditions: $t = 0$ and SST concentration of 100 nM per SST type.

It is instructive to compare how the reaction evolves over time at different temperatures. One difficulty of interpreting the simulations, however, is that the critical size can increase over time as SSTs become gradually depleted. To determine the critical size at a given time t we take a snapshot of the instantaneous SST concentration, $c_i(t) = N_i/(N_A V)$ (where i specifies a SST type), and simulate independently the growth of one fragment under an equivalent background concentration of SSTs, using the algorithm described in section 5.1. We first determine the free-energy function $F_{(t)}(N) = -\beta^{-1} \ln p_{\mathrm{eq}}(N)_{(t)}$ using the biasing described in section 5.1.2, where $p_{\mathrm{eq}}(N)_{(t)}$ is the probability at equilibrium of finding the fragment with size N,[4] from which we deduce the critical size, $N^{\ddagger}_{(t)}$, for which $F_{(t)}(N)$ has a maximum (Fig. 6.4). The concentration of nuclei in the system is determined as $c_{\mathrm{nuc}}(t) = \sum_{N>N^{\ddagger}_{(t)}} c(N,t)$ (Fig. 6.4); for reference, the concentration of fragments with sizes above the initial critical size was also determined, as $c^{(0)}_{\mathrm{nuc}}(t) = \sum_{N>N^{\ddagger}_{(0)}} c(N,t)$.

[4]To determine $p_{\mathrm{eq}}(N)_{(t)}$ we followed the procedure outlined in the first paragraph of section 5.2.2

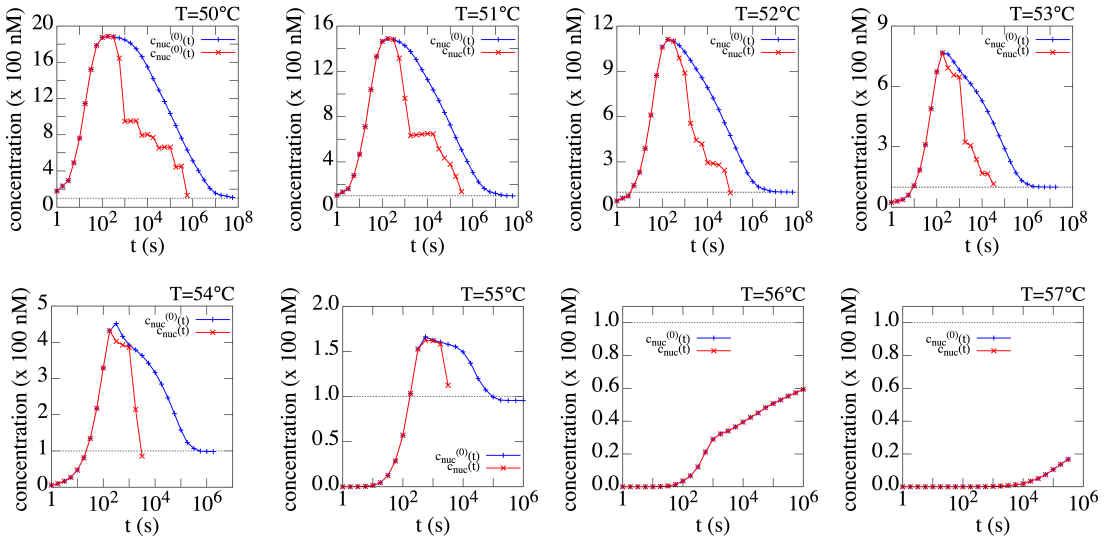


Figure 6.5: Isothermal self assembly - concentration of nuclei (red curves) and fragments with sizes above the initial critical size (blue curves) as a function of time.

At lower temperatures ($T \leq 54°C$) there is a rapid proliferation of nuclei in the system at early times ($t \lesssim 10^3$ s) and a near total depletion of SSTs. These nuclei are typically unable to grow past a few dozen subunits in size as SSTs become depleted far too quickly. To resolve the depletion trap that ensues, the system evolves slowly through the Ostwald-ripening analogue where larger fragments grow further at the expense of smaller ones. The near-exponential increase in lag times with decreasing temperature results mainly from the near-exponential decrease of the SST-dissociation propensities with decreasing temperature[5].

At $T = 56°C$ nucleation is rate-controlling but the nucleation barrier is sufficiently small that the timescale required for the first fragments to reach the target band approaches the elongation time (the minimum overall timescale of self assembly). For $T = 57°C$, an excessively higher nucleation barrier imposes an additional timescale to the system (hence the longer lag time relative to $T = 56°C$). For temperatures 56°C-57°C (both above the onset of monomer depletion), the consumption of SSTs over time raises the nucleation barrier making it increasingly

[5]In our model, dissociation propensities are determined as $k_+ e^{\beta \Delta G^{\ominus}(\mu)}$, where $\mu = 1, \cdots, 32$ is a type of binding site. As can be seen in Annex, the free-energies $\beta G^{\ominus}(\mu)$ show near-linear decreases with decreasing temperature (hence the near-exponential decrease of the dissociation propensities).

difficult for new nuclei to form (a steady state is therefore only reached asymptotically in the simulations).

At $T = 55°$C, there is a mild proliferation of nuclei at early times (up to ~ 160 nM); however, nucleation is sufficiently slow that many of these nuclei (~ 40 nM) have reached the target band by the time SSTs become depleted. The reaction then slows down, and proceeds (slowly) through the Ostwald-ripening analogue until a steady state is reached. The timescale required for the first fragments to reach the target band (by which time SST concentrations are at about 60% of their peak values) is again close to the elongation time.

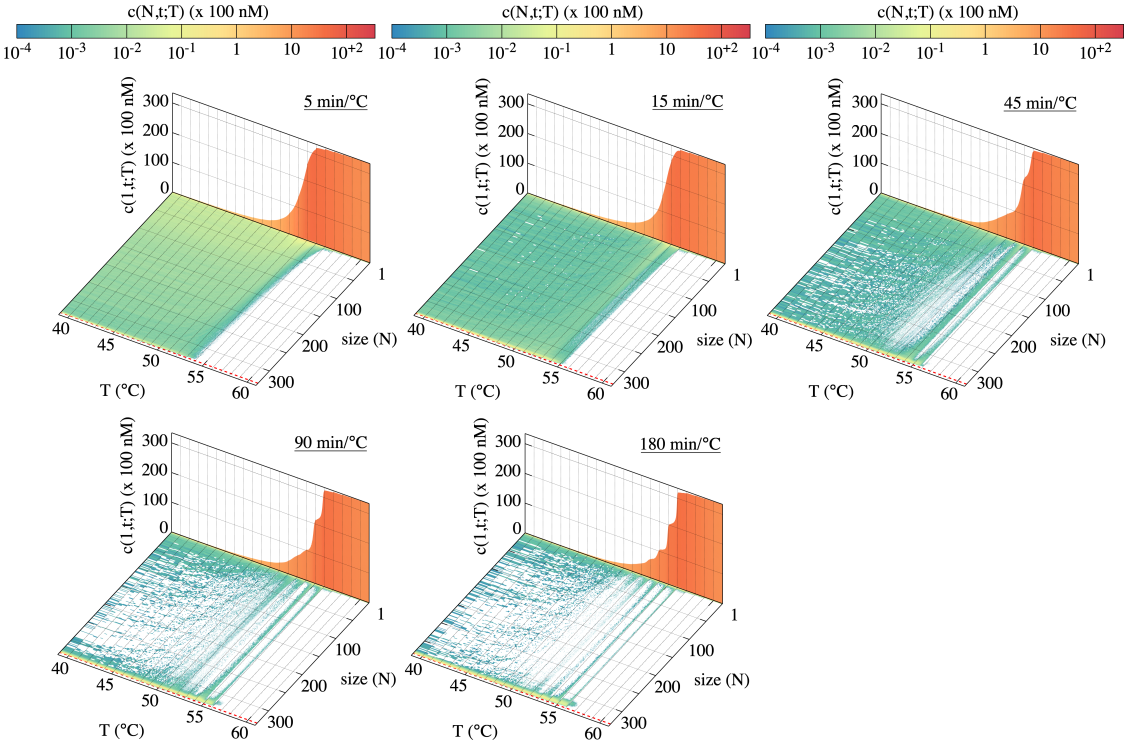


Figure 6.6: Simulation of a temperature ramp - mean concentration of fragments as a function of fragment size (N) and time (t). To make the figures easier to interpret, we plot the results in terms of a temperature variable $T = 65 - t/t_{\text{step}}$, where t_{step} is the cooling rate, and 65 corresponds, in Celsius, to the temperature at the start of the simulations. Note that in the simulations the temperature was decreased stepwise ($\Delta T = 1°$ C). Each data point in the figure corresponds to a mean value over $n_{\text{sim}} = 100$ simulations. The dashed red lines at $N = 334$ indicate the size of R6.

The width of the temperature window where productive self assembly can occur (depletion-trap free and accessible within experimental timescales) is expected to

change according to the size of the target structure. A small decrease in nucleation times with increasing target size is expected because less SST types will correspond to sites at the outer border of the target structure (which have a reduced number of binding partners) relative to SST types that corresponds to internal sites. To verify this we determine mean nucleation times for target structures similar in shape to R6 (with similar height-to-width ratios) but larger sizes (see annex J). Elongation times are expected to scale as $t_{\text{elon}} \sim N_{\text{target}}^{1/2}$, where N_{target} is the target structure size. An explanation for this dependence follows. For a perfectly circular fragment, the commutative propensity for a binding event, Γ_+, scales with the perimeter of the circle; hence $\Gamma_+ \sim r$, where r is the radius of the circle. The area of the fragment, $A \sim r^2$, is proportional to the number of subunits contained in it (as R6 is two-dimensional); therefore we have $\Gamma_+ \sim N^{1/2}$. Next, we assume intermediates to be all circular, and make a continuous approximation to the variable N. The elongation time is determined as $t_{\text{elon}} \sim \int_{N^\ddagger+1}^{N_{\text{target}}} N^{-1/2} \sim N_{\text{target}}^{1/2}$. To verify this formula we determine mean elongation times for target structures similar in shape to R6 (with similar height-to-width ratios) but larger sizes (see annex J).

We use the criterion $N_{\text{e}} t_{\text{elon}}/t_{\text{nuc}} \geq 1$ suggested in Refs. [205, 206] to determine the temperature threshold below which a depletion trap will occur (Fig. 6.7 - left). We make the approximation that the elongation size (N_{e}) scales with target size as $\sim \frac{N_{\text{target}}}{N_{\text{R6}}} N_{\text{e}}$ where $N_{\text{R6}} = 334$ is the size of R6, that nucleation times are independent from target size, and that elongation times scale as $t_{\text{elon}} \sim N_{\text{target}}^{1/2}$. We determine a nucleation timescale for the NVT system as $t_{\text{nuc}}/N_{\text{target}}$. Note that N_{target} corresponds to the number of different SST types in the system (each of which functions as a potential nucleation seed). We assume that self assembly is experimentally accessible if $t_{\text{nuc}}/N_{\text{target}} < 24$ h (Fig. 6.7 - right). As expected, the temperature window for productive assembly (depletion-trap free and experimentally accessible) shrinks with increasing target size, from a little over 1° C for R6, down to 0.1° C for target sizes $\sim 150 N_{\text{R6}}$ and above.

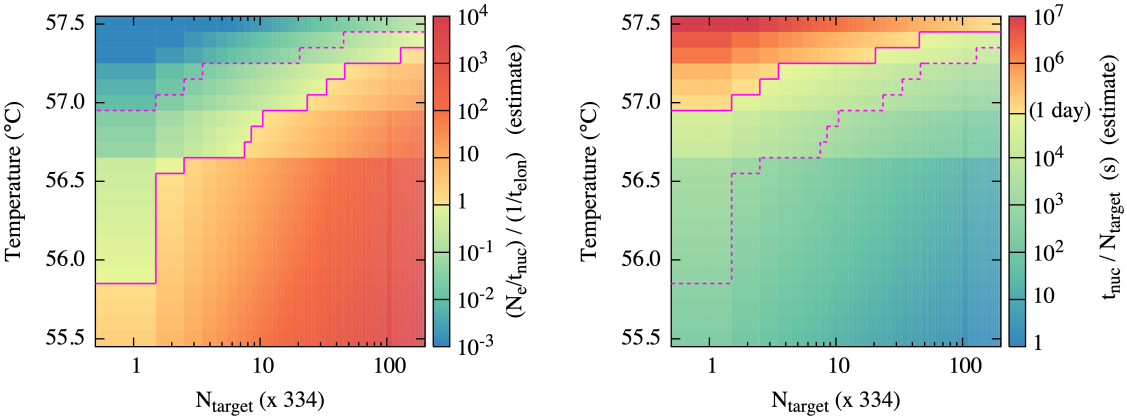


Figure 6.7: Temperature window for productive self assembly as a function of target structure size. Left - value of $N_e t_{elon}/t_{nuc}$ as a function of temperature and target structure size. Right - value of t_{nuc}/N_{target} as a function of temperature and target structure size. We use red and orange to colours regions of the parameters space for which assembly is not productive. The temperature window for productive assembly is shown in both plots (solid and dashed pink lines), and corresponds to the overlap region between both thresholds - monomer depletion and experimental accessibility.

6.3 Simulation of a temperature ramp and comparison with experiments

As discussed in section 4.2, in the experiments of Refs. [2, 3] R6 were produced by subjecting a mixture of SSTs to a step-wise cooling ramp ($\Delta T = 1°$ C). In this section we use the model to simulate a cooling ramp under similar conditions and compare the results with experimental measurements made available.

Ref. [3] presents the result of a fluorometric assay[6] performed during the course of a cooling ramp where a cooling rate of 180 min/°C was applied. They present relative changes in the fluorometric signal measured at the end of each temperature step. A more detailed analysis of the samples (e.g. final assembly yields, types of assembly errors encountered) was not published in Ref. [3]. In Ref. [2] native agarose gel electrophoresis [7] was applied to the samples post-assembly, and assembly yield was determined as the fraction of DNA material that ended up in complexes with sizes compatible with that of R6 (a fluorometric assay, if performed, was not

[6] A technique that measures real time changes in base pairing levels in a sample.

[7] A method for separating a population of DNA molecules and complexes according to their molecular sizes.

mentioned in Ref. [2]). In Ref. [2] measurements of assembly yields for cooling rates ranging from 5 min/°C to 120 min/°C are provided.

Our simulations were initiated from a configuration with 100 unbound SSTs per SST type, at T =65°C. The volume (V) was fixed at 3.321×10^{-15} dm^3 such that each SST type is initially present at a concentration 100 nM, such as in Refs. [2, 3]. The temperature was decreased in discrete steps (assumed instantaneous) of 1°C until it reached T =40°C. Several cooling rates were attempted, ranging from 5 min/°C to 180 min/°C; for a given cooling rate, $n_{\text{sim}} = 100$ independent simulations were carried out.

For the cooling rates tested, a concentration peak is visible for fragment sizes close to that of R6 (Fig. 6.6), which becomes narrower and approaches the size of R6 as the temperature decreases. Shown in Fig. 6.6 are mean concentration values, $\langle c(N,t;T)\rangle_{n_{\text{sim}}}$, where N is a fragment size, t is time, and the angular brackets $\langle\,\rangle_{n_{\text{sim}}}$ denote an average over $n_{\text{sim}} = 100$ simulations. To keep the notation simple, we will omit the angular brackets $\langle\,\rangle_{n_{\text{sim}}}$ when referring to averages between simulations. As in section 6.2, we introduce a minimum size $N_{\text{tb}} = 300$ (∼90% of the size of R6) above which a fragment is assumed to have finished its growth. The size range $N \geq N_{\text{tb}}$ will be referred to as the "target band", and we define $c_{\text{tb}}(t;T) \equiv \sum_{N \geq N_{\text{tb}}} c(N,t;T)$ (Fig. 6.8 - left (inset)). As in 6.2, we determine the the critical fragment size and the height of the nucleation barrier at a given time t by taking a snapshot of the instantaneous SST concentrations, $c_i(t;T) = N_i(t;T)/(N_A V)$ (where i specifies a SST type), and simulating independently the growth of one fragment under an equivalent background concentration of SSTs, using the algorithm described in section 5.1. We first determine the free-energy function $F_{(t;T)}(N) = -\beta^{-1} \ln p_{\text{eq}}(N)_{(t;T)}$ using the biasing described in section 5.1.2, where $p_{\text{eq}}(N)_{(t;T)}$ is the probability at equilibrium of finding the fragment with size N,[8] from which we deduce the critical size, $N^{\ddagger}_{(t;T)}$, for which $F_{(t;T)}(N)$ has a maximum (Fig. 6.9 - right). We determine the concentration of nuclei at a given time t as $c_{\text{nuc}}(t;T) = \sum_{N > N^{\ddagger}_{(t;T)}} c(N,t;T)$.

[8]To determine $p_{\text{eq}}(N)_{(t;T)}$ we followed the procedure outlined in the first paragraph of section 5.2.2

To make a comparison with the assembly yields published in Ref. [2], we determine the fraction of SSTs that ended up in fragments with sizes in the target band ($N \geq 300$) (Fig. 6.8 - right). To make a comparison with the fluorometric assay of Ref. [3] we determine the number of formed base pairs at the end of each temperature step as $n_b(T) = n_{10}(T) \times 10 + n_{11}(T) \times 11$, where $n_{10}(T)$ and $n_{11}(T)$ are the total number of bonded domains with lengths 10 and 11 among all fragments in the system (Fig. 6.8 - left). Note that the fluometric assay simply measure the number of bonds, not the number of fully formed assemblies.

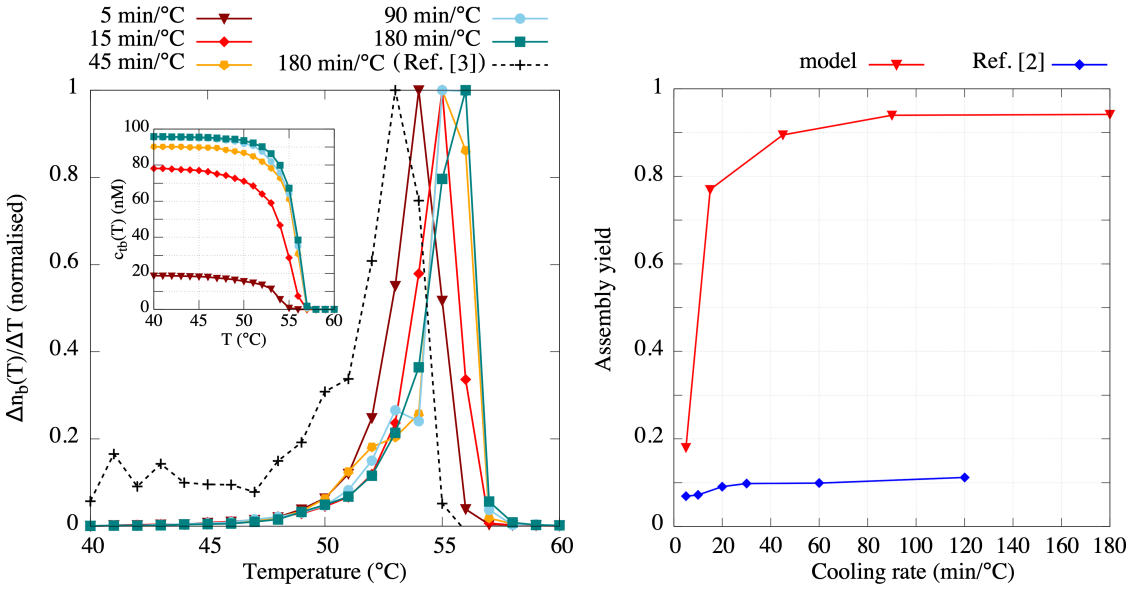


Figure 6.8: Simulation of a step-wise cooling ramp ($\Delta T = 1°$ C). Left, main plot - relative changes in base pairing levels with decreasing temperature, $\Delta n_b(T)/\Delta T$, where $n_b(T)$ is the number of base pairs formed (or a proxy for it - Ref. [3]) at the end of a temperature step (at temperature T). The coloured curves are the results of simulations; the dashed black curve is an experimental measurement published in Ref. [3]. A multiplicative constant was applied to each data set such that the peaks in the figure appear with the same heights. Left, inset - mean concentration of fragments with sizes in the target band ($N \geq 300$) at the end of each temperature step. Right - final assembly yield as a function of cooling rate predicted by the model (red curve), and measured experimentally in Ref. [2].

