ENS de Lyon

LIP

Doctoral School **ED InfoMaths - 512**

University Department **Laboratoire de l'Informatique du Parallélisme**

Thesis defended by **Pierre Marcus**

Defended on **July 18, 2024**

In order to become Doctor from ENS de Lyon and from LIP

Academic Field **Computer Science**

Scalable DNA algorithms

Thesis supervised by Nicolas Schabanel

Committee members

Referees	Damien Woods	Professor at Hamilton Institute
	Yann Ponty	Professor at LIX
Examiners	Cendrine Moskalenko	Professor at LabPhys
	Raluca Tiron	Professor at CEA LETI
	Jérémie Chalopin	Professor at LIS
Supervisor	Nicolas Schabanel	Professor at LIP

La plupart du temps, les Zertes zertillonnent. Ils courent, ils grimpent, s'empilent, s'effondrent, sautent et recommencent. Le reste du temps, ils font autre chose.
Claude Ponti

ENS de Lyon

LIP

Doctoral School **ED InfoMaths - 512**

University Department **Laboratoire de l'Informatique du Parallélisme**

Thesis defended by **Pierre Marcus**

Defended on **July 18, 2024**

In order to become Doctor from ENS de Lyon and from LIP

Academic Field **Computer Science**

Scalable DNA algorithms

Thesis supervised by Nicolas Schabanel

Committee members

Referees	Damien Woods	Professor at Hamilton Institute
	Yann Ponty	Professor at LIX
Examiners	Cendrine Moskalenko	Professor at LabPhys
	Raluca Tiron	Professor at CEA LETI
	Jérémie Chalopin	Professor at LIS
Supervisor	Nicolas Schabanel	Professor at LIP

ENS de Lyon

LIP

École doctorale **ED InfoMaths - 512**

Unité de recherche **Laboratoire de l'Informatique du Parallélisme**

Thèse présentée par **Pierre Marcus**

Soutenue le **18 juillet 2024**

En vue de l'obtention du grade de docteur de l'ENS de Lyon et de l'LIP

Discipline **Informatique**

Scalable DNA algorithms

Thèse dirigée par Nicolas Schabanel

Composition du jury

Rapporteurs	Damien Woods	professeur au Hamilton Institute
	Yann Ponty	professeur au LIX
Examinateurs	Cendrine Moskalenko	professeur au LabPhys
	Raluca Tiron	professeur au CEA LETI
	Jérémie Chalopin	professeur au LIS
Directeur de thèse	Nicolas Schabanel	professeur au LIP

Remerciements

SCALABLE DNA ALGORITHMS

Abstract

The *DNA computing* field consists in using DNA as dynamic building blocks. By interacting together, they can implement small algorithms and effectively compute.
Many successful approaches were made. For instance, by implementing logical circuits where reconfigurations of DNA complexes progressively evaluate the network. Another approach is to attach DNA strands according to defined rules to a substrate made of large DNA objects called *DNA origami*. However, all the current approaches face the challenge of scalability. In most designs, the size of the input is linked to either the DNA origami or the number of strands. The number of strands, is limited not only technically but also theoretically, as there is an inherent chance of hybridization error between two strands that are not fully complementary.
In this thesis, we want to solve this scalability issue on the particular problem of *maze solving*. This problem was already solved in both in a non-reversible and non-scalable fashion. We propose to implement a reversible random walk walker on a DNA origami. Our point is twofold. First, we can make a design with only four different strands, no matter the size of the maze. Most importantly, using reversibility is a key factor, as it can harness randomness to reverse hybridization errors.
In the first part, we conducted experiments where we attached static paths made of DNA strands on a DNA origami. We will validate our ability to both conduct, observe and process these experiments. In the second part, we propose an implementation of a reversible random walk using a variation of the *toehold mediated strand displacement* technique. We have conducted and developed experiments on this variation using a bottom-up approach. Our experiments led to preliminary results of the technique on a DNA origami.

Keywords: dna computing, dna origami, strand displacemement, scalable, random walk

---

Laboratoire de l'Informatique du Parallélisme
ENS de Lyon, – 46, allée d'Italie, – 69364 LYON CEDEX 07, FRANCE

Contents

Thesis Outline

The *DNA computing* field consists in using DNA as dynamic building blocks. DNA (single) strands can form a *double-helix duplex* when they possess complementary sequences. One can implement binding rules between objects by carefully choosing the set of sequences and then effectively run small algorithms. We can identify two categories of approaches to *DNA computing*. The first approach is the use of a substrate DNA object (*DNA origami*), which is duplicated many times in the experiment solution. The computation state is therefore encoded by the configuration of the *DNA origami*. The computations are run independently on the many origami instances. The *DNA origami technique*[Rot], invented by Paul Rothemund, gives a reliable and flexible way to design and assemble large (about 100 nanometers wide) objects. The second approach to DNA computing is to encode the state of computation in the concentration of DNA complexes. The *toehold mediated strand displacement* technique allows implementing reconfiguration rules between complexes. Despite astonishing results and many experiments known, one of the main challenges in *DNA computing* is its scalability as the size of the input increases. In most designs, the size of the input is related to either the actual design of the *DNA origami*, or the number of strands to mix in the experiment. In the first case, it is impossible to easily modify the origami design. Regarding the second case, the number of strands is limited in practice because of hybridization errors. Indeed, there is an inevitable chance of *hybridization* errors between two strands that are not fully complementary.

In this thesis, we propose to investigate techniques to solve this scalability issue on the particular problem of *maze-solving*. Maze-solving on *DNA origami*, consists in encoding a maze with an entrance and an exit on a 2D *DNA origami*. The goal is to design a DNA system that explores and identifies the path between the entry and the exit. This problem was already solved in a both non-reversible and non-scalable fashion [Cha+19]. In this design, *hairpins*, single strands folded on themselves, dynamically attach to the origami surface, propagating a path on the *DNA origami*. This design produces a non-reversible but random navigator in the maze. It is non-reversible because of the particular use of the *hairpins*, which cannot fold back to their initial state. It is random because at each intersection, there is an equal distance between the different next cells to go, which leads to a random direction taken each time. The navigators, at the end, attach a path on each *DNA origami*, going from the entrance to a dead end. Using magnetic beads, only the "correct" origami, with a path between the entrance and the exit remain in the tube, which solves the maze problem. However, this design is not scalable because, as the maze grows, the concentration of such "correct" *DNA origami* decreases exponentially. Our idea is to use a reversible system to turn this non-scalable design into a scalable one. Indeed, if the navigator were to be able to go back on their path and stop upon reaching the exit, it would have two properties. First, all the navigators on the *DNA origami* would reach the exit. Secondly, the mean time to reach the exit would be quadratic to the maze size. It means that, as the size of the maze grows, the same design will be usable.

The first part of this thesis introduces the reader to DNA nanotechnology. We show how the characteristics of DNA, and the current technology, allow for self-assembling DNA structures of increasingly larger sizes (chapter 1). In a second time (chapter 2), we explore how the rules that allow for the self-assembly of (static) structures can actually be used in dynamic ways, to compute. We focus on two techniques, *tile assembly*[EW17] and *strand displacement*[ZW09]. We also describe how these dynamic systems can be observed, for instance using Atomic force microscopy (AFM) or fluoresence reporting. Finally, after the description of several examples of *DNA computing* algorithms, we discuss the problem of scalability and introduce the solution we propose for the scalable solving of mazes.

The objective of the second part is to validate our origami designs, to statically assemble a path made of several strands on a *DNA origami*, to validate the geometry of our path designs, and to check the ability to visualize a path with the AFM. After the elaboration of 2D origami with a regular grid and paths attached to it (chapter 3), we describe our material and methods (chapter 4), and then present the experimental results (chapter 5). We will assess two versions of *DNA origami* made to avoid torsions. We will successfully assemble a path made of a dozen strands, whose attachments, one to another, are hard-coded. Removing the hard-coding will prove to also assemble. It will show that a set of four "universal" strands that can form any path on a *DNA origami*. Then we will investigate how to control the dynamics of this universal set using the temperature. Because we did not succeed in controlling this assembly, we took another approach, which led to part 3.

In the third and last part, we propose a random walk design using a variation of the *strand displacement* technique. The control of the balance of the random walk is crucial to make the paths reversible. This balance will be tunable through the concentration of the strands in the buffer, and through the *toehold* energies, which is a strong strength of our design. We describe and discuss our design in the first roadmap section (chapter 6). To lead to the full experiment, we design a bottom-up approach that studies our core mechanism, from a simplified system composed of a few strands, to a more complex one supported on a *DNA origami*. The second section is dedicated to the materials and methods (chapter 7) for the experiments we ran. The third section (chapter 8) shows the results for the three *toy models* of the bottom-up approach. The *Toy Model A* develops protocols to balance a two-state system made with *strand displacement*. In particular, we will remove the quencher modification for fluorescence reporting, to avoid an increase in duplexes stability due to direct contact quenching. Our solution uses guanine quenching with fluorescein. We will also use sequence mismatch in the *toeholds* to easily change the energy of the *toeholds* and try to balance the systems. The *Toy Model B* introduces an energy barrier to the *strand displacement* using a small complex made of three strands instead of a *DNA origami*. By observing the behavior of a set of four *toehold* energies, we will validate the effectiveness of *strand displacement*, despite this energy barrier. Finally, the *Toy Model C* is designed to explore the impact of this energy barrier when it is supported on the surface of a *DNA origami*. The preliminary results will not conclude to the effectiveness of *strand displacement* on this *toy model*. However, because of the knowledge gained from the first *toy models*, we will be able discuss the ways one can pursue the experiments of this *toy model C*.

Bibliography of the current part

[Cha+19] Jie Chao et al. “Solving Mazes with Single-Molecule DNA Navigators”. In: *Nature Materials* 18.3 (Mar. 2019), pp. 273–279. ISSN: 1476-1122, 1476-4660. DOI: `10.1038/s41563-018-0205-3`. (Visited on 05/01/2021).

[EW17] Constantine G. Evans and Erik Winfree. “Physical Principles for DNA Tile Self-Assembly”. In: *Chemical Society Reviews* 46.12 (2017), pp. 3808–3829. ISSN: 0306-0012, 1460-4744. DOI: `10.1039/C6CS00745G`. (Visited on 10/09/2023).

[Rot] Paul W K Rothemund. “Folding DNA to Create Nanoscale Shapes and Patterns Supplementary Notes 1–1”. In: (), p. 82.

[ZW09] David Yu Zhang and Erik Winfree. “Control of DNA Strand Displacement Kinetics Using Toehold Exchange”. In: *Journal of the American Chemical Society* 131.47 (Dec. 2009), pp. 17303–17314. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja906987s`. (Visited on 05/01/2021).

Part I

Introduction

Chapter 1

Introduction to DNA Nanotechnology

Outline of the current chapter

This chapter introduces the basic principles of DNA nanotechnology. We will examine how the chemical structure of DNA, physical properties, and thermodynamics allow for the self-assembly of structures with sizes of several orders of magnitudes.

1.1 Chemical Structure of DNA

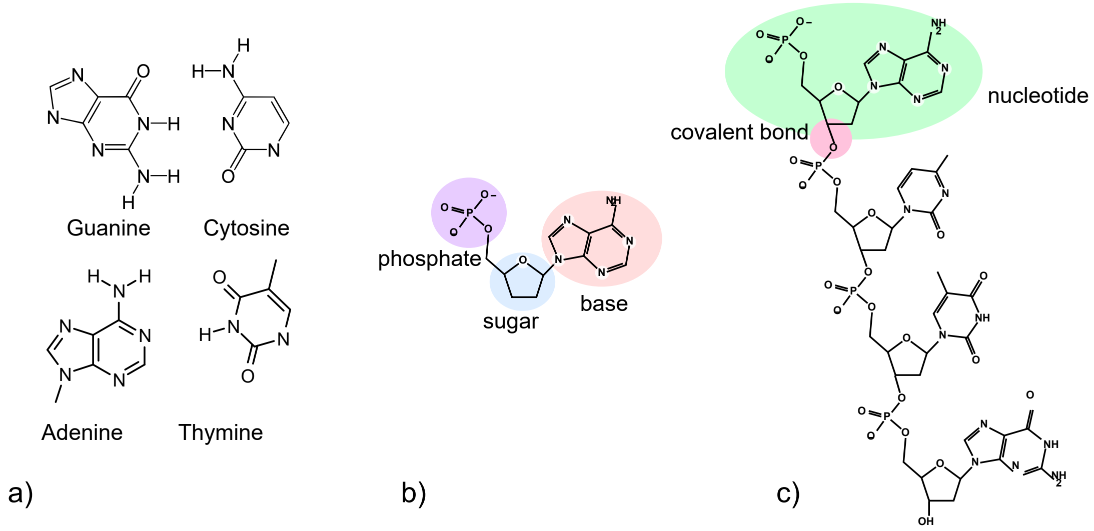
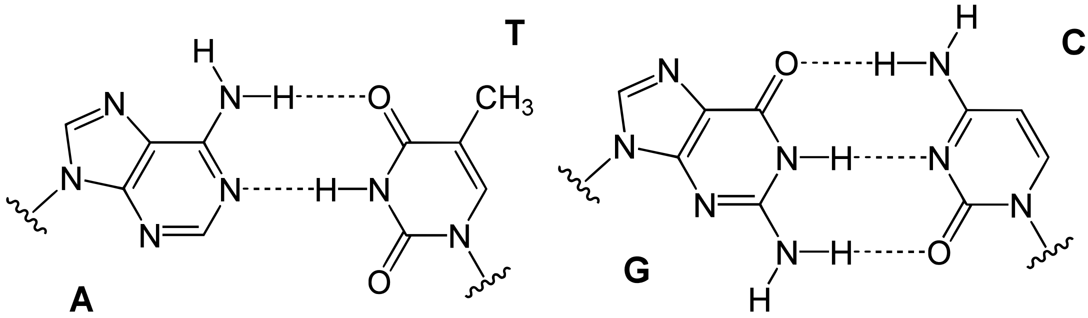


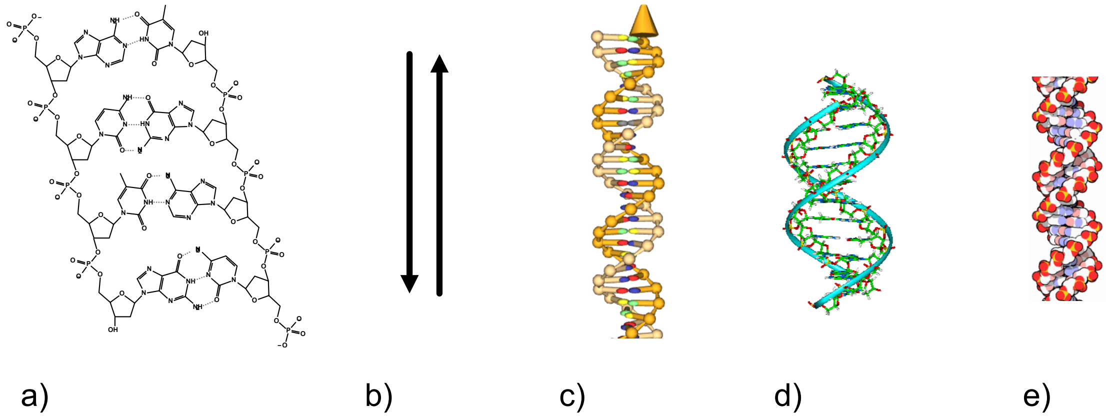
Figure 1.1: The chemical structure of a DNA molecule. Four bases (a) can be included in a nucleotide (b) which are bound together to form a single-stranded DNA (c). Image adapted from by Madprime (CC BY-SA 3.0)

Single-stranded DNA. The DNA molecule (Figure 1.1) is a polymer made of a succession of nucleotides. Each nucleotide is composed of a base, a sugar and a phosphate group. There are four possible bases : Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). By extension, we name each nucleotide by the base it contains. The nucleotides are covalently bonded together between the phosphate group and the sugar. DNA is a directed molecule because of this phosphate-sugar bond. The direction from phosphate to sugar is called 5' to 3'. The sequence of base, usually enumerated from 5' to 3' is called the sequence of the DNA. For instance the sequence 5'-ATGC-3' refers to the DNA with the sequence of bases (Adenine, Thymine, Guanine, Cytosine). When a DNA molecule is alone, we call it single-stranded DNA (ssDNA).

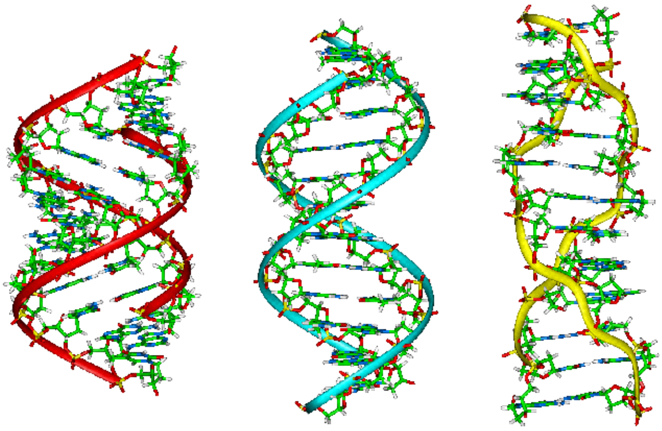
Synthesis and sequencing of DNA The current technology can both easily synthesize and sequence (read) DNA. The cost of DNA sequencing has plummeted over the last few decades. In particular, high-throughput methods such as Illumina of Oxford Nanopores can make readings of about 100$ per billion bases. When it comes to DNA synthesis, progress was also important. However the nature of the tasks is fondamentally different, and more challenging. The synthesis is usually made iteratively by adding one base at each cycle. With an efficiency at each cycle near 99% (depending on the technology), the percentage of (entirel) correct strands drops to about 50% for 80 base-length strands. Further purification can isolate correctly synthesized strands, but it ultimately raises the total cost per base. DNA synthesis (without purification) can be as low as 0.40$ per base. These lowered costs of both synthesis and sequencing drove the progress of the DNA computing field.



(a) Possible hydrogen bonds between bases Adenine Thymine and Guanine Cytosine. The wigly lines represent the connections to the sugars of the nucleotides. The pair A-G can form one more hydrogen bond, which enhances its stability.



(b) Representations of double-stranded DNA. From left to right : **Chemical structure** of a double strand with four nucleotides. **Schematic representation of complementary domains.** Each line represents a DNA strands. The arrow indicates the 5′ to 3′ direction. **3D DNA ball-and-stick representation.** A pair of adjacent ball and stick represents a nucleotide. The inner colored oval ball at the end of a stick represents a base. **Chemical ball-and-stick representation.** The added ribbons represents the sugar-phosphate backbone which are alligned on the phosphate atoms. **Space filling representation.**



(c) Three possible DNA double-helix structures. From left to right, A, B and Z forms. The ribbons represents the phospho-sugar backbone.

Figure 1.2

Double-stranded DNA. DNA can be found in a double-stranded DNA (dsDNA) structure, which has the famous double helix structure. For this to happen, the strands must be arranged in opposit directions, and each pair of facing nucleotides must have an affinity (Figure 1.2b). The base pairs with the most affinity are A-T and G-C because they can form hydrogen bonds (Figure 1.2a). For instance if a strand's sequence is `ATGC`, then it can make double helix structure with a strand with the sequence `GCAT`. `GCAT` is called the reverse complement of `ATGC`. By extension we call *complementary strand*, the strand having the reverse complement sequence The length of a double-stranded DNA in expressed in base pair (bp). Depending of the environment, several double helix structures are possible (Figure 1.2c). The *B-form* is the most common in liquid [FG53].

DNA in practice. After ordering, DNA is conditionned in microtubes. Two forms can be possible, either in a buffer, or dried. A common concentration we usually order is 2.5 nmol, at 100 µM for a volume of 25 µL. At this concentration, in 1 µL (1 mm^3) there are 6×10^{21} individual DNA strands.

Buffers for DNA. Although DNA can be stored in pure salted water, it is common to store it in Tris-EDTA buffer (TE) buffer. The Ethylenediaminetetraacetic acid (EDTA) present in TE buffer is a chelator of magnesium, which means it has high affinity and prevent further reaction when complexed. Its presence inihibits enzymes activity that could digest DNA after a contamination. Buffers with salts like magnesium and sodium are often used when working with dsDNA. Indeed the presence of cation screens the negative charges of the sugar-phosphate backbone, which enhances the stability of the duplex.

DNA in cells. DNA is used as a support for genetic information in cells. However the sizes and the environments are extremely different to what we use in the DNA computing field. First we use strands of 100 bp to 10 kbp. On the other hand, a human chromosome can contains hundreds of million of base-pairs Moreover most of DNA computing technique uses only DNA. Many additionnal proteins packs chromosomal DNA into many level of structures (Figure 1.3). These structures are known modulate the expression of genes, and thus contains additionnal information that are not written in the genetic sequences.

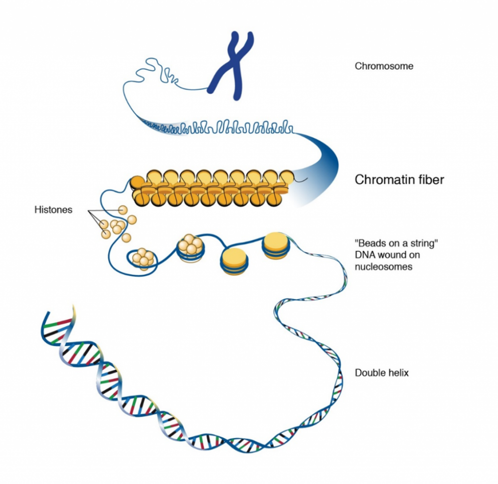
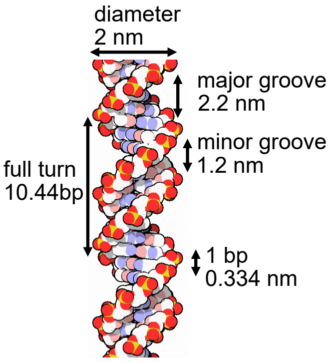
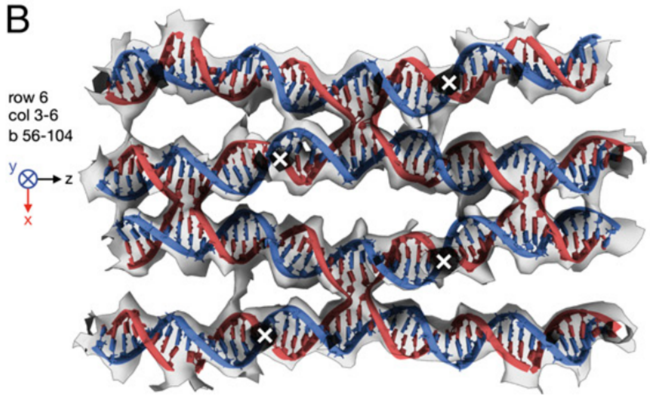


Figure 1.3: Schematic representation of DNA packed into a chromosome (National Human Genome Research Institute)

1.2 Physical Parameters



(a) Space filling representation of a DNA double helix, with size annotations.



(b) Cryo-EM density map of a DNA origami with DNA model fitted on it. [Bai+12]

Figure 1.4

Geometry of double-stranded DNA The double-stranded B-form helix has a radius of 1 nm, and a rise of 0.334 nm per basepair. It takes on average 10.44 bp to make a full turn. Because the two strands are not diametrically opposed, we can observe one minor and one major groove in the helix. *Stacking interactions* between consecutive basepairs induces rigidity to the double helix structure. The persistence length in high salt concentration is around 45nm [Por91], while the persistence length of ssDNA is close to 1 nm [Tin+97]. Figure 1.4b shows a large-scale DNA structure called *DNA origami* imaged with the *cryo-EM* technique. We see that this structure, made of multiple DNA strands uses the precise geometry of DNA to form a mesh.

1.3 Thermodynamics

1.3.1 Particular case of DNA duplex

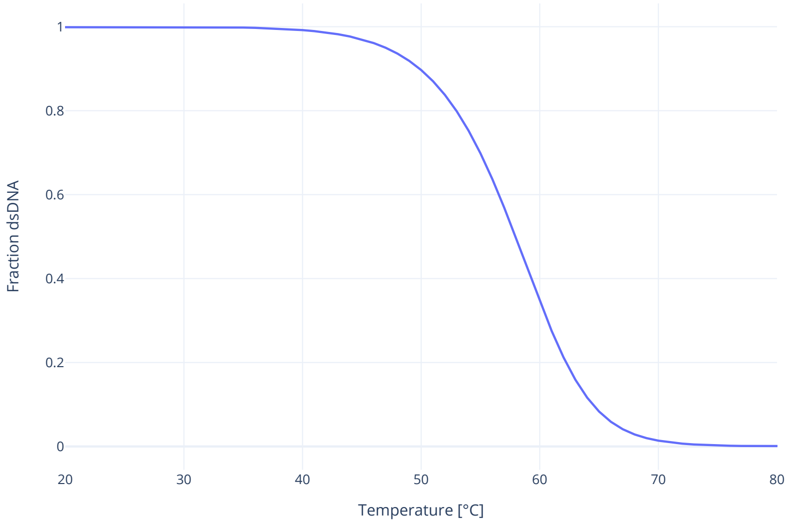


Figure 1.5: Fraction of dsDNA at equilibrium in function of the temperature. Prediction was made with Nupack software with 1 µM of strand with sequence `ATGCAGGATCAGTATA` and 1 µM of its complementary strand, in 12.5 mM Magnesium.

Thermodynamic equilibrium. We illustrate thermodynamics on a pair of complementary strands (denoted A and A*), but it can be generalized to an arbitrary number of strands. Figure 1.5 represents the fraction of dsDNA for a given sequence (concentration of dsDNA divided by the maximal possible concentration of dsDNA). The fraction represented is the one at the thermodynamic equilibrium, which means that the system will converge to this behavior over time. We observe that at high temperatures almost no dsDNA are formed. We say that high temperatures *denature* DNA. Then decreasing the temperature progressively assemble the dsDNA. This distribution can be computed through thermodynamics. The law of mass action states that $\frac{[A][A*]}{[AA*]} = K$ where K is the dsDNA complexation constant. Often more indicative metric is the free energy of the duplex ΔG°, expressed in kcal mol^{-1}. Both constants are related through the formula $K = RT\ln(\Delta G^{\circ})$. The free energy depends on the particular sequence, on possible secondary structures, and salt concentration in the buffer. The temperature at which half the duplex are formed is called the melting temperature T_m.

Melting curves and differences between experimental and thermodynamic equilibrium. The melting temperature can lead back to the free energy through the formula $T_m = -\frac{\Delta G^{\circ}}{R\ln[AA*]_{\text{initial}}/2})$ (with A and A* in equimolar concentrations). It can be measured experimentally by measuring a melting curve (progressive temperature increase starting from assembled duplexes, to complete denaturation), or a hybridization curve (progressive temperature decrease starting from full denaturation). In reality depending of the system, we can question whether or not the empirical melting curve obtained by heating and decreasing the temperature reproduces well the theoretical melting curve. For short oligos, we can expect that the system reach a state close to the thermodynamic equilbrium in few seconds, and that the at this rate no hysteris phenomenon show a difference between the melting curve and the hybridization curve. On the other hand, larger systems, can have more than two states (ssDNA and dsDNA), with kinetics barriers, between them. Kinetics and possible additionnal states (secondary structures) are necessary to grasp these behaviors.

Kinetic modeling. DNA hybridization can modeled using a continuous time markov process, with two states. Between these two states, a there is a **b**inding transition, with an associated bimolecular rate $k_{\mathbf{b}}$ in $\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. There is also an **u**nbinding transition, with an associated unimolecular rate $k_{\mathbf{u}}$ in s^{-1}. The law of detailed balance $k_b/k_u = \exp(-\Delta G^{\circ}/k_B T)$ reunites the kinetic model with the thermodynamic equilibrium. We see that at a given free energy there is a choice to make. In practice the binding rate is often considerer is constant in first approximation [Win06]. Experiments and simulations support this approximation, because most of the decrease of free energy is transfered as an increase of the unbinding rate [Oul+13]. However determining a precise value for these rates is difficult. Indeed these rates actually depends on secondary structures, sequence dependent stacking. The best prediction uses a weight voted neighbour voting algorithm [Zha+] (using four features computed on the sequence) to predict the rates within a factor of 3 with 91% accuracy. A system with more states can be similarly modeled, by respecting the law of detailed balance on every pair of transition. In particular the way the detailed balance is respected can be indepenently chosen on each pair of transition.

Melting temperature and PCR technique. Prediction of duplex energy and melting temperature is particularly crucial in Polymerase Chain Reaction (PCR). PCR is a technique that transformed molecular biology. The goal is to amplify the concentration of a known strand whose sequence extremity is present in solution. This amplification can be used to obtain a high concentration of the desired sequence, for instance to run on a gel, or to do sequencing. Alternatively the amplification in a fluorescent PCR machine, applied on a known range of dillutions can be used to infer the initial concentration in a sample. The princple is to use a polymerase, and short oligos that are complementary to the extremities of the desired sequences called primers. Then the sample is subjects to temperature cycles of hybridization of the primers, extension by the polymerase enzymes, and denaturations by heating. If the temperatures are well chosen, (and that the primer has high affinity only to the sequence of interest), then each cycle will double the concentration. By extenstion, the machine that can cycle the temperatures of the tubes is called a PCR machine. Also the machine which can addition read fluorescence is called a Quantitative Polymerase Chain Reaction (qPCR) machine.

1.3.2 Thermodynamic predictions

Predictions of hybridization are crucial in PCR, where cycles of hybridizations and denaturations are performed. Rough rules for the prediction of the hybridization were established, based on the G/C content of the strands. [SAS96] improves those rules by using a Nearest-neighbor (NN) model. This model and (all further models) is a contribution model where each local pattern add an energetic contribution. The NN model was further refined by adding energies for secondary structures motifs [San98] [SH04].

```
       ↓   ↓   ↓
5′ C-G-T-T-G-A 3′
   * * * * * *
3′ G-C-A-A-C-T 5′
     ↑   ↑
```

$$\Delta G^\circ_{37}(\text{pred.}) = \Delta G^\circ(\text{CG/GC}) + \Delta G^\circ(\text{GT/CA}) + \Delta G^\circ(\text{TT/AA})$$

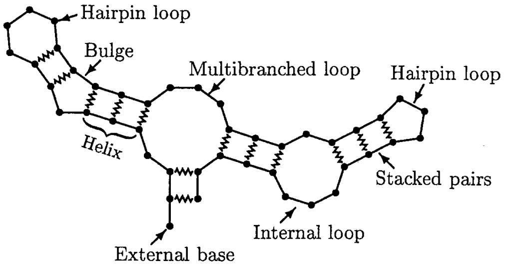
$$+ \Delta G^\circ(\text{TG/AC}) + \Delta G^\circ(\text{GA/CT}) + \Delta G^\circ(\text{init.})$$

$$= -2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.98 + 1.03$$

$$\Delta G^\circ_{37}(\text{pred.}) = -5.35 \text{ kcal/mol}$$

$$\Delta G^\circ_{37}(\text{obs.}) = -5.20 \text{ kcal/mol}$$

(a) Schematic representation of [San98] NN model on two complementary strands.



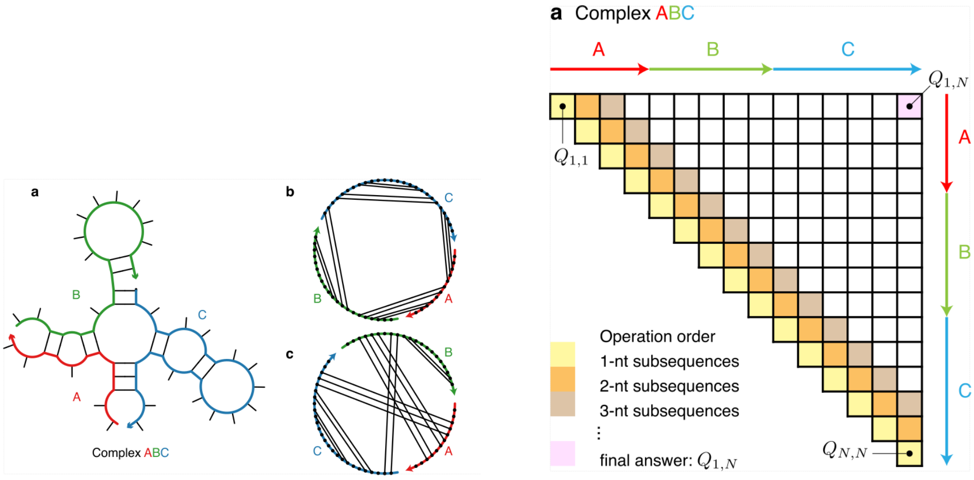
(b) Example of (RNA) secondary structure with labeled motifs, from [LP00a]

Figure 1.6

Nearest-neighbor model and its extension to secondary structures. The Nearest-neighbor model predicts the free energy of a DNA duplex by summing three types of contributions : the initiation energy (constant for all sequences), the symmetry penalty if the duplex is self-complementary, the neighbors energies (by pairs), and the terminal A/T penalty (accounting for base fraying) [San98]. To predict the free energy ΔG° at a given temperature, ΔG° can be decomposed as $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$. This must also true for the energies asssociated to the individual local patterns. The goal is therefore to measure ΔH° and ΔS° for each local pattern. The thermodynamic parameters of the model were fitted based on measurements of 108 sequences by UV melting. The validity was tested against a different set of 264 sequences, leading for a standard deviation of the prediction of the melting temperature of 2.3°C. The presence of salt improves the stability of dsDNA. To account for it, phenomenological salt correction are established [San98] [SH04] [DZ04] [KP05], which treat the magnesium the same ways as sodium (up to concentrations). [SH04] improved the initial NN model with motifs like loop, hairpin loops, internal mismatch, terminal mismatch, and bulges. This allows for evaluation of a large set of secondary structures.

UV melting. UV melting is a technique to measure the melting curve of oligos. It consists of melting double-stranded DNA by progressively raising the temperature. It particularity is to quantifiy the amount of dsDNA and ssDNA by measuring the UV absorbance of the sample. Indeed UV absorbance in different for dsDNA and ssDNA.

Secondary structures. Secondary structures refers to the set of (hybridized) basepairs between two or more strands. The extended NN model allows to consider secondary structures, instead of only perfectly complementary strands. In this context, we call a complex a set of strands, which are all assembled by some basepairs. Secondary structures of a complex can be thought of micro-states of statistical physics, for which there is a distribution that follows the Boltzmann equation.



(a) An example of a DNA complex without pseudoknot. Two strands ordering (ABC and ACB) for the associated polymer graph are represented. The ordering ABC proves that this structure does not contains pseudoknots. [FPP20].

(b) Graphic decomposition of the computation of thermodynamics of an ordered complex. Each "cell" computation requires the knowledge of the cells on the other side of the corresponding diagonal. This allows the use of dynamic programming. [FPP20]

Figure 1.7

Algorithms and softwares. Prediction the most probable secondary structure is challenging because of the exponential number of configurations of a given complex. For instance, the prediction of the secondary structure according to this model is NP-complete in the general case [LP00b]. This is why it is common to restrict to pseudoknot-free structures which lead to dynamic algorithms with polynomial time complexities.