For a cooling rate of 180 min/°C, the model predicts a peak in base-pairing formation at T =56°C, in good agreement with the experimental peak (T =53°C) reported in Ref. [3]. The location of the peak does not decrease much if a faster cooling rate is used (down to T =54°C for 5 min/°C) as fragment growth occurs

over a similar range of temperatures. The size distributions (Fig. 6.6) show, however, large differences between cooling rate. For cooling rates 90 min/°C and 180 min/°C, for which the nucleation barrier (Fig. 6.9) stayed the highest during the cooling ramp (as SSTs were consumed the most during each temperature step), the size distributions look closest to the bimodal distributions observed in isothermal assembly at long reaction times (Fig. 6.1). Faster cooling rates resulted in a proliferation of nuclei which is more extreme the faster the cooling rate, and consequently less fragments were able to reach the target band.

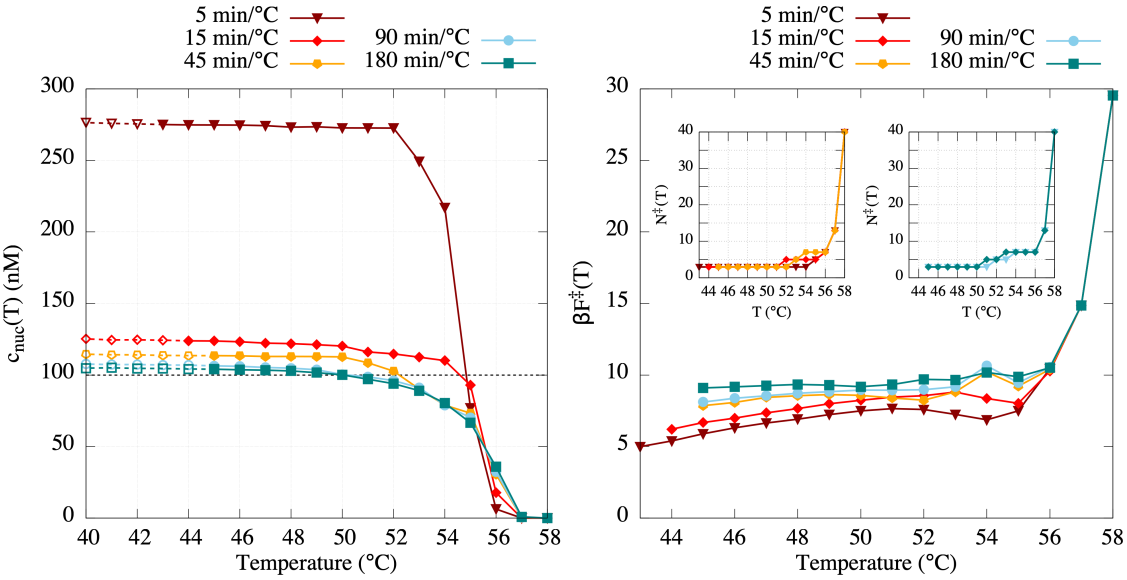


Figure 6.9: Simulation of a step-wise cooling ramp ($\Delta T = 1°$ C). Left - concentration of nuclei at the end of each temperature step. Right, main plot - height of the nucleation barrier at the start of each temperature step. Right, inset - size of the critical nuclei at the start of each temperature step. In both figures we use solid points if a stable nuclei can still form from unbound SSTs in the system, and open points (left figure only) if the (single-target) simulations yield $F(N) > F(1)$ for all fragment sizes above one, in which case a stable nucleus can no longer form from unbound SSTs in the system. In the latter case (open points) we estimate anyway, a concentration of nuclei $c_{\mathrm{nuc}}(t;T) = \sum_{N>N^{\ddagger}_{(\mathrm{last})}} c(N,t;T)$, where $N^{\ddagger}_{(\mathrm{last})}$ is the critical size at the lowest temperature where a stable nuclei could still form (for all cooling rates tested, we have $N^{\ddagger}_{(\mathrm{last})} = 3$).

The model predicts a much larger increase in assembly yield with decreasing cooling rate than that observed in Ref. [2]. It should be noted that in Ref. [2], and for all cooling rates tested, a large amount of DNA material ended up in aggregates with sizes ranging from just above the size of R6 to up to about four times the size

of R6. This indicates that non-specific contacts between fragments (not currently described in the model) are a critical source of assembly errors.

6.4 Summary and discussion

In this chapter, we introduce a model that describes the simultaneous growth of multiple fragments of R6 in a system where SSTs are present in finite supply. The model predicts that there is only a narrow temperature window 55°C$\sim$ 56°C where lag times are close to the elongation time - the minimum overall timescale required for fragments to complete their growth. The reaction becomes stalled at temperatures above due to an excessively high nucleation barrier, and at temperatures below due to increasingly severe depletion traps.

Performing an assembly reaction under isothermal conditions, even if a temperature in the window 55°C$\sim$ 56°C is used, is never an optimal strategy. We see that after a short phase ($\sim$ 1 hour) where a sizeable number of fragments complete their growth, the reaction slows down substantially - at 55°C$\sim$ due to a depletion trap, and at 56°C$\sim$ due to the rise of the nucleation barrier. A well designed cooling ramp, where the rise of nucleation barrier is compensated by gradual decreases in temperature, can offset these slowdowns and speed up the reaction significantly. For instance, a cooling ramp from 57°C to 51°C at 90 min/°C ($\sim$10 hours reaction time) achieved in the simulations a similar outcome ($\sim$90 fragments completing growth) to a $\sim$30 hours-long incubation at 55°C.

The model predicts a peak in base-pairing formation that is in close agreement with the experimental peak reported in Ref. [3]. However, the models predicts assembly yields that are significantly above those achieved in experiments (Ref. [2]). This difference is expected to be the result of non-specific bonds between fragments that occur during the reaction, a type of assembly error that is not yet well understood and thus described by the model. We leave the treatment of such effects to further work.

7
Summary

In this thesis, we introduced a novel approach for modelling self-assembling systems that use the single-stranded-tile (SST) assembly method. For the model system chosen from Refs. [2, 3] (the R6 structure) we first focused on the dominant nucleation pathways, finding that structures with a high number of bonds between SSTs and compact shapes are typically the most likely to form. However, we also found that for certain (recurrent) sizes, maximally-bonded structures are actually disfavoured due to statistical factors. We then developed a simple model that describes the bonding of SSTs with incorrect sequences to (otherwise correctly formed) structures, finding that these bonds are typically very weak, and have a relatively minor impact on the assembly kinetics and thermodynamics.

By performing simulations in the NVT ensemble, we found that there is only a narrow temperature window ($\sim$2°C wide) where the lag period before the first structures finish their growth approaches the elongation timescale (the minimum overall timescale of self assembly). The reaction becomes stalled at temperatures above due to an excessively high nucleation barrier, and at temperatures below due to increasingly severe depletion traps. By performing simulations at conditions similar to experiments, we find that the model predicts a peak in base-pairing formation at a temperature close (3°C difference) to that observed in the experiments. However, the model predicts assembly yields that are significantly higher than those observed in

experiments. This difference is attributed to non-native contacts between partially formed structures, a type of assembly error not yet well understood, and thus not described in the model.

The key ingredient that using oxDNA as a parametrisation tool provides is that it allows to capture, in a physically motivated way, the complex entropic barriers associated with the initiation of secondary domains upon SST association. Previous models of SST assembly [177, 180, 181] describe SSTs as patchy-particle with rigid patches (where patches represent binding domain), and do exhibit entropic barriers associated with forming secondary domains. However, in these models the entropic barrier arises from the loss of rotational freedom of the patchy particle (e.g., a patchy particle with one bond formed can rotate around the axis defined by the bond, but becomes constrained once a second bond forms) and not from a realistic description of DNA's chain-like nature. Moreover, because these models use lattice representations of the system, each SST can only rotate in a discrete number of ways. A consequence of these simplifications is that these models predict entropic barriers associated with forming secondary domains that are significantly smaller than those predicted by oxDNA. This leads to visible qualitative differences in behaviour; for instance in the models of Refs. [180, 181] there is substantially stronger bias towards maximally-bonded structures than in our model.

Our approach can in principle be applied to different assembly schemes provided subunits bind exclusively with local neighbours in the target structure. For instance, extending our approach to describe three-dimensional SST structures [17, 18] should be straightforward. Making different shapes, as done in Ref. [2] by leaving out certain SST types from the system, would also be straightforward to investigate, and it would be interesting to see how the nucleation pathway might change for a more complex shape.

The lack of a central organising element makes the SST method very susceptible to depletion traps (unlike, e.g., DNA origami which uses a scaffold strand). The use of a well-designed assembly seed could in principle ameliorate this problem and make the SST method more robust. The dependence of the parameter space for

productive assembly (depletion-trap free and experimentally accessible) on the type of assembly used is something we could easily study using our model.

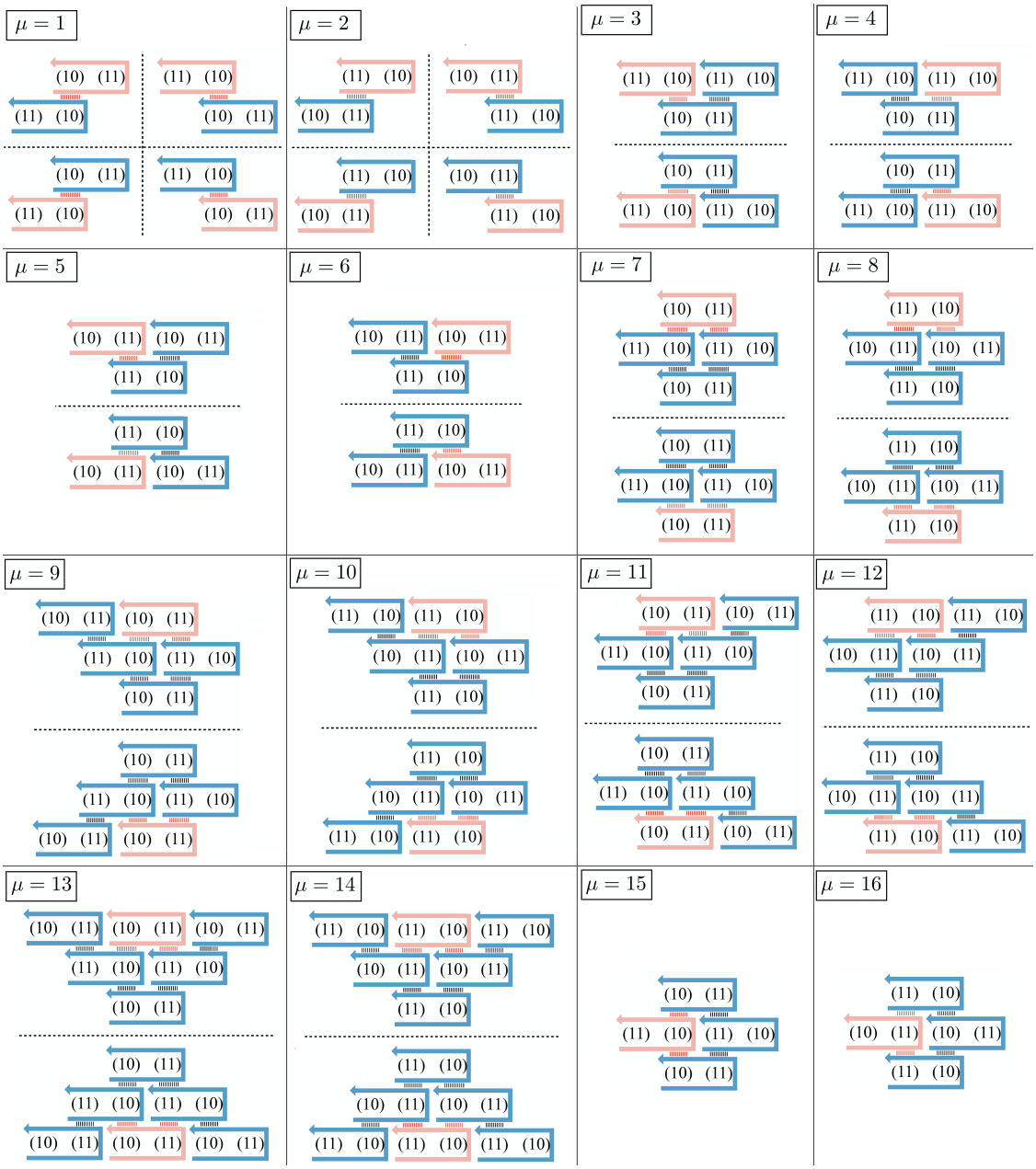
A possible extension to the model is to use the NN model [31] to estimate how binding free energies change according to the base sequences in the binding domains of a SST. Such a reparametrisation would be helpful, for instance, in shining light on conflicting predictions that currently exist in the literature regarding the effect of heterogeneities in the binding free energies (arising here from sequence differences among SST types) on the stability of intermediates. Some models predict that heterogeneities in the binding free energies have a stabilising effect on intermediates [180], while other models predict that such heterogeneities destabilise intermediates [210]. Also, we could make it more likely for nuclei containing SSTs from selected regions of the target structure to form (e.g. by increasing the G-C content of these SSTs), and potentially increase the region of the parameter space where productive assembly can occur.

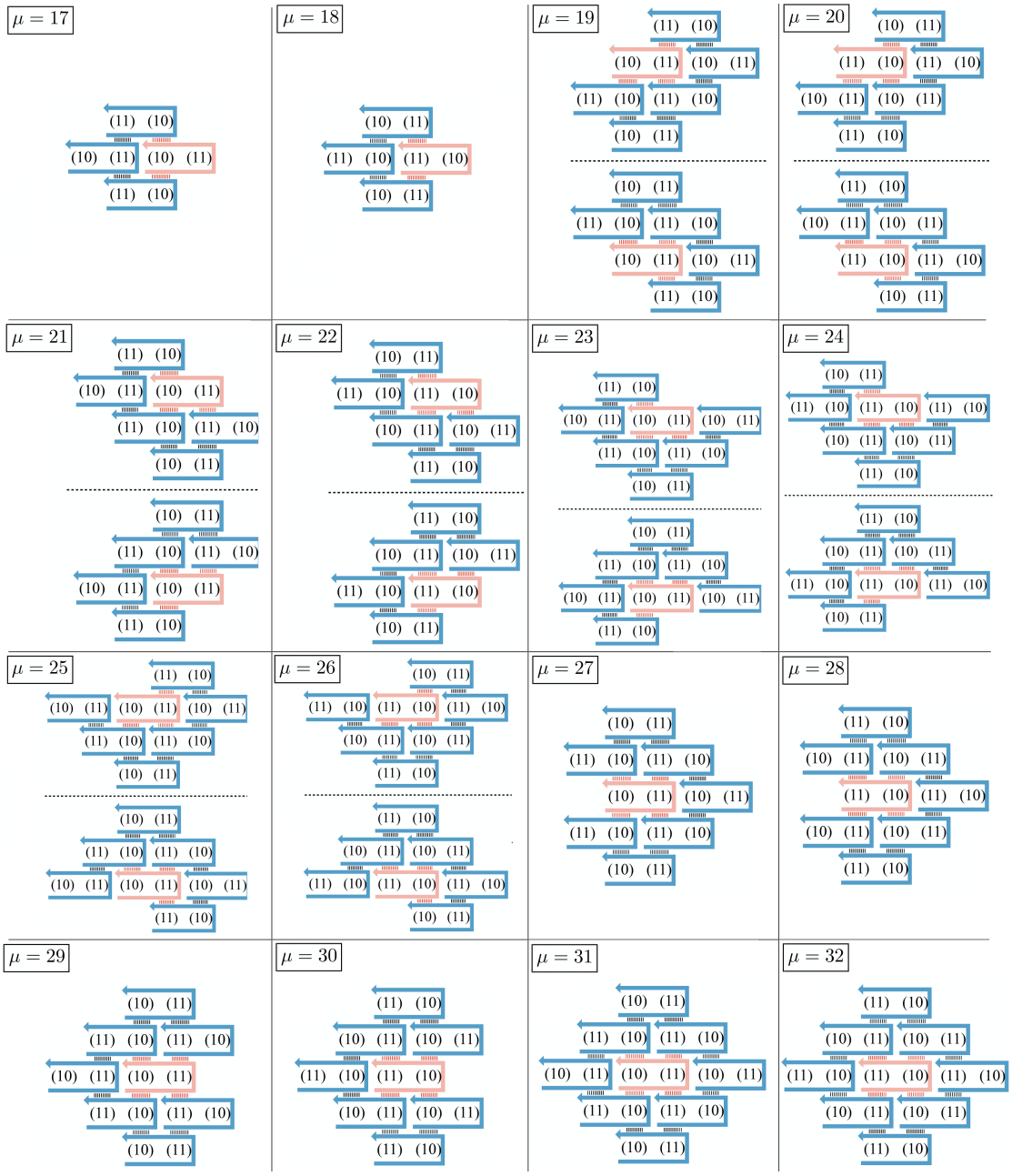
Appendices

A

Types of binding sites (μ) identified for R6

In this appendix we show all 32 types of binding site identified for R6. In the schemes below, the binding site is coloured pink. Two binding sites are considered different if there is at least one first- or second-neighbour that is different between the two. Each SST is represented with an arrow that specifies the directionality of the respective DNA chain (from the 5'-end to the 3'-end). As discussed in section 4.3.2, for each type of binding site μ, we used simulations with oxDNA1 to determine standard-state free-energy changes $\Delta G^{\ominus}(\mu)$ associated with the binding of a SST to that binding site. As discussed in section 1.3.3, oxDNA1 describes B-form DNA as a fully-symmetrical structure, without minor or major grooves. As a consequence, two binding sites that differ only by the directionality of the DNA chains (see, e.g., the case $\mu = 1$) yield identical free-energy changes $\Delta G^{\ominus}(\mu)$ in oxDNA1. Such cases will therefore be grouped together into a single type of binding site.