Figure 1.7b, describes the dynamic subsequence algorithm, where each cell corresponds to a linear or quadratic time complexity depending on the particular implementation, resulting in N^4 or N^3 time algorithms [Dir+07] [DP03] [San98]. Nupack

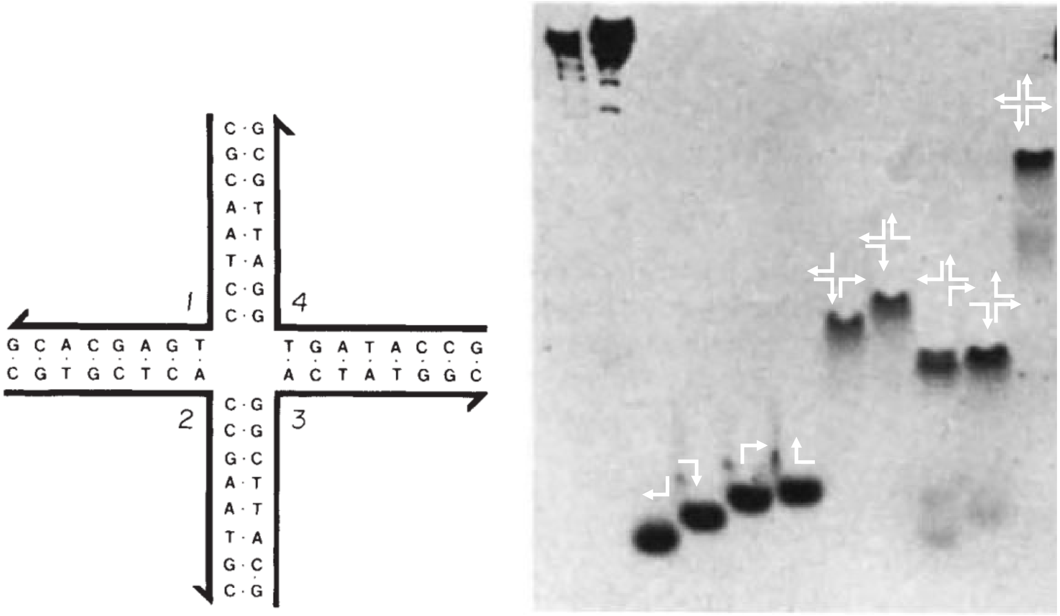
[FPP20] and ViennaRNA [Lor+11] are two widely used implementations Both possess an additionnal large range of features. In particular they can also compute the energy of a particular structure, the structure having the Minimum Free Energy (MFE) and the basepairing probability matrix.

Pseudoknot-free secondary structures. The absence of pseudoknot in a secondary structure can be characterized in two ways : **Dot-parens** Using a well-parenthesized string representation which expresses the basepairing in 5' to 3' order. There are two additionnal symbols `.` for unpaired bases and + to separate strands. For instance a dot-parens reprensentation for two complementary strands of length 4 is `((((+))))`. If the last nuclotide of the first strand is unpaired, it becomes `(((.+.)))`. The second characterization is the existence of a the existence of a **planar polymer graph**. It consists in setting a graph with nucleotides as nodes dispered on a circle, and edges for basepairing. We can remark that both characterisations works by the existence of a correct representation for a given ordering of the strands. For instance, some wrong ordering of a pseudoknot-free might not exhibit pseudoknot-free representation. It can be shown that valid permutations are shared between the two characterizations.

1.4 Towards Self Assembled Structures

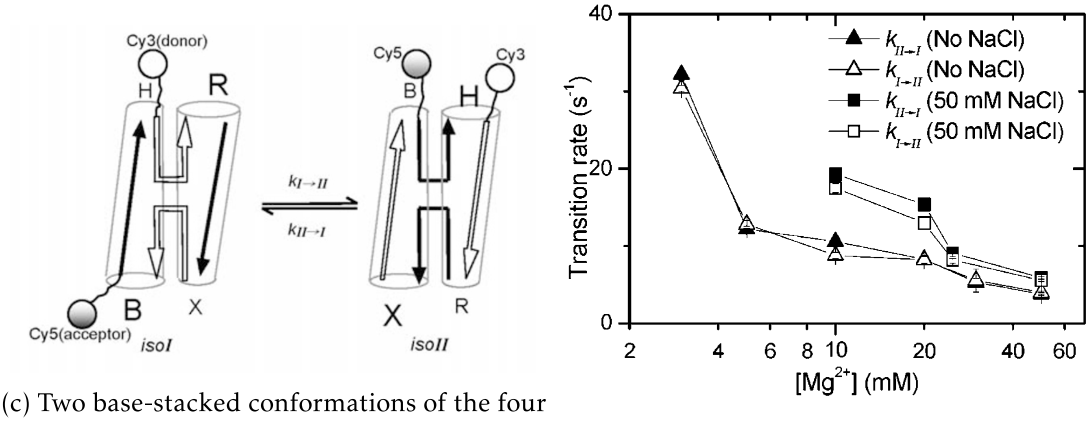
We have seen previously the formation of double-stranded helix when two complementary strand meet. In particular we discussed the fact that this was generally fast, and that at room temperature all the duplexed are formed (for not too small oligos). Simlarly structures made of more than two different strands can be designed. Then the simple mixing, denaturation by heating, and progressive decreasing of the temperature can assemble the desired structure. This phenomenon is called self-assembly. Self-assembly of large structure can be difficult : geometrical constraints of DNA must be satisfied . Also unintended basepairing must be limited. Even in this case, the assembly of a large structure can be a slow process, and kinetic barriers must be controlled. In this section we present four examples that are relevent for understanding DNA computing : the assembly of a four way DNA junction, the crystalization of DNA tiles, the DNA Origami technique, and larger structures.

Holliday Junction. Ned Seeman goal was to perform protein crystallography, where proteins are attached on a DNA lattice [See82]. The first step of this journey was to synthesis an immobile Holliday junction (or four way junction) (Figure 1.8a). The immobility is to be compared to mobile holliday junction happening during the homologous recombination of chromosomes. Figure 1.8b shows an acrylamide gel, which discriminates structures according to their size and conformation. The high molecular weight of the last lane proves the assembly of the four strands structures. In presence of high salt concentrations one of two of the four strands are likely to form straight helices because of base stacking. Two conformations are therefore possibles (Figure 1.8c). Recent work have studied individual dynamic of such conformations [McK+03]. Even if the fourway junction is immobile in a base pairing sense, there is a constant transition between the two conformations. At 25°C there are from 30 transitions per seconds in 3 mM of Magnesium to 5 transitions per second at 5 mM of Magnesium.



(a) Sequence of [KMS83] for immobile four way junction

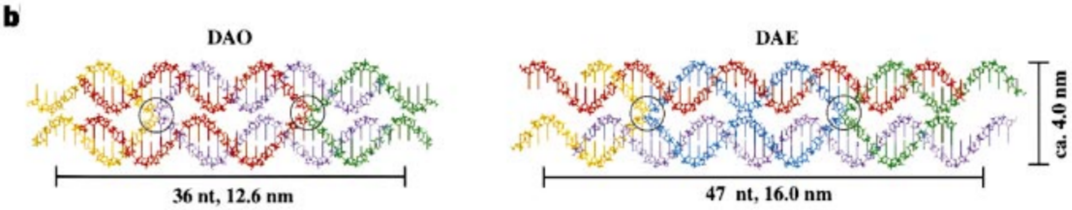
(b) Acrylamide gel of the self assmebled four way junction structure [KMS83]



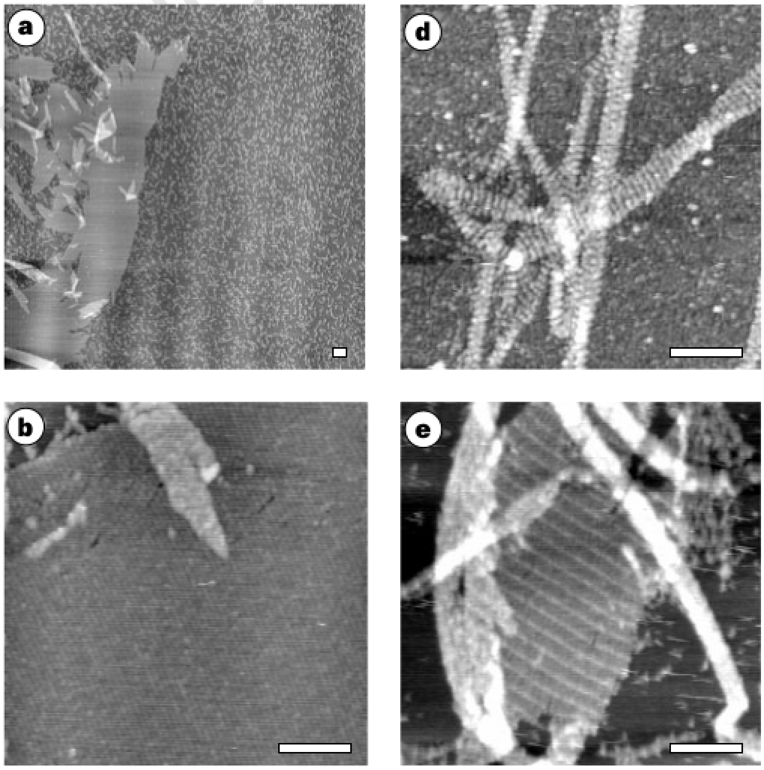
(c) Two base-stacked conformations of the four ways junction, included a fluorescent reporting, from [McK+03]

(d) Rates of conformation transitions at 25°C, from [McK+03]

Figure 1.8



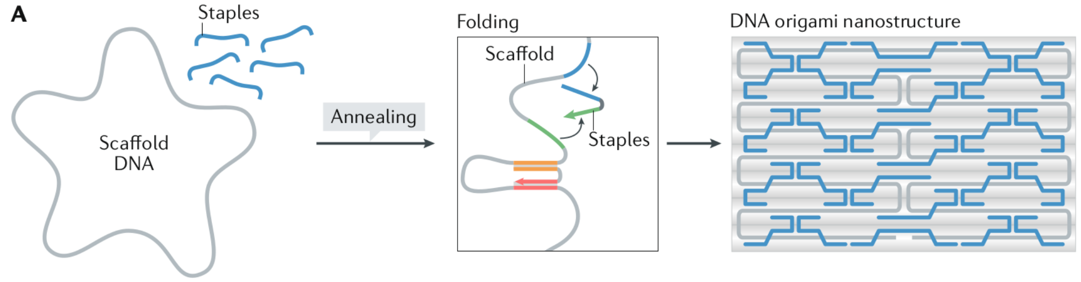
(a) Two types of tiles designs



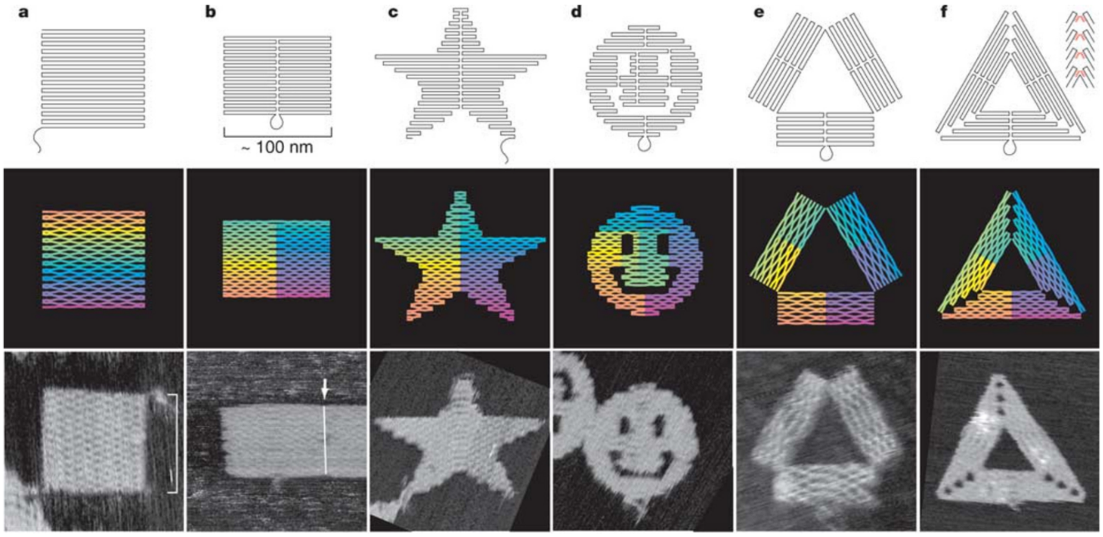
(b) AFM imaging of tile crystalization. 300nm scale bars.

Figure 1.9: DNA crystal made from DNA tiles [Win+98]

Tile Assembly. In tile assembly, tiles are made of small DNA complexes (Figure 1.9a). The four single-stranded ends, called sticky ends can be designed so that they assemble according to specific rules (Figure 1.9). DNA self assembly has computing power, linked to the Wang tiling. Examples of successful experiment are Sierpinsky motifs [RPW04], ,binary counter [Eva] or iterations of a cellular autmata [Woo+19]. However the technique has a inherent possibilities of error, and optimization have to be found to favor a growth without misplaced tiles. In the next chapter we explains the thermodynamical considerations of tile assembly.



(a) Assembly process of a DNA origami using a temperature annealing [Dey+21].



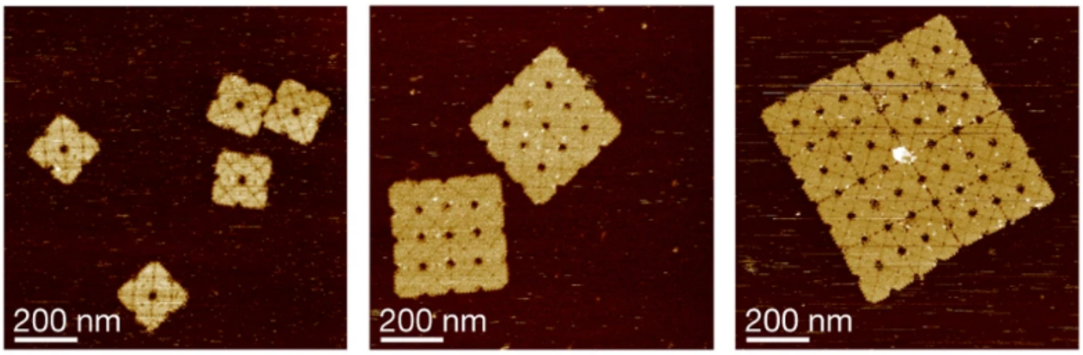
(b) Various DNA origami designs[Rot].

Figure 1.10: DNA origami

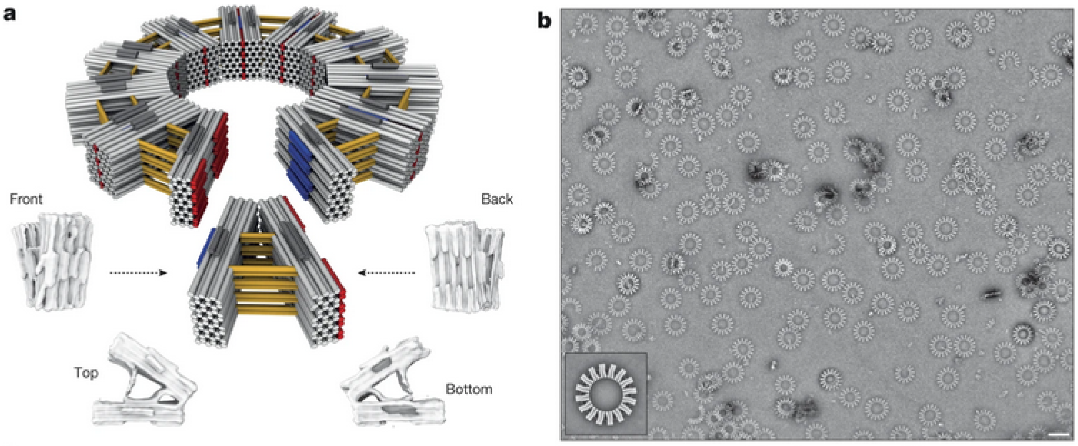
DNA Origami. DNA Origami [RPW04] is a technique invented by Paul Rothemund which allow make large 3D structure made of dsDNA. It is composed of one long circular single strande of about 7000 nucleotides, called scaffold. The scaffold is folded into a desired shape using about 200 small strands called staples. Each staple is complementary by part to the scaffold (Figure 1.10a) It has many advantages which explains its extensive use in DNA computing. The technique has often high yield and the conception of an origami is relatively easy. Also it can the functionnalized by attaching proteins or making single-strands go out of the origami.

Structures of size of higher magnitude DNA origami can be assembled together to maked structures of higher sizes of magnitude. Two principles can be used to make interactions between origami. First, the use of sticky ends, similarly to the sticky ends on DNA tiles. A second way is the use of *pi-stacking*. We saw in the NN model *stacking*, the interaction between consecutive nucleotides imposes rigidity to the double helix. Some of this stacking can also occurs between two nucleotides that are not bonded together in a same DNA strand. Cross-overs of DNA origami exposes the section of helices, where such *pi-stacking* occurs. By controlling the possibity of makesstacking between DNA origamis, one can impose rules of assembly between DNA origamis. Figure 1.11 shows three examples of such higher magnitude structures.

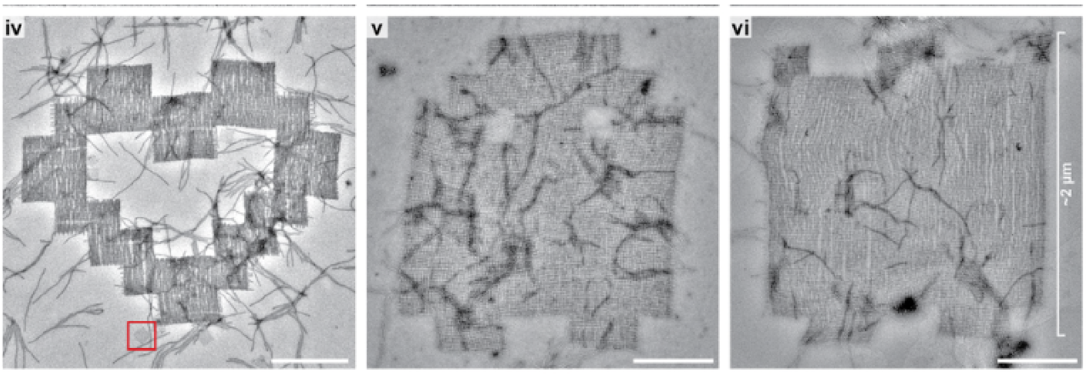
. Structures with higher magnitudes sizes were successufly acheived, mostly by assembling toghether Origamis using specific rules up to micrometer scale



(a) Assembly of square DNA origami into micrometer scale squares. The edges of the origamis uses a combination of both pi-stacking, an two nucleotides hybridization. A staged assembly, where each step *n* mixes squares with *n* origami wide. [TPQ17]



(b) Self-assembly of DNA origamis. This self assembly is inspired by the assembly proteins to form viral capsids. 3D designs can assemble sphere-like shapes involving 50,000 nucleotides. [WSD17]



(c) Self assembly of DNA origamis to form micrometer scale shapes. Slat origamis (line shaped), that are assembled using sticky ends binding. The particularity is to use a DNA origami to nucleate the assembly. This allows for the assembly of these large scale structures without a manual staged assembly. [Win+23]

Figure 1.11

In this chapter we saw that DNA the complementary nature of DNA makes it a convenient programmable material. The thermodynamics of DNA are known, and it allows for some broad understanding of the kinetics. The phyisical properties of DNA, its persistence length, and its double0helix geometry allows for self-assembling structures, from dozens of nucleotides to several thousands.

Bibliography of the current part

[Bai+12] Xiao-chen Bai et al. "Cryo-EM Structure of a 3D DNA-origami Object". In: *Proceedings of the National Academy of Sciences* 109.49 (Dec. 2012), pp. 20012–20017. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1215713109. (Visited on 11/01/2023).

[Dey+21] Swarup Dey et al. "DNA Origami". In: *Nature Reviews Methods Primers* 1.1 (Jan. 2021), p. 13. ISSN: 2662-8449. DOI: 10.1038/s43586-020-00009-8. (Visited on 12/24/2023).

[Dir+07] Robert M. Dirks et al. "Thermodynamic Analysis of Interacting Nucleic Acid Strands". In: *SIAM Review* 49.1 (Jan. 2007), pp. 65–88. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/060651100. (Visited on 05/24/2021).

[DP03] Robert M. Dirks and Niles A. Pierce. "A Partition Function Algorithm for Nucleic Acid Secondary Structure Including Pseudoknots". In: *Journal of Computational Chemistry* 24.13 (Oct. 2003), pp. 1664–1677. ISSN: 01928651, 1096987X. DOI: 10.1002/jcc.10296. (Visited on 11/03/2021).

[DZ04] Roumen A. Dimitrov and Michael Zuker. "Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids". In: *Biophysical Journal* 87.1 (July 2004), pp. 215–226. ISSN: 00063495. DOI: 10.1529/biophysj.103.020743. (Visited on 05/25/2021).

[Eva] Constantine Evans. "Crystals That Count!" In: ().

[FG53] Rosalind E. Franklin and R. G. Gosling. "Molecular Configuration in Sodium Thymonucleate". In: *Nature* 171.4356 (Apr. 1953), pp. 740–741. ISSN: 1476-4687. DOI: 10.1038/171740a0. (Visited on 01/06/2024).

[FPP20] Mark E. Fornace, Nicholas J. Porubsky, and Niles A. Pierce. "A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed". In: *ACS Synthetic Biology* 9.10 (Oct. 2020), pp. 2665–2678. ISSN: 2161-5063, 2161-5063. DOI: 10.1021/acssynbio.9b00523. (Visited on 05/01/2021).

[KMS83] Neville R. Kallenbach, Rong-Ine Ma, and Nadrian C. Seeman. "An Immobile Nucleic Acid Junction Constructed from Oligonucleotides". In: *Nature* 305.5937 (Oct. 1983), pp. 829–831. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/305829a0. (Visited on 12/25/2023).

[KP05] R. T. Koehler and N. Peyret. "Thermodynamic Properties of DNA Sequences: Characteristic Values for the Human Genome". In: *Bioinformatics* 21.16 (Aug. 2005), pp. 3333–3339. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/bti530. (Visited on 01/06/2024).

[Lor+11] Ronny Lorenz et al. "ViennaRNA Package 2.0". In: *Algorithms for Molecular Biology* 6.1 (Dec. 2011), p. 26. ISSN: 1748-7188. DOI: `10.1186/1748-7188-6-26`. (Visited on 01/16/2024).

[LP00a] Rune B. Lyngsø and Christian N. S. Pedersen. "RNA Pseudoknot Prediction in Energy-Based Models". In: *Journal of Computational Biology* 7.3-4 (Aug. 2000), pp. 409–427. ISSN: 1066-5277, 1557-8666. DOI: `10.1089/106652700750050862`. (Visited on 01/06/2024).

[LP00b] Rune B. Lyngsø and Christian N. S. Pedersen. "RNA Pseudoknot Prediction in Energy-Based Models". In: *Journal of Computational Biology* 7.3-4 (Aug. 2000), pp. 409–427. ISSN: 1066-5277, 1557-8666. DOI: `10.1089/106652700750050862`. (Visited on 01/08/2024).

[McK+03] Sean A. McKinney et al. "Structural Dynamics of Individual Holliday Junctions". In: *Nature Structural Biology* 10.2 (Feb. 2003), pp. 93–97. ISSN: 10728368. DOI: `10.1038/nsb883`. (Visited on 09/12/2023).

[Oul+13] Thomas E. Ouldridge et al. "DNA Hybridization Kinetics: Zippering, Internal Displacement and Sequence Dependence". In: *Nucleic Acids Research* 41.19 (Oct. 2013), pp. 8886–8895. ISSN: 1362-4962, 0305-1048. DOI: `10.1093/nar/gkt687`. (Visited on 06/24/2021).

[Por91] Dietmar Porschke. "Persistence Length and Bending Dynamics of DNA from Electrooptical Measurements at High Salt Concentrations". In: *Biophysical Chemistry* 40.2 (May 1991), pp. 169–179. ISSN: 03014622. DOI: `10.1016/0301-4622(91)87006-Q`. (Visited on 01/16/2024).

[Rot] Paul W K Rothemund. "Folding DNA to Create Nanoscale Shapes and Patterns Supplementary Notes 1–1". In: (), p. 82.

[RPW04] Paul W. K Rothemund, Nick Papadakis, and Erik Winfree. "Algorithmic Self-Assembly of DNA Sierpinski Triangles". In: *PLoS Biology* 2.12 (Dec. 2004). Ed. by Anne Condon, e424. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.0020424`. (Visited on 10/09/2023).

[San98] J. SantaLucia. "A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics". In: *Proceedings of the National Academy of Sciences* 95.4 (Feb. 1998), pp. 1460–1465. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.95.4.1460`. (Visited on 05/23/2021).

[SAS96] John SantaLucia, Hatim T. Allawi, and P. Ananda Seneviratne. "Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability". In: *Biochemistry* 35.11 (Jan. 1996), pp. 3555–3562. ISSN: 0006-2960, 1520-4995. DOI: `10.1021/bi951907q`. (Visited on 01/05/2024).

[See82] Nadrian C. Seeman. "Nucleic Acid Junctions and Lattices". In: *Journal of Theoretical Biology* 99.2 (Nov. 1982), pp. 237–247. ISSN: 00225193. DOI: `10.1016/0022-5193(82)90002-9`. (Visited on 02/23/2024).

[SH04] John SantaLucia and Donald Hicks. "The Thermodynamics of DNA Structural Motifs". In: *Annual Review of Biophysics and Biomolecular Structure* 33.1 (June 2004), pp. 415–440. ISSN: 1056-8700, 1545-4266. DOI: `10.1146/annurev.biophys.32.110601.141800`. (Visited on 05/23/2021).

[Tin+97] Bernard Tinland et al. “Persistence Length of Single-Stranded DNA”. In: *Macromolecules* 30.19 (Sept. 1997), pp. 5763–5765. ISSN: 0024-9297, 1520-5835. DOI: `10.1021/ma970381+`. (Visited on 01/16/2024).

[TPQ17] Grigory Tikhomirov, Philip Petersen, and Lulu Qian. “Fractal Assembly of Micrometre-Scale DNA Origami Arrays with Arbitrary Patterns”. In: *Nature* 552.7683 (Dec. 2017), pp. 67–71. ISSN: 1476-4687. DOI: `10.1038/nature24655`. (Visited on 01/18/2024).

[Win+23] Christopher M. Wintersinger et al. “Multi-Micron Crisscross Structures Grown from DNA-origami Slats”. In: *Nature Nanotechnology* 18.3 (Mar. 2023), pp. 281–289. ISSN: 1748-3395. DOI: `10.1038/s41565-022-01283-1`. (Visited on 04/03/2024).

[Win+98] Erik Winfree et al. “Design and Self-Assembly of Two-Dimensional DNA Crystals”. In: *Nature* 394.6693 (Aug. 1998), pp. 539–544. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/28998`. (Visited on 12/25/2023).

[Win06] E. Winfree. “Algorithmic Self-Assembly of DNA”. In: *2006 International Conference on Microtechnologies in Medicine and Biology*. Okinawa: IEEE, May 2006, pp. 4–4. ISBN: 978-1-4244-0337-0 978-1-4244-0338-7. DOI: `10.1109/MMB.2006.251471`. (Visited on 10/09/2023).

[Woo+19] Damien Woods et al. “Diverse and Robust Molecular Algorithms Using Reprogrammable DNA Self-Assembly”. In: *Nature* 567.7748 (Mar. 2019), pp. 366–372. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/s41586-019-1014-9`. (Visited on 02/06/2022).

[WSD17] Klaus F. Wagenbauer, Christian Sigl, and Hendrik Dietz. “Gigadalton-Scale Shape-Programmable DNA Assemblies”. In: *Nature* 552.7683 (Dec. 2017), pp. 78–83. ISSN: 1476-4687. DOI: `10.1038/nature24651`. (Visited on 01/08/2024).

[Zha+] Jinny X Zhang et al. “Predicting DNA Hybridization Kinetics from Sequence”. In: (), p. 10.

Chapter 2

DNA Computing

Outline of the current chapter

In this chapter we describe how experiments with DNA system can actually compute. First we describe the process of making DNA origami, which are often used as a substrate for computations. Then we describe how dynamic mechanisms can be implemented with DNA using tile assembly and strand displacement techniques. We discuss techniques used to observe the result of the computations. Finally we point out the scalability issues of the current state of DNA computing.

2.1 The process of Making DNA Origami

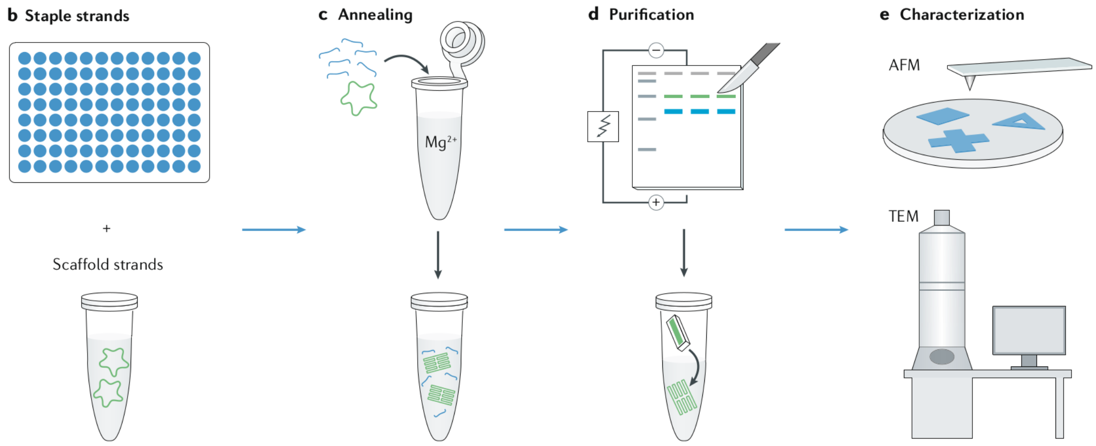


Figure 2.1: Process of making DNA origamis [Dey+21].

Process to make a DNA origami The making of a DNA origami is a 6 steps process :

1. **Design.** DNA origami are designed with the help of a software. These softwares let the user first choose the folding of the scaffold strand, and then define the topology of the staples that will hopefully keep the scaffold folded as desired. CadNano and Scadnano (web-based rewriting of CadNano) uses only a 2D view, where squared or honeycomb lattice can be attached. However the 3D geometry of the designd structure is not easily accessible. Because of it cross-over lengths are not visible. Because these softwares only worked a single grid, it was also difficult to acheive design and simulate 3D designs. To overcome these issues the software ENSNano[LS], proposes a split view of both a 2D grid representation, and a 3D view of the DNA helices.

2. **(Optional) Simulations.** Simulations can be performed to assess if the structure is twisted, or if the staples are maintaining the baispairings. Different simulation methods are available, from coarses grain model (CanDo, MrDNA [MA20]) to molecular dynamic (OxDNA [Sen+21]).

3. **Staples synthesis.** Although no strand purification is necessary, the high number of strands (around 200) totals to near 700$.

4. **Anneal the origami.** Mixing the staples and the scaffold in the appropiate buffer and proceed to an annealing. The standard excess of staples is 10X, but can be reduced to 2X in some cases [RPW04]. The scaffold concentration can vary from 2nM to 40nM. There is no absolute rule for determining the annealing ramps.

5. **(Optional) Assess the folding.** The verification of correct assembly can be assesed by running an agarose gel. Nanoscoping imaging techniques like AFM or Transmission electron microscopy (TEM) can be used.

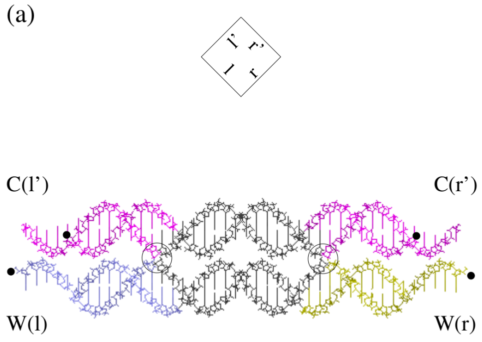
6. **Purification.** The purification is used to separate the excess staples from the origami, and remove the misshaped origami. Different methods are available

among agarose gel electrophoresis, centrifugation filter, or Polyethylene glycol (PEG) precipitation [Wag+17]

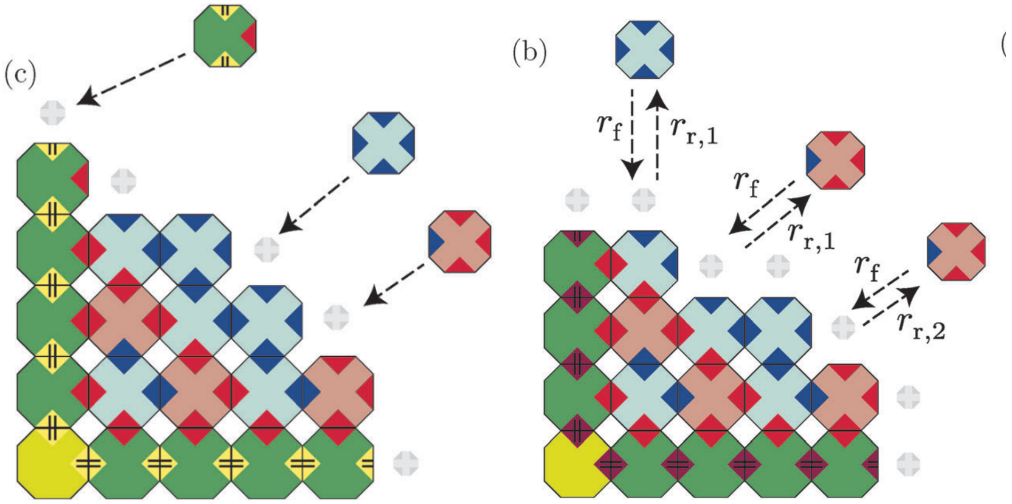
2.2 Implementing dynamic behavior with DNA

In this section we ellaborate on two mechanisms able to implement dynamic rules for DNA systems. The first one is tile assembly, that we saw in the previous chapter. The second is strand displacement.

2.2.1 Tile Assembly



(a) DNA implementation of a Wang tile [Win06]



(b) One step of the Sierpinsky assembly in aTAM at $\tau = 2$[EW17]

(c) One step of the Sierpinsky tiling assembly in κTAM [EW17]

Figure 2.2

DNA Gel electrophoresis DNA Gel electrophoresis is a technique used to separate or analyse the DNA fragments of a sample. The sample is charged on a gel matrix, on which an electric field is applied. Negatively charged molecule, such as DNA, migrate towards the positive electrode. At the end molecules are spatially sorted by fragment sizes. There are two common types of gels matrices for DNA, agarose or acrylamide. Agarose is more suited for DNA fragments of several hundreds of basepairs. Smaller fragments are better seperated with acrylalide gel. Once run, the gels can be imaged to analyse the sample. Then bands of interest can also be cut and extracted to be further used in experiments.

Kinetic of tile assembly and modelisations. abstract tile assembly model (aTAM) is the simplest model of tile assembly. Each side of a tile has a single-stranded sticky end, either weak (accounting for one bond) or strong (accounting for two bonds). We observe all the configurations that are reachable using a succession of assembly steps. At each assembly step, one tile may attach to existing tiles, only if the sum of bonds is greater than τ (refered as the temperature). $\tau = 2$ is the most useful model in practice, and allow for theoretical simulation of arbitrary Turing Machine [Win06]. The kinetic tile assembly model (kTAM) is closer to the real physical process. It describes the assembly over time, with rates for assembly and disassembly. The assembly rate r_f (in s^{-1}) is depends only on the concentration of the incoming, no matter the number of bonds made. The detachement rate depends of the number of correct bonds made $r_{r,b}$. Because of it, misplaced tiles are supposed to detach fast. Equation 2.1 defines both rates. k_f is the bimolecular forward rate (in $\mathrm{s}^{-1}\,\mathrm{M}^{-1}$). $[c]$ is the tile concentration. ΔG_{se} is the free energy of single bond. α is a constant accounting for the loss of rotationnal entropy.

$$r_f = k_f[c] \text{ and } r_{r,b} = k_f exp(b\Delta G_{se})/RT + \alpha \tag{2.1}$$

[Eva].

The analysis of the system is made easier by introducing two dimensionless energy : $G_{se} = \Delta G_{se}/RT$ and G_{mc}, and $\hat{k_f} = k_f e^{\alpha}$ such that $[c] = \exp(-G_{mc}) + \alpha$, so that the two rates have unform respresentation.

$$r_f = \hat{k}_f \exp(-G_{mc}) \text{ and } r_{r,b} = \hat{k}_f \exp(-bG_{se})$$

.