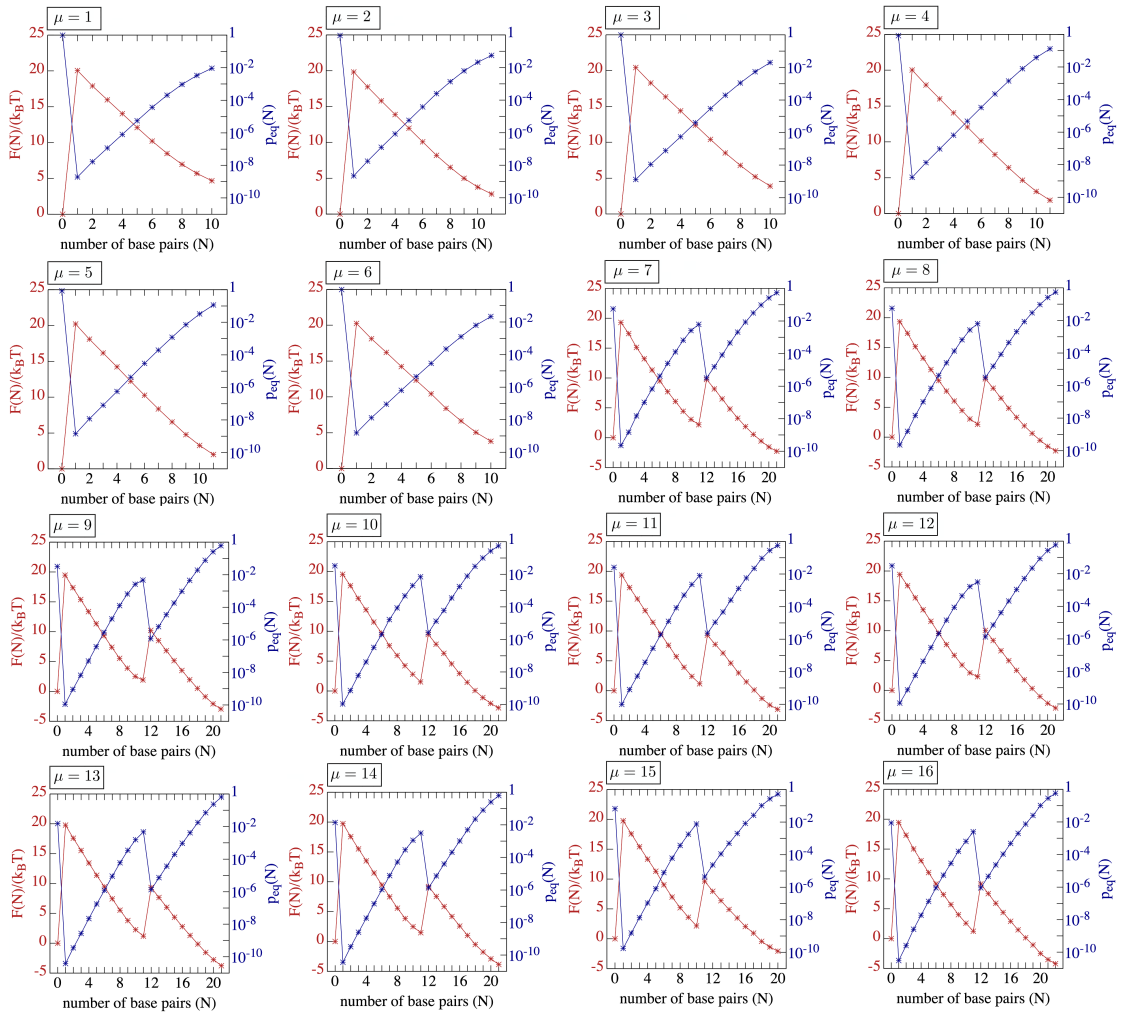

μ = 1
μ = 2
μ = 3
μ = 4
μ = 5
μ = 6
μ = 7
μ = 8
μ = 9
μ = 10
μ = 11
μ = 12
μ = 13
μ = 14
μ = 15
μ = 16

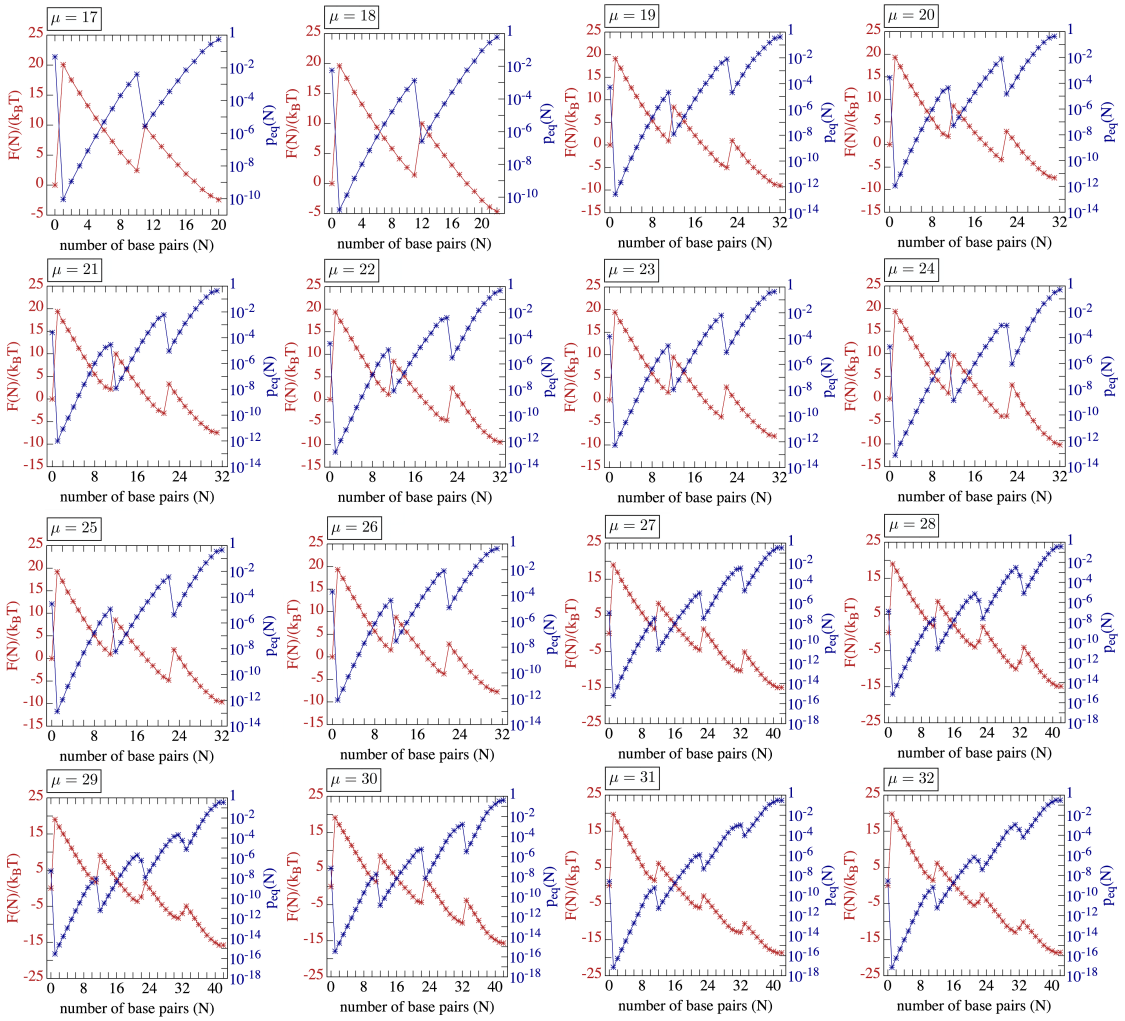

$\mu = 17$
$\mu = 18$
$\mu = 19$
$\mu = 20$
$\mu = 21$
$\mu = 22$
$\mu = 23$
$\mu = 24$
$\mu = 25$
$\mu = 26$
$\mu = 27$
$\mu = 28$
$\mu = 29$
$\mu = 30$
$\mu = 31$
$\mu = 32$

B

Binding probabilities $P_{eq}^{(\mu)}(N)$ (oxDNA simulations)

In this appendix we present equilibrium probability distributions obtained from simulations with oxDNA. Shown below are probability distributions obtained for T =50°C (the temperature for which the simulations were performed), for a SST concentration of 100 nM. The probability distributions are given as a function of the number of base pairs formed between SST and binding site, $P_{\mathrm{eq}}(N)$. Also shown are the free-energy functions, defined as $\beta F(N) = -\ln P_{\mathrm{eq}}(N) + c$, where the constant c is fixed such that $\beta F(0) = 0$.

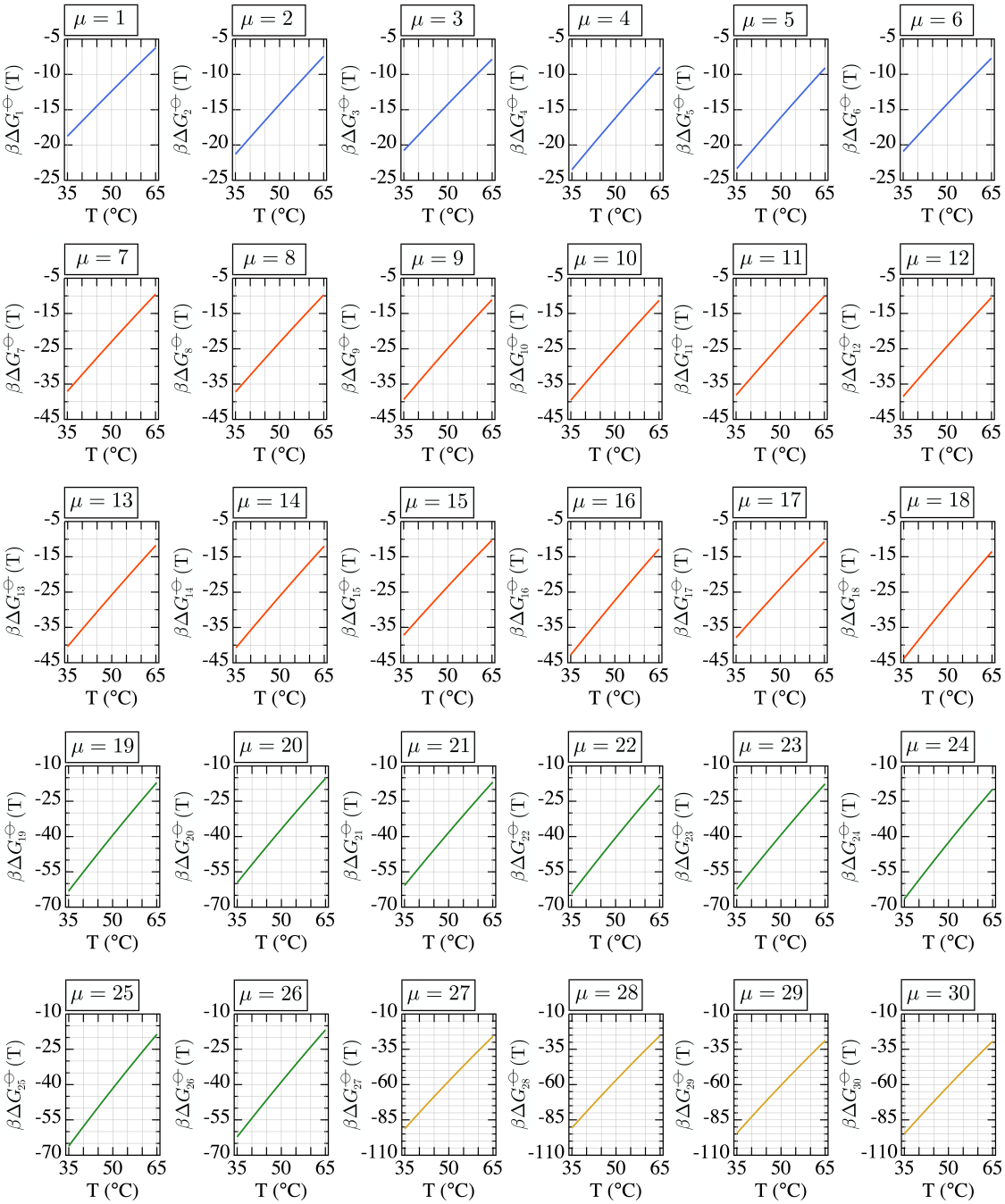

$\mu = 1$
$\mu = 2$
$\mu = 3$
$\mu = 4$
$\mu = 5$
$\mu = 6$
$\mu = 7$
$\mu = 8$
$\mu = 9$
$\mu = 10$
$\mu = 11$
$\mu = 12$
$\mu = 13$
$\mu = 14$
$\mu = 15$
$\mu = 16$
F(N)/(k$_B$T)
p$_{eq}$(N)
number of base pairs (N)

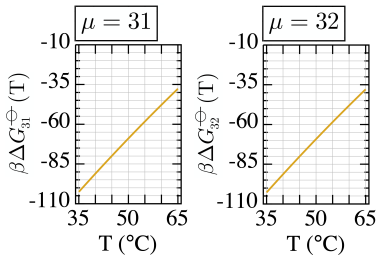

$\mu = 17$
$\mu = 18$
$\mu = 19$
$\mu = 20$
$\mu = 21$
$\mu = 22$
$\mu = 23$
$\mu = 24$
$\mu = 25$
$\mu = 26$
$\mu = 27$
$\mu = 28$
$\mu = 29$
$\mu = 30$
$\mu = 31$
$\mu = 32$
F(N)/(k_BT)
P_eq(N)
number of base pairs (N)

C

Standard-state binding free energies $\Delta G^{\ominus}(\mu)$ (plots)

In this appendix we present the standard-state free-energy changes $\Delta G^{\ominus}(\mu)$, associated with the binding of a SST to a binding site of type μ, determined via simulations with oxDNA. In the figures below, we use different colours according to the number of binding domains in the binding site: one binding domain ($\mu = 1, \cdots, 6$) - blue; two binding domains ($\mu = 7, \cdots, 18$) - red; three binding domains ($\mu = 19, \cdots, 26$) - green; four binding domains ($\mu = 27, \cdots, 32$) - gold.


$\mu = 1$
$\mu = 2$
$\mu = 3$
$\mu = 4$
$\mu = 5$
$\mu = 6$
$\mu = 7$
$\mu = 8$
$\mu = 9$
$\mu = 10$
$\mu = 11$
$\mu = 12$
$\mu = 13$
$\mu = 14$
$\mu = 15$
$\mu = 16$
$\mu = 17$
$\mu = 18$
$\mu = 19$
$\mu = 20$
$\mu = 21$
$\mu = 22$
$\mu = 23$
$\mu = 24$
$\mu = 25$
$\mu = 26$
$\mu = 27$
$\mu = 28$
$\mu = 29$
$\mu = 30$
$\beta\Delta G^{\ominus}_{\mu}$ (T)
T (°C)


$\mu = 31$
$\mu = 32$
-10
-35
-60
-85
-110
35
50
65
T (°C)
$\beta\Delta G^{\ominus}_{31}$ (T)
$\beta\Delta G^{\ominus}_{32}$ (T)

D

$\beta\Delta G^{\circ\!\!-\!\!\bigcirc\!\!-}(\mu)$ (numeric values)

T(°C)	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	$\mu = 5$	$\mu = 6$	$\mu = 7$	$\mu = 8$	$\mu = 9$	$\mu = 10$	$\mu = 11$	$\mu = 12$	$\mu = 13$	$\mu = 14$	$\mu = 15$	$\mu = 16$
35	-18.752	-21.300	-20.776	-23.492	-23.334	-20.920	-37.073	-37.225	-39.361	-39.589	-38.136	-38.504	-40.426	-40.755	-37.220	-42.811
36	-18.323	-20.824	-20.332	-22.989	-22.842	-20.465	-36.120	-36.271	-38.374	-38.601	-37.166	-37.533	-39.433	-39.752	-36.280	-41.762
37	-17.894	-20.349	-19.888	-22.487	-22.351	-20.010	-35.168	-35.320	-37.389	-37.617	-36.198	-36.566	-38.444	-38.753	-35.343	-40.716
38	-17.467	-19.874	-19.445	-21.986	-21.861	-19.557	-34.219	-34.370	-36.408	-36.635	-35.233	-35.600	-37.456	-37.756	-34.408	-39.673
39	-17.039	-19.401	-19.003	-21.486	-21.372	-19.104	-33.272	-33.424	-35.430	-35.657	-34.270	-34.638	-36.471	-36.762	-33.476	-38.634
40	-16.613	-18.928	-18.562	-20.987	-20.885	-18.652	-32.328	-32.480	-34.455	-34.681	-33.310	-33.677	-35.488	-35.770	-32.548	-37.598
41	-16.187	-18.456	-18.122	-20.490	-20.398	-18.202	-31.386	-31.538	-33.482	-33.709	-32.352	-32.719	-34.508	-34.782	-31.622	-36.566
42	-15.761	-17.985	-17.682	-19.994	-19.912	-17.752	-30.446	-30.598	-32.513	-32.739	-31.396	-31.763	-33.530	-33.797	-30.699	-35.536
43	-15.337	-17.514	-17.243	-19.499	-19.428	-17.303	-29.509	-29.661	-31.546	-31.772	-30.442	-30.810	-32.555	-32.814	-29.779	-34.510
44	-14.913	-17.045	-16.806	-19.005	-18.944	-16.855	-28.575	-28.727	-30.583	-30.808	-29.491	-29.860	-31.582	-31.834	-28.861	-33.488
45	-14.489	-16.576	-16.368	-18.512	-18.462	-16.409	-27.643	-27.795	-29.622	-29.848	-28.543	-28.911	-30.611	-30.857	-27.947	-32.469
46	-14.067	-16.109	-15.932	-18.021	-17.980	-15.963	-26.713	-26.865	-28.664	-28.890	-27.597	-27.966	-29.643	-29.883	-27.036	-31.453
47	-13.645	-15.642	-15.497	-17.531	-17.500	-15.518	-25.786	-25.938	-27.710	-27.935	-26.653	-27.022	-28.678	-28.912	-26.127	-30.441
48	-13.224	-15.176	-15.063	-17.042	-17.020	-15.075	-24.861	-25.014	-26.758	-26.984	-25.712	-26.082	-27.715	-27.943	-25.222	-29.432
49	-12.804	-14.712	-14.629	-16.555	-16.542	-14.632	-23.939	-24.092	-25.809	-26.035	-24.774	-25.144	-26.754	-26.978	-24.320	-28.426
50	-12.385	-14.248	-14.197	-16.069	-16.065	-14.191	-23.019	-23.173	-24.864	-25.090	-23.838	-24.208	-25.796	-26.016	-23.420	-27.424
51	-11.966	-13.785	-13.765	-15.584	-15.590	-13.751	-22.102	-22.256	-23.921	-24.147	-22.904	-23.275	-24.841	-25.056	-22.524	-26.425
52	-11.549	-13.324	-13.335	-15.101	-15.115	-13.312	-21.188	-21.342	-22.982	-23.208	-21.974	-22.345	-23.888	-24.100	-21.631	-25.430
53	-11.133	-12.863	-12.905	-14.619	-14.641	-12.875	-20.277	-20.431	-22.046	-22.272	-21.045	-21.417	-22.938	-23.146	-20.741	-24.439
54	-10.718	-12.404	-12.477	-14.138	-14.169	-12.438	-19.368	-19.522	-21.113	-21.340	-20.120	-20.493	-21.990	-22.196	-19.854	-23.451
55	-10.304	-11.946	-12.050	-13.659	-13.698	-12.003	-18.462	-18.617	-20.183	-20.410	-19.197	-19.571	-21.045	-21.249	-18.970	-22.466
56	-9.891	-11.489	-11.623	-13.181	-13.229	-11.570	-17.559	-17.714	-19.257	-19.484	-18.277	-18.651	-20.103	-20.305	-18.090	-21.486
57	-9.479	-11.033	-11.198	-12.705	-12.760	-11.137	-16.658	-16.814	-18.334	-18.561	-17.360	-17.735	-19.164	-19.364	-17.212	-20.509
58	-9.069	-10.579	-10.775	-12.231	-12.293	-10.706	-15.761	-15.917	-17.414	-17.642	-16.446	-16.822	-18.228	-18.426	-16.338	-19.535
59	-8.659	-10.126	-10.352	-11.758	-11.828	-10.277	-14.867	-15.023	-16.498	-16.726	-15.534	-15.911	-17.294	-17.492	-15.468	-18.566
60	-8.252	-9.674	-9.931	-11.286	-11.363	-9.849	-13.975	-14.132	-15.586	-15.814	-14.626	-15.004	-16.363	-16.561	-14.601	-17.600
61	-7.845	-9.224	-9.511	-10.817	-10.901	-9.422	-13.087	-13.244	-14.677	-14.905	-13.721	-14.099	-15.436	-15.633	-13.737	-16.638
62	-7.441	-8.775	-9.092	-10.349	-10.439	-8.997	-12.202	-12.359	-13.771	-14.000	-12.818	-13.198	-14.511	-14.709	-12.877	-15.680
63	-7.037	-8.328	-8.675	-9.882	-9.979	-8.574	-11.320	-11.477	-12.869	-13.098	-11.919	-12.300	-13.589	-13.788	-12.020	-14.726
64	-6.636	-7.882	-8.259	-9.418	-9.521	-8.152	-10.441	-10.599	-11.971	-12.200	-11.023	-11.405	-12.670	-12.871	-11.167	-13.776
65	-6.236	-7.438	-7.844	-8.955	-9.064	-7.732	-9.566	-9.724	-11.077	-11.306	-10.130	-10.513	-11.755	-11.957	-10.317	-12.831