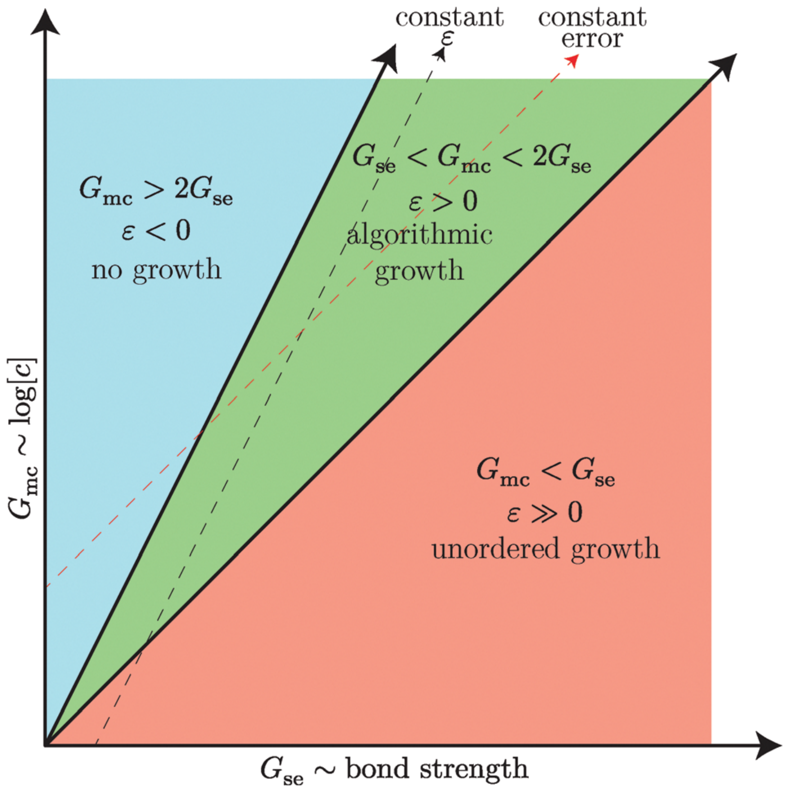


Figure 2.3: κTAM phase diagram of growth and error compared to ATAM ($\tau = 2$)

Control of the assembly error. We express conditions on G_{se} and G_{mc} to compare κTAM and ATAM ($\tau = 2$)behaviors. We focus on a site where a tile can attach with two bonds. If $G_{mc} = 2G_{se}$, then the rates of attachment and detachement are equal. This means hat there is no preferable growth direction. We can deduces three regimes : **An inhibited growth regime** when $G_{mc} > 2G_{se}$. **A correct growth regime** when $G_{se} < G_{mc} < 2G_{se}$. Weakly attached tiles detach at a rates faster than global the tile attachement rate. On the other hand, strongly attached tiles detach at rate smaller than the attachement rate. **An unordered growth** when $G_{mc} < G_{se}$, as all tiles attach faster than they detach. The presence of error when tuning the parameters decrease significantly the assembly time, because of the need to disassemble errors. The *kinetic trap model* [Win06] formalizes this idea and show that near equilibrium, the growth rate evolves quadratically with the error rate. This implies that smarter tile sets that circunvent error are necessary to be able to assemble large crystals without errors. A strategy is to use *proofreading* tiles where one original tile is subdivided into four smaller tiles [WB04].

2.2.2 Strand Displacement

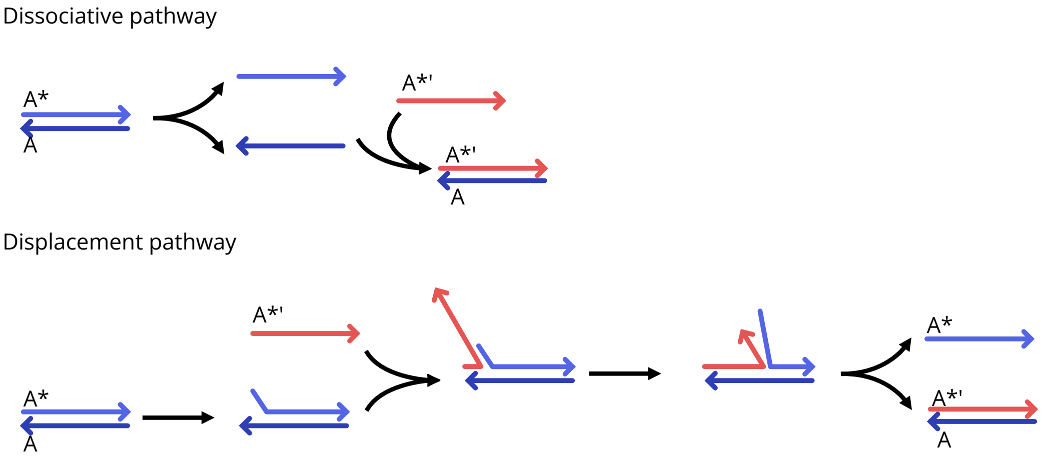


Figure 2.4: Two pathways occuring in strand replacement experiments [Rey+00]. A*, and A*' are two distinguishable strands, both complementary to A.

Strand Replacement and displacement. Strand displacement occurs when two strand compete for hybridization on a third substrate strand. A branch migration motif appears, which migrates randomly over the subtrate strand. At then end of the two strands ends up unziping completely the other one (Displacement pathway of Figure 2.4). Strand replacement experiments exhibit this phenomenon[Rey+00]. One oligo of a duplex is replaced by an excess of a third oligo having same sequence (but distinguishable through fluorescence reporting). Two pathways compete (Figure 2.4). A **dissociation pathway**, where the initial duplex dissociates due to thermal fluctiations, followed by a hybridization of the other oligo. A **displacement pathway** where strand displacement actually occurs. A high temperature, the dissociation pathway predominates over strand displacement while it is the contrary at low temperature. In particular the bimolecular displacement rates for respectively 12 nt, 14 nt and 16 nt 16nt oligo were measured to be from $4\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ to $13\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$.

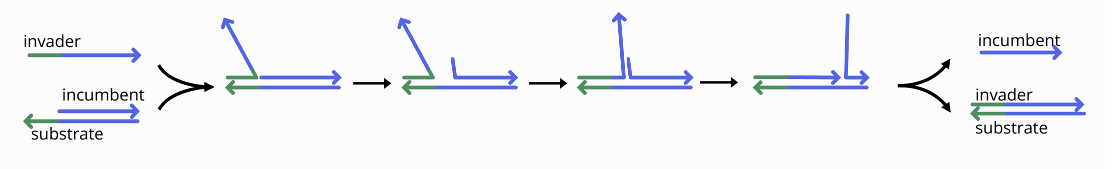


Figure 2.5: Transitions of toehold mediated strand displacement. Some key steps of unimolecular tranisitions are represented.

toehold mediated strand displacement (TMSD). toehold mediated strand displacement (TMSD) (Figure 2.5),is a technique where strand displacement initiated by the presence (often 6 nt long) single-stranded domain called *toehold*. It usually works at room temperature, to reduce spontaneous dissociations pathways we have seen earlier. We describe the process over time. The *substrate strand* is initially duplexed with an *incumbent strand*. The *invader strand* may to bind its *toehold* to the complementary one of the substrate. At this point, a strand displacement process occurs. The branch migration driven by a succession of basefraying and binding. Finally, either the invader wins the displacement, or it comes back at the toehold binding step. When toeolds are longs, the invader may not unbind, and a new branch migration is initiated. On the other hand for small toeholds, it has chances to either unbind or initiate a new branch migration. We see that toeholds *mediates* and initiate the branch migration. Also it biases the branch migration random process in the displacemetn direction.

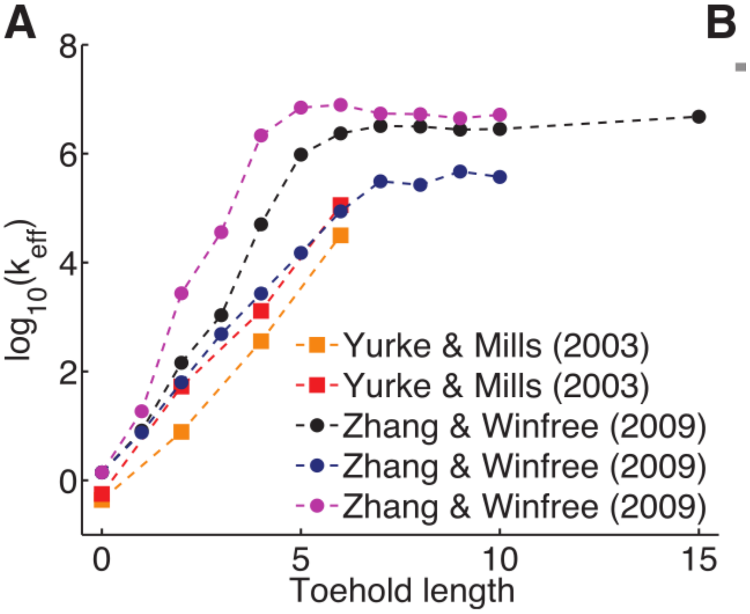
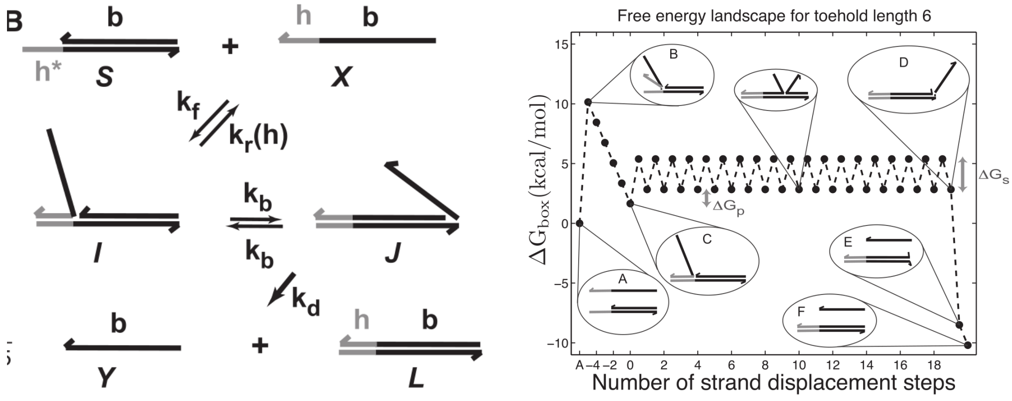


Figure 2.6: Effective displacement rates combined from several experimental setups [YM03] [ZW09] [Sri+13]. Each curve corresponds to the same branch migration domain and toehold prefix.

Rates of TMSD The TMSD experiments can, at fixed conditions, be approximated as a second order reaction, where effective (or apparent) rates k_{eff} can be measured. Figure 2.6 sums up rates from [ZW09] and [YM03]. Rates evolve exponentially with respect to the toehold length, from $10 \times 10^2\,\text{s}^{-1}\,\text{M}^{-1}$ for 2 nt toehold to as high as $3 \times 10^6\,\text{s}^{-1}\,\text{M}^{-1}$ for 6 nt. The rate plateau occurs because for strong toehold, they allmost never unbind. Therefore the rate is determined by the bimolecular meeting rate [ZW09]. Depending of the experiments, the curves have different characteristics even when looking at the rate in function of the toehold energy . The reason is that the curves are obtained with only one strand having suffixes of the complete toehold. The particularity of the toehold has sequence effects, that are successfuly modelized in the next paragraph.



(a) three step model (b) Intuitive Energy landscape model

Figure 2.7: Two models for strand displacement

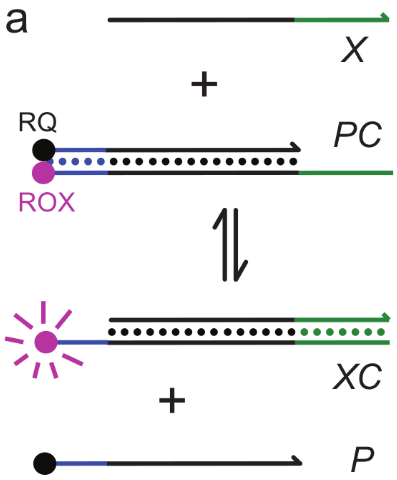
Modelisation of TMSD A phenomeological *three steps model* helps understanding the process of TMSD [ZW09] (Figure 2.7a). The three steps are **toehold association**, **two states for branch-migration**, and **toehold dissociation**. It can be reduced to a two states model under quasi-steady assumptions. This assumption implies bound on the concentrations. For instance for teohold of less than 6 nucleoties, they approximate the concentration bound to 200nM of each strand. This model has pitfalls, despite describing well the experiments.The toehold association rate is observed to be sequence dependent. Indeed factor 20 between fastest and slowest sequence was measured. However no prediction method is given. This makes the precise prediction for strand displacement a single never seen toehold sequence impossible. Also the monomolecular displacement rate of the *three steps model* accounts not only for the branch random walk but also for many physical phenomenon. This implies that the impact of different salt conditions of the rates on the *four states model* is not well understood, because of the many physical processes these rates aggregate.

To improve this model [Sri+13] proposes an intuitive energy landscape model (IEL) (Figure 2.7b). The idea is to make the assumption that the branch migration is driven by two consecutive events, basepair fraying and hybridization. Each one of this step is a state of the model. This gives a linear model, for which we can observe the energetic profile (Figure 2.7b). From left two right. The attachement of one basepair of the toehold first destabilizes the complex (B). The binding of the teohold progressively decreases the energy (C). Then the fraying/hybridization events makes a sawtooth profile. Finally one strand is ejected which decreases the energy a lot. This model introduces two new energies : the sawtooth amplitude ΔGs which accounts for the slowness of branch migration compared to base fraying. The branch migration initiation penalty ΔGp. This branch migration intiation penalty is not accounted into standard models of secondary structures, but is supported by molecular dynamic simulations. It is claimed that this model is captures all the necessary processes of strand displacement. However simiarly to the ike in the three steps models they are no known tables or methods for evaluating the underlying parameters, especially with different salt conditions or temperatures.

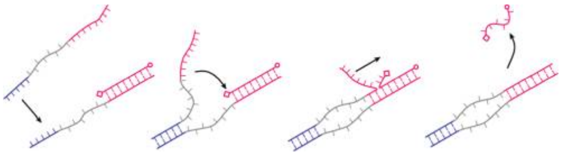
Methods for strand displacement experiments. Concerning the strands purification, litterature show a various approaches. Depending ont the experiment the purification of all strands[23], a subset of the strands [QW11][Che12][23] , or no purification [Gen+11] [Mac+14] is possible.

Now we detail the reporting of TMSD. The reporting can be made either directly when the fluoroescence intensity is linked to the progress of displacement, or indirecttly when the studied displacement is used as the input to a second displacement system equipped with a fluorophore/quencher pair. The advantage of indirect reporting is that there is not possibility of fluorophore/quencher energetic impact that could modifiy the energies and the rates. However indirect reporting it makes the system slightly more complex, and the sequences could cross-talk between the two systems.

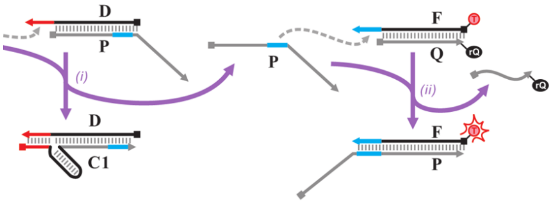
Figure 2.8a shows an example of direct reporting where the initial duplex have quenched fluorophore, and the displaced structure does not quench the fluorophore. Figure 2.8b describes an alternative method of direct reporting where both fluorophore and quencher are on the same strand. Only the single stranded structure of this strand inhibits the fluorescence emission, by coiling which transilantly brings the two modifications closer. Figure 2.8c is an example of indirect reporting, where the displacement of interest leads to displacment of a second system which have a fluorophore-quencher pair. In this paper the large excess of the reporting displacement system makes the fluorescence mimick the displacmeent of interest with a delay of few seconds. Indirect reporting (Figure 2.8d) [ZW09] can also be used without excess which forces to fit the reporter rate first. About the signal aquisition regular spectro-fluorometers can be used [Gen+11] [ZW09]. , as well as a plate readers [Che12] or superresolution fluorescence microscopy [BK16] (for single molecule study). Finally, the strand displacement experiments are made in a variety of buffers containing Tris and Magnesium of Sodium salt. For instance TAE buffer with 12.5 mM $MgCl_2$ [Gen+11], or only TE buffer with $MgCl_2$ added [23]. [Che12] uses a more complex TNaKMg buffer (Tris , NaCl, KCl, $MgCl_2$, EDTA). The working temperature is often near room temperature (25°C [23], 28°C [Gen+11]), but can be higher, for instance 37°C [Che12].



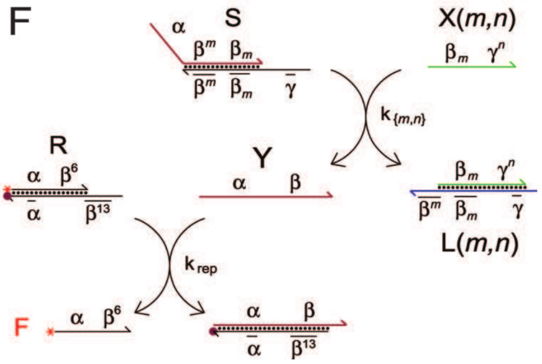
(a) Direct reporting with fluorophore/quencher on the *isubstrate/incumbent strands* [23]



(b) Direct reporting with fluorophore/quencher on the *incumbent strand* [Gen+11]



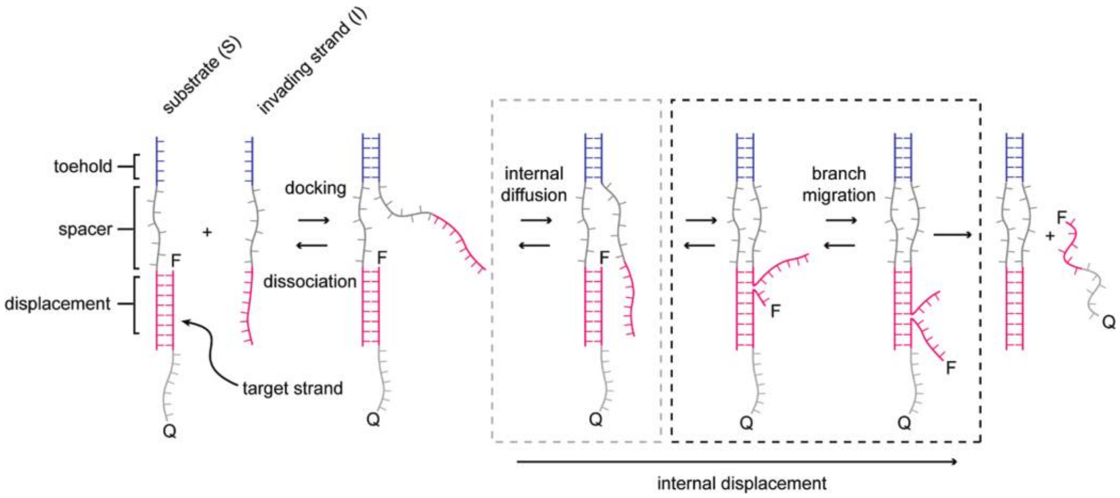
(c) Indirect reporting with fluorophore/quencher strand displacement, used in excess [Che12]



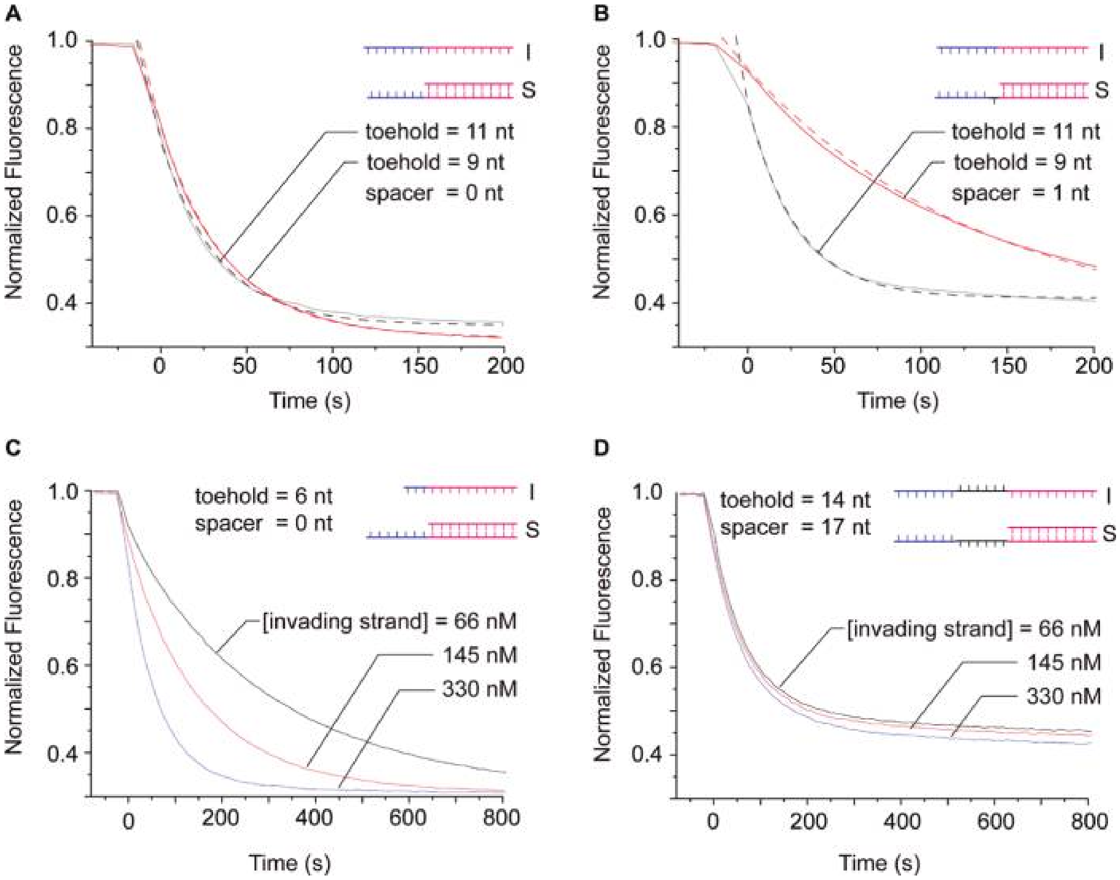
(d) Indirect reporting with fluorophore/quencher strand displacement [ZW09]

Figure 2.8: Several fluorescence reporting systems

Variations



(a) Principle of remote toehold



(b) Two applications of remote toehold. A,B : proofchecking regime. C,D : concentration indepenent reactions

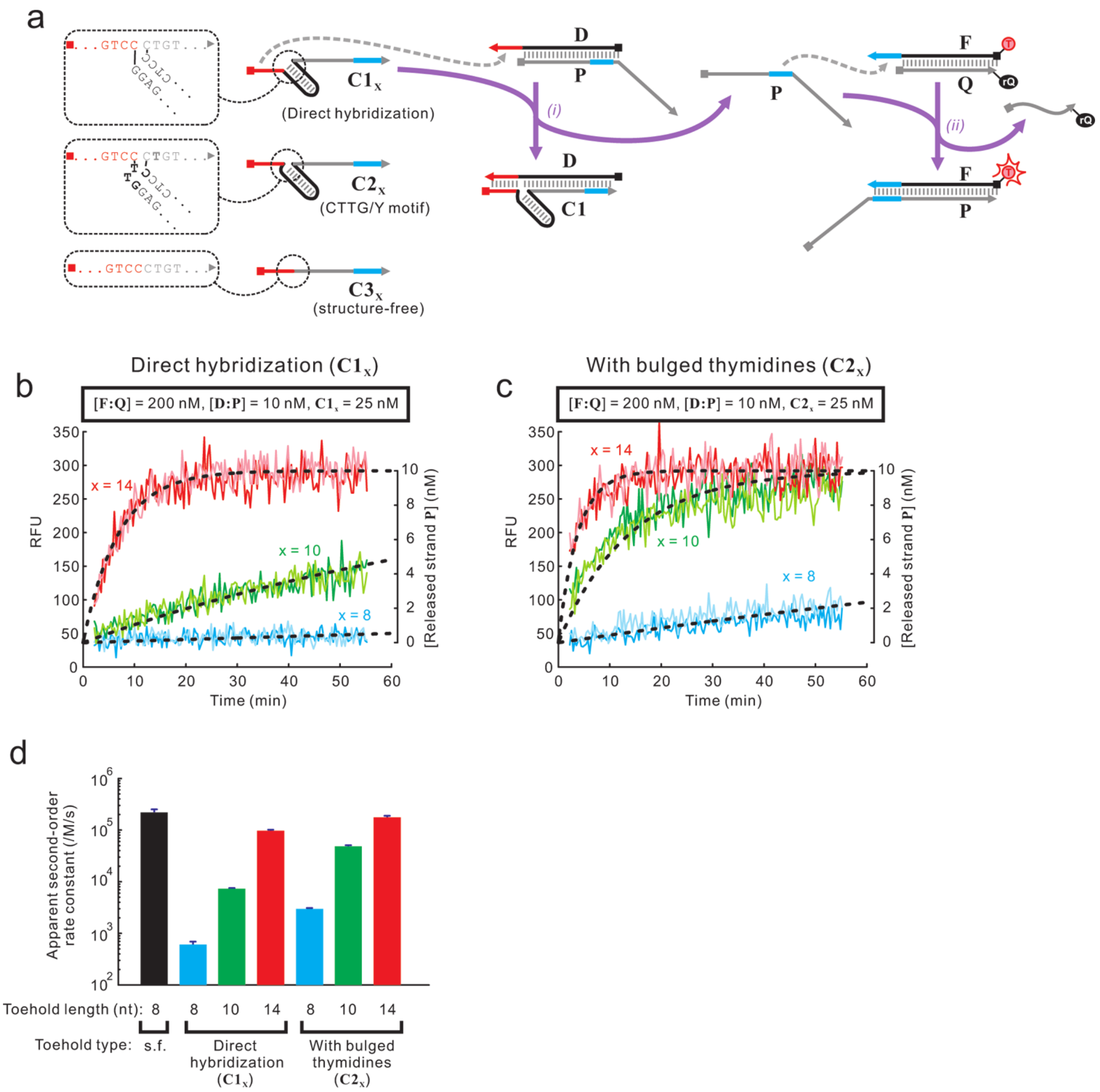
Figure 2.9: Remote toehold

Remote toehold. The intention of remote toehold is to add variability to strand displacement that is not (toehold) sequence dependent[Gen+11]. There are two motivations

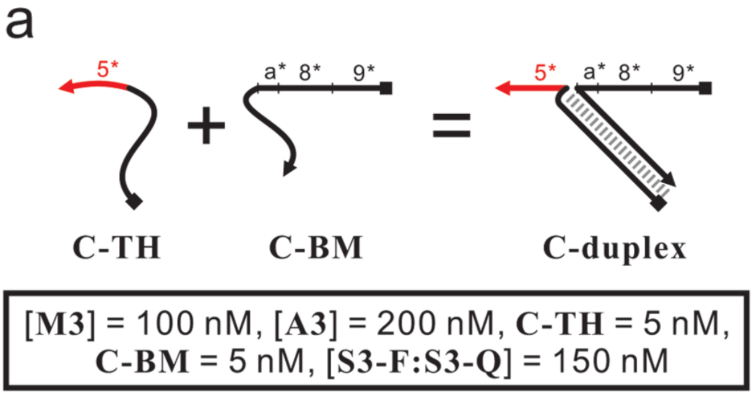
: the set of sequence for small toehold sequences is combinatorially small, and it is difficult to accurately adjust small rate variations due to the exponential relationship with the binding energy. This variation is added through a spacing between the toehold and the displacement domains, either to one or both oligos containting toeholds (Figure 2.9a). Three kind of spacing are explored, single-stranded, double stranded, and with an internal PEG modification. It is shown that the rate decrease due to spacing can lead to more precise adjustement than with modifying the toehold sequene. For instance extending the ssDNA remote section from 17 nt to 23 nt only divides the rate by 2.

[Gen+11] proposes two applications (Figure 2.9b): a proof-reading regime, where the branch migration delay helps discriminate strand displacement or long toehold. On regular TMSD such long toehold would produce about same rates. The second application is to obtain TMSD that behaves the same ways at any concentration. This is made by limiting the rate using the branch migration rates, which makes it the limiting process over toehold association.

Three-way junction separated toehold. [Che12] motivation is to make DNA circuits that are easily reconfigurable *after* DNA synthesis. It shows an associative toehold activation strategy, where the branch migration domain and the toehold are on two different strands. They are united and activated with the hybridization with a third strand. Strand displacement on this duplex must overcome the three way junction between toehold and branch migration. The presence of the bulge reduces massively the second order apparent rate when the toehold has the ability to unbind. For a 8 nt toehold ($-6.2\,\mathrm{kcal\,mol^{-1}}$) the apparent second order rates goes from $6 \times 10^{2}\,\mathrm{s^{-1}\,M^{-1}}$ without the bulge, to $2 \times 10^{5}\,\mathrm{s^{-1}\,M^{-1}}$ with the bulge. The 10 nt toehold $-10\,\mathrm{kcal\,mol^{-1}}$ is still unpractical with the bulge ($7 \times 10^{3}\,\mathrm{s^{-1}\,M^{-1}}$). On the other hand the 14 nt toehold ($-14\,\mathrm{kcal\,mol^{-1}}$), having an irreverible binding shows almost no reduction. It is shown that the impact of the bulge canbe reduced using Thymidines bulge in the three-way branch. It is a pattern known for its stability because it allows for coaxial stacking of the strand.

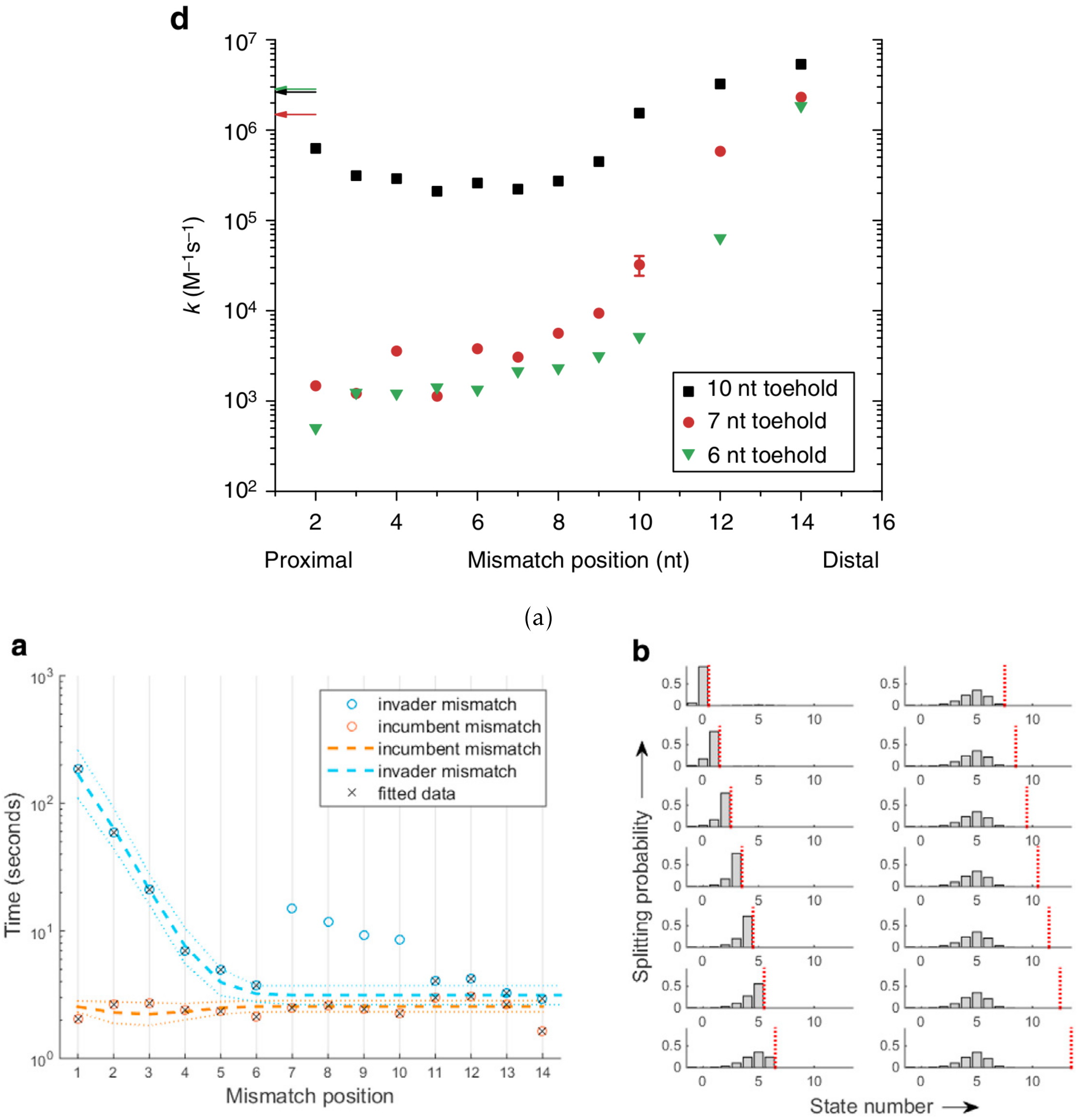


(a) Experimental results of three way toehold



(b) Principle of associative toehold

Figure 2.10: Three way junction toehold



(a)

(b)

Figure 2.11: Experimental results of mismatch in the branch migration domain

Basepair mismatch in branch migration domain. The introduction of a mismatch in the branch migration domain can slow down the bimolecular displacement rates from one to three order of magnitudes depending of the mismatch position and the toehold energy. Both bulk studies [Mac+14] and individual molecules studies [BK16] find the rate reduction most impactfull near the toehold (proximal). A mismatch near the end of the branch migration has almost no impact. The reason is that the incumbent dissociation often occurs spontaneously when only attached by few basis. OxDNA simulations also support the idea that the more proximal the mismatch is, the more it biases the branch migration random walk towards invader ejection. The high sensitivy to mismatch applied combined with toehold exchange strand displacement to create highly specific probes [23].

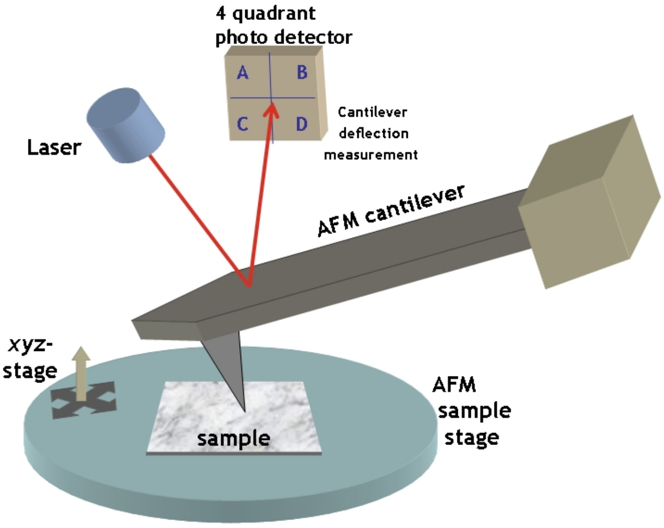
toehold exchange strand displacement (TESD). toehold exchange strand displacement (TESD) consists in two competing toehold mediated strand displacement (TMSD) by adding one toehold domain on each side of the branch migration domain. This results in a system which can transition reversibly between two states. The equilibrium between the two states is determined by the initial concentration of each strand. The reversibility of the TESD, can be used to discrimitate mismatches [23], create catalytic cycle where a small quantity of input is converted into a large quantity of output [QW11], or make reversible walkers [Thu+17]. The rate of TESD can be fast as TMSD, with one direction rate up to $3 \times 10^6\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ [ZW09]. [23] the (apparent) rate of the (full system) reaction is measured to $1 \times 10^6\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. The phenomeological four states model, also fits TESD. However, in some conditions [QW11] smaller magnitudes of TESD have been measured ($5 \times 10^4\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ for TESD, compared to $2 \times 10^6\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ for TMSD on the same toeholds).

Leaking prevention. Despite teohold-less strand displacement rate being considered to be of the order of $1\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$, in application spurious reactions (leaking) often happens with larger orders of magnitude. [QW11] found that stacking between two complexes could initiate undesired strand displacement are higher rate of about $20\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. This is solved by introducing assymetry in the toeholds to prevent displacement that happens by the only simple pi-stacking of complexes. Leaking can be also be amplified by the particular mechanisms as in catalytic amplification mechanism [QW11] which is sensible to spurious hybridization. This is solved by a thresholding mechanism, constantly trying to eliminate a bounded concentration of leaked intermediary input complexes.

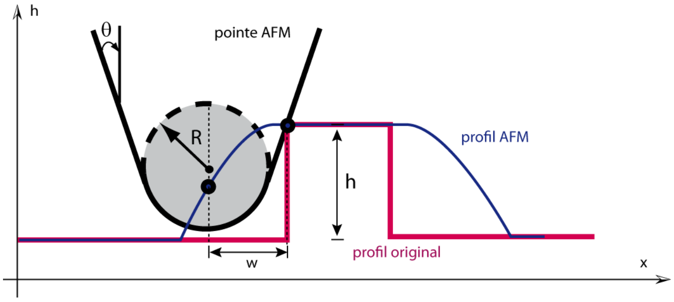
2.3 Measurement of the Computation

Depending on the system the computation can measured when the computing occurs, or at the end. Because the objects are of the nanometers scale, it is impossible to use conventionnal optical microscopes. We discuss two common ways of measuring the result of DNA computing systems : Atomic force microscopy (AFM) which scan the topology of samples, and fluorescence that can both be used for averaged or for single molecule measurements.

2.3.1 Atomic Force Microscopy



(a) Principle of AFM scanning (CC-BY-2.5 Opensource Handbook of Nanoscience and Nanotechnology)



(b) Convolution with the probe [Mon]. The dillation effect depends on the size of the tip R, and the height of the object h.

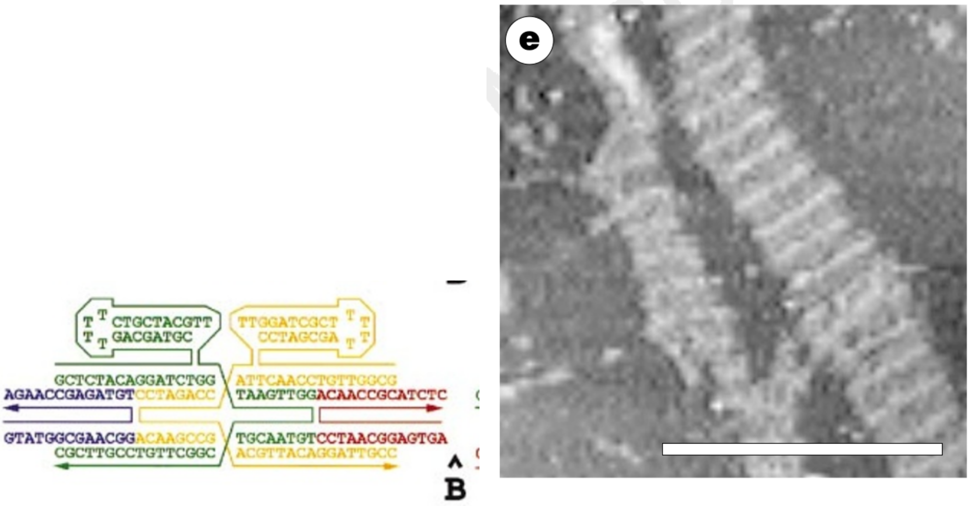
Figure 2.12: Atomic Force Microscopy

General principle Atomic force microscopy approach the sample with a probe with a sharp tip of few atoms. At each point the interactions between the tip and the sample gives indications about the surface topography. This interactions deflects the probe, which measured by reflecting a laser on a 4 quadrants photo-detector(Figure 2.12a). The probe can be physically controlled by a piezzoelectric crystal. Three measuring modes are possible:

- *Contact mode* where at each pixel, the tip is in contact with the surface. The feedback loop is adjusted to keep the cantilever deflect at a constant fixed value.
- *Tapping mode* where the probe is excited with oscillation, near its resonance frequency. A retroaction cycle adjust the height of the probe to keep a constant amplitude of oscillation, thereby a fixed force.
- *The peak force mode*, which is a combination of contact and tapping mode. The tip is oscillated at low frequency to perform force curve cycles abothe the sample, thereby measuring the surface position at each pixel. The setpoint curves correspon in this case to the maximum force applied in the force curves.

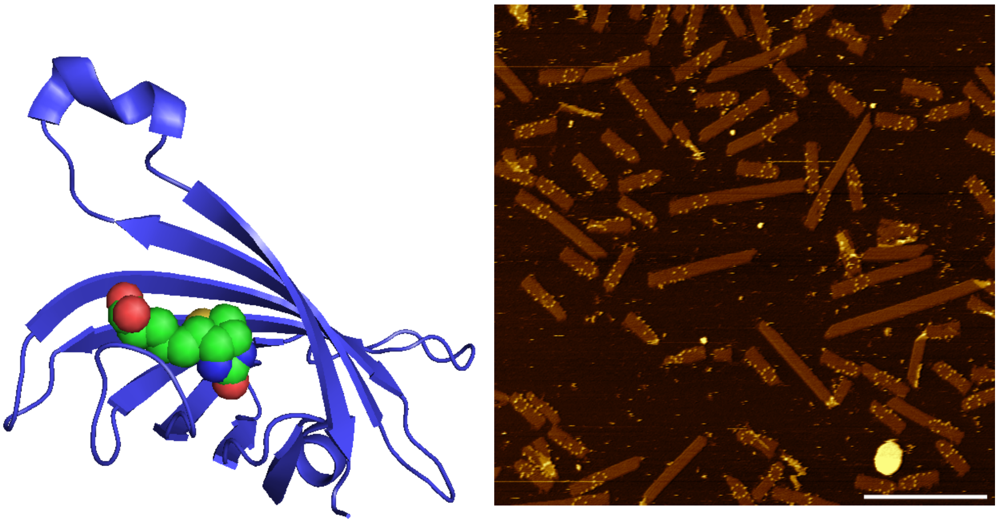
Made in each point, AFM records topological map with nanometer resolution. However the resolution is limited by the width of the tip (Figure 2.12b) because the image is dilated by the tip finite size. the convolution between the tip and the sample is measured The sample can be imaged either in ambiant or or in a buffer.

Identifying labeled strands. DNA Strands can be labeled with a ball-like object in order to indentify on the topography map of the AFM which are placed on a DNA crystal or on a DNA origami. One possibility is to add a secondary structure (dumbell or hairpin) at the end of the strand (Figures 2.13a and 2.13b). A more visible technique is to attach a streptavidin molecule (abound 5 nmwide) (Figure 2.13d).



(a) Tile with hairpin structures [Win+98]

(b) AFM observation of the tiles a hairpin. Scale bar :300 nM.



(c) Streptavidin bounded with a biotin (spheres)

(d) AFM image of biotin-streptravidin labelling on DNA origamis [Woo+19]. Scale bar :500 nM.

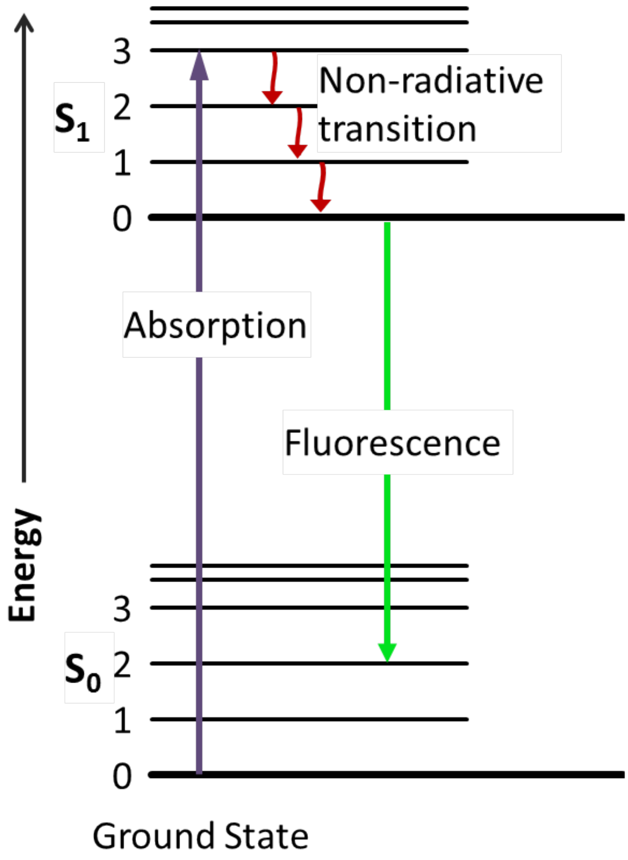
Figure 2.13

2.3.2 Fluorescence Reporting

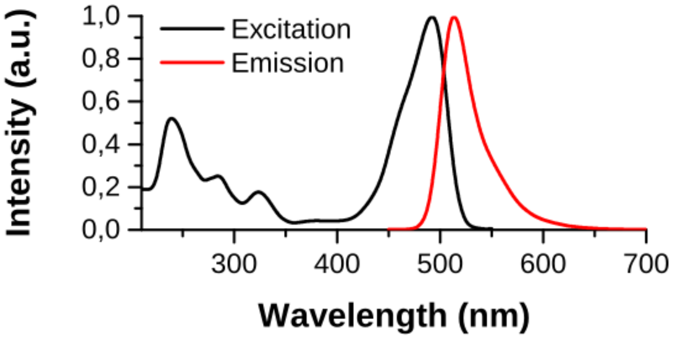
General Principle. DNA molecules can be synthesized with fluorescence modifications at the 5′ or 3′ end. These fluorescent modifications change its the vibrational energy by the absorption of high-energy photons and emit in return lower-energy photons when transitioning back to ground state of vibrational energy. This can be represented as a spectrum with one absorption curve and an emission curve at higher wavelengths. The maximum of emission and absorption wavelengths vary from one flurophores to another. The emitted photons can be transferred to nearby molecules through Förster resonance energy (FRET). This is used to get a signal that depends on the secondary structures of DNA complex, by using a couple of fluorophore modifications with overlapping emission and absorption curves so that the complex of interest have the two modifications close by. In some situations, it can be easier to use a dark quencher which dissipates heat instead of emitting another photo.

Ensemble (bulk) reading using a spectrofluorometer Using fluorescence reporting, we can obtain averaged information about the complex present in the solution at a given time. This is generally made using a spectrofluorometer (Figure 2.15c). It consists in a broad range light source, a monochromator with a tunable slit to isolate a specific excitation wavelength range. This excitation light goes through a four window cuvette. The light emitted by all the fluoroscope is read along another direction, filtered with another monochromator, and read by a photosensor. The parameters that are tunable are the excitation and emission slit (wavelength center and width), the integration time (duration of which emission light is accumulated by the sensor). Optionally, an automatic shutter can only excite the sample during the reading, to prevent bleaching (loss of the fluorescent capability of the fluorophore), which can happen over time and cause signal to diminish. The intensity of bleaching depends on the stability of the fluorophore and the intensity of the emission.

Unique molecule Fluorescence can also be used to observe a single molecule or complex. The impact of photobleaching is much more problematic for unique molecule fluorescence, where fluorophore can easily bleach in the order of ten seconds, making impossible the analysis of a single complex over a long period of time. A possibility is to design experiments where the fluorophore-modified oligo are continuously refreshed. This is for instance a strategy used in DNA paint, where a super resolution camera observe the transition binding of an excess of fluorophore-modified oligos.

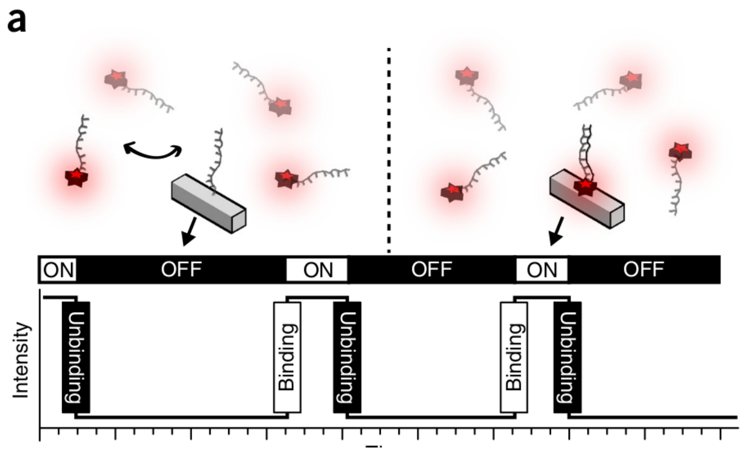


(a) Example of fluorescence emission associated with the transition of vibrational energy levels

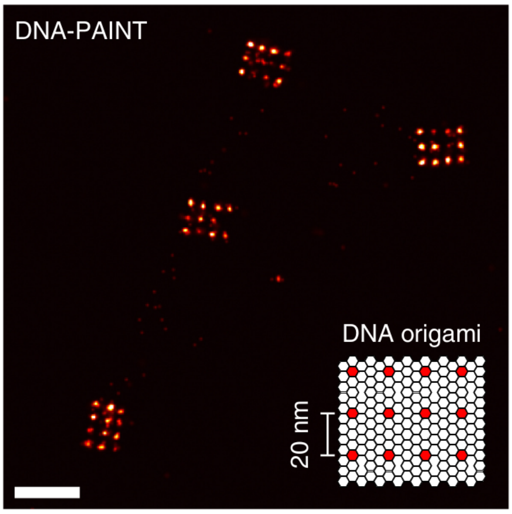


(b) Excitation and emission curve for 6-Carboxyfluorescein (6-FAM)

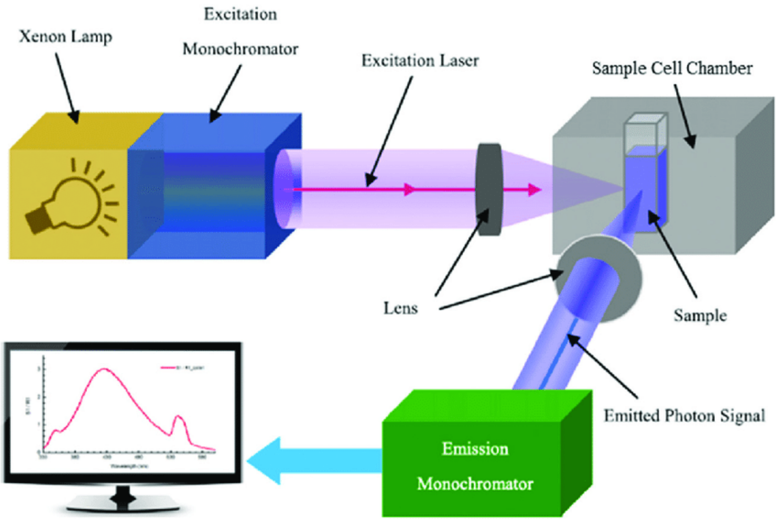
Figure 2.14



(a) DNA paint principle with transiant binding of excess fluoroescent oligos [Sch+17]



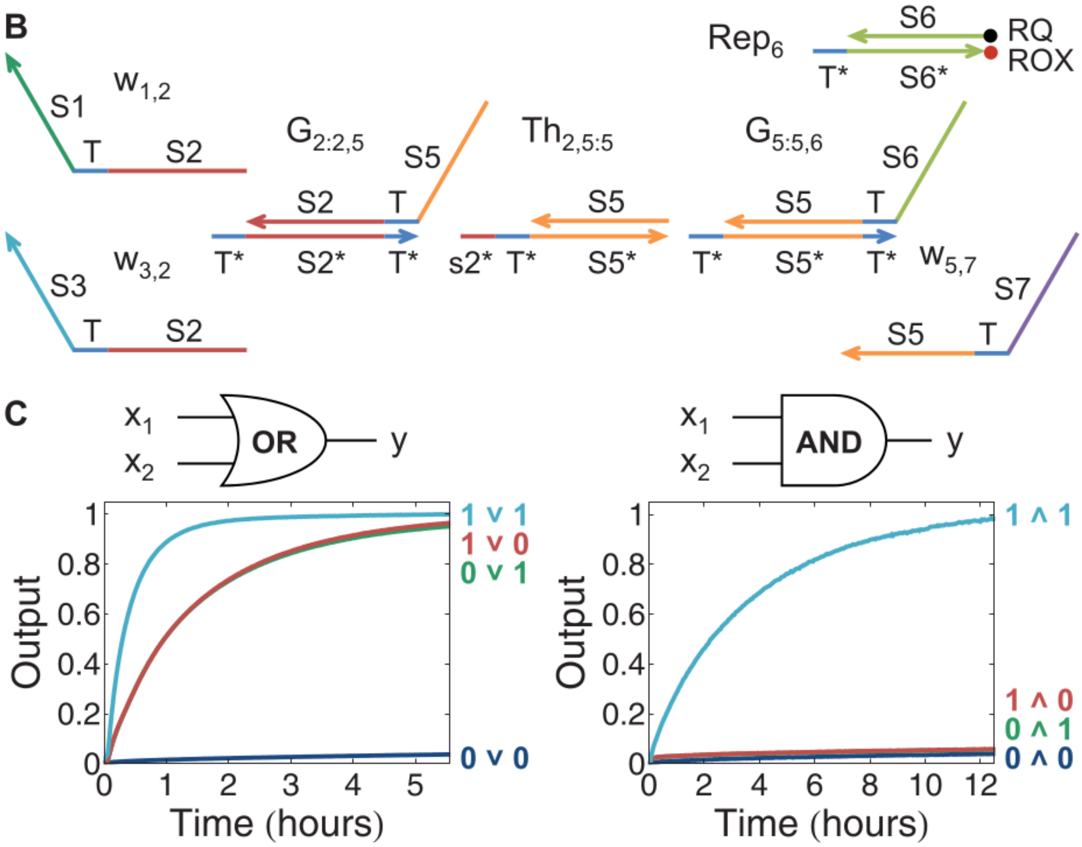
(b) DNA paint image of DNA origamis [Sch+17]



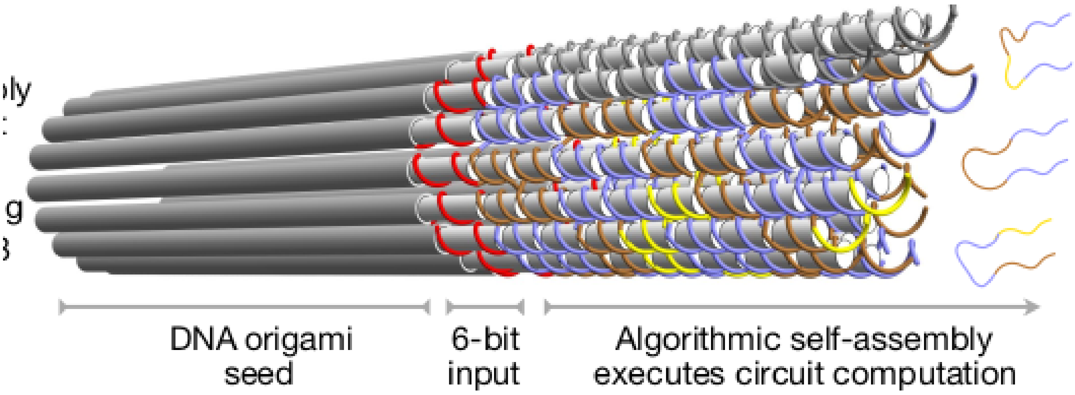
(c) Schematic configuration of a spectrofluorometer. CC BY-SA IOP Publishing Ltd

Figure 2.15

2.4 Examples of DNA Computing



(a) ssDNA



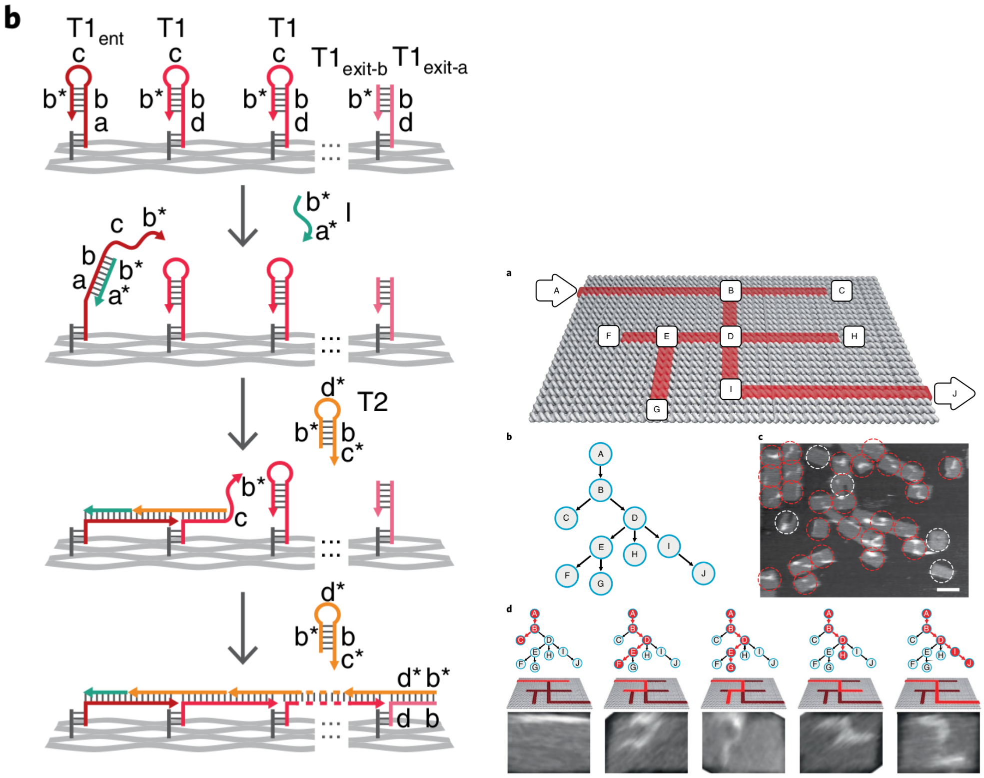
(b) ssDNA

Figure 2.16: Examples of DNA computing systems

Rules we can design carefully systems of DNA complexes which behave according to rules and at the end perform computations. We can indentify two ways of practically doing DNA computing. A local way where the information is computed locally on substrate. For instance iterations of a celular automaton are computed through the crystalising of DNA tile on a DNA origmai [Woo+19]. For global computing, the states of the system is usually the concentrationa of the complexes, which evolves over time as the computation proceed. It is often used combined with TMSD [QW11].

2.5 Scaling up DNA Algorithms

Limitations of current techniques Current techniques limit the scalable abilities of DNA computing. First because an increase of the problem instance size often would translate in an increase the number of strands, which ultimately increases the error rates. There are also time scalability issues,as increasing the size of the instance would mean extremelty long computation time. For instance, gates in [QW11] takes 30-60min to process. The final experiment time is up to 10 hours for a depth 7 circuit. Similarly, a 5 bits counter [Eva] can successfluy count from 0 to 31 in 97 hours incubation time. Finally there are more fundamental constraints to the particular implementation DNA systems. For instance the origami seeds to grow nanotubes [Woo+19] and the size of the instance and of the seed is limited by size of a DNA origami. Another example is the solving of a maze using a DNA walker that takes random direction but walk in an irreversible trend, to computes all possible path betweem start and a dead end. Then origamis where the walker has reached the correct dead-end are manually filtered with magnetic beads. oThe number of such instancdes deacrease expoentnially with the size of the instance, resulting in a non scalable design.



(a) Walker principle

(b) Maze encoded on a DNA origami

Figure 2.17: Maze solving using a irreversible DNA walker

Introducing reversibility to maze solving. Maze solving using DNA is already a challenged problem [Cha+19]. They successully perform an *irreversible walk* accross a graph with 10 nodes, and then extract only the instances that traversed correctly the graph from start to exit. The experiment works in three steps :

1. assemble and filter the origami
2. make the walker traverse the origami. The walker is made by extenting a double stranded path. At the end of the path, two domains are exposed (c,b* figure 2.17a). Hairpins in solution can open itself mediated by toehold c, and attach to the end of the path. The unfolding of this hairpins this time mediates the unfolding of dock hairpin
3. filter using magnetic beads to keep only instance that has reached the exit

They show that their walker is fast (with a propagation of 4 min per cell). Also it selects the next direction randomly with allmost equiprobable choice.

We propose to challenge this problem of maze solving using this time a *reversible* DNA walker. We know the the mean time first arrival in a tree for a balanced random walk is polynomial. This means the such a solution would be scalable in time. Although many reversible walkers [Thu+17] was achieved, they do not write their trace on the surface, which is important for DNA computing. Our goal is to illustrate on the maze solving problem, the capacities of reversible DNA systems for scaling and absorbing error using randomness.

In this chapter, we saw two examples of dynamic behavior implemented with DNA: tile assembly and strand displacement. In particular, strand displacement offers the possibility for many design variations. The result of such dynamic system can be measured or observed through different techniques, in single molecules or in bulk. We observed how these technical capacities can be united to make DNA systems that compute dedicated problems. Finaly we exposed the scalability issues of the current state of the art of DNA computing. We proposed to work on the maze-solving problem, which was solved in a non-scalable and non-reversible way. Our solution aims to show, in this particular case, how introducing reversibility can help overcome this problem of scalability.

Bibliography of the current part

[23] *Optimizing the Specificity of Nucleic Acid Hybridization - PMC*. https://www.ncbi.nlm.nih.gov/pmc/a Jan. 2023. (Visited on 01/26/2023).

[BK16] D. W. Bo Broadwater and Harold D. Kim. "The Effect of Basepair Mismatch on DNA Strand Displacement". In: *Biophysical Journal* 110.7 (Apr. 2016), pp. 1476–1484. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.02.027. (Visited on 02/06/2023).

[Cha+19] Jie Chao et al. "Solving Mazes with Single-Molecule DNA Navigators". In: *Nature Materials* 18.3 (Mar. 2019), pp. 273–279. ISSN: 1476-1122, 1476-4660. DOI: 10.1038/s41563-018-0205-3. (Visited on 05/01/2021).

[Che12] Xi Chen. “Expanding the Rule Set of DNA Circuitry with Associative Toehold Activation”. In: *Journal of the American Chemical Society* 134.1 (Jan. 2012), pp. 263–271. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja206690a`. (Visited on 01/08/2024).

[Dey+21] Swarup Dey et al. “DNA Origami”. In: *Nature Reviews Methods Primers* 1.1 (Jan. 2021), p. 13. ISSN: 2662-8449. DOI: `10.1038/s43586-020-00009-8`. (Visited on 12/24/2023).

[Eva] Constantine Evans. “Crystals That Count!” In: ().

[EW17] Constantine G. Evans and Erik Winfree. “Physical Principles for DNA Tile Self-Assembly”. In: *Chemical Society Reviews* 46.12 (2017), pp. 3808–3829. ISSN: 0306-0012, 1460-4744. DOI: `10.1039/C6CS00745G`. (Visited on 10/09/2023).

[Gen+11] Anthony J. Genot et al. “Remote Toehold: A Mechanism for Flexible Control of DNA Hybridization Kinetics”. In: *Journal of the American Chemical Society* 133.7 (Feb. 2011), pp. 2177–2182. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja1073239`. (Visited on 09/29/2021).

[LS] Nicolas Levy and Nicolas Schabanel. “ENSnano: A 3D Modeling Software for DNA Nanostructures”. In: ().

[MA20] Christopher Maffeo and Aleksei Aksimentiev. “MrDNA: A Multi-Resolution Model for Predicting the Structure and Dynamics of DNA Systems”. In: *Nucleic Acids Research* 48.9 (May 2020), pp. 5135–5146. ISSN: 0305-1048, 1362-4962. DOI: `10.1093/nar/gkaa200`. (Visited on 01/19/2024).

[Mac+14] Robert R. F. Machinek et al. “Programmable Energy Landscapes for Kinetic Control of DNA Strand Displacement”. In: *Nature Communications* 5.1 (Nov. 2014), p. 5324. ISSN: 2041-1723. DOI: `10.1038/ncomms6324`. (Visited on 01/10/2024).

[Mon] Fabien Montel. “Dynamique à l’équilibre et hors d’équilibre de la chromatine visualisée par microscopie de force atomique: effet des variants d’histones et des facteurs de remodelage”. In: ().

[QW11] Lulu Qian and Erik Winfree. “Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades”. In: *Science* 332.6034 (June 2011), pp. 1196–1201. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1200520`. (Visited on 10/20/2022).

[Rey+00] Luis P Reynaldo et al. “The Kinetics of Oligonucleotide Replacements”. In: *Journal of Molecular Biology* 297.2 (Mar. 2000), pp. 511–520. ISSN: 00222836. DOI: `10.1006/jmbi.2000.3573`. (Visited on 01/11/2024).

[RPW04] Paul W. K Rothemund, Nick Papadakis, and Erik Winfree. “Algorithmic Self-Assembly of DNA Sierpinski Triangles”. In: *PLoS Biology* 2.12 (Dec. 2004). Ed. by Anne Condon, e424. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.0020424`. (Visited on 10/09/2023).

[Sch+17] Joerg Schnitzbauer et al. “Super-Resolution Microscopy with DNA-PAINT”. In: *Nature Protocols* 12.6 (June 2017), pp. 1198–1228. ISSN: 1754-2189, 1750-2799. DOI: `10.1038/nprot.2017.024`. (Visited on 01/17/2024).

[Sen+21] A. Sengar et al. "A Primer on the oxDNA Model of DNA: When to Use It, How to Simulate It and How to Interpret the Results". In: *Frontiers in Molecular Biosciences* 8 (June 2021), p. 693710. ISSN: 2296-889X. DOI: `10.3389/fmolb.2021.693710`. (Visited on 09/07/2023).

[Sri+13] Niranjan Srinivas et al. "On the Biophysics and Kinetics of Toehold-Mediated DNA Strand Displacement". In: *Nucleic Acids Research* 41.22 (Dec. 2013), pp. 10641–10658. ISSN: 1362-4962, 0305-1048. DOI: `10.1093/nar/gkt801`. (Visited on 09/29/2021).

[Thu+17] Anupama J. Thubagere et al. "A Cargo-Sorting DNA Robot". In: *Science* 357.6356 (Sept. 2017), eaan6558. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aan6558`. (Visited on 05/01/2021).

[Wag+17] Klaus F. Wagenbauer et al. "How We Make DNA Origami". In: *ChemBioChem* 18.19 (Oct. 2017), pp. 1873–1885. ISSN: 14394227. DOI: `10.1002/cbic.201700377`. (Visited on 05/01/2021).

[WB04] Erik Winfree and Renat Bekbolatov. "Proofreading Tile Sets: Error Correction for Algorithmic Self-Assembly". In: *DNA Computing*. Ed. by Gerhard Goos et al. Vol. 2943. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 126–144. ISBN: 978-3-540-20930-0 978-3-540-24628-2. DOI: `10.1007/978-3-540-24628-2_13`. (Visited on 01/17/2024).

[Win+98] Erik Winfree et al. "Design and Self-Assembly of Two-Dimensional DNA Crystals". In: *Nature* 394.6693 (Aug. 1998), pp. 539–544. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/28998`. (Visited on 12/25/2023).

[Win06] E. Winfree. "Algorithmic Self-Assembly of DNA". In: *2006 International Conference on Microtechnologies in Medicine and Biology*. Okinawa: IEEE, May 2006, pp. 4–4. ISBN: 978-1-4244-0337-0 978-1-4244-0338-7. DOI: `10.1109/MMB.2006.251471`. (Visited on 10/09/2023).

[Woo+19] Damien Woods et al. "Diverse and Robust Molecular Algorithms Using Reprogrammable DNA Self-Assembly". In: *Nature* 567.7748 (Mar. 2019), pp. 366–372. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/s41586-019-1014-9`. (Visited on 02/06/2022).

[YM03] Bernard Yurke and Allen P. Mills. "Using DNA to Power Nanostructures". In: *Genetic Programming and Evolvable Machines* 4.2 (June 2003), pp. 111–122. ISSN: 1573-7632. DOI: `10.1023/A:1023928811651`. (Visited on 09/15/2023).

[ZW09] David Yu Zhang and Erik Winfree. "Control of DNA Strand Displacement Kinetics Using Toehold Exchange". In: *Journal of the American Chemical Society* 131.47 (Dec. 2009), pp. 17303–17314. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja906987s`. (Visited on 05/01/2021).

Part II

Path Self-Assembly on DNA Origami

Chapter 3

Roadmap for Origami, Mazes, and Path Assembly Designs

Outline of the current chapter

In this chapter, we explain how we can functionalize a DNA origami to encode a maze. Our idea is to later be able to implement a random walk made of multiple DNA strands that will attach and detach dynamically from the maze-assembled pathway. We first address the problem of attaching such a path, but in a static version. We define possible experimental approaches and our experimental goals.

3.1 Design description

3.1.1 DNA Origami Functionnalization and Grid

Functionalization with outgoing staples. The functionalization consists of DNA origami staples with an extended single-stranded domain going out of the DNA origami (Figure 3.1). The position at which the staples go out of the origami must be carefully chosen. To achieve this, we make sure the staple leaves at a position that is orthogonal to the origami plane. Hence, a small error in the prediction of the angle of the supporting helix is not likely to make the single-stranded domain to come out on the wrong side. The 3D view of CodeNano and ENSNano are very useful for choosing such an outgoing position.

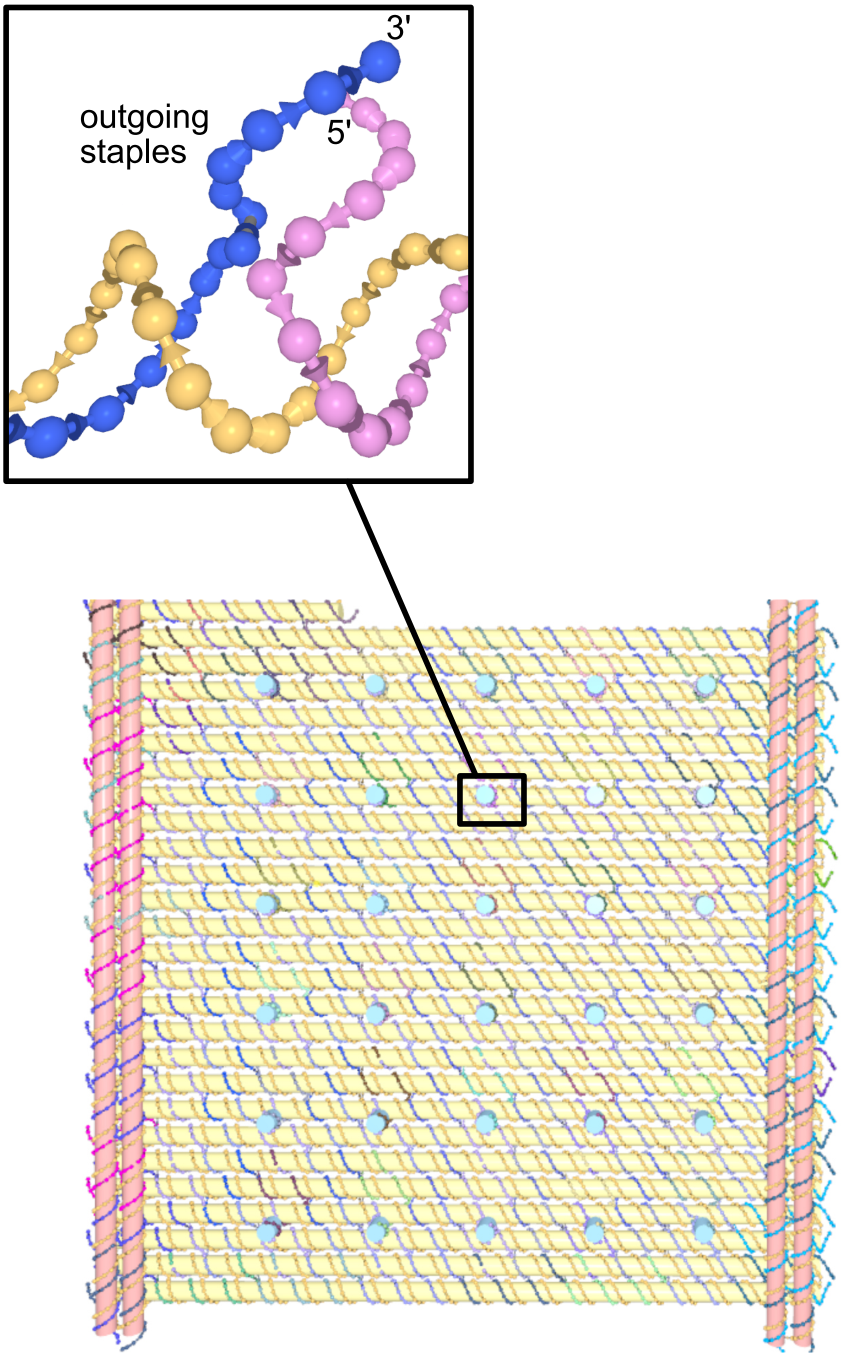


Figure 3.1: Top view of a 5 × 6 grid on our DNA origami. Here two staples are outgoing: one exiting in the 5’ sense, and another one in the 3’ sense. The squared zoom represents two of the outgoing staples as a side view.

helices gap	distance (nm)	number of turns	distance (nm)
0	2.65	0	0
1	5.30	1	3.49
2	7.95	2	6.97
3	10.60	3	10.46

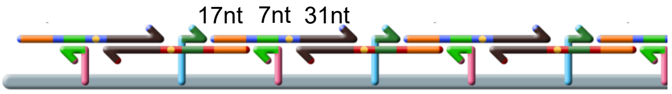
Table 3.1: Distances between two consecutive outgoing staples in vertical and horizontal directions. The formulas are distance horizontal = helices gap × (helix diameter + inter-helix gap), and distance vertical = number of turn × bp/turn × distance/bp.