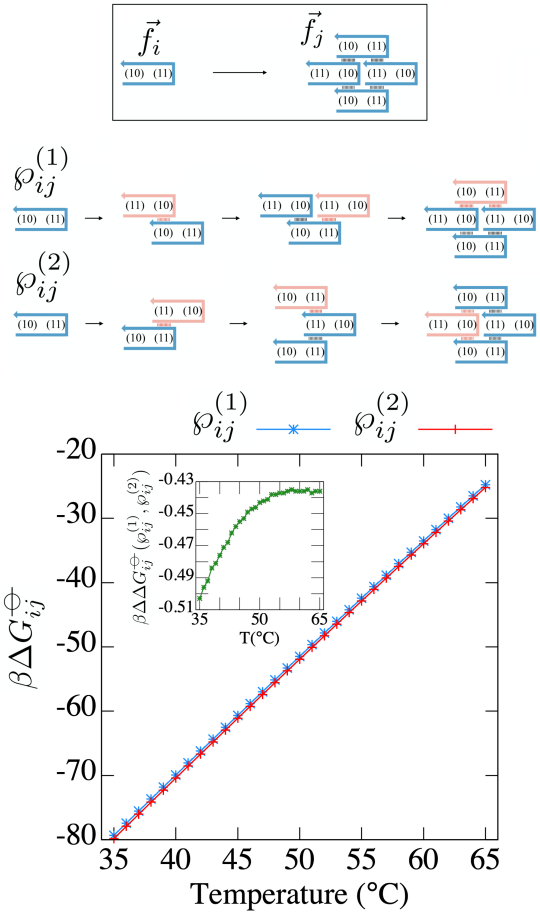
T(°C)	$\mu = 17$	$\mu = 18$	$\mu = 19$	$\mu = 20$	$\mu = 21$	$\mu = 22$	$\mu = 23$	$\mu = 24$	$\mu = 25$	$\mu = 26$	$\mu = 27$	$\mu = 28$	$\mu = 29$	$\mu = 30$	$\mu = 31$	$\mu = 32$
35	-37.903	-43.898	-63.234	-59.571	-60.806	-64.252	-62.334	-66.521	-66.135	-62.222	-91.272	-90.681	-94.438	-94.828	-102.552	-102.763
36	-36.961	-42.835	-61.609	-58.017	-59.271	-62.627	-60.761	-64.878	-64.458	-60.631	-88.955	-88.361	-92.109	-92.503	-100.242	-100.454
37	-36.021	-41.776	-59.991	-56.468	-57.739	-61.008	-59.194	-63.240	-62.788	-59.045	-86.648	-86.050	-89.790	-90.189	-97.943	-98.155
38	-35.083	-40.721	-58.378	-54.925	-56.213	-59.395	-57.632	-61.608	-61.123	-57.464	-84.349	-83.748	-87.481	-87.883	-95.654	-95.866
39	-34.148	-39.669	-56.770	-53.387	-54.692	-57.788	-56.076	-59.982	-59.465	-55.888	-82.060	-81.456	-85.181	-85.587	-93.375	-93.587
40	-33.215	-38.620	-55.169	-51.854	-53.175	-56.187	-54.525	-58.363	-57.813	-54.317	-79.780	-79.174	-82.891	-83.301	-91.105	-91.318
41	-32.285	-37.574	-53.573	-50.326	-51.663	-54.592	-52.980	-56.749	-56.167	-52.751	-77.510	-76.901	-80.611	-81.024	-88.846	-89.059
42	-31.357	-36.532	-51.983	-48.803	-50.155	-53.003	-51.440	-55.141	-54.527	-51.191	-75.249	-74.638	-78.340	-78.757	-86.597	-86.810
43	-30.432	-35.493	-50.399	-47.286	-48.653	-51.420	-49.906	-53.540	-52.894	-49.636	-72.997	-72.384	-76.080	-76.500	-84.359	-84.572
44	-29.509	-34.457	-48.821	-45.775	-47.155	-49.843	-48.378	-51.944	-51.267	-48.086	-70.754	-70.139	-73.829	-74.252	-82.130	-82.343
45	-28.588	-33.424	-47.248	-44.268	-45.662	-48.272	-46.855	-50.355	-49.647	-46.542	-68.521	-67.905	-71.588	-72.013	-79.911	-80.125
46	-27.670	-32.395	-45.682	-42.767	-44.174	-46.707	-45.338	-48.771	-48.032	-45.002	-66.297	-65.680	-69.356	-69.785	-77.702	-77.916
47	-26.754	-31.370	-44.121	-41.271	-42.691	-45.148	-43.826	-47.194	-46.424	-43.468	-64.083	-63.464	-67.135	-67.565	-75.504	-75.718
48	-25.841	-30.347	-42.566	-39.781	-41.212	-43.595	-42.320	-45.623	-44.823	-41.939	-61.878	-61.258	-64.923	-65.356	-73.315	-73.530
49	-24.931	-29.328	-41.017	-38.296	-39.739	-42.048	-40.820	-44.058	-43.228	-40.416	-59.682	-59.062	-62.722	-63.156	-71.137	-71.352
50	-24.023	-28.313	-39.474	-36.817	-38.270	-40.508	-39.325	-42.499	-41.639	-38.898	-57.496	-56.875	-60.530	-60.966	-68.969	-69.183
51	-23.117	-27.301	-37.938	-35.343	-36.806	-38.973	-37.837	-40.946	-40.057	-37.385	-55.319	-54.698	-58.348	-58.785	-66.811	-67.026
52	-22.214	-26.292	-36.407	-33.875	-35.348	-37.445	-36.353	-39.399	-38.481	-35.878	-53.152	-52.531	-56.175	-56.614	-64.663	-64.878
53	-21.314	-25.287	-34.882	-32.412	-33.894	-35.923	-34.876	-37.859	-36.912	-34.376	-50.994	-50.374	-54.013	-54.453	-62.525	-62.740
54	-20.417	-24.285	-33.364	-30.955	-32.445	-34.408	-33.405	-36.325	-35.350	-32.879	-48.846	-48.227	-51.861	-52.302	-60.397	-60.613
55	-19.522	-23.287	-31.851	-29.504	-31.002	-32.899	-31.939	-34.797	-33.794	-31.389	-46.708	-46.089	-49.719	-50.161	-58.280	-58.496
56	-18.630	-22.293	-30.345	-28.058	-29.564	-31.396	-30.480	-33.276	-32.245	-29.904	-44.580	-43.962	-47.587	-48.029	-56.173	-56.389
57	-17.741	-21.302	-28.845	-26.619	-28.131	-29.900	-29.026	-31.761	-30.702	-28.424	-42.461	-41.844	-45.465	-45.908	-54.076	-54.292
58	-16.855	-20.315	-27.352	-25.185	-26.703	-28.410	-27.579	-30.253	-29.167	-26.951	-40.352	-39.737	-43.353	-43.797	-51.989	-52.206
59	-15.971	-19.332	-25.865	-23.758	-25.280	-26.927	-26.137	-28.751	-27.638	-25.483	-38.253	-37.640	-41.251	-41.695	-49.913	-50.130
60	-15.090	-18.352	-24.385	-22.336	-23.863	-25.451	-24.702	-27.256	-26.117	-24.021	-36.164	-35.553	-39.160	-39.604	-47.847	-48.064
61	-14.213	-17.376	-22.911	-20.920	-22.452	-23.981	-23.273	-25.767	-24.602	-22.565	-34.085	-33.477	-37.079	-37.523	-45.792	-46.009
62	-13.338	-16.404	-21.444	-19.511	-21.046	-22.518	-21.850	-24.285	-23.095	-21.115	-32.017	-31.410	-35.008	-35.452	-43.747	-43.964
63	-12.467	-15.436	-19.984	-18.108	-19.645	-21.062	-20.434	-22.810	-21.594	-19.671	-29.958	-29.355	-32.948	-33.392	-41.713	-41.930
64	-11.598	-14.472	-18.530	-16.711	-18.251	-19.613	-19.024	-21.342	-20.101	-18.233	-27.910	-27.309	-30.898	-31.342	-39.689	-39.906
65	-10.733	-13.512	-17.083	-15.320	-16.862	-18.171	-17.620	-19.880	-18.615	-16.801	-25.872	-25.275	-28.859	-29.302	-37.676	-37.893

E

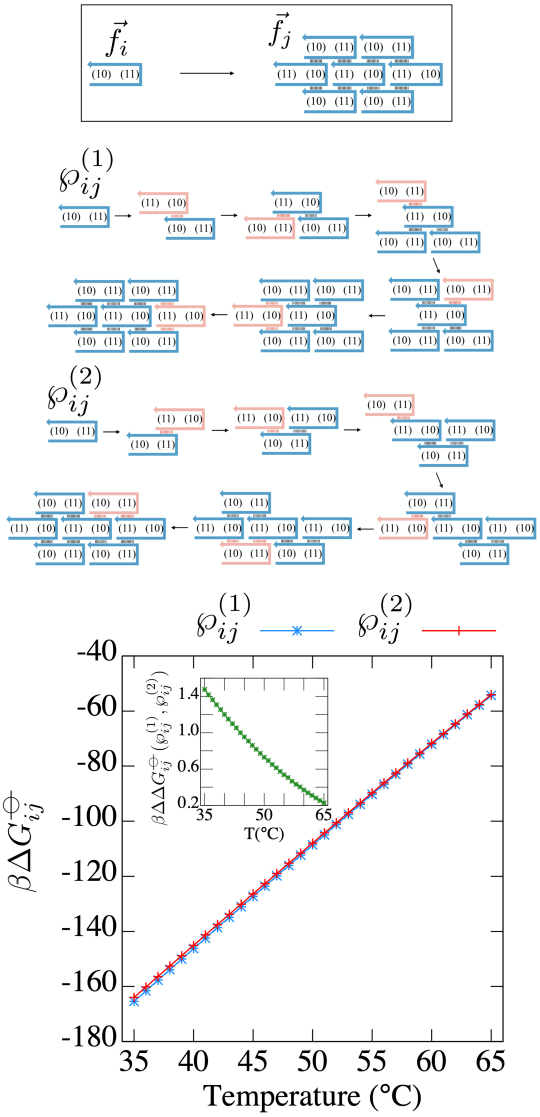
Deviations from detailed balance

In this appendix we consider a few pairs of fragment states $\vec{f}_i, \vec{f}_j$ and compute free-energy differences $\beta\Delta G_{ij}^{\ominus}$ for different paths $\wp_{ij} \equiv \vec{f}_i \rightarrow \vec{f}_{i+1} \rightarrow \vec{f}_{i+2} \rightarrow \cdots \rightarrow \vec{f}_{j-1} \rightarrow \vec{f}_j$.