Grid on a single-layered DNA origami. We want to make a regular grid from outgoing staples. These outgoings staples, with a appropriate domains to attach strands will be called *docking staples*. We only consider grids that are orthogonal to the origami helices, although it is possible to have a diagonal grid [Thu+17]. We discuss the possible parameters for such a grid. The difficulty is choosing the parameters well so that the distances between two consecutive docks are similar both vertically and horizontally. Table 3.1 shows the distances between two consecutive docks 1) horizontally as a function of the number of turns and 2) vertically as a function of the number of helices. For the first values, the smallest differences along the two axis are achieved for the pairs with helices gap = number of turns. The value helices gap = number of turns = 1 does not allow for a decent stapling of the DNA origami. Then the value 3 leads to a smaller difference in the two axes than the value 2. Figure 3.1 shows the possible docking positions of a 6 × 5 grid with values helices gap = number of turns = 2.

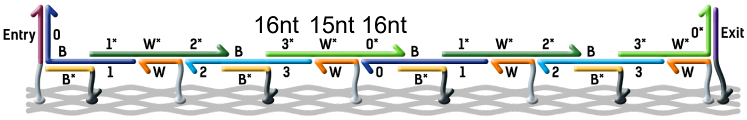
3.1.2 Multi-strands Path on Origami

We discuss how we can implement a single-helix wide path on the origami, made of multiple strands attached one to another. First, we describe three designs propositions that will be experimentally tested. Then we focus on the way we can introduce flexibility in them to make turns.

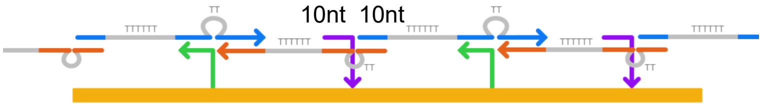
Three path design propositons. The three path designs we propose share common particularities (Figure 3.2). Each path strand attaches to a functionalized "docking" staple, and to the previous and following path strands. The distance between two docks is reached by 31 nt. The entry and exit docks contain an extended domain, mimicking the presence of a nearby path attached. The first design has docks that go out the origami in 3' direction, but this 3' direction goes up and down every other two strands. On the other hand, in the second design, the docks go out the origami in alternation of 5' and 3' directions. But the extremities of the staples always wind up the path. Finally, the third design is similar to the second one, but with single-stranded parts, to reach the same distance between two consecutive docks. These iterations will be justified experimentally in the result section.



(a) Origami buoy design.



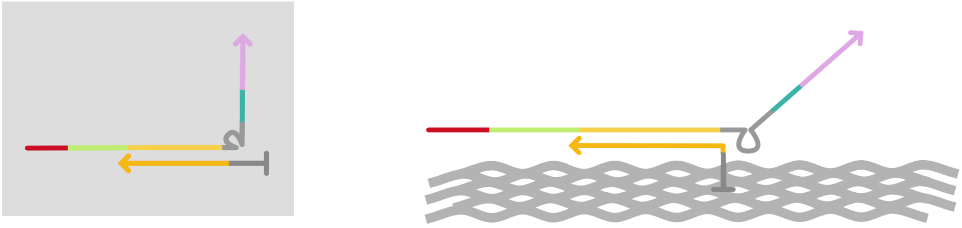
(b) Origami 2-helices rods designs.



(c) Origami 2-helices rods designs.

Figure 3.2: Three feasible designs for our path attached on a DNA origami.

Flexible turns. In the three presented designs, there is a nick (no covalent bond along the DNA backbone) between the docking domain and the next or previous path. To make a turn at this position, we need the strand covering this nick not to be entirely double-stranded. Our solution is to introduce a poly-T sequence at this position. Our experiments were done with a short 2-T sequence. At the design time, we decided to use the sequence 2-T (TT). With this sequence and simulations using rigid helices in ENSNano, the result was visibly feasible. After the experiments was done, we think this this 2-T sequence might be too short.



top view side view

Figure 3.3: Turn of a path assembled on to the DNA origami. The bulge represent a flexible poly-T sequence.

Practical ways of assembling the paths strands. The experiments presented in the result section will only use a one-pot annealing. This means that the origami and the path are assembled in the same tube, simply by decreasing the temperature. Controlling the assembly by making sure the origami assembly temperatures and the path assembly temperatures are different, similarly to the one-pot annealing of [Thu+17].

In this chapter, we have investigated strategies on how a path made of multiple DNA strands could be assembled on a DNA origami. We choose the settings of a grid, composed of *docking* staples, coming out of the DNA origami. The parameters of the grid define the length of the strands of the path. We also discussed the possibility of making turns. Now the next result chapter was driven by two goals: First, to try assembling a path on a DNA origami, to validate the geometrical constraints. Then, to validate our ability to visualize such one helix-wide path with AFM imaging.

Bibliography of the current part

[Thu+17] Anupama J. Thubagere et al. “A Cargo-Sorting DNA Robot”. In: *Science* 357.6356 (Sept. 2017), eaan6558. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aan6558`. (Visited on 05/01/2021).

Chapter 4

Material and Methods

Outline of the current chapter

4.1 Design Tools

Origami Design with CodeNano and ENSNano. The result section will describe two designs : the first one has orthogonal rods made of six-helices bundles. The second one has orthogonal rods made of two adjacent helices. The first DNA origami (with six-helices bundles) was designed with the CodeNano software. On the other hand, the second one (with two adjacent helices) was designed with ENSNano. We describe the process to obtain the sequences of the DNA origami. We name the dock staples in ENSNano according to its position on the grid, and extend it in single-strand. After exporting the staple sequences into a spreadsheet, the stringle-stranded dock extension contains "?"characters (because the nucleotides are not paired to the scaffold). We can either remove the ? for unused docks, or replace it with an appropriate sequence. To achieve this, we wrote a Python script to load the sequence table with the Pandas library and apply the necessary transformations.

Sequence Design. The sequences for the path were generated using the Nupack library. For the first hard-coded path, we used the Nupack 1 version with a custom Python script that calls the program version of Nupack. The other designs we conceived with the Nupack 2 version and its native Python module. We designed the sequences by first generating a pool of possibles sequence for each domain. We check that no secondary structures are predicted to exists at 25 °C. Then we randomly select an appropriate subset,whose domains have with similar bindind energies, and for which there is no predominant cross-talk.

4.2 Experimental Tools

DNA Origami Preparation and Annealing. The DNA origami are annealed in TAE buffer, with 12.5 mM $MgCl_2$, at a concentration of 2 nM to 10 nM scaffold, with a 10 times excess of staples. The staples are ordered from the IDT manufacturer, as custom oligos in 96-well plates, at100 µM concentration, in 1X TE buffer. They are then mixed, and assembled with a temperature ramp annealing. We found that a 3 hour long annealing at fixed rate, starting from 90 °C to 25 °C has yields with our designs. The PCR used to program such temperature range annealing is an Eppendorf Mastercycler.

AFM imaging. AFM images are produced using a JPK Nanowizard with a Fastscan module with Nanoworld tips. The imaging is Operated in liquid, in tapping mode. The samples are incubated at 2 nM in 50 µL for 10 minutes. A rinsing to remove excess staples on the mica surface is optional done. Our protocol changes the buffer three times with 50 µL of NaCl 3 times, followed by another change to the original annealing buffer. The imaging is done at 512 px or 1024 px, with window sizes from 10 µm to 500 nm.

Measurement of melting temperature with qPCR. Annealing curves of pairs of strands are measured in a *CFX connect* qPCR machine. We found 1 µM of each strand to be a good concentration with a high and reproducible signal. For duplexes without fluorophore and quenchers, we add a double-strand intercalating dye (Evagreen at 1 X concentration), For the annealing curve of DNA origami that will be described in the result section, we add Evagreen at 2 X concentration to compensate for the high amount of DNA in the tubes.

4.3 Analysis Tools

AFM processing AFM images are processed with the Gwyddion software. To process the image, we use the following steps: 1) Load the height channel. 2) Flatten with a polynomial of degree 2 on both axes. 3) If possible, select zones with origami with a height threshold value to remove noise or empty zones. 4) Align rows on the selected grains. 5) Select the background. 6) With the polynomial tool with degree zero on the background, set the mean of the background to zero. 7) Set the scale theme to gold, and set the Z-color scale manually from -1 nm to 3 nm.

4.4 DNA Sequences

Non hardcoded Path The sequences of the path strands, for the non hard-coded path, are detailed in the following table.

name	sequence
path 0	`CTTCCATACTCTTAC CTTAAACATCTAACAC TTTT CACTACTTACTCTAC`
path 1	`GAGGTAAAGTTAGTG TTTT GTGTTAAGTTTAGATG GTAGAGTAAGTAGTG`
path 2	`CACTAACTTTACCTC CTTAAACATCTAACAC TTTT CATTCTCACACTATC`
path 3	`GTAAGAGTATGGAAG TTTT GTGTTAAGTTTAGATG GATAGTGTGAGAATG`

Non Hardcoded Path with single-stranded domains The sequences of the path strands are detailed in the following table.

name	sequence
path 0	`CACTTAAATC TTTTTT CTCTCTATAC TT CACTAATTTC`
path 1	`GAAATAAGTG TT GAATGTTTAG TTTTTT GAAATTAGTG`
path 2	`CACTTATTTC TTTTTT CTCTCTATAC TT CACTTTATTC`
path 3	`GATTTAAGTG TT GAATGTTTAG TTTTTT GAATAAAGTG`

Chapter 5

Results

Outline of the current chapter

In this chapter, we observe two versions of the DNA origami that will support the path. We will see that our more complex versions actually raise problems, such as the inability to prevent pi-stacking. Then we will assemble a path, by encoding the position of each path strand with the sequence uniquely. We will see that a 'bad' choice of design will lead to unbalanced domain lengths and ultimately bad assemblies. Finally, motivated by the success of the path experiments, we will try to relax the system, and use only sequences for path strands.

5.1 DNA Origami Substrate

The DNA origami will be used as a substrate for the path assembly. Careful design and production are crucial, which we will detail in this section. First, we detail experimental results with two DNA origami designs. Then we study with several experimental methods the assembly behavior during a temperature annealing. Finally, we focus on the pi-stacking of our origami, causing them to aggregate.

5.1.1 Rigid 2D Origami Using Rods

Single layer DNA origami are expected to be able to twist. This will be especially unfortunate if it allows for a strand to attach between two non-adjacent docks. This kind of problem is solved [Thu+17] by using a double-layered DNA origami. With the access of CodeNano and then ENSNano, we wanted to try a simpler alternative. The idea is to have a regular single-layer origami and several rods in an orthogonal direction to prevent bending. We iterated over two designs for the rods, which are described in the two next paragraphs.

Buoy rods design with 6-helices bundles. In this design, each rod is made of a six-helices bundle. These rods are placed so that the center of the bundle aligns with the plane of the origami. One difficulty we encountered was to properly connect the staples between the rods and the 2D plane. The presence of these "connection" staples often forces them to break the pattern of the center part staples. The AFM images (Figure 5.2) obtained in AC mode in buffer, shows that we managed to correctly assemble it.

2-Helices rods designs. After the problem encountered in the buoy design, we decided to opt for a simpler approach. Two helices are attached on two opposite edges of the flat central part. On Figure 5.4 we can measure the height of the origami to be 1.5 nm. The rods raise to an additional 0.7 nm in height. A problem we encountered with this design was the landing side on mica for AFM observation. When the origami lands on the side where the docks are outgoing, it is difficult to observe the eventual strands attached to it. We observed that the design produced a bias towards landing on the "opposite" side. The probability of landing side was random across the experiments.

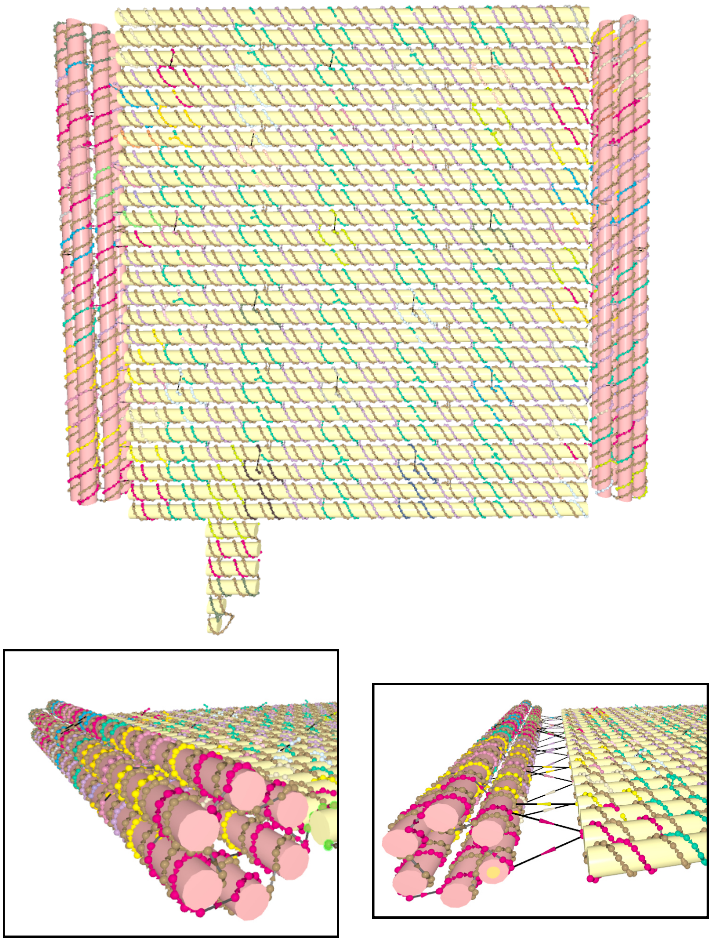


Figure 5.1: Origami buoy design : to prevent the origami to twist, the edges of the DNA origami are linked to six-helix bundles

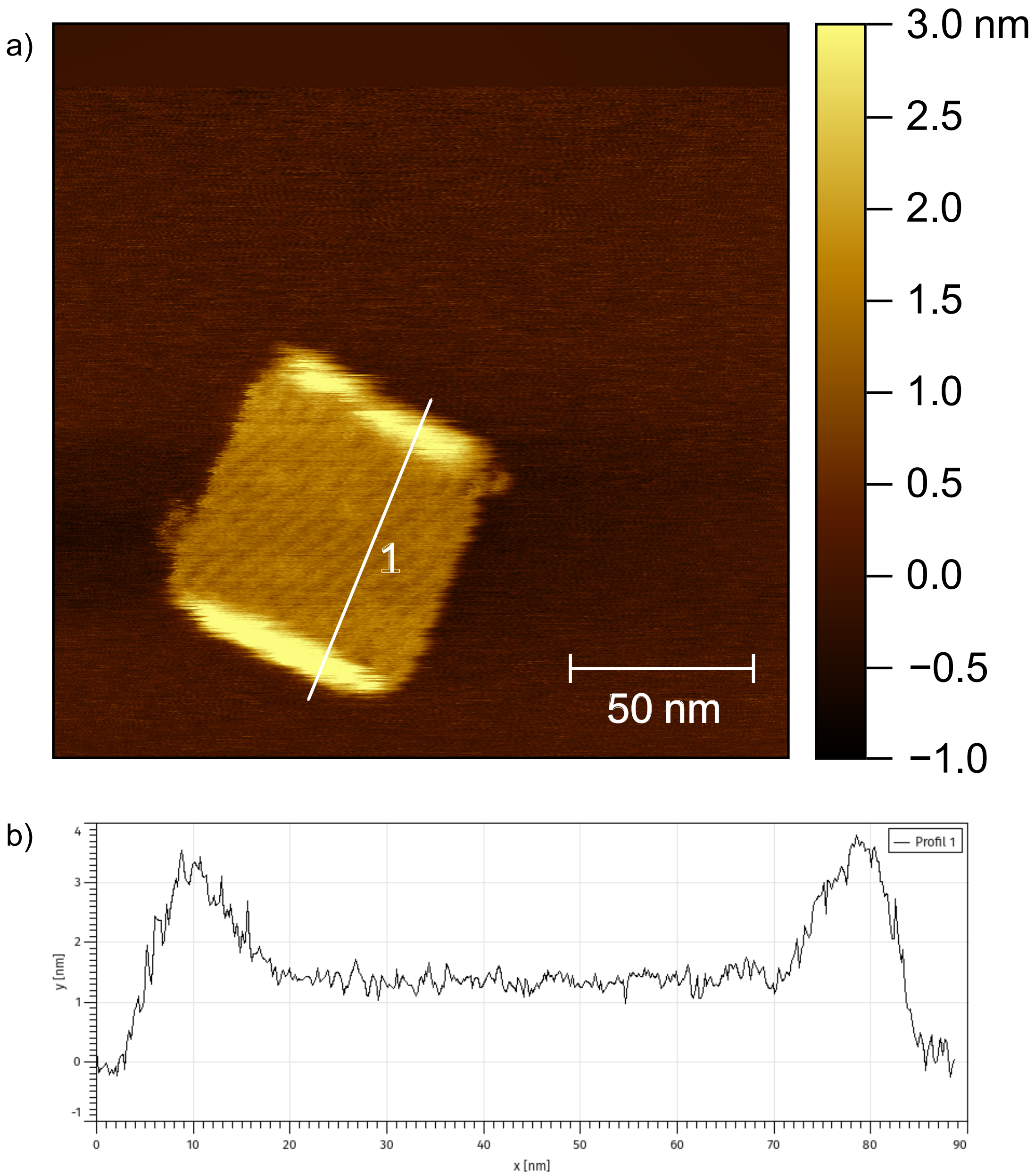


Figure 5.2: a) AFM image of the buoy rods origami. b) Height profile along the line showing, showing the 2 rods at the edges of the origami.

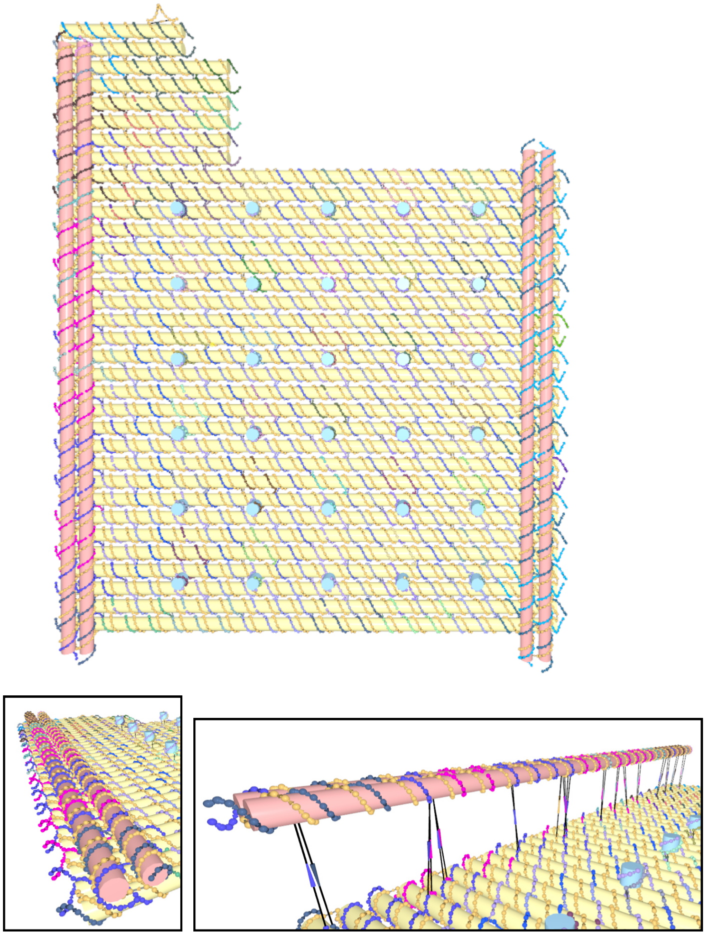


Figure 5.3: Origami buoy design : to prevent the origami to twist, the edges of the DNA origami are linked to two-helices rods

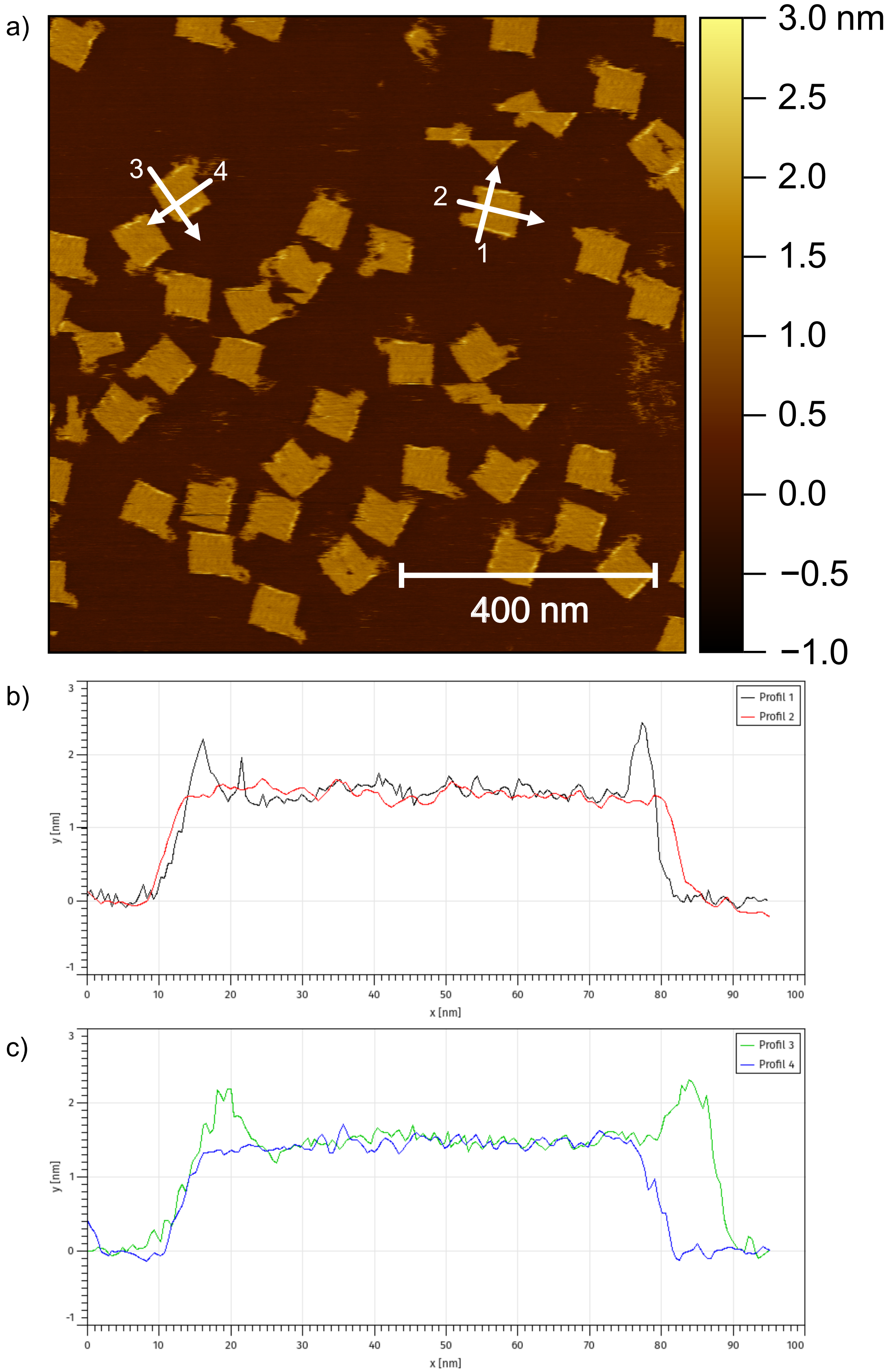


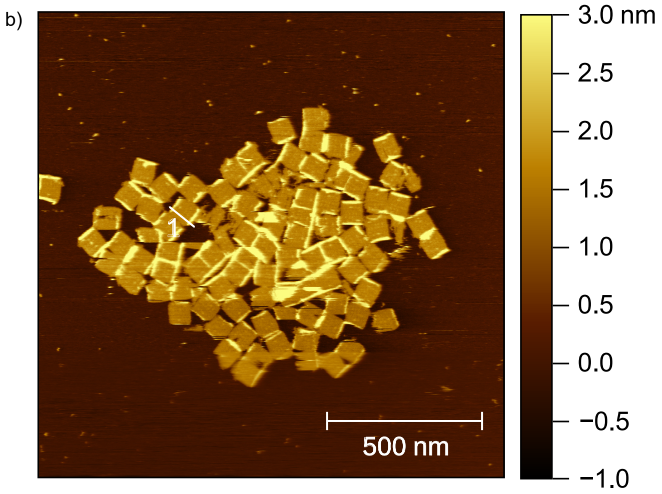
Figure 5.4: a) AFM image of 2-helices rods DNA Origami design. b) Height profiles along the two cross-sections a DNA origami (1 and 2 on the subfigure a) c) Height profiles along the two cross-sections a DNA origami (3 and 4 on the subfigure a)

5.1.2 Pi-stacking of DNA origami

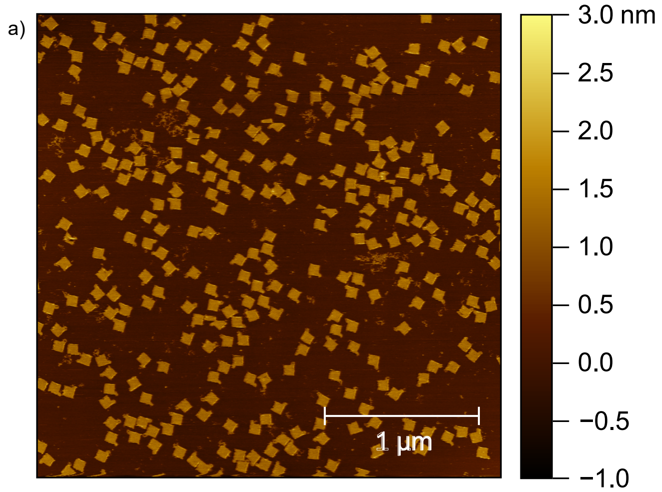
Pi-stacking between the helices ends at the edges of the DNA origami produces a interaction between origami. In such conditions, groups of origami can be observed on the mica surface using AFM. Such interaction could also happens in solution, particularly in the context of Part II, where the experiments are intended to be done near room temperature.

Our first 6-helices bundle rods design was supposed to be resistant to pi-stacking. Our argument was twofold: The 6-helices bundles physically shield the ends of the helices of the main part. The only exposed helices ends are those of the 6-helices bundles. However, this design happened to produce pi-stacking on every edge of the origami. First, the 6-helices bundles accounts for 12 possible pi-stackings. And on the other side, the bundle can move up or down, particularly when landed on a surface, which exposes all the helices of the 2D plane. A usual technique is to put hairpins or poly-T sequences at the end of staples on the sides of the DNA origami. Using this technique was not straightforward, as it meant removing some staples that attach the 6-helices bundles to the plane, and also drastically impacts the quality of the stapling at the end of the bundles.

To prevent pi-stacking with poly-T sequences, we had to change the structure of the rods. We opted for simpler rods, made of 2 helices on each side. This time the rods are attached on one side of the origami. It allowed to used 6-T sequences on every helices ends. As expected, this completely prevented pi-stacking.



(a) AFM imaging of the 6-helices bundles rods design.



(b) AFM imaging of the 2-helices rods design.

Figure 5.5: Illustration of the pi-stacking gathering on effect on the 2-helices design (a), and the absence of pi-stacking for the buoy rods design (b).

> We have designed and successfully assembled two DNA origami with rods to prevent torsion. The more complexed design, with 6-helices bundles, proved to induce pi-stacking that is hard to fix. On the other hand, the simpler design with 2-helices rods, allowed the use of poly-T sequences to effectively solve the problem of pi-stacking of origami.

5.1.3 DNA Origami Assembly

We will see that measuring the temperature at which the origami assembles will be useful to set-up the dynamic assembly of paths. To obtain the relationship between temperature and assembly yield, we ran two experiments. In the first one, we stop the assembly at a defined temperature, and observe the result with AFM and agarose gel.

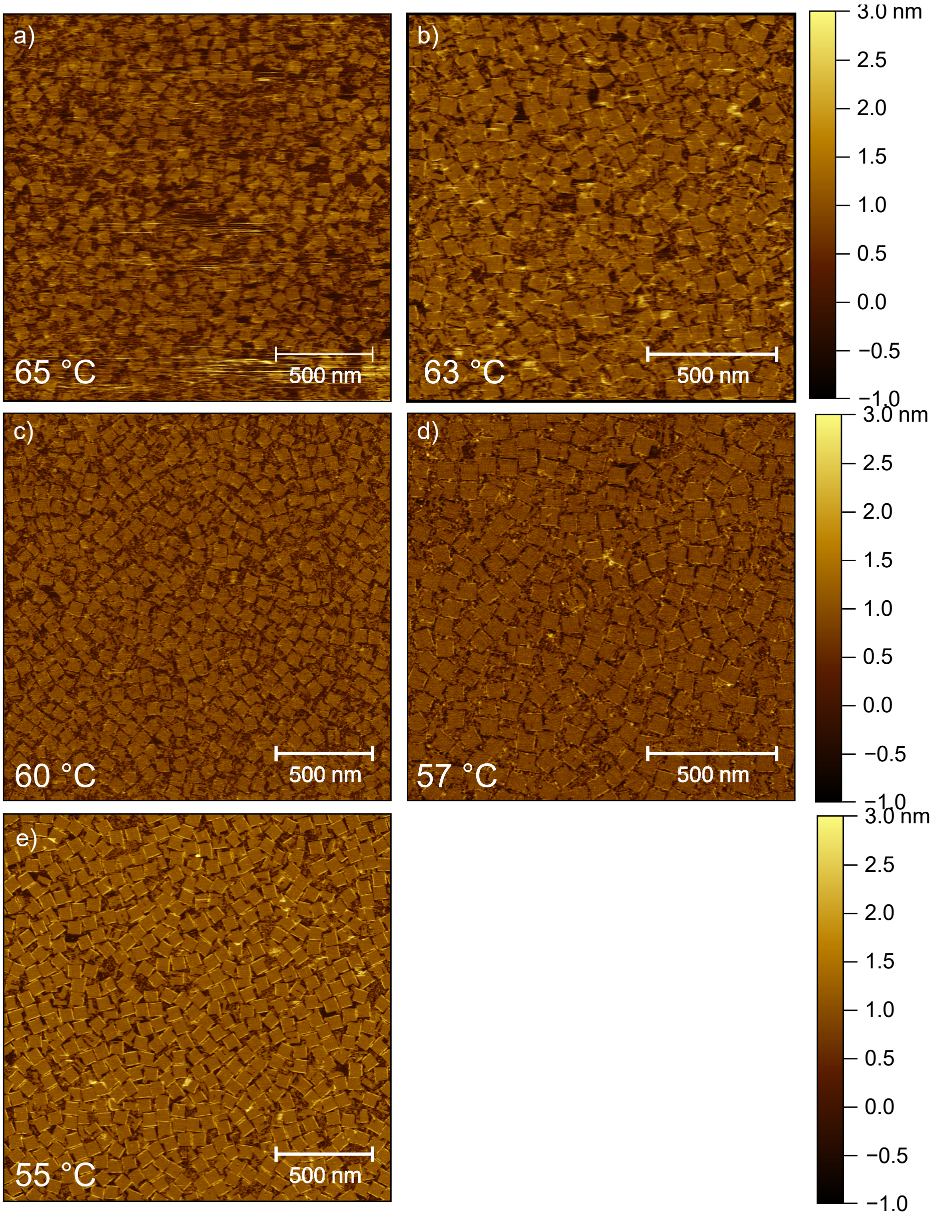


Figure 5.6: AFM of assembly of DNA origami structures present in the samples as a function of temperature stop.

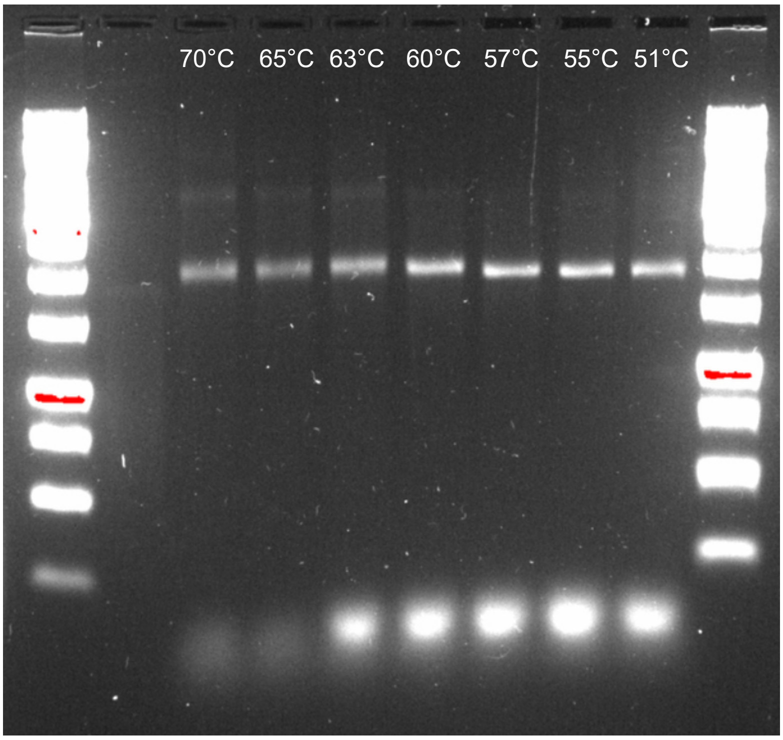


Figure 5.7: Agarose gel of assembly of DNA origami along a temperature annealing. The first and last lanes contains Generuler 1kb ladder.

Evolution of the origami assembly along a temperature annealing. We wanted to observe the evolution of the assembly of the origami during a temperature annealing. Along with annealing, we collect every so often a tube sample from the PCR machine, cool it in an ice bath, and store it at 2 °C. [Wag+17] proposes a similar protocol, but does a fast freezing to stop further assembly in liquid nitrogen. Each tube is assessed both on an agarose gel (Figure 5.7), and using an AFM image (Figure 5.6). The AFM shows a significant progression at 57 °C. There are with still some unformed origami, but the one with the expected squared shape are mostly formed. At 55 °C we see only good assembled origami. On the agarose gel, we can look at the intensity of the most intense band, its location, and the presence of bands of higher molecular weight. Based on these three indicators, we conlude to a significant improvement in the assembly between 60 °C and 57 °C, in good agreement with AFM imaging.

qPCR observation of the DNA origami assembly. We propose a protocol to observe the assembly of our DNA origami using a qPCR hybridization curve. Similarly to a qPCR hybridization or melting curve for a DNA duplex, we measure the signal of the sample with a double-stranded intercalating dye (Evagreen). Extracting the assembly of the origami not straightforward because the excess staples contribute to a large part of the measured signal. To be able to isolate only the origami contribution from the staples contribution, we have two samples: 1) The *origami sample*, containing the scaffold and the staples in excess. 2) The *staples sample*, containing only the staples, in the same concentration as the *origami sample*. Each sample is replicated three times, in three wells of a qPCR strip. These strips are annealed in the qPCR machine with the following protocol: 1) The strands are denaturated 95 °C, for 5 minutes. 2) The fluorescence is read at this temperature 8 times, with 1 minute intervals (plus reading time of about 30 seconds). 3) The temperature is decreased by one degree, and the reading of step 2) is done. This step 3) is repeated until 25 °C.

Figure 5.8 shows the temperature ramp, and the measured signal. The minimum and maximum signals is the same for both curves on this figure, which is not always the case for each well in the same conditions. First we can see that there is not much signal evolution when staying at the same temperature in this conditions. Therefore we only consider the last measurement at each temperature. For each of the three wells who replicate the same sample, we normalized the curves to have the same minimum and maximum values (between 0 and 1). Then we averaged the signals of each sample, after checking there is no outliers. We obtain normalized and averaged curves in function of the temperature for each sample (Figure 5.9). The two curves already exhibit differences, which means that our protocol is able to retreive some information from the origami assembly.

To try to extract more precisely the contribution from the assembly of the origami, we discuss what influences the measured signals. There are three factors : 1) The quantity of double-stranded DNA. 2) The temperature sensitivity of the dye. 3) The photobleaching of the dye along the readings. We have checked using a hairpin strand, that when the amount of doubled-stranded DNA does not change, both the temperature sensitivity of the dye, and the photobleaching leads to an affine profile. We used this phenomenon to try to remove the excess staples signal, because at low temperature the origami is not expected to make further double-strands. We make the simplification that $\text{signal}_{\text{origami sample}} = \text{signal}_{origami} + c \times \text{signal}_{\text{staples sample}}$. We optimized the coefficient c to obtain a linear profile between 25 and 50 °C. It gives $c = 0.687$. We observe the curve for $\text{signal}_{origami}$, which increases gradually when the temperature decreases. Below 57 °C the profile is linear. The fact that the linear profile is on a larger temperature range that on what the constant c was optimized gives credibility to the linear assumption. The temeprature at which the profile is not linear anymore is 57 °C, which suggest than the end of the origami assembly is at 57 °C.

This method could be useful to optimize annealing protocols for complicated DNA origami. The fact that the processing only relies on the profile at low temperature, where no further assembly occurs, gives a lot of flexibility. For instance, one could try to change the temperature curve between 90 and 50 °C in many possible ways, while still keeping the temperature curve below 50 °C. The same samples could be reused many times for many different protocols. Smaller temperature steps, and a external monitor using a temperature probe could give more precise results.

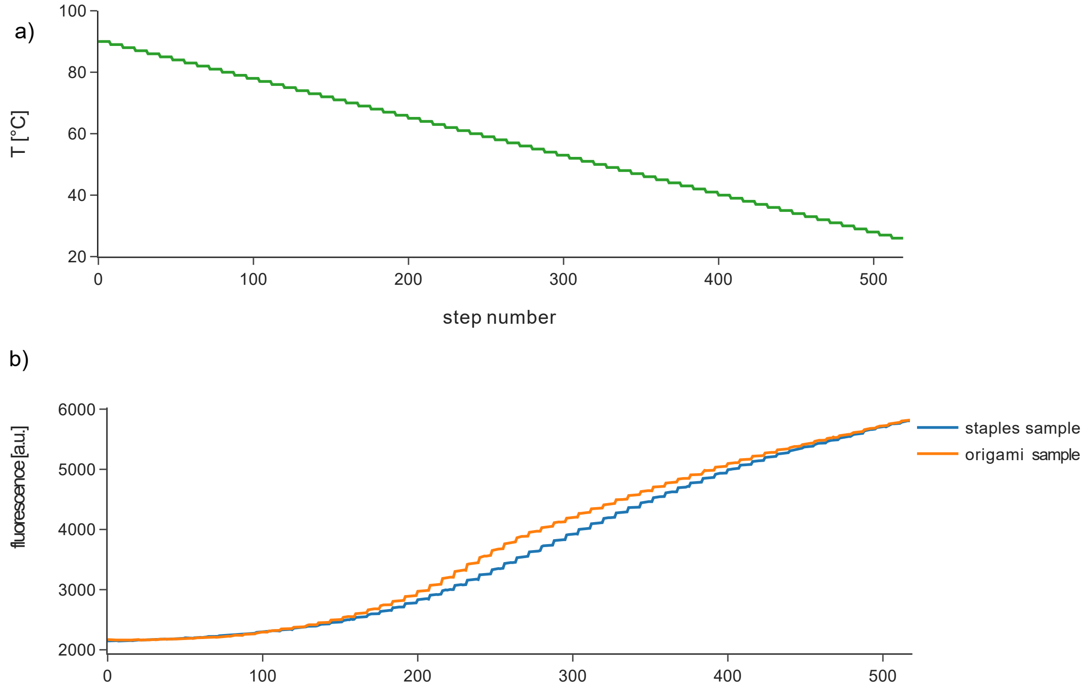


Figure 5.8: a) Temperature of the qPCR at each step, in the chronological order. The temperature starts at 90 °C and stops at 25 °C. b) Fluorescence measurement for two particular wells. One well one contains the *origami sample*, and one well contains the *staple sample*.

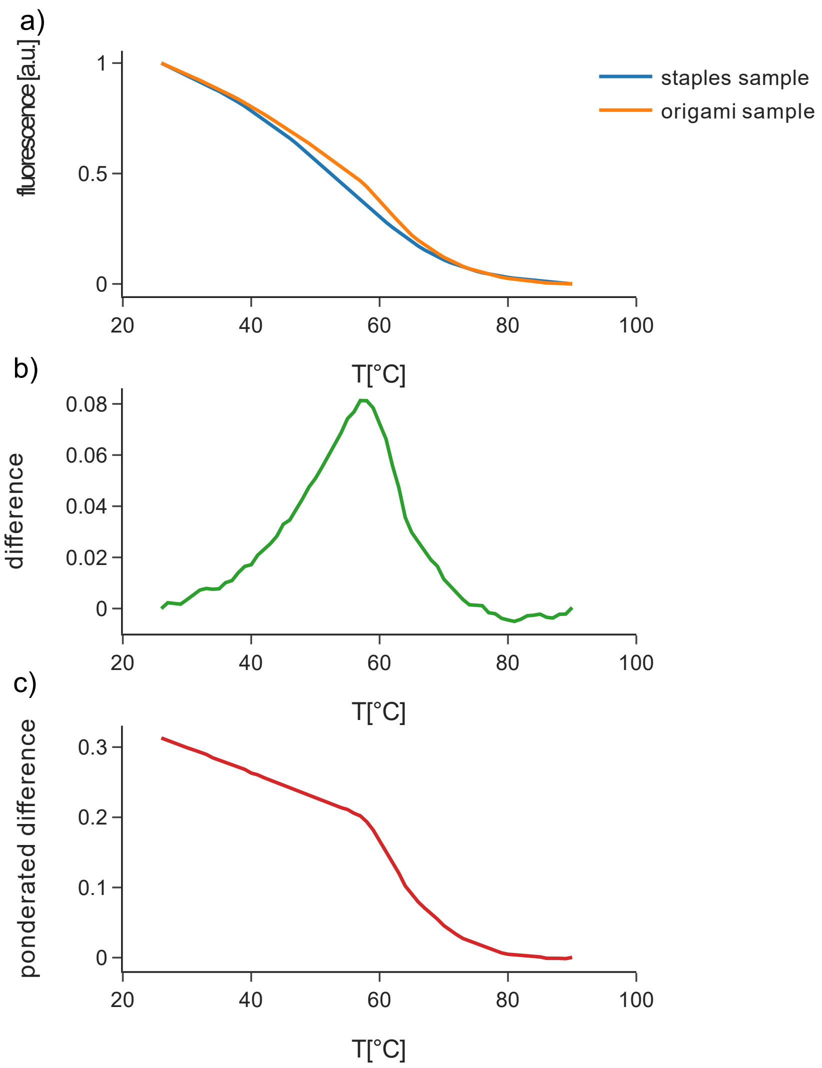


Figure 5.9: a) Fluorescence signal of the last measurement for each temperature step. For each sample, all replicants are normalized between minimum and maximum values, and then averaged. b) Difference of the averaged curves between the two samples. c) Ponderated difference of the averaged curves between the two samples ($\text{signal}_{\text{origami sample}} - c \times \text{signal}_{\text{staples sample}}$).

We have studied the evolution of the assembly of the 6-helices bundles origami when subjected to a temperature annealing. We used two methods, one method already described [Wag+17] using AFM and agarose gel, and other one based on fluorescence readings with a qPCR machine. Our two methods agree with 57 °C for the end of the origami assembly at, despite not having the exactly same annealing rate (about −1 °C / 3 min for the AFM, and about −1 °C / 10 min for the qPCR method).

5.2 Assembly of Hard-coded Paths

In the previous section, we studied the assembly of the DNA origami that will be our substrate. In this new section, we will start to assemble paths on this substrate by attaching hard-coded paths. We will define two alternative versions, with respectively asymmetric and symmetric lengths of the sequence domains composing the strands. We will see that the symmetric length version, because of the more uniform domain lengths, provides a successful path assembly onto the origami.

5.2.1 Asymmetric paths version

Our asymmetrical form path has two particularities (Figure 5.10) : 1) The docking strands are all oriented 5’ to 3’ going out the DNA origami. 2) The lengths of the domains are diverse (17 nt,7 nt,31 nt). This implfies different assembly temperature for each domain. For instance, the 7 nt length of the docking strands might cause the path to assemble last, at room temperature. As expected, this design does not assemble well (Figure 5.11). We can observe many partial paths on the origami, but no complete ones. A profile aling one line in the AFM topology image shows the presence of the path visible as a single peak with height compatible with a DNA helix. The height of the path on the AFM is about 1 nm, which is smaller than the expected 2 nm for a DNA helix.This difference can be explained horizontally by the width of the tip.

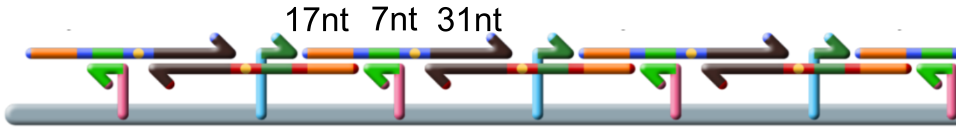


Figure 5.10: Assymetrical path design. The horizontal grey bar represents the top surface of the DNA origami. Vertical strands coming from this gray bar are "docking" staples. Yellow dots, represents the presence of a single-stranded flexible sequence TT allowing for turns.

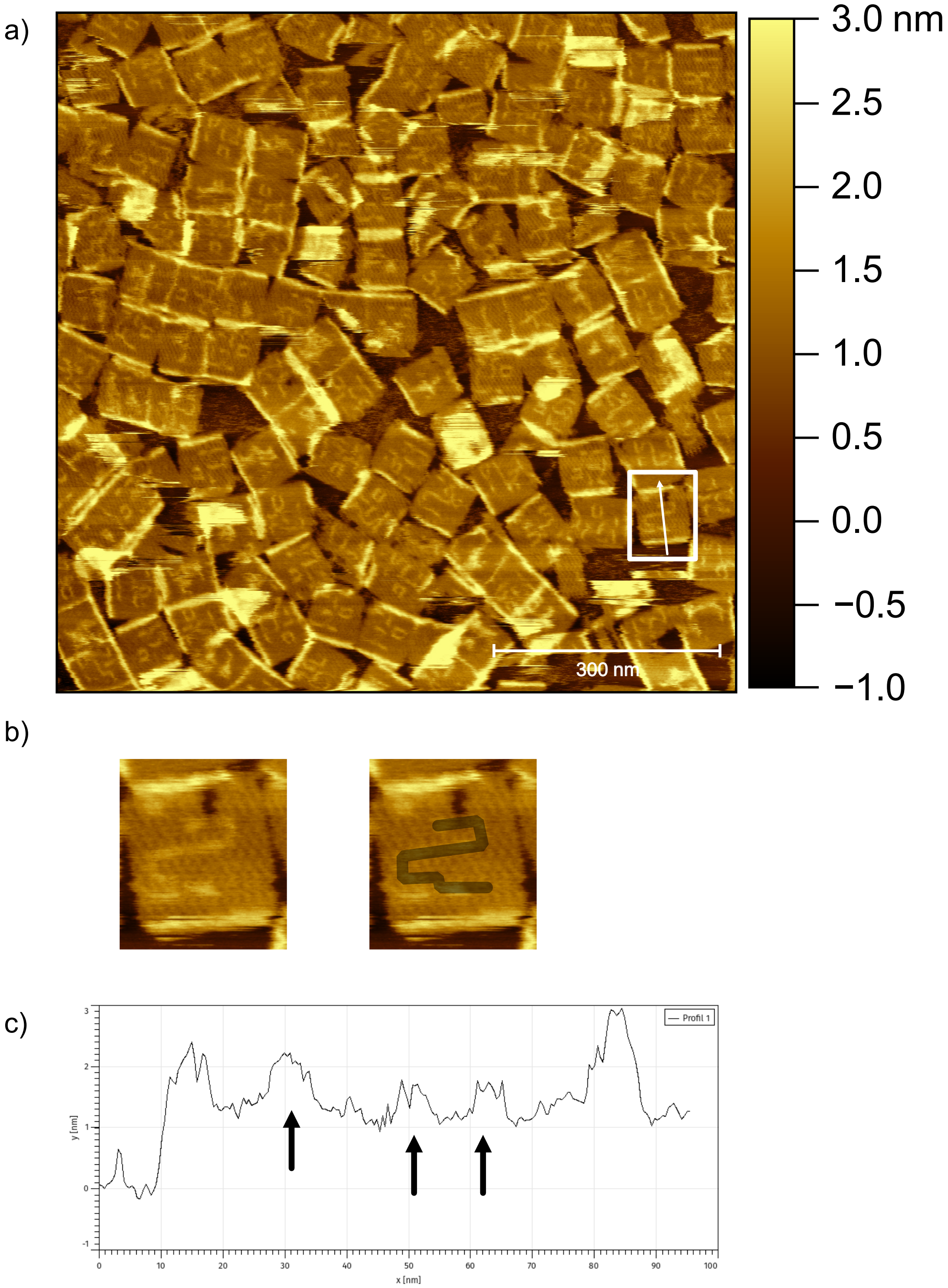


Figure 5.11: a) AFM image of the the assembly of the asymmetric path. The background shows the excess staples and path strands, as no rinsing was done after the incubation on the mica. b) Zoom on the assembly. The right panel emphasizes the intended path. c) Height profile of the origami (line of figure a). The black arrows on the profile indicate the section of the paths helices.

5.2.2 Symmetric paths version

The symmetric path version solves the domain unbalanced difficulties identified in the previous asymetric versions. This is allowed by the alternating direction of dock staples. As a results the three domains that are attached per path strand (Figure 5.12) have similar lengths, with 16 nt, 15 nt, and 16 nt. As we can see on Figure 5.13, this led to the perfect assembly of path with a good yield. However, a large majority of the origami are landed on the wrong side (with the indicator on the top right corner), as explained previously. We can see on the profile (Figure 5.13) the presence of two path helices (indicated by the black arrows). The height of these path helices are abnormally small, which is an artefact of the large size of the AFM window. When the origami lands on the wrong side, being under the surface of the origami, we see a merged peak. This merged peak has a height closer to the expected one because of its extended width Profile 2 on Figure 5.13.

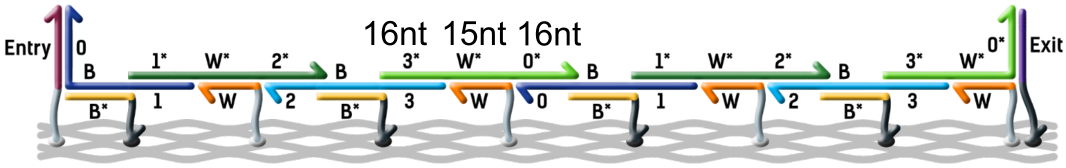


Figure 5.12: Symetrical path design. Alternating direction of dock staples allow the three domains attached to a single path stand to have similar lengths (16 nt, 15 nt, 16 nt).

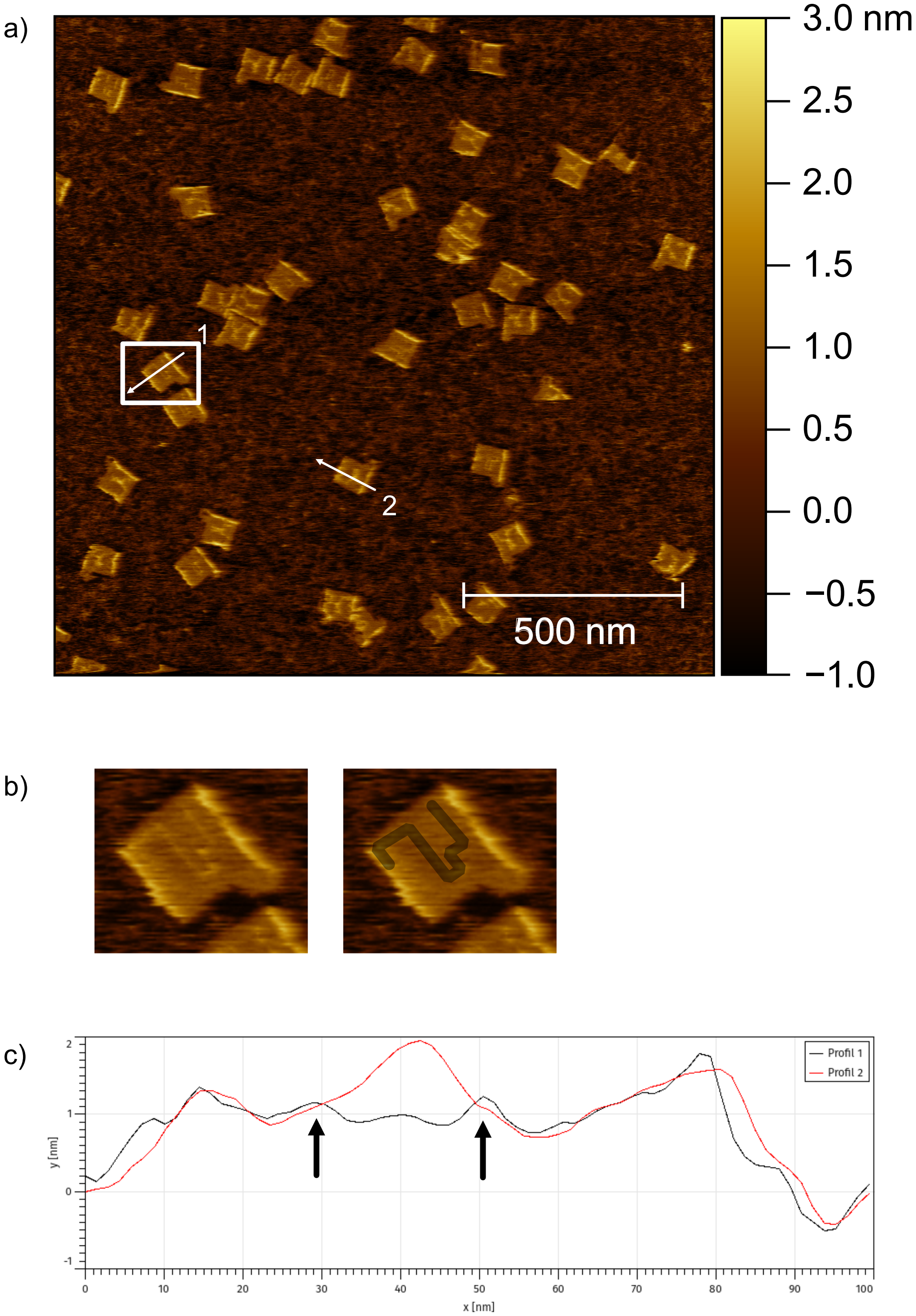


Figure 5.13: a) AFM image of the the assembly of the symmetric path. The background shows the excess staples and path strands as no rising were done after the incubation on the mica. b) Zoom of one assembly. The right panel emphasizes the intended path. c) Height profiles of origami landed on the correct side (profile 1) and on the incorrect side (profile 2). The twp sections of the path (black arrows on profile 1), appear merged when hidden under the origami (profile 2).

In this section, we saw that the asymmetrical path version , due to a different assembly temperatures, has a huge impact on the yield when proceeding to one-pot annealing. By alternating the dock directions, we have balanced the domain lengths, and successfully assembled a hardcoded segmented path on our DNA origami. This validates the functionalization of our DNA origami, and the ability to visualize the attached paths. Our results also highlighted the importance of controlling the landing face of the DNA origami to allow for a good AFM image analysis.

5.3 Assembly of non-hardcoded path : set of 4 universal strands

5.3.1 Removing the hardcoding of the symetric path version

After the validation of the previous design, we tried to relax it by keeping only four sequence for four path strands. The same annealing protocol produced a good assembly. Figure 5.14 shows the AFM of the path attached to the origami.

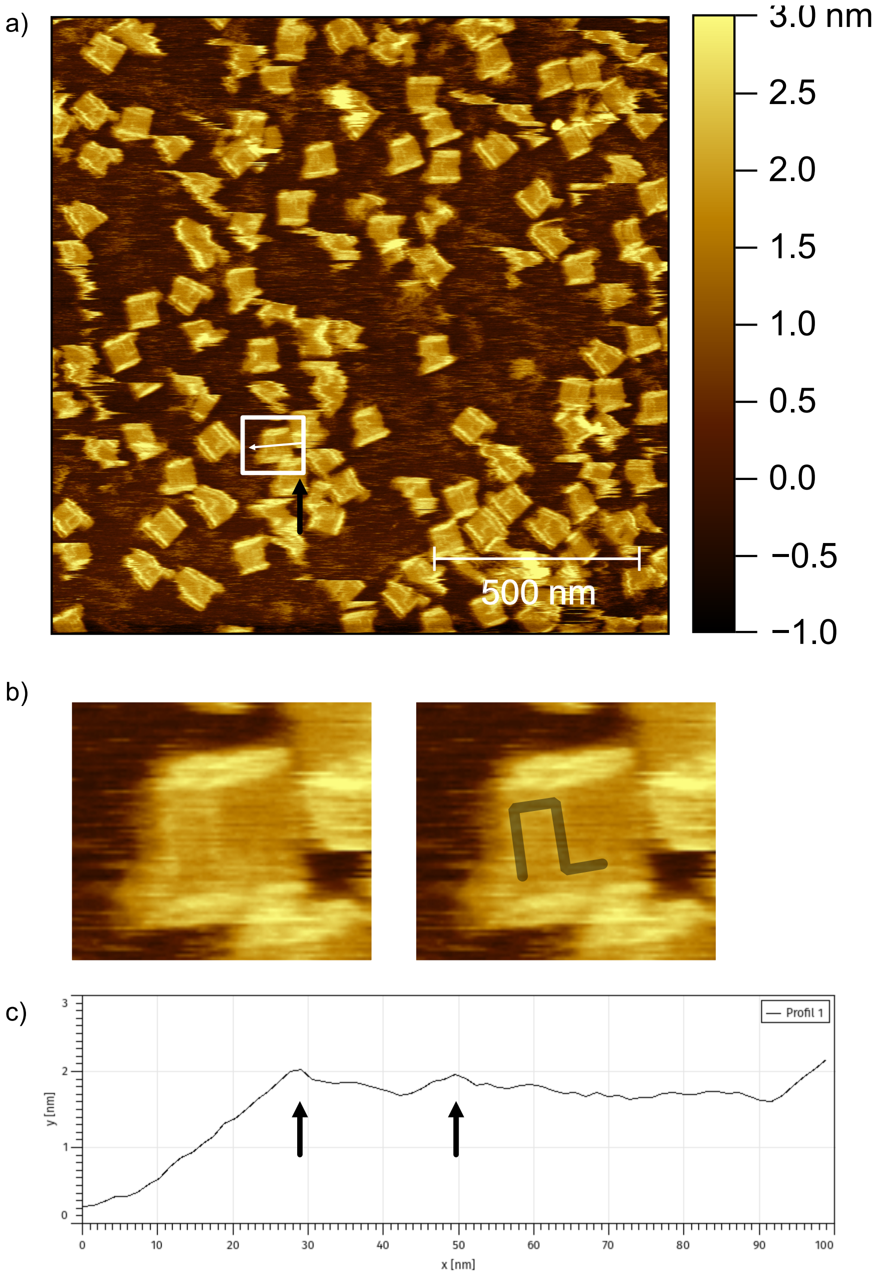


Figure 5.14: a) AFM image of the the assembly of the 4-universal symmetric path. The AFM tip was double-ended which produced a visual duplicating artefact. b) Zoom on one assembly. The right panel emphasizes the intended path. c) Height profile of the origami (line of figure a). The black arrows on the profile indicate the section of the paths helices.

5.3.2 Towards a dynamic assembly

With the evidence that we can assemble a non fully hardcoded path onta a DNA origami, we tried to control the kinetics of its assembly. We thought that before trying to implement a random walk behavior by controlling the temperature, we could try to selectively assemble the path, due to the presence of the initiating "start" dock: This required path strands with smaller domains to reduce the assembly temeprature. Indeed, we could measure the DNA melting temperature of our path near 55 °C close to the assembly of our DNA origami. As a result, the path partially assemble during the assembly temprature of the DNA origami, excluding any hope of a control of the dynamics. This is why we chose for smaller and less energetic domains of 10 nucleotides (Figure 5.15). The reduction of the domains induced single-stranded section, that we fill with a poly-T sequence to avoid secondary structures.

Metling curves of pairs of path strands. To check that the path assembled in the right temperature region, we measured qPCR hybridization curves with pairs of paths that attach together. Technical limitations forced us to set the DNA concentration the curves at 1 µM. The curves are interpreted in terms of duplex concentrations, and then adapted using the law of mass action to a different range of concentration. We can see that at 100 nM, that is a 10X excess for 10 nM origamis, all the pairs start to assemble near 50 °C. This means that we can try to induce assembly between 50 °C and 55 °C.

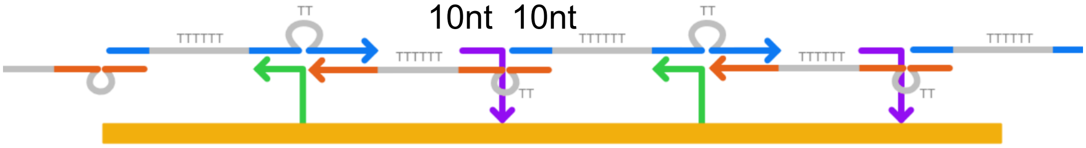


Figure 5.15: Weaker binding symetrical path design.

Testing the impact of the start dock. We want to control and selectively induce the assembly of the path using a "start" dock. To achieve this, we compare the assembly of a line-shaped path, in two situations. In the first situation, a regular dock is used. In the second situation, the first dock is a "start" dock, with an extended domain (20 nt). We expect the extension of the "start" dock to nucleate the assembly of the path. We proceeded to an annealing of the origami to 55 °C, left the temperature for 30 minutes, and then added the path strand to the tube, followed by an annealing from 55 °C to 40 °C. In both cases the yield is quite low. The experiment is therefore difficult to interpret. As a first answer, we tried to count the number of origami (landed on the right side) on the first AFM image, with at least one strand attached in both versions. The version with the start dock is more favored (43%, $n = 54$), compared to the control one (20%, $n = 31$). This confirms that the "start" dock extension encourage some assembly. However it does not prove that the assembly order goes from the "start" dock to then end of the path. More importantly the absence of the "start" dock does not prevent the path assembly.

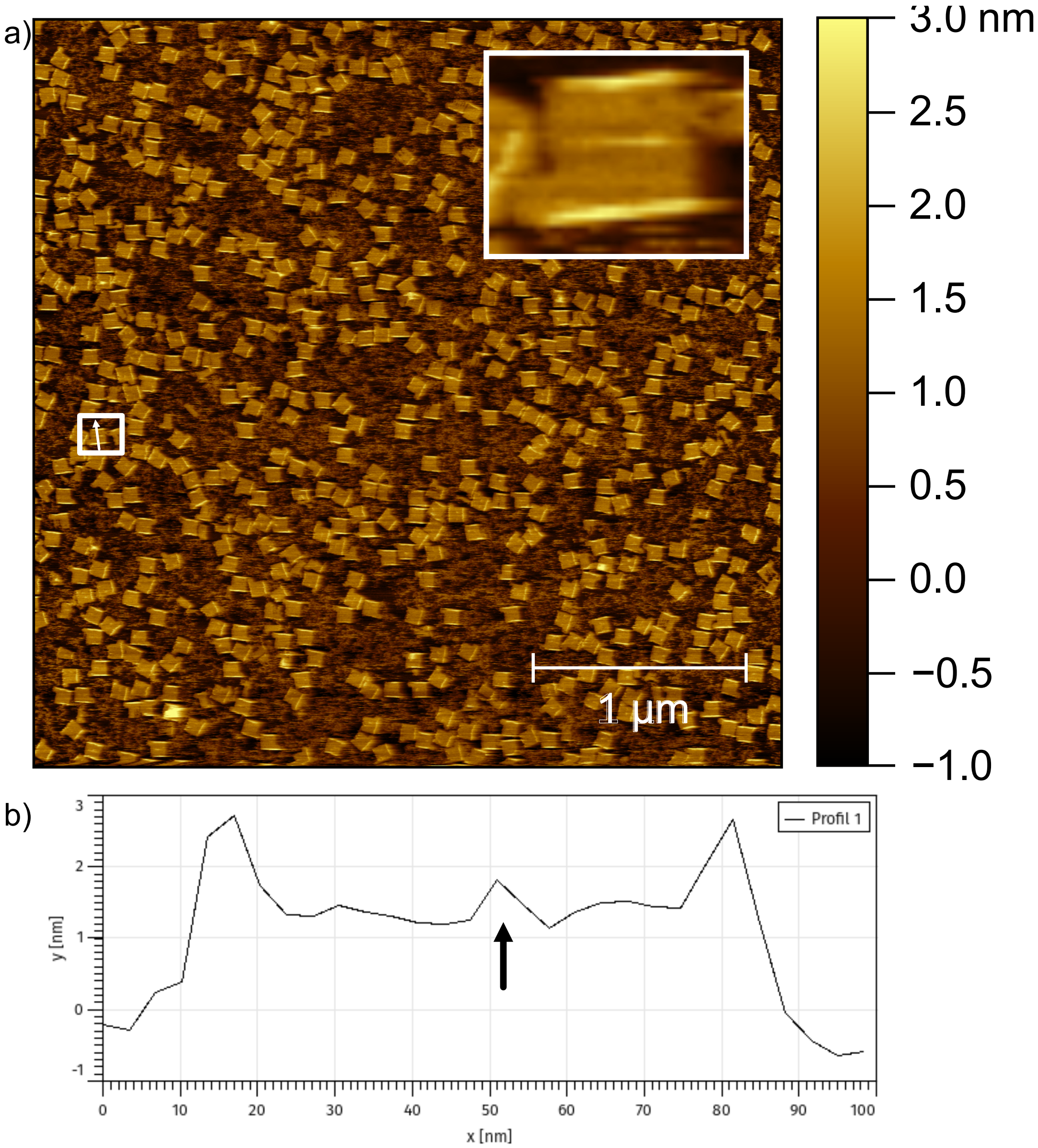


Figure 5.16: a) AFM image of origami with the 4-universal asymmetric path. The origami does not have a start dock with an extended domain. b) Height profile of one origami. The black arrow points to the section of the line-shaped path.

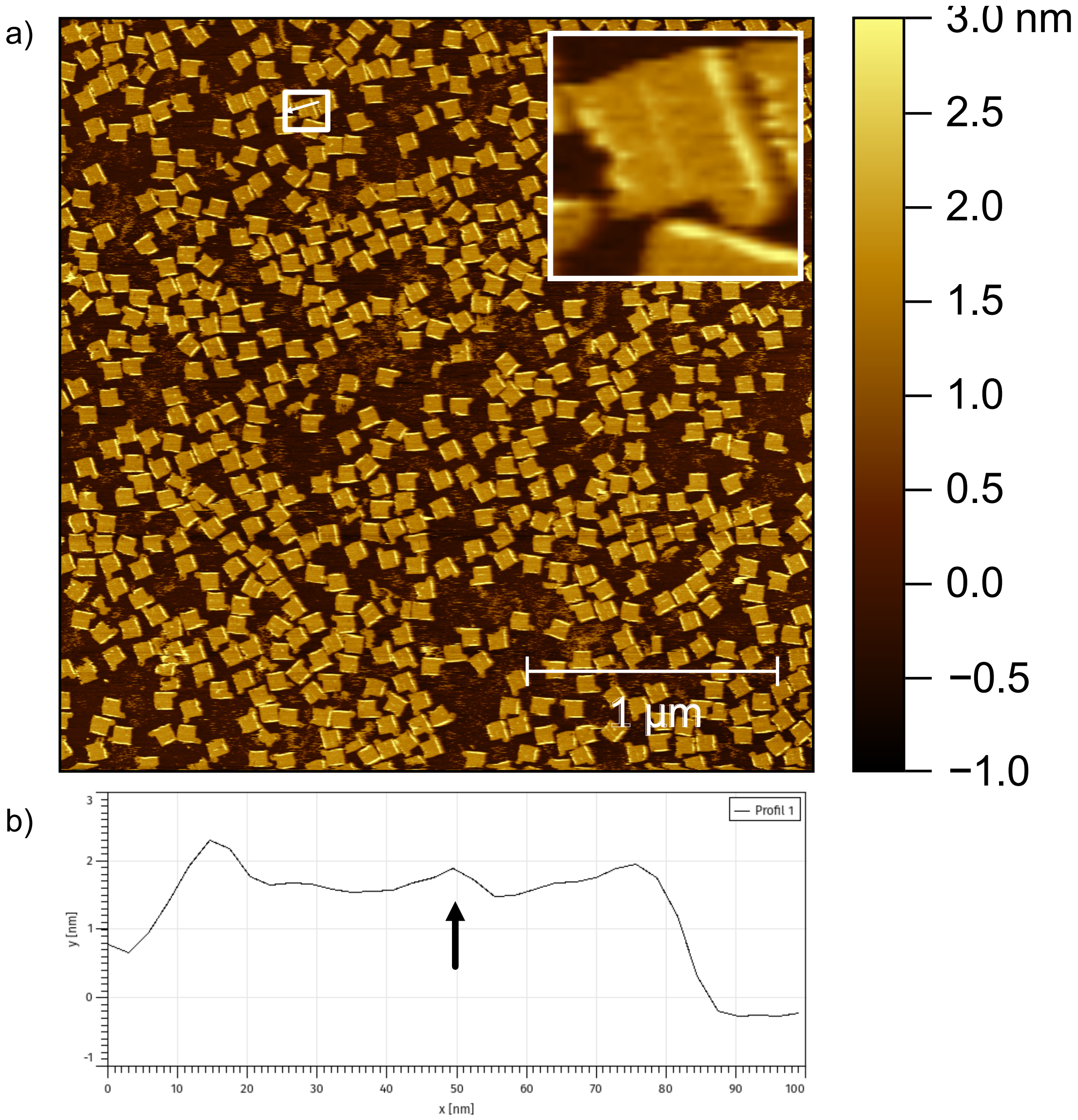


Figure 5.17: a) AFM image of origami with the 4-universal asymmetric path. The origami does have a start dock with an extended domain to try to nucleate the path assembly. b) Height profile of one origami. The black arrow points to the section of the line-shaped path.

In this section, we tried to relax the hardcoded segmented path design. We selected four sequences for four DNA path strands. A regular one-pot annealing did not assemble the path. We successfully assembled it, by annealing first the DNA origami, and then adding the path strands, and cooling to a experimentally determined temperature. However and the assembly of the DNA origami and of the path share a common assembly temperature range, which prevents further experiments where the assembly order is precisely controlled. To be able to control the kinetics, we designed a path design with weaker domains to assemble at lower temperatures. First, the assembly with a temperature analysis was reproduced. Then, our goal was to find a protocol where only the presence of the "start" dock would induce the assembly. Preliminary results indicate that the "start" dock leads to more path strands attached to the origami, but it does not guarantee a full assembly. Also, in our condition, the absence of the "start" dock did not prevent any path strand to attach. Finally, we did not manage to find a convenient way of comparing the very similar looking AFM images. Without such analysis metric, we did not know how to compare further experiments.