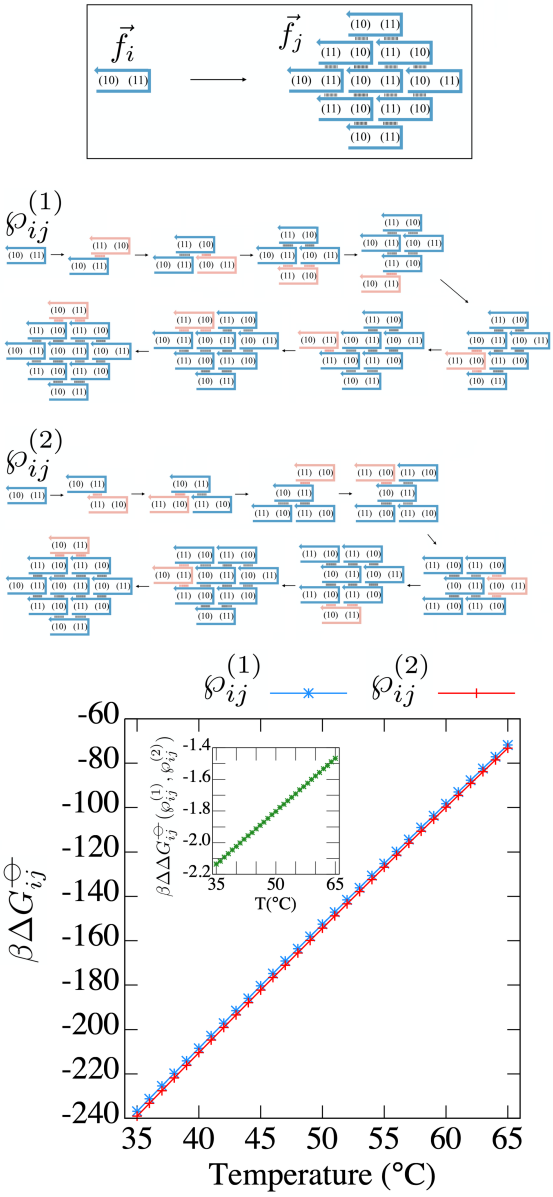
Example 1



Example 2



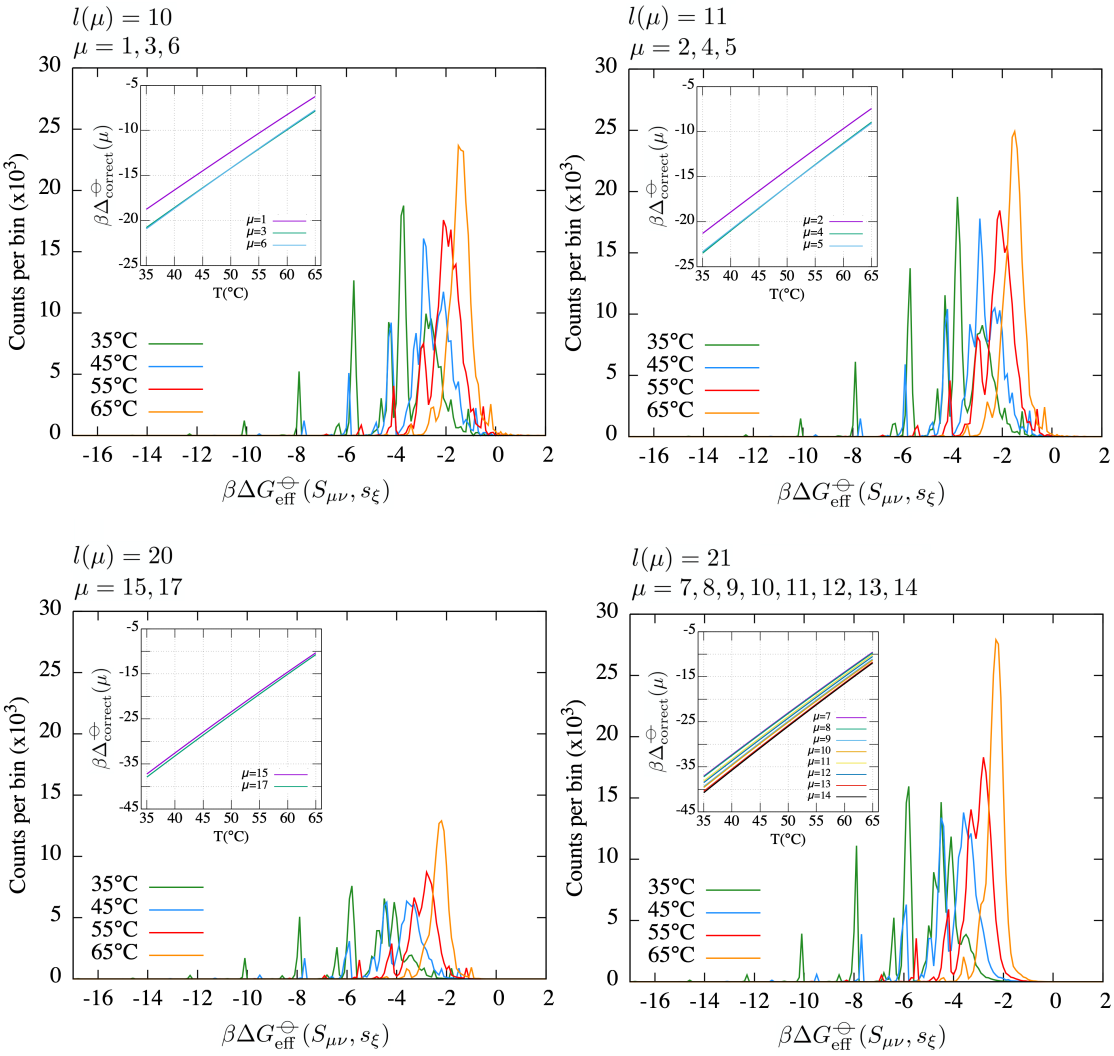
Example 3

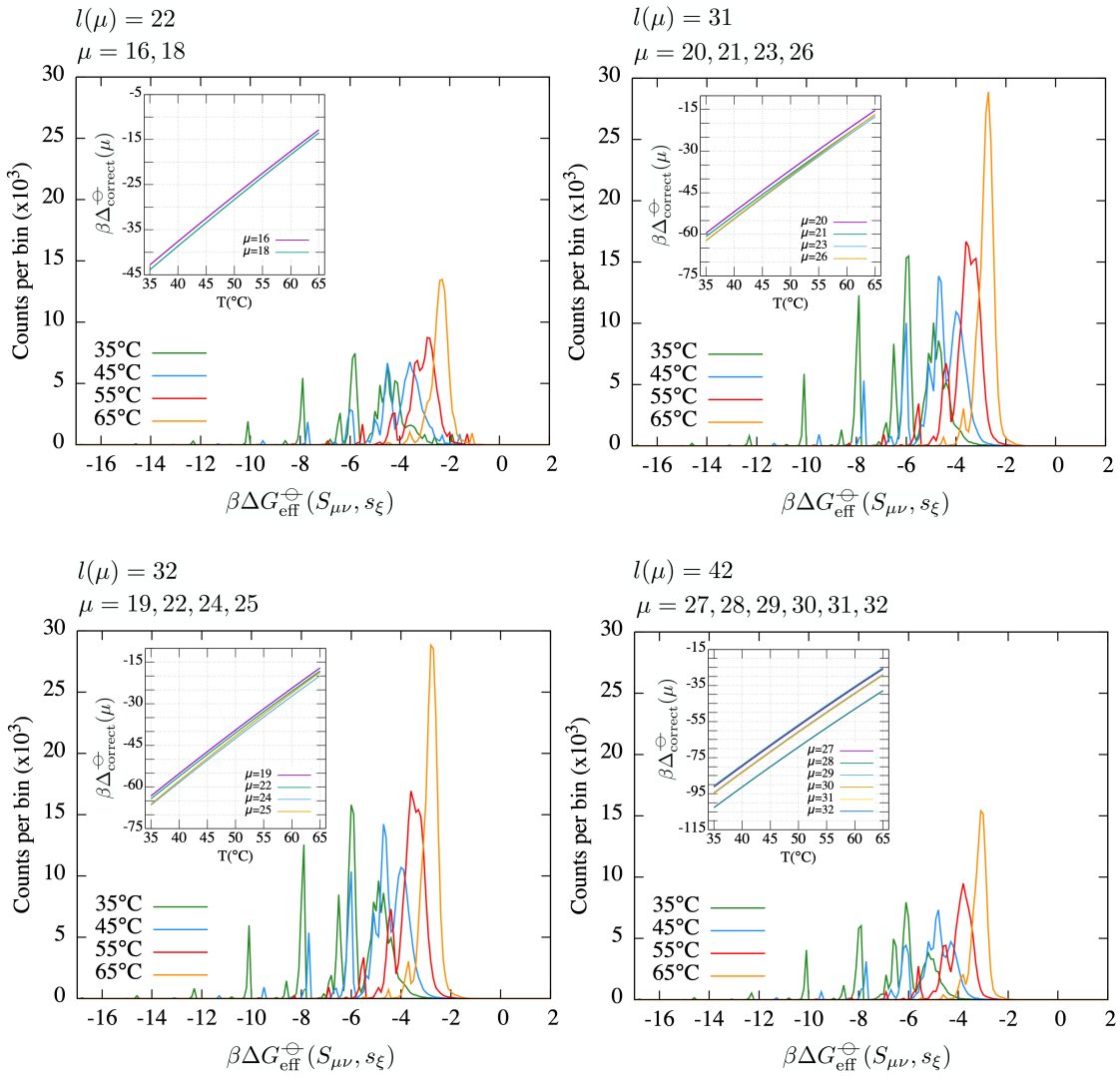

$\vec{f}_i$
$\vec{f}_j$
$\wp_{ij}^{(1)}$
$\wp_{ij}^{(2)}$
$\wp_{ij}^{(1)}$
$\wp_{ij}^{(2)}$
$\beta\Delta G_{ij}^{\ominus}$
-60
-80
-100
-120
-140
-160
-180
-200
-220
-240
35
40
45
50
55
60
65
Temperature (°C)
$\beta\Delta\Delta G_{ij}^{\ominus}(\wp_{ij}^{(1)},\wp_{ij}^{(2)})$
-1.4
-1.6
-1.8
-2.0
-2.2
35
50
65
T(°C)

F

Effective binding free energies for misbonds

In this appendix we present the effective free energies of misbond formation, $\beta\Delta G^{\ominus}_{\text{eff}}(S_{\mu\nu}, s_{\xi})$, determined for all combinations of sequences $S_{\mu\nu}$ and s_{ξ}. In each plot we present results obtained for a certain $l(\mu)$, which gives the cumulative length of the binding domains of a binding site μ (hence each plot covers various types of binding site μ that yield the same $l(\mu)$). Also shown in the plots (see insets) are binding free-energies for SSTs with correct sequences. In the main plots, we use a bin of $\beta\Delta\Delta G^{\ominus}_{\text{eff}} = 0.1$.

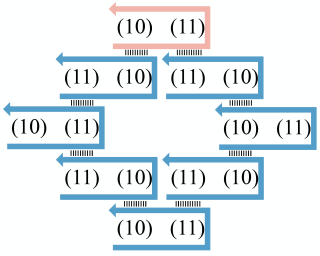

$l(\mu) = 10$
$\mu = 1, 3, 6$
$l(\mu) = 11$
$\mu = 2, 4, 5$
$l(\mu) = 20$
$\mu = 15, 17$
$l(\mu) = 21$
$\mu = 7, 8, 9, 10, 11, 12, 13, 14$
Counts per bin (x10^3)
$\beta\Delta G^{\ominus}_{\text{eff}}(S_{\mu\nu}, s_\xi)$
$\beta\Delta^{\ominus}_{\text{correct}}(\mu)$
T(°C)
35°C
45°C
55°C
65°C


$l(\mu) = 22$
$\mu = 16, 18$
$l(\mu) = 31$
$\mu = 20, 21, 23, 26$
$l(\mu) = 32$
$\mu = 19, 22, 24, 25$
$l(\mu) = 42$
$\mu = 27, 28, 29, 30, 31, 32$
Counts per bin (x10^3)
$\beta\Delta G^{\ominus}_{\text{eff}}(S_{\mu\nu}, s_\xi)$
$\beta\Delta^{\ominus}_{\text{correct}}(\mu)$
T(°C)
35°C
45°C
55°C
65°C

G

Bridging event (example)

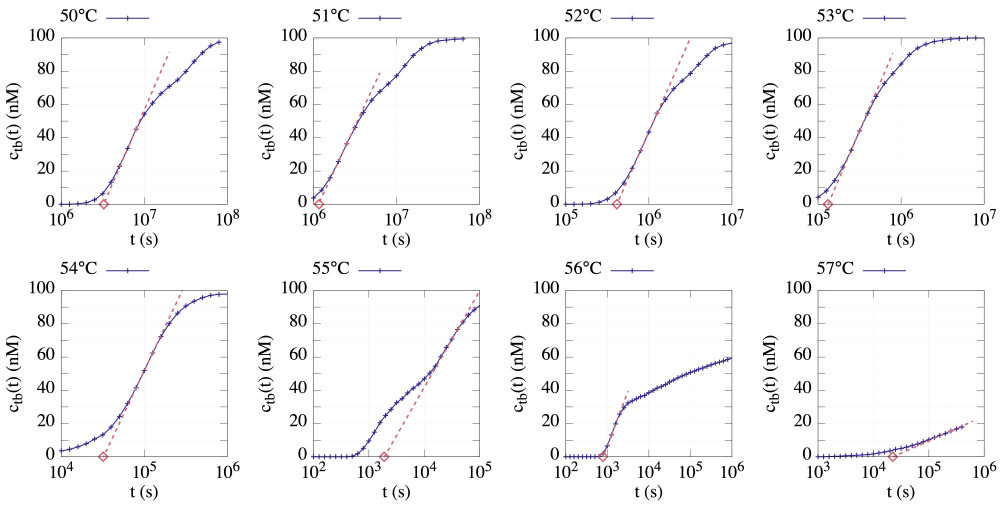
In this annex we give an example of a bridging event (see section 5.6.1 for context and details).



H

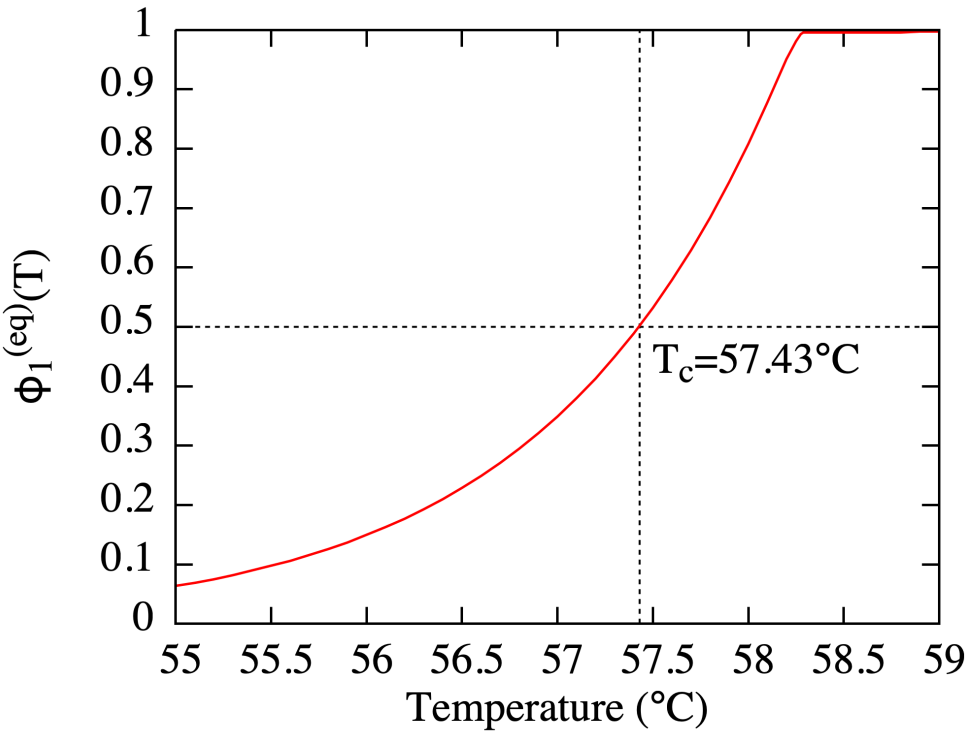
Lag time calculation

As described in the main text (section 6.2) we determine lag time by making a linear fit to $c_{\mathrm{tb}}(\log t)$ at the point where the difference $\Delta_{\{\log t\}} c_{\mathrm{tb}}(\log t)$ is maximal (see figure below). We define lag time as the point where the linear fit (red dashed curve) crosses the (log-)time axis (red diamond).



I

$\phi_1^{\mathrm{eq}}(T)$



J

t_{nuc} and t_{elo} as a function of target structure size (simulations).

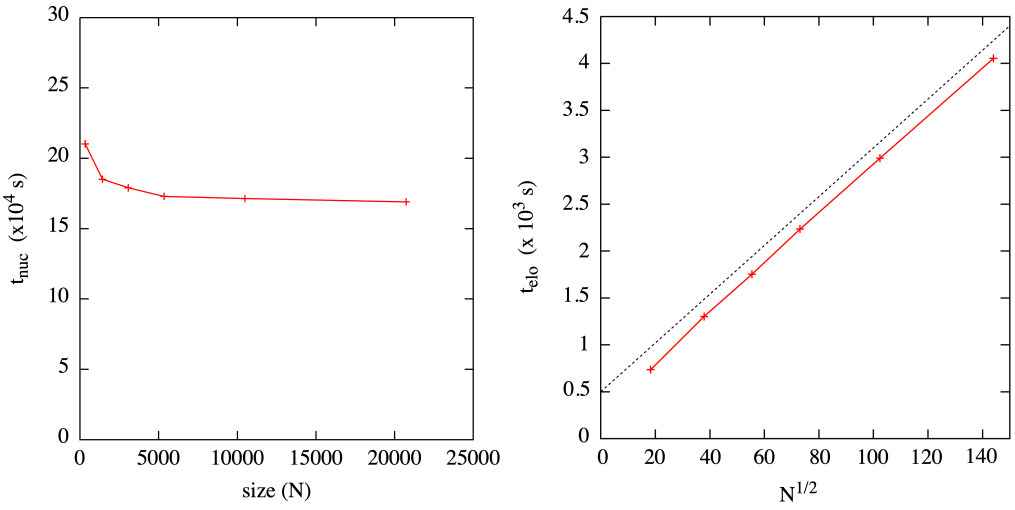


Figure J.1: Mean nucleation (left) and elongation (right) times as a function of target structure size. The simulations were performed at T =56°C. The sizes (N) and aspect ratios (height/weight) (AR) of the target structures used in the simulations are: $N = 334$ and $AR = 23/15$; $N = 1434$ and $AR = 47/31$; $N = 3071$ and $AR = 69/45$; $N = 5324$ and $AR = 91/59$; $N = 10478$ and $AR = 127/83$; $N = 20651$ and $AR = 177/117$. The dashed black curve in the right plot is a guide to the eye.

Bibliography

[1] URL dna.physics.ox.ac.uk.

[2] Bryan Wei, Mingjie Dai, and Peng Yin. Complex shapes self-assembled from single-stranded dna tiles. *Nature*, 485(7400):623–626, 2012. doi: 10.1038/nature11075. URL https://doi.org/10.1038/nature11075.

[3] Jean-Philippe J. Sobczak, Thomas G. Martin, Thomas Gerling, and Hendrik Dietz. Rapid folding of dna into nanoscale shapes at constant temperature. *Science*, 338(6113):1458, 12 2012. doi: 10.1126/science.1229919. URL http://science.sciencemag.org/content/338/6113/1458.abstract.

[4] Nadrian C. Seeman and Hanadi F. Sleiman. Dna nanotechnology. *Nature Reviews Materials*, 3(1):17068, 2017. doi: 10.1038/natrevmats.2017.68. URL https://doi.org/10.1038/natrevmats.2017.68.

[5] Thomas E. Ouldridge, Petr Šulc, Flavio Romano, Jonathan P. K. Doye, and Ard A. Louis. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic Acids Research*, 41(19):8886–8895, 08 2013. ISSN 0305-1048. doi: 10.1093/nar/gkt687. URL https://doi.org/10.1093/nar/gkt687.

[6] John S Schreck, Thomas E Ouldridge, Flavio Romano, Petr Šulc, Liam P Shaw, Ard A Louis, and Jonathan P K Doye. Dna hairpins destabilize duplexes primarily by promoting melting rather than by inhibiting hybridization. *Nucleic Acids Res*, 43(13):6181–6190, Jul 2015. ISSN 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). doi: 10.1093/nar/gkv582.

[7] Niranjan Srinivas, Thomas E Ouldridge, Petr Šulc, Joseph M Schaeffer, Bernard Yurke, Ard A Louis, Jonathan P K Doye, and Erik Winfree. On the biophysics and kinetics of toehold-mediated dna strand displacement. *Nucleic Acids Res.*, 41(22):10641–10658, 2013. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3905871/.

[8] Benedict E. K. Snodin, Flavio Romano, Lorenzo Rovigatti, Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. Direct simulation of the self-assembly of a small dna origami. *ACS Nano*, 10(2):1724–1737, 02 2016. doi: 10.1021/acsnano.5b05865. URL https://doi.org/10.1021/acsnano.5b05865.

[9] Vid Kočar, John S. Schreck, Slavko Čeru, Helena Gradišar, Nino Bašić, Tomaž Pisanski, Jonathan P. K. Doye, and Roman Jerala. Design principles for rapid folding of knotted dna nanostructures. *Nature Communications*, 7(1): 10803, 2016. doi: 10.1038/ncomms10803. URL https://doi.org/10.1038/ncomms10803.

[10] Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. Dna nanotweezers studied with a coarse-grained model of dna. *Physical Review Letters*, 104(17):178101–, 04 2010. doi: 10.1103/PhysRevLett.104.178101. URL https://link.aps.org/doi/10.1103/PhysRevLett.104.178101.

[11] Thomas E. Ouldridge, Rollo L. Hoare, Ard A. Louis, Jonathan P. K. Doye, Jonathan Bath, and Andrew J. Turberfield. Optimizing dna nanotechnology through coarse-grained modeling: A two-footed dna walker. *ACS Nano*, 7 (3):2479–2490, 03 2013. doi: 10.1021/nn3058483. URL https://doi.org/10.1021/nn3058483.

[12] Petr Šulc, Thomas E. Ouldridge, Flavio Romano, Jonathan P. K. Doye, and Ard A. Louis. Simulating a burnt-bridges dna motor with a coarse-grained dna model. *Natural Computing*, 13(4):535–547, 2014. doi: 10.1007/s11047-013-9391-8. URL https://doi.org/10.1007/s11047-013-9391-8.

[13] Iong Ying Loh, Juan Cheng, Shern Ren Tee, Artem Efremov, and Zhisong Wang. From bistate molecular switches to self-directed track-walking nanomotors. *ACS Nano*, 8(10):10293–10304, 10 2014. doi: 10.1021/nn5034983. URL https://doi.org/10.1021/nn5034983.

[14] Meihan Liu, Juan Cheng, Shern Ren Tee, Sarangapani Sreelatha, Iong Ying Loh, and Zhisong Wang. Biomimetic autonomous enzymatic nanowalker of high fuel efficiency. *ACS Nano*, 10(6):5882–5890, 06 2016. doi: 10.1021/acsnano.6b01035. URL https://doi.org/10.1021/acsnano.6b01035.

[15] Fan Hong, Fei Zhang, Yan Liu, and Hao Yan. Dna origami: Scaffolds for creating higher order structures. *Chemical Reviews*, 117(20):12584–12640, 10 2017. doi: 10.1021/acs.chemrev.6b00825. URL https://doi.org/10.1021/acs.chemrev.6b00825.

[16] Pengfei Wang, Travis A. Meyer, Victor Pan, Palash K. Dutta, and Yonggang Ke. The beauty and utility of dna origami. *Chem*, 2(3):359–382, 2017. doi: https://doi.org/10.1016/j.chempr.2017.02.009. URL http://www.sciencedirect.com/science/article/pii/S2451929417300785.

[17] Yonggang Ke, Luvena L. Ong, William M. Shih, and Peng Yin. Three-dimensional structures self-assembled from dna bricks. *Science*, 338(6111):1177, 11 2012. doi: 10.1126/science.1227268. URL http://science.sciencemag.org/content/338/6111/1177.abstract.

[18] Luvena L. Ong, Nikita Hanikel, Omar K. Yaghi, Casey Grun, Maximilian T. Strauss, Patrick Bron, Josephine Lai-Kee-Him, Florian Schueder, Bei Wang, Pengfei Wang, Jocelyn Y. Kishi, Cameron Myhrvold, Allen Zhu, Ralf Jungmann, Gaetan Bellot, Yonggang Ke, and Peng Yin. Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature*, 552(7683):72–77, 2017. doi: 10.1038/nature24648. URL https://doi.org/10.1038/nature24648.

[19] Damien Woods, David Doty, Cameron Myhrvold, Joy Hui, Felix Zhou, Peng Yin, and Erik Winfree. Diverse and robust molecular algorithms using reprogrammable dna self-assembly. *Nature*, 567(7748):366–372, 2019. doi: 10.1038/s41586-019-1014-9. URL https://doi.org/10.1038/s41586-019-1014-9.

[20] Sha Sun, Yuxuan Yang, Dongmin Li, and Jin Zhu. Large chiral nanotubes self-assembled by dna bricks. *Journal of the American Chemical Society*, 141(50):19524–19528, 12 2019. doi: 10.1021/jacs.9b08737. URL https://doi.org/10.1021/jacs.9b08737.

[21] . URL https://en.wikipedia.org/wiki/Nucleobase.

[22] James D Watson and Francis F H C Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953. URL https://doi.org/10.1038/171737a0.

[23] . URL en.wikipedia.org/wiki/DNA.