5.4 Conclusion

To conclude, this part taught us valuable information for future experiments with DNA origami. First we designed two origami with different rods designs on the edges preventing bending, that were successfully assembled. Then we validated the technical capacity to observe a single helix-wide path attached to a DNA origami. We were also introduced to the issue of controlling the landing side, independently of what is attached to the origami. Finally we managed to assemble a hardcoded path our DNA origami, and then generalize it by reducing to a number of four strands. We tried to iterate the design to further control the assembly, but unsuccessfully. Even if further experiments could have been conducted on this system, we chose to switch to a more promising approach. In am going to describe in part II an alternative path design using a variation *toehold exchange strand displacement technique*.

Bibliography of the current part

[Thu+17] Anupama J. Thubagere et al. “A Cargo-Sorting DNA Robot”. In: *Science* 357.6356 (Sept. 2017), eaan6558. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aan6558`. (Visited on 05/01/2021).

[Wag+17] Klaus F. Wagenbauer et al. “How We Make DNA Origami”. In: *ChemBioChem* 18.19 (Oct. 2017), pp. 1873–1885. ISSN: 14394227. DOI: `10.1002/cbic.201700377`. (Visited on 05/01/2021).

Part III

Strand Displacemement Random Walk

Chapter 6

Roadmap for a DNA Random Walk

Outline of the current chapter

This chapter defines our roadmap for building random walk assembly behavior on the DNA origami using DNA strand displacement. After describing the principle of the design, we propose two configurations to test and validate this design experimentally. Finally, we will end up with three simplified designs that could be iterated following experiments presented in the next two chapters.

6.1 Design Description

Motivations. The random walk will be encoded as multiple *step* strands bound one to another (Figure 6.1).

Here we want to use toehold mediated strand displacement (TMSD) and more precisely a variation of toehold exchange strand displacement (TESD) to control the progression of the random walk. We see two benefits. First, TESD has minimal leaking, particularly in situations with small sets of strands like ours. More importantly, the fact that displacement controls the rate of both assembly and disassembly should ease the balance of the walk.

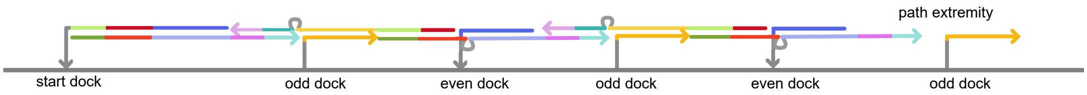


Figure 6.1: Schematic sideview of a random walk path. The greyline represents the top surface of a DNA origami.

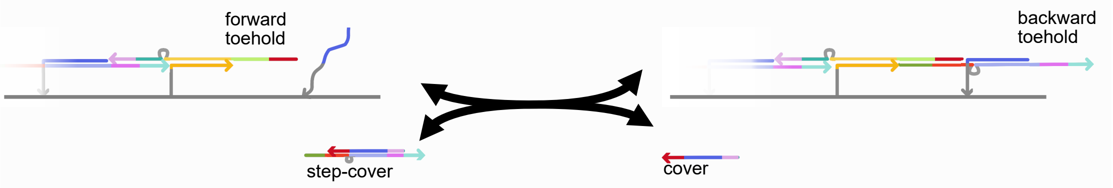


Figure 6.2: Random walk mechanism using strand displacement. The grey line represents the top surface of the DNA origami.

The reversible strand displacement mechanism. The core mechanism of the random walk is composed of two entities (Figure 6.2) : a *step strand* , hydridized with a protection to form the *step cover* complex; and the *cover strand* (as a ssDNA monomer). The purpose of the *step-cover duplex* will be to move the random walk forward. On the other hand, the purpose of the *cover strand* will be to move the random walk backward.

Now we explain the dyamics that implement these purposes. We consider a path composed of several *step strands* assembled on the origami. The last *step strand* (refered to as the *terminal step*) exposes the *forward toehold*. When a *step-cover duplex* encounters this toehold, a strand displacement mechanism can occurs. If the strand displacement succeeds *cover strand* is displaced and freed up in solution. The *step* therefore remains hybridized both to the former *terminal step*, and the *dock*. However, this process is reversible: A *cover strand* may encounter the now exposed *backward toehold*. The same displacement, but this time in reverse order detaches the path, and free up a *step-cover duplex* in solution.

We can notice that this process is indeed a variation of TESD, with the branch migration domain cut in two strands. We know that this will add a large energy barrier to the transitions. This is expect to slowdown the (forward) displacement rate by several orders of magnitudes compared to regular strand displacement.

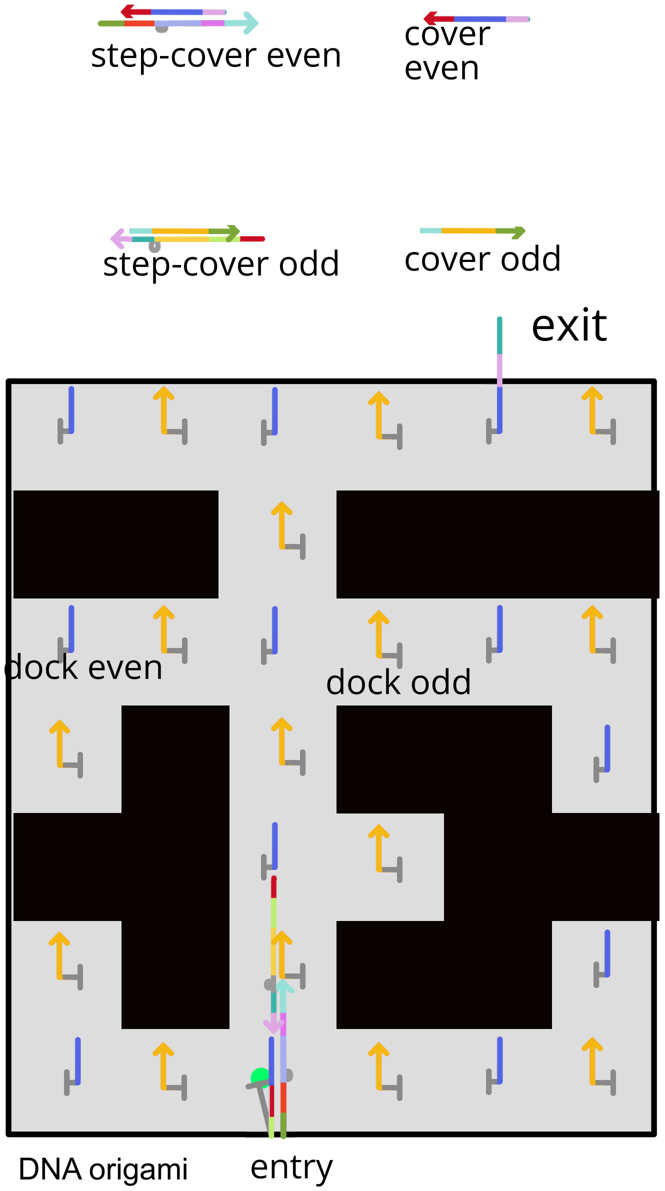


Figure 6.3: Checkerboard coloring of the steps, covers and the docks

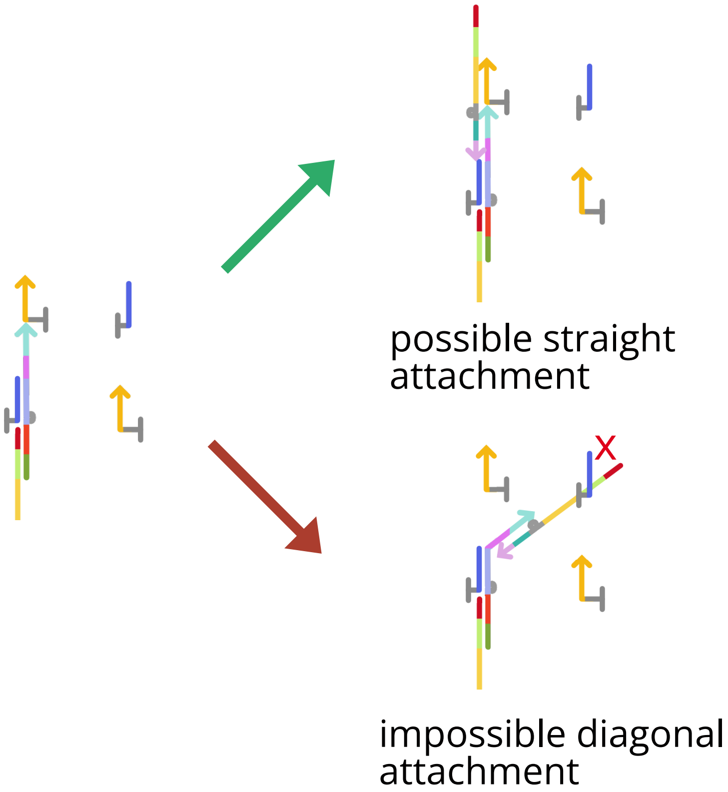


Figure 6.4: Due to the checkerboard coloring, there is no possiblity of attachement of path in a diagonal way.

Checkerboard paths and docks The design with *step strand* bound one to another needs the *step strands* to be arranged in an alternating sense. This imposes to make two versions (colors) of both the *cover strand* and *step strand* (Figure 6.3). As an additional benefit, it prevents the path to skip a cell, by attaching a path in diagonal (Figure 6.4). Have could notice that the two *step* versions share the same pair of toeholds. But the toeholds play an opposite roles for each: the forward toehold of the even *step strand* is the backward toehold of the odd *step strand* . Therefore, the only differences between the two versions are the sense of the duplexes, and the sequences appart from the toeholds.

Turns As in part I, we choose to use TT for the sequence for these "flexible" domains. Similaly to part I, the dock complementary to the *step* makes a nick (gap) to the previous *step*, which also allows for a turn. Figure 6.5 shows that no matter the parity of the *step* located a a turn, this TT (round gray ball) allows for the path to make turns without inducing a physical offset.

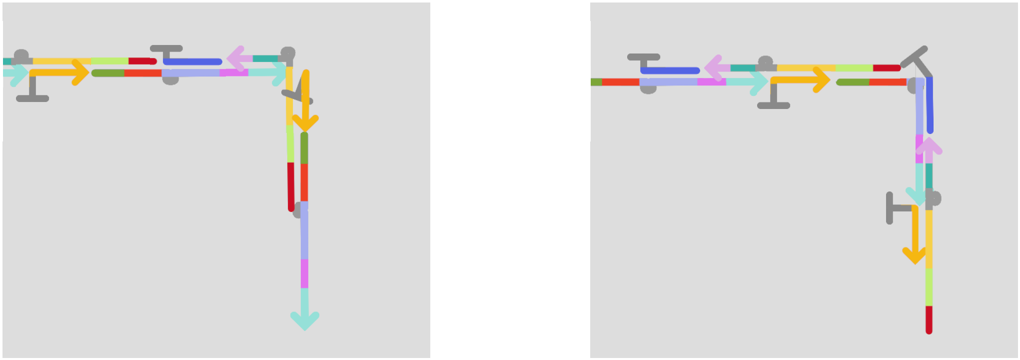
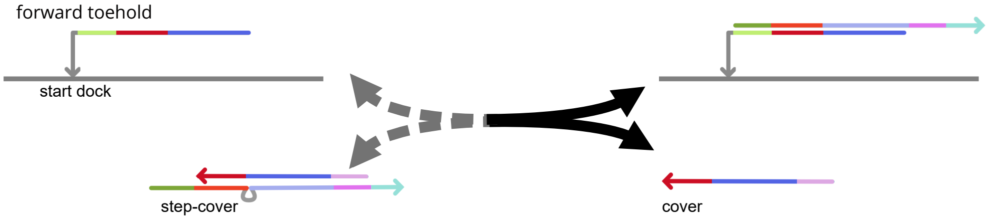
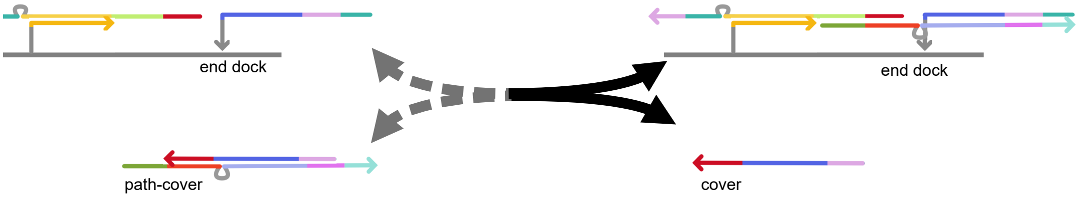


Figure 6.5: Path turns using TT sequence. a) and b) reprents the situation for both *step* parities.



(a) Entry dock strand for even cells.



(b) Exit dock strand for even cells

Figure 6.6: Propositions for entry and exit dock variations.

Entry and exit Specific *dock* strands are necessary to encode the entry that will initiate the path progression, and the exit that will stop the path from going backward. Figures 6.6a and 6.6b show these strands for even cells, but the one for odd cells can be constructed similarly. We can note that speaking about energy, the binding of a *step* to the "start" dock is more stable than the binding to both regular dock and a previous *step*. For the "exit" dock, it binds to the backward toehold, which should inhibit path removal by the invasion of a cover strand. However, this implementation of the exit is not correct because it can initiate a backward path assembly. An alternative implementation using a hairpin dock should both let the path assemble, and prevent the use of the extended domain for the assembly of non-terminal paths. Combining these two phenomena could probably solve the problem but requires energetic fine-tuning, which we will not explore in this work.

Concentration of path-cover and cover The free floating *step-cover duplex* and *cover strand* , will need to be in excess in solution, for two reasons. We need enough path strands to explore all accessible cells on each origami. Also, the progressive assembly of the path will deplete the concentration of *step-cover duplex* , and increase the concentration of *cover strand* . This will in turn favor *step* removal over time. Adding an excess of both *step cover* and *cover* will mitigate the impact of this concentration unbalance.

6.2 Experiments Design and Constraints

6.2.1 Random Walk Rates Adjustment

Forward and backwad rates are unlikely to be exactly the same. We discuss how to experimentally change conditions in order to balance these rates.

Toehold sequence energy The energy of the toehold sequence has a significant impact on the strand displacement rate (from $1 \times 10^3\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ to $3 \times 10^6\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ [ZW09]). In our case, we have two toehold sequences, each one playing both roles of path attachment and path detachment for, respectively, odd and even cells. Because of this, it is unclear whether changing the sequence of one toehold will help to balance all rates. However, in the experimental section, we will see that we ran experiments with a sequence mismatch in the toehold of only the *step strand* . This technique could be used to reduce the toehold energy of a specific pair if we observe a large imbalance.

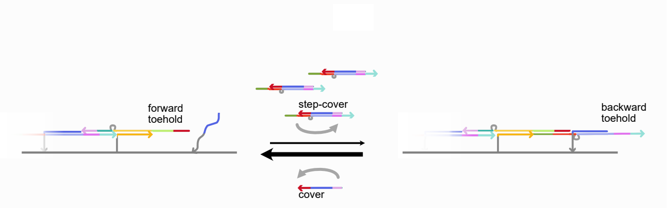


Figure 6.7: Example of balancing of the random walk by increasing the concentration of *step cover*.

Concentration of step-cover and Cover Using the concentration imbalance of step-cover and cover can help to balance the random walk. For instance, Figure 6.7 illustrates a situation where the random walk mechanism is biased towards *step strand* removal. The presence of more *step-cover duplex* compensates this bias, and can result in a balanced random walk.

6.2.2 Leaking on DNA Origami

WWe discuss here the consequences of duplex rearrangements without the proper use of a toehold, called leaking. Because our design is reversible, there is some chance of recovery. Here we map all the possible consequences of two types of leaks. The first one is the leaking that causes a *step* strand to attach to a dock, despite the *cover* protectiom. The second one is the complete removal of an interor *step* by a cover, despite the connection of two adjacent *steps* strands.

Leaking of a step-cover on a dock : self nucleation of a path Leaking can attach a *step strand* to a single dock domain, without using a toehold (transition from state a) to state b) in Figure 6.8. Let $k_{a\,\mathrm{dock}}$ be this rate. Due to the presence of the cover, the expected order of magnitude is $1\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. The (unexpected) presence of *step* as monomers could also lead to this situation. In the result section, we will see that the measurement of this rate $1.2 \times 10^{5}\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. This behavior could be explained by monomer *step* in solution. If this happens, there are three cases.

- Case c) : because the backward toehold is exposed, the *cover strand* of this *step strand* can remove the step. The associated rate, $k_{r\,\mathrm{dock}}$ should be faster than k_b.
- Case d): because the step is not attached to a previous step, it exposes one toehold and one domain. The associated rate $k_{a\,\mathrm{long\ toehold}}$ is expected to be greater than k_f, but of similar magnitude..
- Case e) : As a terminal step, it exposes the toeholds and domains for path extension. There is a possibility of path extension, with the regular rate k_f.

We can expect the k_f rate having a slowdown of a factor ten compared to regular strand displacement with the same toehold. Then the probability of recovery by removing the *step* can be estimated to $(10/(10+1+1)) = 0.83$. To summarize, the rate of path nucleation should be very low, and has a high probability of recovery if it happens.

Now we discuss the consequences of the removal of an interior *step strand* by a *cover strand* (transition from state a) to state b) in Figure 6.9). The associated rate $k_{r\,\mathrm{interior}}$ is expected to be on order of $1\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. From this situation, there are three possible cases:

Leaking of cover removing an interior step.

- Case c) : as the previous path exposes toeholds for path extension, the hole can be filled with a new *step-cover duplex* . The associated rate k_{repair} is expected to be faster than k_f, and of the same magnitude.
- Case d): the previous *step strand* is exactly in the same situation as a *terminal step*. The associated rate is the regular *step* detachement rate k_b.
- Case e): finally, the next *step* exposes a toehold, and an additional domain. These two sequences are of 9 nt. The two will play the role of a longer toehold to detach the *step*. The associated rate $k_{r\mathrm{long\ toehold}}$ is expected to be higher than k_f, possibly of one or two orders of magnitudes.

In this situation, the presence of the long toehold in the case d) will not favor the recovery of the situation b). However, the interior step removal rate $k_{r\,\mathrm{interior}}$ is expected to be very low $1\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. In particular, compared to the nucleation of a path, here we do not expect experimental problems that would speed up this rate.

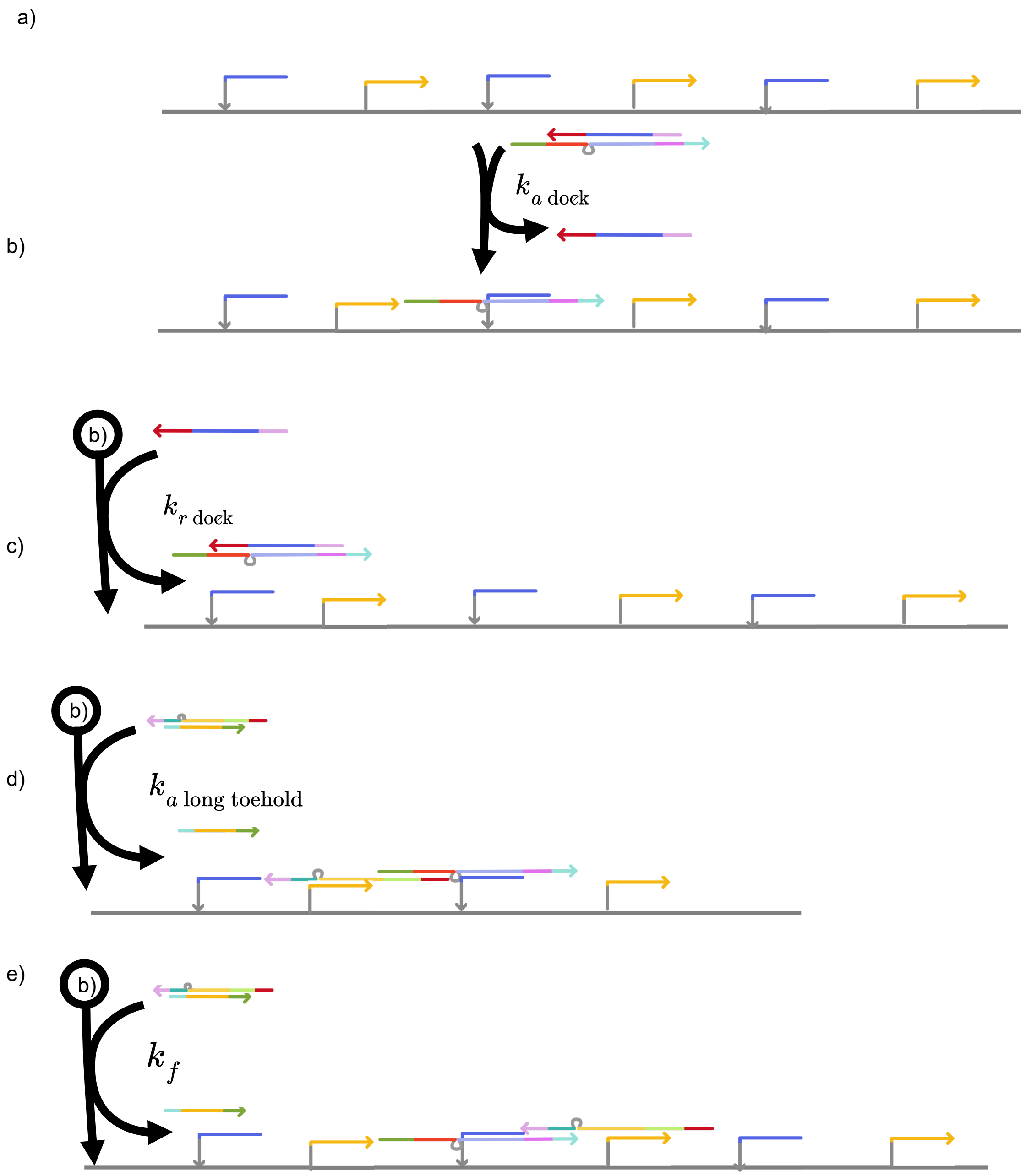


Figure 6.8: Leaking leading to self nucleation of a path on a dock

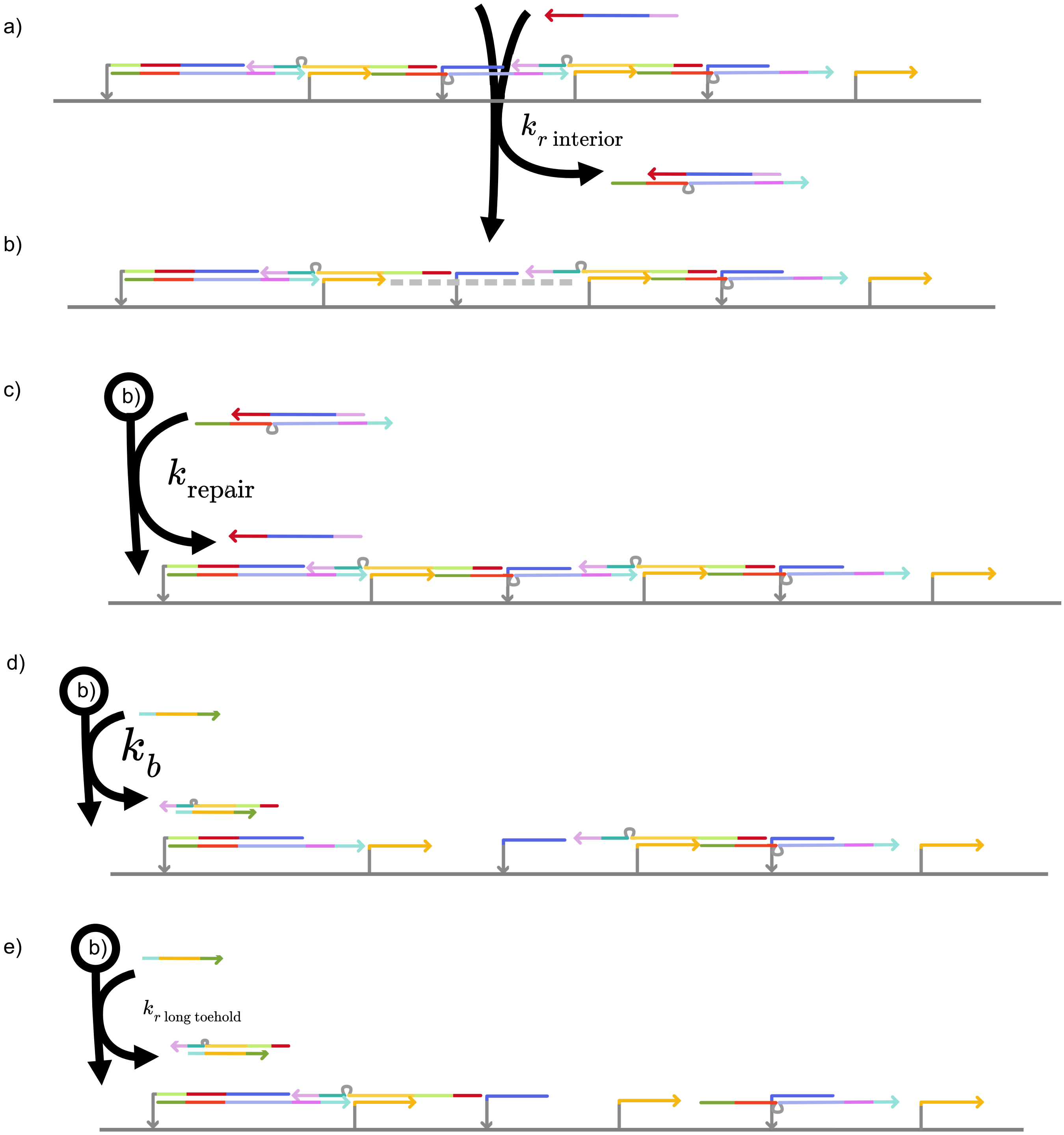


Figure 6.9: Leaking leading to the removal of an interior step

6.2.3 Path Self-Nucleation

Because of the excess of the two types of *step-cover duplex* in solution, they could polymerize one to another. However we expect the presence of *cover strand* in the about the same concentration to prevent this polymerization.

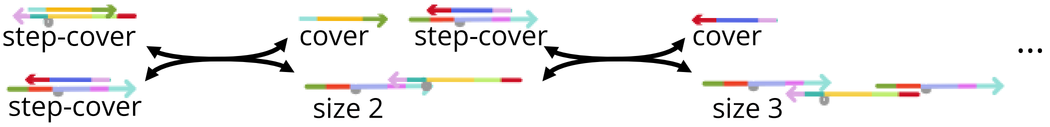


Figure 6.10: Spurious path polymerization

6.3 Experiments for Validation

We thought about two experimental designs, that would validate the princple of the random walk for our system. The first one, *branching*, is designed to be more qualitative, while the second one, *1D random walk* is more quantitative.

6.3.1 Branching

The branching experiment will mainly prove that the path has the ability to go forward and backward, and stop at the end dock. Information about the rates of the walk progression could also be infered. We propose three versions of the same maze with two branches of the same size (Figure 6.11). At the end, we expect to see that :

- Version 1 : with only the exit docking 1, only paths reaching the left end are visible.
- Version 2 : with only the exit docking 2, only paths reaching the left end are visible.
- Version 3 : with two exit dockings. There is a equal distribution of paths taking the left and right turns.

It would prove that the path have the ability to reverse backwards, and that at an intersection the probabilities are equal.

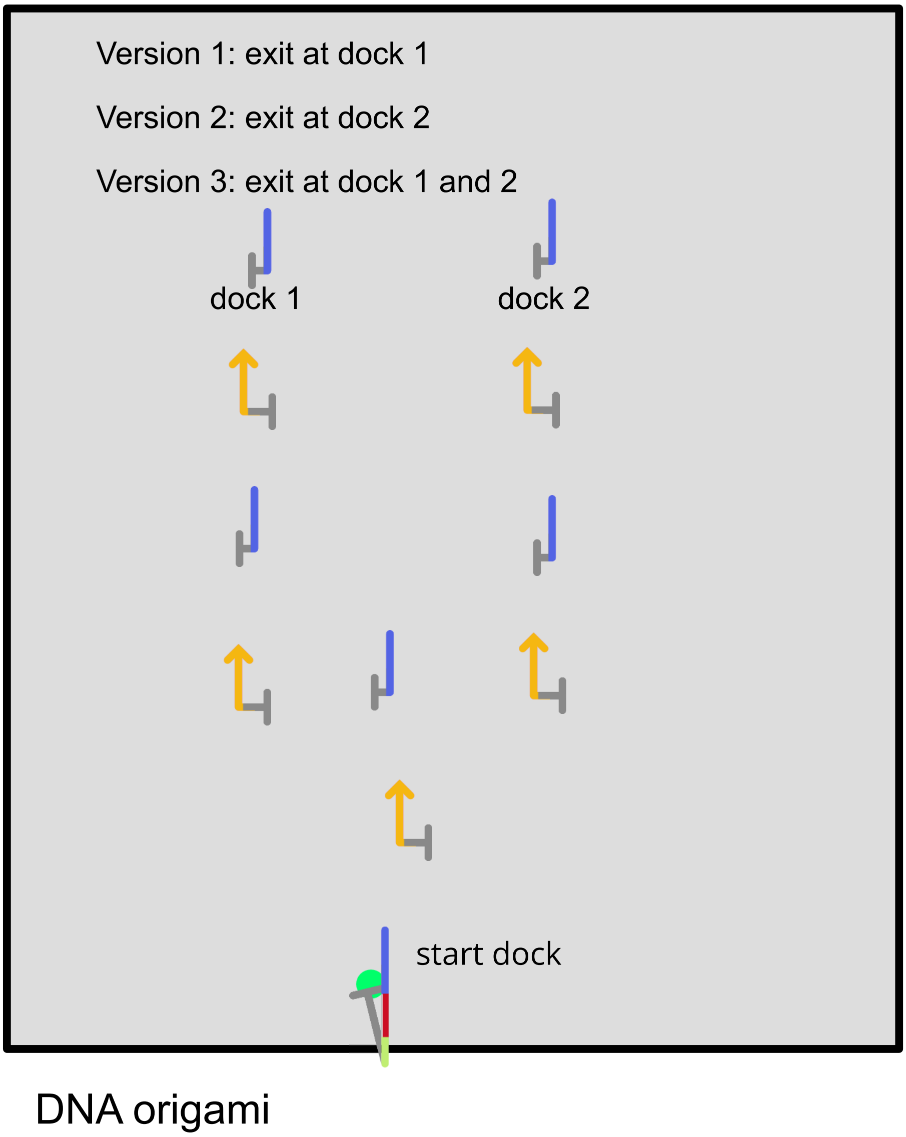


Figure 6.11: Branching experiment

6.3.2 One Dimentional Random Walk

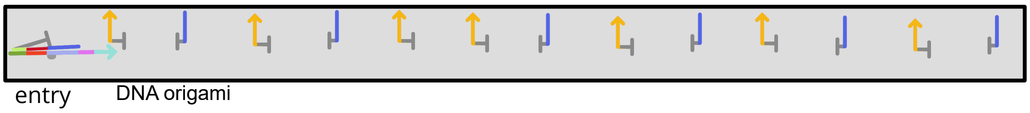


Figure 6.12: 1D random walk experiment design.

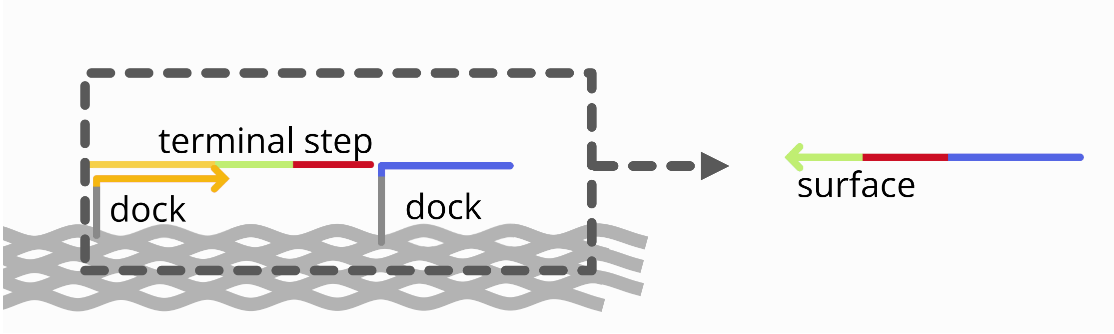
In this experimenal design, we make a simple *line maze* without any forks, of different length n. There can be two variations, with and without end dock. The mean time of first arrival to the end, that will be measured by adding an exit on the maze, is quadratic to length of the maze. This measure, possible with fluorescence or AFM could validate the presence of a random walk. Indirect measurements of the forward and backward rates could also be done.

6.4 Experimental Bottom-up Approach

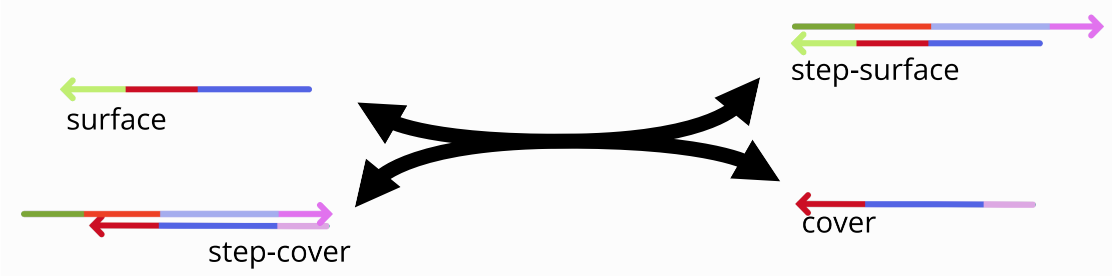
In order to simplify the system to be able to tune and assess all the strand displacement rates, we decided to propose a bottom-up approach with three toy models. They all model strand displacement *step* attachment and detachment to obtain insight for the associated kinetics and developp our experimental tools. The last one, Toy Model C, is extremely close to the final random walk design.

6.4.1 Toy Model A : One-strand Substrate

In the random walk design, at the end of the terminal *step strand* , three domains are colocalized (Figure 6.13a): the green toehold , the red domain (called previous path), and the blue dock. This Toy Model A consists of unifying these three domains into a single ssDNA molecule. It corresponds to regular TESD. Figure 6.13b describes the *step* assembly and removal in this toy model.



(a) Simplification for Toy Model A

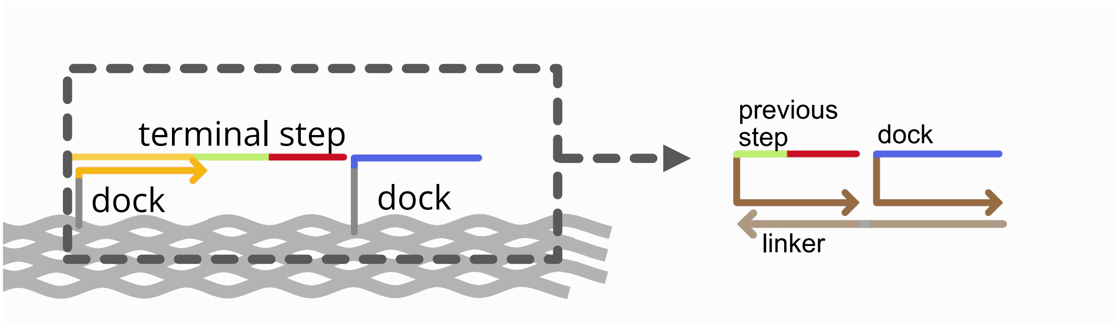


(b) Strand displacement path assembly and removal, expressed in the Toy model A

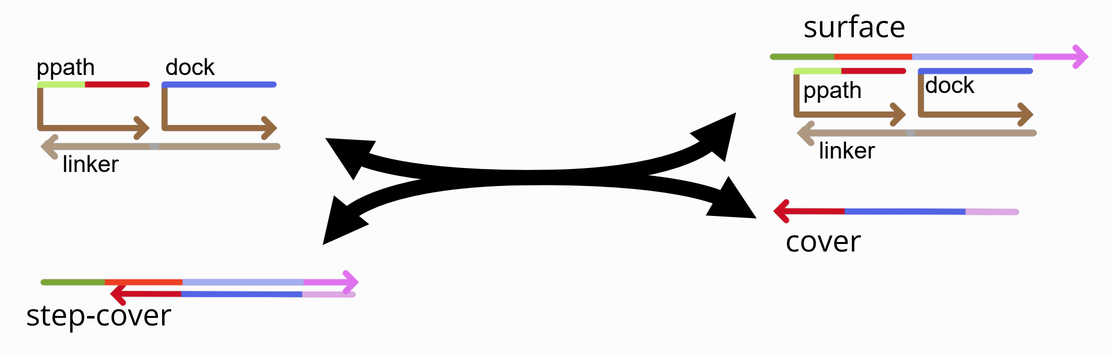
Figure 6.13: Toy Model A design : the origami is removed and its role is reduced the *surface* DNA strand.