[24] Maxim D. Frank-Kamenetskii. Biophysics of the dna molecule. *Physics Reports*, 288(1):13–60, 1997. doi: https://doi.org/10.1016/S0370-1573(97)00020-3. URL https://www.sciencedirect.com/science/article/pii/S0370157397000203.

[25] Peter Yakovchuk, Ekaterina Protozanova, and Maxim D. Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the dna double helix. *Nucleic Acids Research*, 34(2):564–574, 01 2006. ISSN 0305-1048. doi: 10.1093/nar/gkj454. URL https://doi.org/10.1093/nar/gkj454.

[26] Hatim T. Allawi and John SantaLucia. Thermodynamics and nmr of internal g · t mismatches in dna. *Biochemistry*, 36(34):10581–10594, 08 1997. doi: 10.1021/bi962590c. URL https://doi.org/10.1021/bi962590c.

[27] Hatim T. Allawi and John SantaLucia. Nearest neighbor thermodynamic parameters for internal g · a mismatches in dna. *Biochemistry*, 37(8):2170–2179, 02 1998. doi: 10.1021/bi9724873. URL https://doi.org/10.1021/bi9724873.

[28] Hatim T. Allawi and Jr SantaLucia, John. Thermodynamics of internal C · T mismatches in DNA. *Nucleic Acids Research*, 26(11):2694–2701, 06 1998. ISSN 0305-1048. doi: 10.1093/nar/26.11.2694. URL https://doi.org/10.1093/nar/26.11.2694.

[29] Hatim T. Allawi and John SantaLucia. Nearest-neighbor thermodynamics of internal a · c mismatches in dna: Sequence dependence and ph effects. *Biochemistry*, 37(26):9435–9444, 06 1998. doi: 10.1021/bi9803729. URL https://doi.org/10.1021/bi9803729.

[30] Nicolas Peyret, P. Ananda Seneviratne, Hatim T. Allawi, and John SantaLucia. Nearest-neighbor thermodynamics and nmr of dna sequences with internal a · a, c · c, g · g, and t · t mismatches. *Biochemistry*, 38(12):3468–3477, 03 1999. doi: 10.1021/bi9825091. URL https://doi.org/10.1021/bi9825091.

[31] John SantaLucia and Donald Hicks. The thermodynamics of dna structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33(1):415–440, 2020/07/13 2004. doi: 10.1146/annurev.biophys.32.110601.141800. URL https://doi.org/10.1146/annurev.biophys.32.110601.141800.

[32] Jean-Philip Piquemal, Lalith Perera, G. Andrés Cisneros, Pengyu Ren, Lee G. Pedersen, and Thomas A. Darden. Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *The Journal of Chemical Physics*, 125(5):054511, 2020/07/13 2006. doi: 10.1063/1.2234774. URL https://doi.org/10.1063/1.2234774.

[33] Dian Jiao, Christopher King, Alan Grossfield, Thomas A. Darden, and Pengyu Ren. Simulation of ca2+ and mg2+ solvation using polarizable atomic multipole potential. *The Journal of Physical Chemistry B*, 110(37):18553–18559, 09 2006. doi: 10.1021/jp062230r. URL `https://doi.org/10.1021/jp062230r`.

[34] Weifeng Li, Lars Nordenskiöld, and Yuguang Mu. Sequence-specific mg2+–dna interactions: A molecular dynamics simulation study. *The Journal of Physical Chemistry B*, 115(49):14713–14720, 12 2011. doi: 10.1021/jp2052568. URL `https://doi.org/10.1021/jp2052568`.

[35] J. Anastassopoulou and T. Theophanides. Magnesium–dna interactions and the possible relation of magnesium to carcinogenesis. irradiation and free radicals. *Critical Reviews in Oncology/Hematology*, 42(1):79–91, 2002. doi: https://doi.org/10.1016/S1040-8428(02)00006-9. URL `https://www.sciencedirect.com/science/article/pii/S1040842802000069`.

[36] Donna R Whelan, Thomas J Hiscox, Julian I Rood, Keith R Bambery, Don McNaughton, and Bayden R Wood. Detection of an en masse and reversible b-to a-dna conformational transition in prokaryotes in response to desiccation. *J R Soc Interface*, 11(97):20140454, Aug 2014. ISSN 1742-5662 (Electronic); 1742-5689 (Print); 1742-5662 (Linking). doi: 10.1098/rsif.2014.0454.

[37] S C Mohr, N V Sokolov, C M He, and P Setlow. Binding of small acid-soluble spore proteins from bacillus subtilis changes the conformation of dna from b to a. *Proc Natl Acad Sci U S A*, 88(1):77–81, Jan 1991. ISSN 0027-8424 (Print); 1091-6490 (Electronic); 0027-8424 (Linking). doi: 10.1073/pnas.88.1.77.

[38] Xiang-Jun Lu, Zippora Shakked, and Wilma K. Olson. A-form conformational motifs in ligand-bound dna structures. *Journal of Molecular Biology*, 300 (4):819–840, 2000. doi: https://doi.org/10.1006/jmbi.2000.3690. URL `http://www.sciencedirect.com/science/article/pii/S0022283600936908`.

[39] Donna R Whelan, Keith R Bambery, Philip Heraud, Mark J Tobin, Max Diem, Don McNaughton, and Bayden R Wood. Monitoring the reversible b to a-like transition of dna in eukaryotic cells using fourier transform infrared spectroscopy. *Nucleic Acids Res*, 39(13):5439–5448, Jul 2011. ISSN 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). doi: 10.1093/nar/gkr175.

[40] Y G Kim, K Lowenhaupt, S Maas, A Herbert, T Schwartz, and A Rich. The zab domain of the human rna editing enzyme adar1 recognizes z-dna when surrounded by b-dna. *J Biol Chem*, 275(35):26828–26833, Sep 2000. ISSN 0021-9258 (Print); 0021-9258 (Linking). doi: 10.1074/jbc.M003477200.

[41] Sung Chul Ha, Ky Lowenhaupt, Alexander Rich, Yang-Gyun Kim, and Kyeong Kyu Kim. Crystal structure of a junction between b-dna and z-dna reveals two extruded bases. *Nature*, 437(7062):1183–1186, 2005. doi: 10.1038/nature04088. URL `https://doi.org/10.1038/nature04088`.

[42] Fritz M. Pohl and Thomas M. Jovin. Salt-induced co-operative conformational change of a synthetic dna: Equilibrium and kinetic studies with poly(dg-dc). *Journal of Molecular Biology*, 67(3):375–396, 1972. doi: https://doi.org/10.1016/0022-2836(72)90457-3. URL `http://www.sciencedirect.com/science/article/pii/0022283672904573`.

[43] T J Thamann, R C Lord, A H Wang, and A Rich. The high salt form of poly(dg-dc).poly(dg-dc) is left-handed z-dna: Raman spectra of crystals and solutions. *Nucleic acids research*, 9(20):5443–5457, 10 1981. doi: 10.1093/nar/9.20.5443. URL `https://pubmed.ncbi.nlm.nih.gov/7301594`.

[44] L J Peck, A Nordheim, A Rich, and J C Wang. Flipping of cloned d(pcpg)n.d(pcpg)n dna sequences from right- to left-handed helical structure by salt, co(iii), or negative supercoiling. *Proc Natl Acad Sci U S A*, 79 (15):4560–4564, Aug 1982. ISSN 0027-8424 (Print); 1091-6490 (Electronic); 0027-8424 (Linking). doi: 10.1073/pnas.79.15.4560.

[45] Nadrian C. Seeman. Nucleic acid junctions and lattices. *Journal of Theoretical Biology*, 99(2):237–247, 1982. doi: https://doi.org/10.1016/0022-5193(82)90002-9. URL http://www.sciencedirect.com/science/article/pii/0022519382900029.

[46] Neville R. Kallenbach, Rong-Ine Ma, and Nadrian C. Seeman. An immobile nucleic acid junction constructed from oligonucleotides. *Nature*, 305(5937):829–831, 1983. doi: 10.1038/305829a0. URL https://doi.org/10.1038/305829a0.

[47] Y L Wang, J E Mueller, B Kemper, and N C Seeman. Assembly and characterization of five-arm and six-arm dna branched junctions. *Biochemistry*, 30(23):5667–5674, Jun 1991. ISSN 0006-2960 (Print); 0006-2960 (Linking). doi: 10.1021/bi00237a005.

[48] Xing Wang and Nadrian C. Seeman. Assembly and characterization of 8-arm and 12-arm dna branched junctions. *Journal of the American Chemical Society*, 129(26):8169–8176, 07 2007. doi: 10.1021/ja0693441. URL https://doi.org/10.1021/ja0693441.

[49] Mary L. Petrillo, Colin J. Newton, Richard P. Cunningham, Rong-Ine Ma, Neville R. Kallenbach, and Nadrian C. Seeman. The ligation and flexibility of four-arm dna junctions. *Biopolymers*, 27(9):1337–1352, 2020/07/20 1988. doi: 10.1002/bip.360270902. URL https://doi.org/10.1002/bip.360270902.

[50] Jung Huei Chen, Neville R. Kallenbach, and Nadrian C. Seeman. A specific quadrilateral synthesized from dna branched junctions. *Journal of the American Chemical Society*, 111(16):6402–6407, 08 1989. doi: 10.1021/ja00198a063. URL https://doi.org/10.1021/ja00198a063.

[51] R I Ma, N R Kallenbach, R D Sheardy, M L Petrillo, and N C Seeman. Three-arm nucleic acid junctions are flexible. *Nucleic acids research*, 14(24):9745–9753, 12 1986. doi: 10.1093/nar/14.24.9745. URL https://pubmed.ncbi.nlm.nih.gov/3808954.

[52] Tsu Ju Fu and Nadrian C. Seeman. Dna double-crossover molecules. *Biochemistry*, 32(13):3211–3220, 04 1993. doi: 10.1021/bi00064a003. URL https://doi.org/10.1021/bi00064a003.

[53] Erik Winfree, Furong Liu, Lisa A. Wenzler, and Nadrian C. Seeman. Design and self-assembly of two-dimensional dna crystals. *Nature*, 394(6693):539–544, 1998. doi: 10.1038/28998. URL https://doi.org/10.1038/28998.

[54] Hao Yan, Thomas H. LaBean, Liping Feng, and John H. Reif. Directed nucleation assembly of dna tile complexes for barcode-patterned lattices. *Proceedings of the National Academy of Sciences*, 100(14):8103, 07 2003. doi: 10.1073/pnas.1032954100. URL http://www.pnas.org/content/100/14/8103.abstract.

[55] Paul W. K. Rothemund, Axel Ekani-Nkodo, Nick Papadakis, Ashish Kumar, Deborah Kuchnir Fygenson, and Erik Winfree. Design and characterization of programmable dna nanotubes. *Journal of the American Chemical Society*, 126(50):16344–16352, 12 2004. doi: 10.1021/ja044319l. URL https://doi.org/10.1021/ja044319l.

[56] Robert D. Barish, Paul W. K. Rothemund, and Erik Winfree. Two computational primitives for algorithmic self-assembly: Copying and counting. *Nano Letters*, 5(12):2586–2592, 12 2005. doi: 10.1021/nl052038l. URL https://doi.org/10.1021/nl052038l.

[57] Paul W. K Rothemund, Nick Papadakis, and Erik Winfree. Algorithmic self-assembly of dna sierpinski triangles. *PLOS Biology*, 2(12):e424–, 12 2004. URL https://doi.org/10.1371/journal.pbio.0020424.

[58] Hao Yan, Sung Ha Park, Gleb Finkelstein, John H. Reif, and Thomas H. LaBean. Dna-templated self-assembly of protein arrays and highly conductive nanowires. *Science*, 301(5641):1882, 09 2003. doi: 10.1126/science.1089389. URL http://science.sciencemag.org/content/301/5641/1882.abstract.

[59] Yu He, Tao Ye, Min Su, Chuan Zhang, Alexander E. Ribbe, Wen Jiang, and Chengde Mao. Hierarchical self-assembly of dna into symmetric supramolecular polyhedra. *Nature*, 452(7184):198–201, 2008. doi: 10.1038/nature06597. URL https://doi.org/10.1038/nature06597.

[60] Chuan Zhang, Min Su, Yu He, Xin Zhao, Ping-an Fang, Alexander E. Ribbe, Wen Jiang, and Chengde Mao. Conformational flexibility facilitates self-assembly of complex dna nanostructures. *Proceedings of the National Academy of Sciences*, 105(31):10665, 08 2008. doi: 10.1073/pnas.0803841105. URL http://www.pnas.org/content/105/31/10665.abstract.

[61] Chuan Zhang, Seung Hyeon Ko, Min Su, Yujun Leng, Alexander E. Ribbe, Wen Jiang, and Chengde Mao. Symmetry controls the face geometry of dna polyhedra. *Journal of the American Chemical Society*, 131(4):1413–1415, 02 2009. doi: 10.1021/ja809666h. URL https://doi.org/10.1021/ja809666h.

[62] Dage Liu, Mingsheng Wang, Zhaoxiang Deng, Richard Walulu, and Chengde Mao. Tensegrity: Construction of rigid dna triangles with flexible four-arm dna junctions. *Journal of the American Chemical Society*, 126(8):2324–2325, 03 2004. doi: 10.1021/ja031754r. URL https://doi.org/10.1021/ja031754r.

[63] Jianping Zheng, Jens J. Birktoft, Yi Chen, Tong Wang, Ruojie Sha, Pamela E. Constantinou, Stephan L. Ginell, Chengde Mao, and Nadrian C. Seeman. From molecular to macroscopic via the rational design of a self-assembled 3d dna crystal. *Nature*, 461(7260):74–77, 2009. doi: 10.1038/nature08274. URL https://doi.org/10.1038/nature08274.

[64] Xiao Wang, Ruojie Sha, Martin Kristiansen, Carina Hernandez, Yudong Hao, Chengde Mao, James W Canary, and Nadrian C Seeman. An organic semiconductor organized into 3d dna arrays by "bottom-up" rational design. *Angew Chem Int Ed Engl*, 56(23):6445–6448, Jun 2017. ISSN 1521-3773 (Electronic); 1433-7851 (Linking). doi: 10.1002/anie.201700462.

[65] Thomas H. LaBean, Hao Yan, Jens Kopatsch, Furong Liu, Erik Winfree, John H. Reif, and Nadrian C. Seeman. Construction, analysis, ligation, and self-assembly of dna triple crossover complexes. *Journal of the American Chemical Society*, 122(9):1848–1860, 03 2000. doi: 10.1021/ja993393e. URL https://doi.org/10.1021/ja993393e.

[66] Chengde Mao, Thomas H. LaBean, John H. Reif, and Nadrian C. Seeman. Logical computation using algorithmic self-assembly of dna triple-crossover molecules. *Nature*, 407(6803):493–496, 2000. doi: 10.1038/35035038. URL https://doi.org/10.1038/35035038.

[67] Dage Liu, Sung Ha Park, John H. Reif, and Thomas H. LaBean. Dna nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. *Proceedings of the National Academy of Sciences of the United States of America*, 101(3):717, 01 2004. doi: 10.1073/pnas.0305860101. URL http://www.pnas.org/content/101/3/717.abstract.

[68] Junghuei Chen and Nadrian C. Seeman. Synthesis from dna of a molecule with the connectivity of a cube. *Nature*, 350(6319):631–633, 1991. doi: 10.1038/350631a0. URL https://doi.org/10.1038/350631a0.

[69] Yuwen Zhang and Nadrian C. Seeman. Construction of a dna-truncated octahedron. *Journal of the American Chemical Society*, 116(5):1661–1669, 03 1994. doi: 10.1021/ja00084a006. URL https://doi.org/10.1021/ja00084a006.

[70] Russell P. Goodman, Richard M. Berry, and Andrew J. Turberfield. The single-step synthesis of a dna tetrahedron. *Chemical Communications*, (12): 1372–1373, 2004. doi: 10.1039/B402293A. URL http://dx.doi.org/10.1039/B402293A.

[71] R. P. Goodman, I. A. T. Schaap, C. F. Tardin, C. M. Erben, R. M. Berry, C. F. Schmidt, and A. J. Turberfield. Rapid chiral assembly of rigid dna building blocks for molecular nanofabrication. *Science*, 310(5754):1661, 12 2005. doi:

10.1126/science.1120367. URL http://science.sciencemag.org/content/310/5754/1661.abstract.

[72] Christoph M. Erben, Russell P. Goodman, and Andrew J. Turberfield. A self-assembled dna bipyramid. *Journal of the American Chemical Society*, 129 (22):6992–6993, 06 2007. doi: 10.1021/ja071493b. URL https://doi.org/10.1021/ja071493b.

[73] Paul W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006. doi: 10.1038/nature04586. URL https://doi.org/10.1038/nature04586.

[74] Ebbe S. Andersen, Mingdong Dong, Morten M. Nielsen, Kasper Jahn, Ramesh Subramani, Wael Mamdouh, Monika M. Golas, Bjoern Sander, Holger Stark, Cristiano L. P. Oliveira, Jan Skov Pedersen, Victoria Birkedal, Flemming Besenbacher, Kurt V. Gothelf, and Jørgen Kjems. Self-assembly of a nanoscale dna box with a controllable lid. *Nature*, 459(7243):73–76, 2009. doi: 10.1038/nature07971. URL https://doi.org/10.1038/nature07971.

[75] Yonggang Ke, Jaswinder Sharma, Minghui Liu, Kasper Jahn, Yan Liu, and Hao Yan. Scaffolded dna origami of a dna tetrahedron molecular container. *Nano Letters*, 9(6):2445–2447, 06 2009. doi: 10.1021/nl901165f. URL https://doi.org/10.1021/nl901165f.

[76] Dongran Han, Suchetan Pal, Jeanette Nangreave, Zhengtao Deng, Yan Liu, and Hao Yan. Dna origami with complex curvatures in three-dimensional space. *Science*, 332(6027):342, 04 2011. doi: 10.1126/science.1202998. URL http://science.sciencemag.org/content/332/6027/342.abstract.

[77] Yonggang Ke, Shawn M. Douglas, Minghui Liu, Jaswinder Sharma, Anchi Cheng, Albert Leung, Yan Liu, William M. Shih, and Hao Yan. Multilayer dna origami packed on a square lattice. *Journal of the American Chemical Society*, 131(43):15903–15908, 11 2009. doi: 10.1021/ja906381y. URL https://doi.org/10.1021/ja906381y.

[78] Shawn M. Douglas, Hendrik Dietz, Tim Liedl, Björn Högberg, Franziska Graf, and William M. Shih. Self-assembly of dna into nanoscale three-dimensional shapes. *Nature*, 459(7245):414–418, 2009. doi: 10.1038/nature08016. URL https://doi.org/10.1038/nature08016.

[79] Yonggang Ke, Niels V. Voigt, Kurt V. Gothelf, and William M. Shih. Multilayer dna origami packed on hexagonal and hybrid lattices. *Journal of the American Chemical Society*, 134(3):1770–1774, 01 2012. doi: 10.1021/ja209719k. URL https://doi.org/10.1021/ja209719k.

[80] Erik Benson, Abdulmelik Mohammed, Johan Gardell, Sergej Masich, Eugen Czeizler, Pekka Orponen, and Björn Högberg. Dna rendering of polyhedral meshes at the nanoscale. *Nature*, 523(7561):441–444, 2015. doi: 10.1038/nature14586. URL https://doi.org/10.1038/nature14586.

[81] Dongran Han, Suchetan Pal, Yang Yang, Shuoxing Jiang, Jeanette Nangreave, Yan Liu, and Hao Yan. Dna gridiron nanostructures based on four-arm junctions. *Science*, 339(6126):1412, 03 2013. doi: 10.1126/science.1232252. URL http://science.sciencemag.org/content/339/6126/1412.abstract.

[82] Fei Zhang, Shuoxing Jiang, Siyu Wu, Yulin Li, Chengde Mao, Yan Liu, and Hao Yan. Complex wireframe dna origami nanostructures with multi-arm junction vertices. *Nature Nanotechnology*, 10(9):779–784, 2015. doi: 10.1038/nnano.2015.162. URL https://doi.org/10.1038/nnano.2015.162.

[83] Wenyan Liu, Hong Zhong, Risheng Wang, and Nadrian C. Seeman. Crystalline two-dimensional dna-origami arrays. *Angewandte Chemie International Edition*, 50(1):264–267, 2020/07/20 2011. doi: 10.1002/anie.201005911. URL https://doi.org/10.1002/anie.201005911.

[84] Pengfei Wang, Stavros Gaitanaros, Seungwoo Lee, Mark Bathe, William M. Shih, and Yonggang Ke. Programming self-assembly of dna origami honeycomb

two-dimensional lattices and plasmonic metamaterials. *Journal of the American Chemical Society*, 138(24):7733–7740, 06 2016. doi: 10.1021/jacs.6b03966. URL https://doi.org/10.1021/jacs.6b03966.

[85] Ryosuke Iinuma, Yonggang Ke, Ralf Jungmann, Thomas Schlichthaerle, Johannes B. Woehrstein, and Peng Yin. Polyhedra self-assembled from dna tripods and characterized with 3d dna-paint. *Science*, 344(6179):65, 04 2014. doi: 10.1126/science.1250944. URL http://science.sciencemag.org/content/344/6179/65.abstract.

[86] .

[87] Peng Yin, Rizal F. Hariadi, Sudheer Sahu, Harry M. T. Choi, Sung Ha Park, Thomas H. LaBean, and John H. Reif. Programming dna tube circumferences. *Science*, 321(5890):824, 08 2008. doi: 10.1126/science.1157312. URL http://science.sciencemag.org/content/321/5890/824.abstract.

[88] Bryan Wei, Mingjie Dai, and Peng Yin. Complex shapes self-assembled from single-stranded dna tiles. *Nature*, 485(7400):623–626, 2012. doi: 10.1038/nature11075. URL https://doi.org/10.1038/nature11075.

[89] Bryan Wei, Mingjie Dai, Cameron Myhrvold, Yonggang Ke, Ralf Jungmann, and Peng Yin. Design space for complex dna structures. *Journal of the American Chemical Society*, 135(48):18080–18088, 12 2013. doi: 10.1021/ja4062294. URL https://doi.org/10.1021/ja4062294.

[90] Yonggang Ke, Luvena L. Ong, William M. Shih, and Peng Yin. Three-dimensional structures self-assembled from dna bricks. *Science*, 338(6111):1177, 11 2012. doi: 10.1126/science.1227268. URL http://science.sciencemag.org/content/338/6111/1177.abstract.

[91] Luvena L. Ong, Nikita Hanikel, Omar K. Yaghi, Casey Grun, Maximilian T. Strauss, Patrick Bron, Josephine Lai-Kee-Him, Florian Schueder, Bei Wang, Pengfei Wang, Jocelyn Y. Kishi, Cameron Myhrvold, Allen Zhu, Ralf

Jungmann, Gaetan Bellot, Yonggang Ke, and Peng Yin. Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature*, 552(7683):72–77, 2017. doi: 10.1038/nature24648. URL https://doi.org/10.1038/nature24648.

[92] Yonggang Ke, Luvena L Ong, Wei Sun, Jie Song, Mingdong Dong, William M Shih, and Peng Yin. Dna brick crystals with prescribed depths. *Nat Chem*, 6(11):994–1002, Nov 2014. ISSN 1755-4349 (Electronic); 1755-4330 (Print); 1755-4330 (Linking). doi: 10.1038/nchem.2083.

[93] Bernard Yurke, Andrew J. Turberfield, Allen P. Mills, Friedrich C. Simmel, and Jennifer L. Neumann. A dna-fuelled molecular machine made of dna. *Nature*, 406(6796):605–608, 2000. doi: 10.1038/35020524. URL https://doi.org/10.1038/35020524.

[94] F C Simmel and B Yurke. Using dna to construct and power a nanoactuator. *Phys Rev E Stat Nonlin Soft Matter Phys*, 63(4 Pt 1):041913, Apr 2001. ISSN 1539-3755 (Print); 1539-3755 (Linking). doi: 10.1103/PhysRevE.63.041913.

[95] Banani Chakraborty, Ruojie Sha, and Nadrian C. Seeman. A dna-based nanomechanical device with three robust states. *Proceedings of the National Academy of Sciences*, 105(45):17245, 11 2008. doi: 10.1073/pnas.0707681105. URL http://www.pnas.org/content/105/45/17245.abstract.

[96] William B. Sherman and Nadrian C. Seeman. A precisely controlled dna biped walking device. *Nano Letters*, 4(7):1203–1207, 07 2004. doi: 10.1021/nl049527q. URL https://doi.org/10.1021/nl049527q.

[97] Jong-Shik Shin and Niles A. Pierce. A synthetic dna walker for molecular transport. *Journal of the American Chemical Society*, 126(35):10834–10835, 09 2004. doi: 10.1021/ja047543j. URL https://doi.org/10.1021/ja047543j.

[98] Hongzhou Gu, Jie Chao, Shou-Jun Xiao, and Nadrian C Seeman. A proximity-based programmable dna nanoscale assembly line. *Nature*, 465(7295):202–205,

05 2010. doi: 10.1038/nature09026. URL https://pubmed.ncbi.nlm.nih.gov/20463734.

[99] Tosan Omabegho, Ruojie Sha, and Nadrian C. Seeman. A bipedal dna brownian motor with coordinated legs. *Science*, 324(5923):67, 04 2009. doi: 10.1126/science.1170336. URL http://science.sciencemag.org/content/324/5923/67.abstract.

[100] Jiří Šponer, Kevin E. Riley, and Pavel Hobza. Nature and magnitude of aromatic stacking of nucleic acid bases. *Physical Chemistry Chemical Physics*, 10(19):2595–2610, 2008. doi: 10.1039/B719370J. URL http://dx.doi.org/10.1039/B719370J.

[101] Jiří Šponer, Petr Jurečka, and Pavel Hobza. Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. *Journal of the American Chemical Society*, 126(32):10142–10151, 08 2004. doi: 10.1021/ja048436s. URL https://doi.org/10.1021/ja048436s.

[102] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 05 1995. doi: 10.1021/ja00124a002. URL https://doi.org/10.1021/ja00124a002.

[103] . URL ambermd.org.

[104] Alexander D. MacKerell Jr. and Nilesh K. Banavali. All-atom empirical force field for nucleic acids: Ii. application to molecular dynamics simulations of dna and rna in solution. *Journal of Computational Chemistry*, 21(2):105–120, 2020/07/27 2000. doi: 10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P. URL https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<105::AID-JCC3>3.0.CO;2-P.

[105] . URL www.charmm.org.

[106] Eric J. Sorin, Young Min Rhee, Bradley J. Nakatani, and Vijay S. Pande. Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophysical Journal*, 85(2):790–803, 2003. doi: https://doi.org/10.1016/S0006-3495(03)74520-2. URL http://www.sciencedirect.com/science/article/pii/S0006349503745202.

[107] Srinivasaraghavan Kannan and Martin Zacharias. Folding of a dna hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophysical Journal*, 93(9):3218–3228, 2007. doi: https://doi.org/10.1529/biophysj.107.108019. URL http://www.sciencedirect.com/science/article/pii/S0006349507715758.

[108] Guillem Portella and Modesto Orozco. Multiple routes to characterize the folding of a small dna hairpin. *Angewandte Chemie International Edition*, 49(42):7673–7676, 2020/07/27 2010. doi: 10.1002/anie.201003816. URL https://doi.org/10.1002/anie.201003816.

[109] Arjan van der Vaart. Coupled binding–bending–folding: The complex conformational dynamics of protein-dna binding studied by atomistic molecular dynamics simulations. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1850(5):1091–1098, 2015. doi: https://doi.org/10.1016/j.bbagen.2014.08.009. URL http://www.sciencedirect.com/science/article/pii/S0304416514002852.

[110] Rodrigo Galindo-Murillo, Daniel R. Roe, and Thomas E. Cheatham. On the absence of intrahelical dna dynamics on the µs to ms timescale. *Nature Communications*, 5(1):5152, 2014. doi: 10.1038/ncomms6152. URL https://doi.org/10.1038/ncomms6152.

[111] Alberto Pérez, F. Javier Luque, and Modesto Orozco. Dynamics of b-dna on the microsecond time scale. *Journal of the American Chemical Society*, 129

(47):14739–14745, 11 2007. doi: 10.1021/ja0753546. URL https://doi.org/10.1021/ja0753546.

[112] Rodrigo Galindo-Murillo, Daniel R Roe, and Thomas E 3rd Cheatham. Convergence and reproducibility in molecular dynamics simulations of the dna duplex d(gcacgaacgaacgaacgc). *Biochim Biophys Acta*, 1850(5):1041–1058, May 2015. ISSN 0006-3002 (Print); 0006-3002 (Linking). doi: 10.1016/j.bbagen.2014.09.007.

[113] Srinivasaraghavan Kannan and Martin Zacharias. Simulation of dna double-strand dissociation and formation during replica-exchange molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, 11:10589–10595, 2009. doi: 10.1039/B910792B. URL http://dx.doi.org/10.1039/B910792B.

[114] Shiyan Xiao, Daniel J. Sharpe, Debayan Chakraborty, and David J. Wales. Energy landscapes and hybridization pathways for dna hexamer duplexes. *The Journal of Physical Chemistry Letters*, 10(21):6771–6779, 11 2019. doi: 10.1021/acs.jpclett.9b02356. URL https://doi.org/10.1021/acs.jpclett.9b02356.

[115] Michael F. Hagan, Aaron R. Dinner, David Chandler, and Arup K. Chakraborty. Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in dna. *Proceedings of the National Academy of Sciences*, 100(24):13922, 11 2003. doi: 10.1073/pnas.2036378100. URL http://www.pnas.org/content/100/24/13922.abstract.

[116] Jejoong Yoo and Aleksei Aksimentiev. In situ structure and dynamics of dna origami determined through molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 110(50):20099, 12 2013. doi: 10.1073/pnas.1316521110. URL http://www.pnas.org/content/110/50/20099.abstract.

[117] Christopher Maffeo, Jejoong Yoo, and Aleksei Aksimentiev. De novo reconstruction of dna origami structures through atomistic molecular dynamics simulation. *Nucleic Acids Res*, 44(7):3013–3019, Apr 2016. ISSN 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). doi: 10.1093/nar/gkw155.

[118] Scott Michael Slone, Chen-Yu Li, Jejoong Yoo, and Aleksei Aksimentiev. Molecular mechanics of dna bricks:in situstructure, mechanical properties and ionic conductivity. 18(5):055012, 2016. doi: 10.1088/1367-2630/18/5/055012. URL `http://dx.doi.org/10.1088/1367-2630/18/5/055012`.

[119] Christoph G. Baumann, Steven B. Smith, Victor A. Bloomfield, and Carlos Bustamante. Ionic effects on the elasticity of single dna molecules. *Proceedings of the National Academy of Sciences*, 94(12):6185, 06 1997. doi: 10.1073/pnas.94.12.6185. URL `http://www.pnas.org/content/94/12/6185.abstract`.

[120] O. Kratky and G. Porod. Röntgenuntersuchung gelöster fadenmoleküle. *Recueil des Travaux Chimiques des Pays-Bas*, 68(12):1106–1122, 2020/07/27 1949. doi: 10.1002/recl.19490681203. URL `https://doi.org/10.1002/recl.19490681203`.

[121] S B Smith, Y Cui, and C Bustamante. Overstretching b-dna: the elastic response of individual double-stranded and single-stranded dna molecules. *Science*, 271(5250):795–799, Feb 1996. ISSN 0036-8075 (Print); 0036-8075 (Linking). doi: 10.1126/science.271.5250.795.

[122] Gijs J. L. Wuite, Steven B. Smith, Mark Young, David Keller, and Carlos Bustamante. Single-molecule studies of the effect of template tension on t7 dna polymerase activity. *Nature*, 404(6773):103–106, 2000. doi: 10.1038/35003614. URL `https://doi.org/10.1038/35003614`.

[123] C. Bouchiat, M. D. Wang, J. F. Allemand, T. Strick, S. M. Block, and V. Croquette. Estimating the persistence length of a worm-like chain molecule

from force-extension measurements. *Biophysical Journal*, 76(1):409–413, 1999. doi: https://doi.org/10.1016/S0006-3495(99)77207-3. URL http://www.sciencedirect.com/science/article/pii/S0006349599772073.

[124] Theo Odijk. Stiff chains and filaments under tension. *Macromolecules*, 28(20): 7016–7018, 09 1995. doi: 10.1021/ma00124a044. URL https://doi.org/10.1021/ma00124a044.

[125] John F. Marko and Eric D. Siggia. Stretching dna. *Macromolecules*, 28(26): 8759–8770, 12 1995. doi: 10.1021/ma00130a008. URL https://doi.org/10.1021/ma00130a008.

[126] M D Wang, H Yin, R Landick, J Gelles, and S M Block. Stretching dna with optical tweezers. *Biophysical journal*, 72(3):1335–1346, 03 1997. doi: 10.1016/S0006-3495(97)78780-0. URL https://pubmed.ncbi.nlm.nih.gov/9138579.

[127] Jeff Gore, Zev Bryant, Marcelo Nöllmann, Mai U Le, Nicholas R Cozzarelli, and Carlos Bustamante. Dna overwinds when stretched. *Nature*, 442(7104): 836–839, Aug 2006. ISSN 1476-4687 (Electronic); 0028-0836 (Linking). doi: 10.1038/nature04974.

[128] R. D Kamien, T. C Lubensky, P Nelson, and C. S O'Hern. Direct determination of dna twist-stretch coupling. 38(3):237–242, 1997. doi: 10.1209/epl/i1997-00231-y. URL http://dx.doi.org/10.1209/epl/i1997-00231-y.

[129] John F. Marko. Dna under high tension: Overstretching, undertwisting, and relaxation dynamics. *Physical Review E*, 57(2):2134–2149, 02 1998. doi: 10.1103/PhysRevE.57.2134. URL https://link.aps.org/doi/10.1103/PhysRevE.57.2134.

[130] Peter Gross, Niels Laurens, Lene B. Oddershede, Ulrich Bockelmann, Erwin J. G. Peterman, and Gijs J. L. Wuite. Quantifying how dna stretches, melts

and changes twist under tension. *Nature Physics*, 7(9):731–736, 2011. doi: 10.1038/nphys2002. URL https://doi.org/10.1038/nphys2002.

[131] Do-Nyun Kim, Fabian Kilchherr, Hendrik Dietz, and Mark Bathe. Quantitative prediction of 3d solution shape and flexibility of nucleic acid nanostructures. *Nucleic Acids Res*, 40(7):2862–2868, Apr 2012. ISSN 1362-4962 (Electronic); 0305-1048 (Print); 0305-1048 (Linking). doi: 10.1093/nar/gkr1173.

[132] Keyao Pan, Do-Nyun Kim, Fei Zhang, Matthew R. Adendorff, Hao Yan, and Mark Bathe. Lattice-free prediction of three-dimensional structure of programmed dna assemblies. *Nature Communications*, 5(1):5578, 2014. doi: 10.1038/ncomms6578. URL https://doi.org/10.1038/ncomms6578.

[133] Shawn M Douglas, Adam H Marblestone, Surat Teerapittayanon, Alejandro Vazquez, George M Church, and William M Shih. Rapid prototyping of 3d dna-origami shapes with cadnano. *Nucleic acids research*, 37(15):5001–5006, 08 2009. doi: 10.1093/nar/gkp436. URL https://pubmed.ncbi.nlm.nih.gov/19531737.

[134] Carlos Ernesto Castro, Fabian Kilchherr, Do-Nyun Kim, Enrique Lin Shiao, Tobias Wauer, Philipp Wortmann, Mark Bathe, and Hendrik Dietz. A primer to scaffolded dna origami. *Nature Methods*, 8(3):221–229, 2011. doi: 10.1038/nmeth.1570. URL https://doi.org/10.1038/nmeth.1570.

[135] Douglas Poland and Harold A. Scheraga. Occurrence of a phase transition in nucleic acid models. *The Journal of Chemical Physics*, 45(5):1464–1469, 2020/07/27 1966. doi: 10.1063/1.1727786. URL https://doi.org/10.1063/1.1727786.

[136] Daniel Jost and Ralf Everaers. A unified poland-scheraga model of oligo- and polynucleotide dna melting: salt effects and predictive power. *Biophysical journal*, 96(3):1056–1067, 02 2009. doi: 10.1529/biophysj.108.134031. URL https://pubmed.ncbi.nlm.nih.gov/18849409.

[137] Alexey Savelyev and Garegin A. Papoian. Molecular renormalization group coarse-graining of polymer chains: Application to double-stranded dna. *Biophysical Journal*, 96(10):4044–4052, 2009. doi: https://doi.org/10.1016/j.bpj.2009.02.067. URL http://www.sciencedirect.com/science/article/pii/S0006349509006729.

[138] Alexey Savelyev and Garegin A. Papoian. Chemically accurate coarse graining of double-stranded dna. *Proceedings of the National Academy of Sciences*, 107 (47):20340, 11 2010. doi: 10.1073/pnas.1001163107. URL http://www.pnas.org/content/107/47/20340.abstract.

[139] Qianqian Cao, Chuncheng Zuo, Yanhong Ma, Lujuan Li, and Zhou Zhang. Interaction of double-stranded dna with a nanosphere: a coarse-grained molecular dynamics simulation study. *Soft Matter*, 7(2):506–514, 2011. doi: 10.1039/C0SM00512F. URL http://dx.doi.org/10.1039/C0SM00512F.

[140] Alexey Savelyev, Christopher K. Materese, and Garegin A. Papoian. Is dna's rigidity dominated by electrostatic or nonelectrostatic interactions? *Journal of the American Chemical Society*, 133(48):19290–19293, 12 2011. doi: 10.1021/ja207984z. URL https://doi.org/10.1021/ja207984z.

[141] Alex Morriss-Andrews, Joerg Rottler, and Steven S. Plotkin. A systematically coarse-grained model for dna and its predictions for persistence length, stacking, twist, and chirality. *The Journal of Chemical Physics*, 132(3): 035105, 2020/07/27 2010. doi: 10.1063/1.3269994. URL https://doi.org/10.1063/1.3269994.

[142] Aymeric Naômé, Aatto Laaksonen, and Daniel P. Vercauteren. 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, 08 2014. doi: 10.1021/ct500222s. URL https://doi.org/10.1021/ct500222s.

[143] Aymeric Naômé, Aatto Laaksonen, and Daniel P. Vercauteren. 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, 06 2015. doi: 10.1021/acs.jctc.5b00113. URL https://doi.org/10.1021/acs.jctc.5b00113.

[144] Thomas A. Knotts, Nitin Rathore, David C. Schwartz, and Juan J. de Pablo. A coarse grain model for dna. *The Journal of Chemical Physics*, 126(8): 084901, 2020/07/27 2007. doi: 10.1063/1.2431804. URL https://doi.org/10.1063/1.2431804.

[145] E J Sambriski, D C Schwartz, and J J de Pablo. A mesoscale model of dna and its renaturation. *Biophysical journal*, 96(5):1675–1690, 03 2009. doi: 10.1016/j.bpj.2008.09.061. URL https://pubmed.ncbi.nlm.nih.gov/19254530.

[146] Daniel M. Hinckley, Gordon S. Freeman, Jonathan K. Whitmer, and Juan J. de Pablo. An experimentally-informed coarse-grained 3-site-per-nucleotide model of dna: Structure, thermodynamics, and dynamics of hybridization. *The Journal of Chemical Physics*, 139(14):144903, 2020/07/27 2013. doi: 10.1063/1.4822042. URL https://doi.org/10.1063/1.4822042.

[147] Andrés Córdoba, Daniel M. Hinckley, Joshua Lequieu, and Juan J. de Pablo. A molecular view of the dynamics of dsdna packing inside viral capsids in the presence of ions. *Biophysical Journal*, 112(7):1302–1315, 2017. doi: https://doi.org/10.1016/j.bpj.2017.02.015. URL http://www.sciencedirect.com/science/article/pii/S0006349517302254.

[148] Wujie Wang, Laura M. Nocka, Brianne Z. Wiemann, Daniel M. Hinckley, Ishita Mukerji, and Francis W. Starr. Holliday junction thermodynamics and structure: Coarse-grained simulations and experiments. *Scientific Reports*, 6 (1):22863, 2016. doi: 10.1038/srep22863. URL https://doi.org/10.1038/srep22863.

[149] Joshua Lequieu, David C. Schwartz, and Juan J. de Pablo. In silico evidence for sequence-dependent nucleosome sliding. *Proceedings of the National Academy of Sciences*, 114(44):E9197, 10 2017. doi: 10.1073/pnas.1705685114. URL http://www.pnas.org/content/114/44/E9197.abstract.

[150] Toru Niina, Giovanni B. Brandani, Cheng Tan, and Shoji Takada. Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations. *PLOS Computational Biology*, 13 (12):e1005880–, 12 2017. URL https://doi.org/10.1371/journal.pcbi.1005880.

[151] Giovanni B Brandani, Toru Niina, Cheng Tan, and Shoji Takada. DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations. *Nucleic Acids Research*, 46(6):2788–2801, 02 2018. ISSN 0305-1048. doi: 10.1093/nar/gky158. URL https://doi.org/10.1093/nar/gky158.

[152] Pablo D. Dans, Ari Zeida, Matías R. Machado, and Sergio Pantano. A coarse grained model for atomic-detailed dna simulations with explicit electrostatics. *Journal of Chemical Theory and Computation*, 6(5):1711–1725, 05 2010. doi: 10.1021/ct900653p. URL https://doi.org/10.1021/ct900653p.

[153] Ari Zeida, Matías Rodrigo Machado, Pablo Daniel Dans, and Sergio Pantano. Breathing, bubbling, and bending: Dna flexibility from multimicrosecond simulations. *Physical Review E*, 86(2):021903–, 08 2012. doi: 10.1103/PhysRevE.86.021903. URL https://link.aps.org/doi/10.1103/PhysRevE.86.021903.

[154] Matias R. Machado and Sergio Pantano. Exploring laci–dna dynamics by multiscale simulations using the sirah force field. *Journal of Chemical Theory and Computation*, 11(10):5012–5023, 10 2015. doi: 10.1021/acs.jctc.5b00575. URL https://doi.org/10.1021/acs.jctc.5b00575.

[155] Astrid Brandner, Andreas Schüller, Francisco Melo, and Sergio Pantano. 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, 2018. doi: https://doi.org/10.1016/j.bbrc.2017.09.086. URL http://www.sciencedirect.com/science/article/pii/S0006291X17318570.

[156] Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. Structural, mechanical, and thermodynamic properties of a coarse-grained dna model. *The Journal of Chemical Physics*, 134(8):085101, 2020/07/27 2011. doi: 10.1063/1.3552946. URL https://doi.org/10.1063/1.3552946.

[157] Petr Šulc, Flavio Romano, Thomas E. Ouldridge, Lorenzo Rovigatti, Jonathan P. K. Doye, and Ard A. Louis. Sequence-dependent thermodynamics of a coarse-grained dna model. *The Journal of Chemical Physics*, 137(13):135101, 2020/07/27 2012. doi: 10.1063/1.4754132. URL https://doi.org/10.1063/1.4754132.

[158] Benedict E. K. Snodin, Ferdinando Randisi, Majid Mosayebi, Petr Šulc, John S. Schreck, Flavio Romano, Thomas E. Ouldridge, Roman Tsukanov, Eyal Nir, Ard A. Louis, and Jonathan P. K. Doye. Introducing improved structural properties and salt dependence into a coarse-grained model of dna. *The Journal of Chemical Physics*, 142(23):234901, 2020/07/27 2015. doi: 10.1063/1.4921957. URL https://doi.org/10.1063/1.4921957.

[159] Oliver Henrich, Yair Augusto Gutiérrez Fosado, Tine Curk, and Thomas E. Ouldridge. Coarse-grained simulation of dna using lammps. *The European Physical Journal E*, 41(5):57, 2018. doi: 10.1140/epje/i2018-11669-8. URL https://doi.org/10.1140/epje/i2018-11669-8.

[160] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, 1999.

[161] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092,

2020/07/31 1953. doi: 10.1063/1.1699114. URL https://doi.org/10.1063/1.1699114.

[162] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in monte carlo simulations. *Physical Review Letters*, 58(2):86–88, 01 1987. doi: 10.1103/PhysRevLett.58.86. URL https://link.aps.org/doi/10.1103/PhysRevLett.58.86.

[163] Jiwen Liu and Erik Luijten. Rejection-free geometric cluster algorithm for complex fluids. *Physical Review Letters*, 92(3):035504–, 01 2004. doi: 10.1103/PhysRevLett.92.035504. URL https://link.aps.org/doi/10.1103/PhysRevLett.92.035504.

[164] Stephen Whitelam and Phillip L. Geissler. Avoiding unphysical kinetic traps in monte carlo simulations of strongly attractive particles. *The Journal of Chemical Physics*, 127(15):154101, 2020/07/31 2007. doi: 10.1063/1.2790421. URL https://doi.org/10.1063/1.2790421.

[165] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977. doi: https://doi.org/10.1016/0021-9991(77)90121-8. URL http://www.sciencedirect.com/science/article/pii/0021999177901218.

[166] Alan M. Ferrenberg and Robert H. Swendsen. New monte carlo technique for studying phase transitions. *Physical Review Letters*, 61(23):2635–2638, 12 1988. doi: 10.1103/PhysRevLett.61.2635. URL https://link.aps.org/doi/10.1103/PhysRevLett.61.2635.

[167] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 2020/07/31 1992. doi: 10.1002/jcc.540130812. URL https://doi.org/10.1002/jcc.540130812.

[168] 2010. URL https://core.ac.uk/download/pdf/22873967.pdf.

[169] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976. doi: https://doi.org/10.1016/0021-9991(76)90041-3. URL http://www.sciencedirect.com/science/article/pii/0021999176900413.

[170] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 12 1977. doi: 10.1021/j100540a008. URL https://doi.org/10.1021/j100540a008.

[171] Santosh K. Upadhyay. *Chemical Kinetics and Reaction Dynamics*. Springer, Dordrecht, 2006.

[172] Daniel T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404–425, 1992. doi: https://doi.org/10.1016/0378-4371(92)90283-V. URL http://www.sciencedirect.com/science/article/pii/037843719290283V.

[173] David J. Warne, Ruth E. Baker, and Matthew J. Simpson. Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *Journal of the Royal Society Interface*, 16, 2019.

[174] Daniel T. Gillespie, Andreas Hellander, and Linda R. Petzold. Perspective: Stochastic algorithms for chemical kinetics. *The Journal of Chemical Physics*, 138(17):170901, 2020/08/04 2013. doi: 10.1063/1.4801941. URL https://doi.org/10.1063/1.4801941.

[175] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for monte carlo simulation of ising spin systems. *Journal of Computational Physics*, 17(1):10–18, 1975. doi: https://doi.org/10.1016/0021-9991(75)90060-1. URL https://www.sciencedirect.com/science/article/pii/0021999175900601.

[176] Daan Frenkel. Order through entropy. *Nature Materials*, 14(1):9–12, 2015. doi: 10.1038/nmat4178. URL https://doi.org/10.1038/nmat4178.

[177] Aleks Reinhardt and Daan Frenkel. Numerical evidence for nucleated self-assembly of dna brick structures. *Physical Review Letters*, 112(23):238103–, 06 2014. doi: 10.1103/PhysRevLett.112.238103. URL https://link.aps.org/doi/10.1103/PhysRevLett.112.238103.

[178] Aleks Reinhardt, Chon Pan Ho, and Daan Frenkel. Effects of co-ordination number on the nucleation behaviour in many-component self-assembly. *Faraday discussions*, 186:215–228, 2016. doi: 10.1039/c5fd00135h. URL http://europepmc.org/abstract/MED/26762705.

[179] Aleks Reinhardt and Daan Frenkel. Dna brick self-assembly with an off-lattice potential. *Soft Matter*, 12(29):6253–6260, 2016. doi: 10.1039/C6SM01031H. URL http://dx.doi.org/10.1039/C6SM01031H.

[180] William M. Jacobs, Aleks Reinhardt, and Daan Frenkel. Communication: Theoretical prediction of free-energy landscapes for complex self-assembly. *The Journal of Chemical Physics*, 142(2):021101, 2020/08/04 2015. doi: 10.1063/1.4905670. URL https://doi.org/10.1063/1.4905670.

[181] William M. Jacobs, Aleks Reinhardt, and Daan Frenkel. Rational design of self-assembly pathways for complex multicomponent structures. *Proceedings of the National Academy of Sciences*, 112(20):6313, 05 2015. doi: 10.1073/pnas.1502210112. URL http://www.pnas.org/content/112/20/6313.abstract.

[182] William M. Jacobs and Daan Frenkel. Self-assembly protocol design for periodic multicomponent structures. *Soft Matter*, 11(46):8930–8938, 2015. doi: 10.1039/C5SM01841B. URL http://dx.doi.org/10.1039/C5SM01841B.

[183] Arvind Murugan, James Zou, and Michael P. Brenner. Undesired usage and the robust self-assembly of heterogeneous structures. *Nature Communications*,

6(1):6203, 2015. doi: 10.1038/ncomms7203. URL https://doi.org/10.1038/ncomms7203.

[184] James G. Wetmur and Norman Davidson. Kinetics of renaturation of dna. *Journal of Molecular Biology*, 31(3):349–370, 1968. doi: https://doi.org/10.1016/0022-2836(68)90414-2. URL http://www.sciencedirect.com/science/article/pii/0022283668904142.

[185] Mary E. Craig, Donald M. Crothers, and Paul Doty. Relaxation kinetics of dimer formation by self complementary oligonucleotides. *Journal of Molecular Biology*, 62(2):383–401, 1971. doi: https://doi.org/10.1016/0022-2836(71)90434-7. URL http://www.sciencedirect.com/science/article/pii/0022283671904347.

[186] Dietmar Pörschke and Manfred Eigen. Co-operative non-enzymatic base recognition iii. kinetics of the helix—coil transition of the oligoribouridylic · oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic ph. *Journal of Molecular Biology*, 62(2):361–381, 1971. doi: https://doi.org/10.1016/0022-2836(71)90433-5. URL http://www.sciencedirect.com/science/article/pii/0022283671904335.

[187] Daniele Andreatta, Sobhan Sen, J. Luis Pérez Lustres, Sergey A. Kovalenko, Nikolaus P. Ernsting, Catherine J. Murphy, Robert S. Coleman, and Mark A. Berg. Ultrafast dynamics in dna: "fraying"at the end of the helix. *Journal of the American Chemical Society*, 128(21):6885–6892, 05 2006. doi: 10.1021/ja0582105. URL https://doi.org/10.1021/ja0582105.

[188] Paul J. Sanstead and Andrei Tokmakoff. Direct observation of activated kinetics and downhill dynamics in dna dehybridization. *The Journal of Physical Chemistry B*, 122(12):3088–3100, 03 2018. doi: 10.1021/acs.jpcb.8b01445. URL https://doi.org/10.1021/acs.jpcb.8b01445.

[189] Michael T. Woodside, Peter C. Anthony, William M. Behnke-Parks, Kevan Larizadeh, Daniel Herschlag, and Steven M. Block. Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science*, 314 (5801):1001, 11 2006. doi: 10.1126/science.1133601. URL `http://science.sciencemag.org/content/314/5801/1001.abstract`.

[190] Gen He, Jie Li, Haina Ci, Chuanmin Qi, and Xuefeng Guo. Direct measurement of single-molecule dna hybridization dynamics with single-base resolution. *Angew Chem Int Ed Engl*, 55(31):9036–9040, Jul 2016. ISSN 1521-3773 (Electronic); 1433-7851 (Linking). doi: 10.1002/anie.201603038.

[191] Xudong Chen, Yan Zhou, Peng Qu, and Xin Sheng Zhao. Base-by-base dynamics in dna hybridization probed by fluorescence correlation spectroscopy. *Journal of the American Chemical Society*, 130(50):16947–16952, 12 2008. doi: 10.1021/ja804628x. URL `https://doi.org/10.1021/ja804628x`.

[192] D. Pörschke, O. C. Uhlenbeck, and F. H. Martin. Thermodynamics and kinetics of the helix-coil transition of oligomers containing gc base pairs. *Biopolymers*, 12(6):1313–1335, 2020/08/06 1973. doi: 10.1002/bip.1973.360120609. URL `https://doi.org/10.1002/bip.1973.360120609`.

[193] Sebastian Sorgenfrei, Chien-yang Chiu, Ruben L. Gonzalez, Young-Jun Yu, Philip Kim, Colin Nuckolls, and Kenneth L. Shepard. Label-free single-molecule detection of dna-hybridization kinetics with a carbon nanotube field-effect transistor. *Nature Nanotechnology*, 6(2):126–132, 2011. doi: 10.1038/nnano.2010.275. URL `https://doi.org/10.1038/nnano.2010.275`.

[194] Shiyan Xiao, Daniel J. Sharpe, Debayan Chakraborty, and David J. Wales. Energy landscapes and hybridization pathways for dna hexamer duplexes. *The Journal of Physical Chemistry Letters*, 10(21):6771–6779, 11 2019. doi: 10.1021/acs.jpclett.9b02356. URL `https://doi.org/10.1021/acs.jpclett.9b02356`.

[195] Michael T. Woodside, William M. Behnke-Parks, Kevan Larizadeh, Kevin Travers, Daniel Herschlag, and Steven M. Block. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. *Proceedings of the National Academy of Sciences*, 103(16):6190, 04 2006. doi: 10.1073/pnas.0511048103. URL `http://www.pnas.org/content/103/16/6190.abstract`.

[196] David Yu Zhang and Erik Winfree. Control of dna strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131(47):17303–17314, 12 2009. doi: 10.1021/ja906987s. URL `https://doi.org/10.1021/ja906987s`.

[197] J. P. Hansen and I. R. McDonald. *Theory of Simple Liquids*. Academic Press, 2006.

[198] Thomas E Ouldridge, Ard A Louis, and Jonathan P K Doye. Extracting bulk properties of self-assembling systems from small simulations. 22(10):104102, 2010. doi: 10.1088/0953-8984/22/10/104102. URL `http://dx.doi.org/10.1088/0953-8984/22/10/104102`.

[199] Frits Dannenberg, Katherine E. Dunn, Jonathan Bath, Marta Kwiatkowska, Andrew J. Turberfield, and Thomas E. Ouldridge. Modelling dna origami self-assembly at the domain level. *The Journal of Chemical Physics*, 143(16):165102, 2020/08/10 2015. doi: 10.1063/1.4933426. URL `https://doi.org/10.1063/1.4933426`.

[200] Katherine E. Dunn, Frits Dannenberg, Thomas E. Ouldridge, Marta Kwiatkowska, Andrew J. Turberfield, and Jonathan Bath. Guiding the folding pathway of dna origami. *Nature*, 525(7567):82–86, 2015. doi: 10.1038/nature14860. URL `https://doi.org/10.1038/nature14860`.

[201] Hanna Vehkamäki. *Classical Nucleation Theory in Multicomponent Systems*. Springer-Verlag Berlin Heidelberg, 2006.

[202] V. A. Chizhikov. Computer simulation of the 2d crystallization of a lipid membrane: Dynamical parquet method. *Crystallography Reports*, 51(1):81–86, 2006. doi: 10.1134/S1063774506010159. URL https://doi.org/10.1134/S1063774506010159.

[203] Craig D Blanchette, Wan-Chen Lin, Christine A Orme, Timothy V Ratto, and Marjorie L Longo. Domain nucleation rates and interfacial line tensions in supported bilayers of ternary mixtures containing galactosylceramide. *Biophysical journal*, 94(7):2691–2697, 04 2008. doi: 10.1529/biophysj.107.122572. URL https://pubmed.ncbi.nlm.nih.gov/18065459.

[204] Siewert J Marrink, Jelger Risselada, and Alan E Mark. Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. *Chem Phys Lipids*, 135(2):223–244, Jun 2005. ISSN 0009-3084 (Print); 0009-3084 (Linking). doi: 10.1016/j.chemphyslip.2005.03.001.

[205] Alexander Yu. Morozov, Robijn F. Bruinsma, and Joseph Rudnick. Assembly of viruses and the pseudo-law of mass action. *The Journal of Chemical Physics*, 131(15):155101, 2020/10/16 2009. doi: 10.1063/1.3212694. URL https://doi.org/10.1063/1.3212694.

[206] Michael F Hagan. Modeling viral capsid assembly. *Advances in chemical physics*, 155:1–68, 2014. doi: 10.1002/9781118755815.ch01. URL https://pubmed.ncbi.nlm.nih.gov/25663722.

[207] Adam Zlotnick, Jennifer M. Johnson, Paul W. Wingfield, Stephen J. Stahl, and Dan Endres. A theoretical model successfully identifies features of hepatitis b virus capsid assembly. *Biochemistry*, 38(44):14644–14652, 11 1999. doi: 10.1021/bi991611a. URL https://doi.org/10.1021/bi991611a.

[208] P. W. Voorhees. The theory of ostwald ripening. *Journal of Statistical Physics*, 38(1):231–252, 1985. doi: 10.1007/BF01017860. URL https://doi.org/10.1007/BF01017860.

[209] Robijn F. Bruinsma, William M. Gelbart, David Reguera, Joseph Rudnick, and Roya Zandi. Viral self-assembly as a thermodynamic process. *Physical Review Letters*, 90(24):248101–, 06 2003. doi: 10.1103/PhysRevLett.90.248101. URL `https://link.aps.org/doi/10.1103/PhysRevLett.90.248101`.

[210] Sahand Hormoz and Michael P. Brenner. Design principles for self-assembly with short-range interactions. *Proceedings of the National Academy of Sciences*, 108(13):5193, 03 2011. doi: 10.1073/pnas.1014094108. URL `http://www.pnas.org/content/108/13/5193.abstract`.