6.4.2 Toy Model B : Two-strands Substrate

In this section, we propose a model where the surface strand of Toy Model A is now replaced by two colocalized strands (using a third linker strand). This toy model will account in some sense to the energy penalty cause by the splitting of the branch migration domain of TESD in two strands. The *linker complex* has two outgoing single strands. The first outgoing strand is the *previous step*, which contains exposed domains of the terminal *step strand* . The second is the *docking strand*, for the *docking strand* pointing out of the origami.



(a) Simplification for Toy Model B

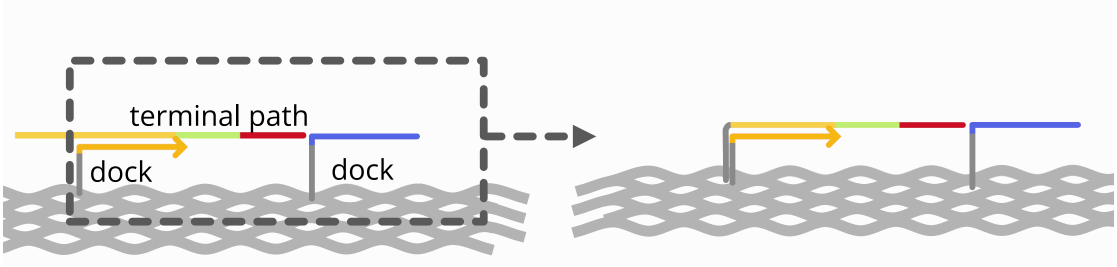


(b) Strand displacement path assembly and removal, expressed in the Toy model B

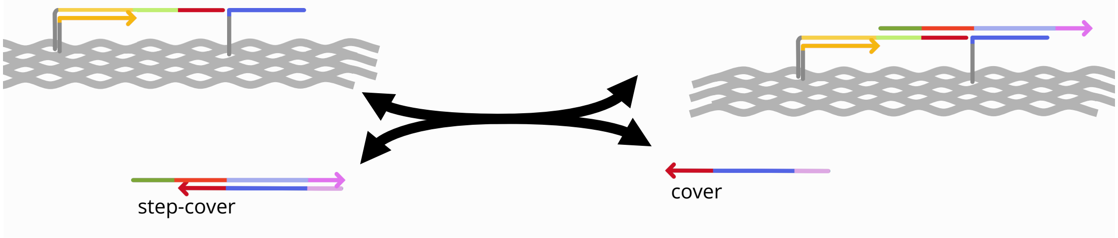
Figure 6.14: Toy Model B design : the origami with a previous *step* and a *dock* is mimicked as two strands combined together by a "linker" DNA strand.

6.4.3 Toy Model C : Origami Substrate

Finally the Toy Model C, exposes the same domains as the Toy Model B. The difference is that the two strands will come out of a DNA origami, instead of being attached on the *linker* strand. This way, the surface interaction, and the overall geometry of the system will be close to what happens in our goal random walk mechanism.



(a) Simplification for Toy Model C



(b) Strand displacement path assembly and removal, expressed in the Toy model C

Figure 6.15: Toy Model C design : the presence of a previous path mimicked by a dedicated staples pointing out the DNA origami.

In this chapter, we have proposed an alternative random walk design that relies on strand displacement mechanisms. The design uses a variation of toehold exchange strand displacement (TESD), with the introduction of a nick in what is refered to the branch migration in the litterature. We have described how the rates of the random walk can be tuned, either by changing the toehold energies or by changing the concentration of the *step* and *cover* strands. We proposed and discussed two experiments that could be useful to experimentally prove that a random walk behavior is obtained when the design will be experimentally achieved. Finally, we proposed an experimental plan for the experimental completion of the design, with a bottom-up approach composed of three toy models. The next chapter describes the material and methods used to design the sequences, run, follow, and analyze the experiments of the three toy models.

Bibliography of the current part

[ZW09] David Yu Zhang and Erik Winfree. "Control of DNA Strand Displacement Kinetics Using Toehold Exchange". In: *Journal of the American Chemical Society* 131.47 (Dec. 2009), pp. 17303–17314. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja906987s. (Visited on 05/01/2021).

Chapter 7

Material and Methods

Outline of the current chapter

7.1 Numerical Analysis

7.1.1 DNA Numeric Thermodynamic Analysis

We describe how we do the numeric DNA thermodynamic predictions and computations. For our first strand sequence generation in Part II, we used Nupack 2 ([Zad+11]) and ViennaRNA ([Lor+11]) called from Python using subprocesses. All subsequent work used the Python API of Nupack 3.0 and 4.0 [FPP20].

Nupack Model Unless explicitly stated, we use the default Nupack model 'dna' without dangle stacking. In most of our experiments, the buffer only contains (divalent) magnesium salt. Even if the Nupack API gives the ability to input sodium salt, it refuses to accept a zero concentration of sodium salt. We think this limitation is a side effect of making sure the minimal correct amount of sodium salt is added. Supplementary material often inputs the minimal required sodium concentration and the intended magnesium concentration to overcome this. Although the impact is minimal, we convert the divalent concentration into an equivalent sodium concentration using the same reference as Nupack [KP05]. Often, we need to compute predictions at a number of different temperatures. This computes a new model with memory allocation and converts the input ΔH° and ΔS° to the wanted temperatures. We often keep the models in a cache to be able to explore many temperatures more conveniently.

Nupack concentration from energy In many situations, we want to use the concentration solver of Nupack, to obtain a given concentration as a function of a continuous energy. Also, we sometimes wante to compute concentrations in an ideal situation where there are no unintended complexes. To to this we need to input the energies of the complexes. It is not possible in the public API of Nupack because the tube analysis and tube concentration functions take sequences as input. However, by looking at the Python code of the API, we managed to access the undocumented solver function. We wrote a wrapper function (Listing 1) around this solver function to use it more easily. We needed to find the `rotational_correction` boolean and `logq` parameters. At first we did not know where to look in the C++ source code, and simply tried to tune the parameters to match the Nupack behavior. By comparing the results of the public API and the direct access of the solver, we found matching parameters to be `rotational_correction` assigned to false, and $k = 0.00198717$. Later, we found the part of the Nupack code with exactly this constant. Along with it, a commentary states that it is not the correct perfect gas constant values, but that it is here for retro-compatibility reasons. All in all, we managed to find a way to access the inner Nupack solver and its parameters, and we have checked that the results are consistent with the sequences to concentration functions. We think that the `rotational_correction` (set to `False`) refers to the symmetry correction of the Nupack documentation.

```
import nupack as npk
def solve_concentrations(
    T_deg : float,
    cplx_concentrations : list[float],
    cplx_DeltaGs : list[float],
    strands_concentrations : list[float],
    complexes_content : list[list[int]],
    rotational_correction= False, debug = False,k = 0.00198717):
        T_k = T_deg + 273.15
        logq = -np.array(cplx_DeltaGs) / (k*T_k)
        return npk.concentration.solve_complex_concentrations(
            indices=self.complexes_content,
            logq=logq,
            x0=strands_concentrations,
            kelvin=T_k,
            rotational_correction=rotational_correction,
            as_strands=True )
```

Listing 1: Python wrapper function to use the Nupack undocumented concentration solver

Nupack Energy The supplementary material of [ZW09] points out that the energies output by Nupack (version 2) should have a correction term. We measured the same phenomenon with Nupack versions 3 and 4. In Figure 7.1 we compute the hybridization yield of two complementary strands as a function of the binding energy at 25°C. We compute it using two methods: 1) with our solver wrapper function, 2) with the *law of mass action*. Between the two methods, there is a difference of 2.38 $\mathrm{kcal\,mol^{-1}}$ to obtain the same behavior. [ZW09] explains that this value is $RT\ln(55)$. The value 55 is related to the water concentration (55 M) because Nupack uses molar fractions. For this reason, each time we give an energy obtained from Nupack, we apply this correction ($(N-1)RT\ln(55)$ where N is the number of strands in the complex).

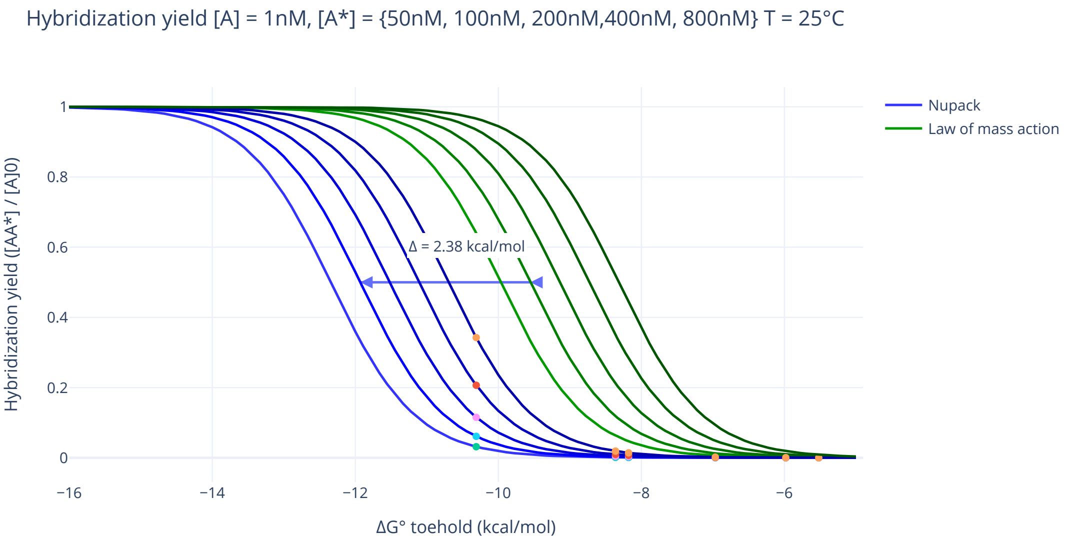


Figure 7.1: Duplex yield in function of ΔG with Nupack's solver and law of mass action. The points are particular case of actual sequences, computed with regular Nupack API to check that this curves are relevant.

7.1.2 Molecular Dynamic Simulations

We conducted Molecular Dynamic Simulations using OxDNA. The design files usually contains too high bonds distances, which can break inital basepairing and never recover it along the simulation. This is why OxDNA simulation are usually made in two parts : first do a relaxation using a fast Montecarlo simulation, by forcing base-pairing. Then in a second step, with actual molecular dynamic simulation, without forcing basepairing.

From ENSNano to OxDNA At the time, ENSNano did not support fully the export in OxDNA format. We modified the export code to suit our needs. This also fixed some edge cases bug that existed. Our new exporter, uses input a json file to set specific DNA sequences for named strands. It outputs an oxDNA constraint file to add mutual trap forces between paired bases for the intial relaxation. An additionnal output file giving the association between OxDNA nucleotide index and the position and name of the associated strand in the ENSNano design. It was expecgted to be useful to add further oxDNA constraints manually be we ended-up not using it. To export, we use run modified exporter, after turning the ENSNAno model mode to "old ENSNANO" model.

This model does not include DNA inclination, but we found the inital geometry closer to the OxDNA model.

Running the OxDNA simulations We ran the occasional origami simulation using the website `oxDNA.org`. For small complexes (Toy Model B), we run both the intial relaxation and the simulation using a local version of OxDNA 3.5 on a regular laptop. The input configuration were inspired by both the one from `oxDNA.org` and [Sen+21]. Both were run on CPU of a laptop, using OxDNA 3.5. We did 1000 iterations for the intitial relaxation and 30000 iterations of 0.003 time step for the molecular dynamic. We observe the energy over time, to assess at which point the trajectory starts to be relevant.

Post-processing Post-processing is done by using the `Oxdna Analysis Tool` as program using a Python code and the standard `subprocess` library. We found that the Python library the `Oxdna Analysis Tool` lacks some functionnalities. For each simulation, we compute aligned trajectories, energy plot over time, and the mean structure. Finally, we capture images in the Oxview webapp.

7.2 Experimental Tools

7.2.1 DNA strands with Fluorescence Reporting

Duplex reporting with fluorophore and quencher. We use fluorescence reporting to observe the evolution of concentration of duplexes (for instance *step-surface* duplex for Toy Model A). One strand representing the "surface" of the DNA origami (depending of the toy model) is modified with 6-Carboxyfluorescein (6-FAM or fluorescein). Fluorescein has a peak emission at 520 nm. The *step* strand is modified with a Iowa Black® FQ, which absorbs from 420 to 620 nm (maximum at 531 nm). When the *step* strand is attached to the "surface", the proximity of the fluorophore to the quencher decreases the fluorescence.

Impact of direct contact quenching and use of Guanine quenching. [MKT02] studies direct quenching for the reporting pair made of fluorophores TET (Tetrachlorofluorescein, similar to fluorosceine) / TAMRA by spacing the fluorophores with 0 , 5 or 10 nt single-stranded mucleotides. It reveals that the 5 nt offset is a intermediary regime between FRET and direct contact quenching; and 10 nt do not show direct contact quenching. It also studies the quenching efficiency by each nucleotide for a large panel of common fluorophores. In general, guanine is the most effective nucleotide for quenching, and fluorophore that emit in he green (like fluorosceine) and yellow wavelength are better quenched with nucleotides. Some fluorophores like fluoresceine have one nucleotide efficiency that stand out (A : 23%,C: 8%, G:32%, T:8%), which allows for a allmost selective nucleotide quenching. However, the analysis of nucleotide quenching is complex, with measured influence all at least the four closest nucleotides [LAS22].

[Zim+19] studies the influence of the fluorophore reporting system on the annealing temperature of a duplex. The fluorophore/quencher pair used is 6-FAM/BHQ2, with different spatial offset between the fluorophore and the quencher. An alternative without quencher is dangling single-stranded guanines (guanine quenching). Compared to the direct contact quenching version, both the FAM/BHQ2 with a 10 nt offset and the guanine quenching, decreases melting temperature by $\simeq 4.5\,°\mathrm{C}$. As a reference, the same sequences without fluorophore but with intercalating dye (SYTTO 82), result in a 5 °C decrease. The quenching efficiencies (percentage of signal reduction when the duplex is formed) are 94.1% for the direct contact version,and 45% efficency for guanine quenching version.

The conclusion is that 1) direct contact quenching stabilizes the duplex by adding a energetic bonus, 2) direct contact quenching can be avoided by physically spacing the fluorophore or quencher. 3) In the case of Fluorescein, guanine quenching is still usable with good quenching efficiency.

Duplex reporting with Guanine quenching. In the result section due, we will observe a duplex bias that we thinked as attributed to direct quneching. To try to counter this, we will rely on guanine quenching for fluoresceine. We will first use the add of uses the add of single-stranded GG dangling nucleotide. Then experiments will show that the double-stranded `GG` sequence present near the flurophore when the duplex is formed is sufficient for guanine quenching.

7.2.2 Fluorescent Aquisitions and Data processing with the Shimadzu fluorometer

In this subsection, we describe two acquisition protocols to measure the fluorescence with a regular fluorometer (Shimadzu RF-5301PC). The machine excites the sample at a narrow wavelength range, selected using a monochromator on a lamp. The sample is placed in a quartz cuvette with the four sides clear, which is necessary to excite in one direction, and read the emitted fluorescence in the other direction. The emission is filtered using another monochromator and then aquired with a sensor. Now we describe our protocol for temporal aquisitions, we will dicuss the way we can mix the content of the cuvette. Finally we describe how we process the data.

Sample Aquisition We do the acquisitions in a Hellma fluorescence cuvette of 500 µL maximum volume and 2 mm × 10 mm optical paths. We use a circulating bath at 25 °C to maintain the temperature constant inside the cuvette. It is crucial because fluorescence of fluorescein is temperature-dependent. Otherwide, the signal would evolve overtime as the cuvette heats up slowly in the machine. A regular experiment follows this scheme:

1. Turn on the lamp fan and the lamp. Wait at least 20 minutes for the lamp to heat up.

2. Set the excitation and emissions slits on respectively 475 nm and 520 nm. Both slits are dixed at 10 nm widths. For this fluorometer, this is the above option after 2 nm, which does not produce enough signal.

3. Adding the buffer in the cuvettes. The experiment are planned to end with 500 µL inside the cuvette. Often, 450 µL of buffer is added at this step.

4. Oppening the shutter. Only a small is supposed to be read. We substact the small signl using the auto-zero button.

5. Start the aquisition, with 2 seconds integration time, and one measurement every 5 seconds.

6. To add a new strand or complex, we pipette it in the buffer. Often, the added volume is 25 µL. We immediately proceed to 20 up-and-down pipetting using a 100 µL pipette set on 50 µL. We discuss the pipette mixing, and the removal of measurements of this step in the two next paragraphs.

7. Measure the fluorescence of the sample as a function of time.

8. Other strands or complexed can be added in the same exeperiment.

Mixing the content of the cuvette Mixing with an up-and-down pipetting is necessary because the cuvette is too narrow for a magnetic stirrer. Also [Gen+11] observes fluorescence loss due to strands adhesion on the magnetic stirrer. We had cases where we observed too high fluorescence level due to not mixing well. But we found 15 to 20 to be a good compromise between speed and mixing quality. The speed is important because the displacement can start to happen without being seen on the measurements. We have found that repeated use of a new cone induced a linear loss that we could measure. We still use a new cone to have a consistant loss over all the experiments. However this loss can be considered neglibile for our experiments where only two mixes are done. In particular, we found that the contact surface of the outer part of the cone is important, so we made sure to used cones that are not too narrow to mitigate losses. As an alternative, using a larger cuvette with 10 mm × 10 mm optical paths, to be able to mix with a 1000 µL pipette could solve some of this mixing issues. However, small piece to rise the cuvette would have to be made, so that the excitation ray reaches volumes smaller than 500 µL.

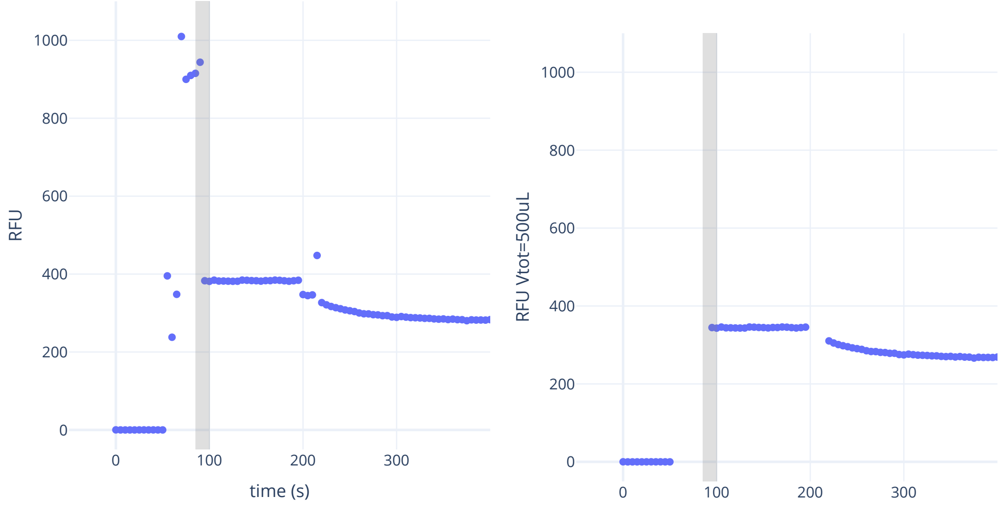
Data Processing The Shimadzu software allows us to copy the fluorescence data into spreadsheet software, where we save it as a .xlsx file. We load the data using the Pandas Python library, which outputs a Pandas DataFrame, where the columns are the individual samples (Figure 7.2a). At this point, the data contained nonsense values when we opened the lid to inject the species in the cuvette (Figure 7.2c) We exclude these points at the same time as computing a normalized signal. Indeed, as we add spieces without fluorophores, the concentration of the strands with fluorophores is decreased. To be able to compare the intial and last fluorescence, we normalize it to the signal that were to be read if the volume was 500 µL at each measurement. First we use the melt function from Pandas to obtain a DataFrame with only three columns, the time, the fluorescence (or RFU for raw fluorescence unit), and the sample name. Then we add a third column volume, which tells at each time, and for each sample, the total volume. In this column we put NaN value for time where lid is opened. Finally we compute the normalized RFU column, which also propagates NaN values (Figures 7.2b, 7.2d).

time	2-s-pc-c-50
..	..
85	915.2
90	943.7
95	382.9
100	381.4
..	..

(a) Initial DataFrame. There is one time column, and one column per sample.

time	sample	RFU	volume	RFU adjusted
..	..	..	..	..
85	2-s-pc-c-50	915.2	NaN	NaN
90	2-s-pc-c-50	943.7	NaN	NaN
95	2-s-pc-c-50	382.9	450.0	344.6
100	2-s-pc-c-50	381.4	450.0	343.3
..	..	..	..	..

(b) Melted DataFrame with one time column, one sample (name) column, and one RFU column. The volume is manually added, The volume is used to compute RFU adjusted column.

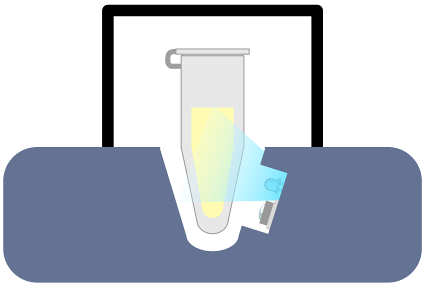


(c) Scatter plot of the (unprocessed) RFU as a function of time. The openning of the lid makes nonsense values. The grey interval corresponds to the time 85 to 100, represented on the figures a) and b).

(d) Scatter plot of the adjusted RFU as a function of time. The grey interval corresponds to the time 85 to 100, represented on the figures a) and b).

Figure 7.2: Data processing on a fluorometer experiment, and an example of the raw, and processed (with normalization and exclusion of nonsense measurements.)

7.2.3 Measurements with the Qubit DNA Quantification Machine



(a) Simplified view of the Qubit 2.0

(b) Photos of the Qubit 2.0 with a tube inside and the lid open

We use the Qubit DNA quantification machine to read fluorescence for both one-time readings and temporal acquisitions. The measurement with the machine was originally intended for DNA quantification, but we decided to use it also to measure strand displacement. In its intended use, a specific buffer with intercalating dye is to be added to a sample. The sample, put in a plastic tube with thin walls (sold by the manufacturer), is illuminated with a blue LED. Finally a receptor to measures the resulting green emission. However, Indeed, the emission excitation and emission wavelengths are compatible with fluorescein, with allowed us to measure the fluoresence similarly to a regular fluorometer. Using this machine has several advantages for one-time readings has several advantages: 1) there is no warm-up time; 2) the sample volume is smaller, it requires only 200 µL of solution; 3) the acquisition of many samples is faster than in a regular fluorometer, as there is no need to clean a cuvette; 4) it is possible to measure the samples several hours later to obtain long-time equilibrium measurements; 5) we can compare the measurements with regular fluorometers One caveat is that, compared to quartz cuvettes, the plastic tubes are prone to ssDNA and DsDNA adhesion on the tubes, which decreases the signal significantly. However, we will describe passivation techniques to limit this phenomenon. We also managed to perform measurements as a function of time. The next paragraphs describe our protocols for one-time measurements, how tube tube passivation can be done, and how to process the machine's data.

Sample Aquisition We managed to sucessfully read samples with a volume of 180 µL to 220 µL and a concentration of 2.5 nM to 10 nM FAM-modified strands. The Qubit is located in an air-conditioned room at 21°C, so we wait a few minutes for our samples to equilibrate thermally. We put the sample in, making sure the orientation of the cap is consistent, and close the Qubit lid. It is important not to write anything on the tube caps, which could impact the measurement. We select the *dsDNA High Sensitivityany* option and measure. It shows an estimated concentration based on the calibration of the previous user, which must be ignored. The relevant data is the *Green RFU value*, found in the Data tab. For temporal data, it is important to remove the tube from the Qubit because doing several acquisitions in a row raises the sample temperature, which affects the fluorescence measurements. We did not observe any photobleaching happening over 30 acquisitions.

Tube surface passivation The fluorescence signal of the same Qubit sample significantly decreased over time. We believe the signal drop is due to strands sticking to the walls of the tubes, which inhibits the fluorescence. We have tested three passivations: 1) adding 20-mer of T -nucleotides in the buffer; 2) adding Bovine Serum Albumine (BSA) (concentrated 0.02%)directly in the buffer; 3) the incubation of the tube surface with BSA concentrated at at 0.02% for 10 minutes prior to the experiments. It is followed by two rinsing with water. BSA in the buffer inhibited the signal decrease,and was used for the first experiments with the Qubit. However, we were worried by the possibility of pipetting errors due to the high viscosity of the buffer. We settled with the passivation of tube surface. However, this process take significant times and manipulation compared to the add of BSA in the buffer.

Data Processing We export the data on a USB key. It writes a CSV file that we load into a DataFrame with the Pandas Python library. No data processing is done other than normalizing the fluorescence similarly to the regular fluorometer (to dillution equivalent for 200 µL). This way we can compare the fluorescence signals before and after dilutions when adding species. It was unclear whether adding a non fluorescent volume modifies the signal only according to the dillution, because the LED might illuminate variable portion of the volume. We have evidence with control samples, that such normalizations are correct, at least within our volume variations.

7.2.4 Experiments

Samples Preparation

All the samples (origami, *step cover* duplexes and *cover*) are prepared or annealed in the same buffer, composed of TAE (at 1X concentration), with $MgCl_2$ 12.5 mM added.

Monomers Monomers (cover strands alone), are denaturated at 1 µM at 95°C for 3 min, and then bring back to 25°C for 5 min. This makes sure there are no duplexes or secondary structures in the stock tubes. We have tried annealing it instead with other samples, and it does not have a measurable impact.

Duplexes and linker complexes For duplexes (like *step-cover duplex*) and bigger complexes (*linker complex of Toy Model B*), we proceed to an annealing with 1 µM of each strand. For simplicity, we assume the differences of concentration from what is claimed by the manufacturer is negligible. For the linker complex, it is particularly important to have each species at the same concentration, which otherwise decreases the yield of the complex, by increasing the yield of the corresponding duplex. For the *step cover* mix, with *step cover* and *cover strand* , we annealed everything in a single tube.

DNA Origami Annealing DNA origami are annealed at 20 nM scaffold. We use a 10X excess of each staple (200 nM for each staple). It is necessary to start from a high concentration because gel purification has small yield.

Contrary to Part I, we proceed to a one hour (instead of 3 hours) long annealing with a temperature range from 95 °C to 25 °C at constant rate. Agarose gel and AFM in the result section will show that the assembly is still correct.

DNA Origami Purification DNA Origami are gel purified on agarose 1%, 50 V for three hours. Both the gel preparation buffer and the running buffer is TBE at 0.5X, 11 mM $MgCl_2$ We put intercalating dye only in the wells (10X SybrSafe) we discuss this at the end of this paragraph. The band is then extracted on a blue light transilluminator. Origamis are extracted using the Montage Gel Extraction kit, which is composed of a nebulizer that splits the agarose into small chunks, and a cellulose membrane that retains the agarose. We start by rinsing the membrane with water and spinning it at 5 krcf for 5 minutes at 4°C. Then we add the extracted agarose band, and spin it at 5 krcf for 5 minutes at 4°C. We immediately put the extracted buffer containing the origami in a BSA-passivated tube, for future fluorescence measurement.

Discussion on intercalating dye for agarose gel We explain why we add dye to the samples. It is more common to add the intercalating dye at 1X in the gel rather than in the sample. It gives better fluorescence when extracting the band. However, since our extraction kit simply extract the buffer of the band of interest, it could extract almost 1X of dye. Then we were worried that when adding the sample to our fluorometer experiment, this dye could induce a fluorescence signal. We see two risks. First that the added fluorescence signal evolves because intercalating dye is much more likely to photobleach than fluorescein. The second risk is that the dye goes into ssDNA and dsDNA of the experiment, and . Because of time constraints, we prefered minimizing the amount of intercalating dye by adding it only in the sample.

7.2.5 Modelisation

We describe here the kinetic models used to model our experiment fluorescence measurements and extract kinetic rates. We start with the simplest 2-states model that accounts for strand displacement in Toy Model A. In the following definition, the "surface" will refer to the surface strand in the Toy Model A, the linker complex in Toy Model B, or the origami with two functionalized staples in Toy Model C.

Description of the Two States Model

We assume that the evolution of the strand displacement is ruled by two bimolecular rates (Figure 7.4). The rate k_f (forward rate) for the conversion of a step-cover complex and one surface monomer to a step-surface duplex and a cover monomer. And the inverse rate k_b (backward rate) for the conversion of a step-surface duplex and a cover monomer to a step-cover complex and one surface monomer.

We denote by $[\mathrm{surf}], [C], [S \bullet \mathrm{surf}], [S \bullet C]$ the concentrations of respectively the surface monomer, the cover monomer, the step-surface duplex, the step-cover duplex.

Using these bimolecular rates, we can explicit the differential equation for the evolution of the concentrations of each constituent (Equation 7.1).

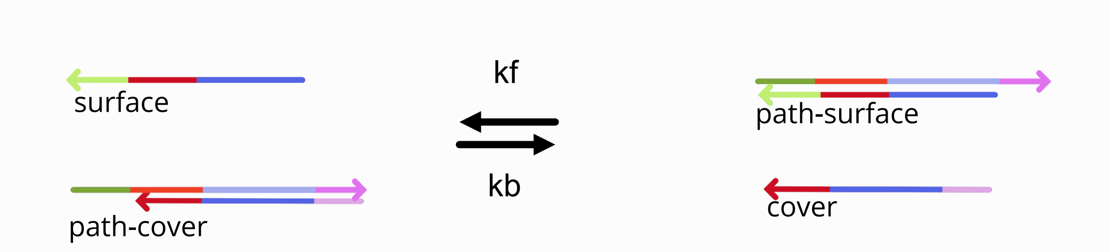


Figure 7.4: Two states rate mode for toy model A

Description

$$\begin{cases} \frac{d[\text{surf}]}{dt} &= k_b[S \bullet \text{surf}][C] - k_f[\text{surf}][S \bullet C] \\ \frac{d[S \bullet C]}{dt} &= \frac{d[\text{surf}]}{dt} \\ \frac{d[C]}{dt} &= -\frac{d[\text{surf}]}{dt} \\ \frac{d[S \bullet \text{surf}]}{dt} &= -\frac{d[\text{surf}]}{dt} \end{cases} \tag{7.1}$$

By conservation of the oligo concentrations we express $\frac{d[\text{surf}]}{dt}$ as a function of $[S]$: :

$$\begin{aligned} \frac{d[\text{surf}]}{dt} = k_b\Big([C]_0[S \bullet \text{surf}]_0 + [C]_0[\text{surf}]_0 + [S \bullet \text{surf}]_0[\text{surf}]_0 + {[\text{surf}]_0}^2\Big) + \\ (k_b - k_f)[\text{surf}]^2 + \\ (-[C]_0 k_b - k_b[S \bullet \text{surf}]_0 - 2k_b[\text{surf}]_0 - k_f[S \bullet C]_0 + k_f[S]_0)[S] \end{aligned} \tag{7.2}$$

Equation 7.2 is a Riccati's equation with constant coefficients. The solution is not exponential unless $k_f = k_b$.

Equilibrium of the two states models Equation 7.1 with $\frac{d[\text{surf}]}{dt} = 0$ gives

$$\frac{k_f}{k_b} = \frac{[S \bullet \text{surf}][C]}{[\text{surf}][S \bullet C]} \tag{7.3}$$

One can show that Equation 7.3 gives

$$\frac{[S \bullet \text{surf}]}{[\text{surf}] + [S \bullet \text{surf}]} = \frac{k_f[S \bullet C]}{k_b[C] + k_f[S \bullet C]} \tag{7.4}$$

Since surf and $S \bullet$ surf are the only complexes containing the surface strand, the quantity $\frac{[S \bullet \text{surf}]}{[\text{surf}]+[S \bullet \text{surf}]}$ of Equation 7.4 is the ratio of conversion of surface to step-surface. We we refer to it as the step assembly ratio.

Limit of the two-states model in with constant step-cover and cover concentrations It is relevant to consider constant concentrations of *step-cover* and *cover*. Indeed, in the experiments for the random walk, the two species are in excess concentration over the

DNA origamis. We denote by $[C]$ and $[S \bullet C]$ the constant concentrations of cover and path-cover. The equations are:

$$\begin{cases} [\mathrm{surf}]'(t) = -k_f[\mathrm{surf}](t) \times [S \bullet C] + k_b[S \bullet \mathrm{surf}](t) \times [C] \\ [S \bullet \mathrm{surf}]'(t) = k_f[\mathrm{surf}](t) \times [S \bullet C] - k_b[S \bullet \mathrm{surf}](t) \times [C] \end{cases}$$

By conservation of the concentration of surface spieces, $[\mathrm{surf}](t) + [S \bullet \mathrm{surf}](t) = [\mathrm{surf}_{\mathrm{total}}]$ is constant. This gives :

$$[\mathrm{surf}]'(t) = k_b[\mathrm{surf}_{\mathrm{total}}] \times [C] - [\mathrm{surf}]\,(t)(k_f \times [S \bullet C] + k_b \times [C])$$

At equilibrium we have,

$$[\mathrm{surf}]'(t) = 0 \Leftrightarrow [\mathrm{surf}]\,(t) = [\mathrm{surf}_{\mathrm{total}}]\frac{k_b[C]}{k_b C + k_f[S \bullet C]}$$

By denoting

$$A = [\mathrm{surf}]'(t = 0) - [\mathrm{surf}_{\mathrm{total}}]\frac{k_b[C]}{k_b[C] + k_f[S \bullet C]}$$

We have :

$$\boxed{\mathrm{surf}(t) = A\exp(-t \times (k_f[S \bullet C] + k_b[C])) + [\mathrm{surf}_{\mathrm{total}}]\frac{k_b[C]}{k_b[C] + k_f[S \bullet C]}}$$

Or, expressed with $[S \bullet \mathrm{surf}](t)$: ,

$$\boxed{[S \bullet \mathrm{surf}](t) = A'\exp(-t \times (k_f[S \bullet C] + k_b[C])) + [\mathrm{surf}_{\mathrm{total}}]\frac{k_f[S \bullet C]}{k_b[C] + k_f[S \bullet C]}}$$

This result, with constant *step* and *step-cover* has two practical advantages. First, the kinetics do not depend on the concentration of *surface*, meaning that there is less possibility of experimental variations when reproducing the experiments. The second advantage is that since the solution is exponential, it is easier to fit and extract the parameters from experimental curves, compared to computing the differential equations.

Equivalence between thermodynamic equilibrium and equilibrium of the two states model A kinetic model is correct thermodynamically if the rates respect the detailed balance (Equation 7.5).

$$\frac{k_f}{k_b} \overset{\text{detailed balance}}{=} e^{-\Delta G^{\circ}/RT} \tag{7.5}$$

We can note that at equilibrium we find again Equation 7.3.

$$\frac{k_f}{k_b} \overset{\text{detailed balance}}{=} e^{-\Delta G^{\circ}/RT} = K \overset{\text{law of mass action}}{=} \frac{[S \bullet \mathrm{surf}][C]}{[\mathrm{surf}][S \bullet C]} \tag{7.6}$$

7.3 DNA Sequences

Figures 7.5, 7.6, and 7.7 represent the DNA sequences used for the Toy Models A, B and C.

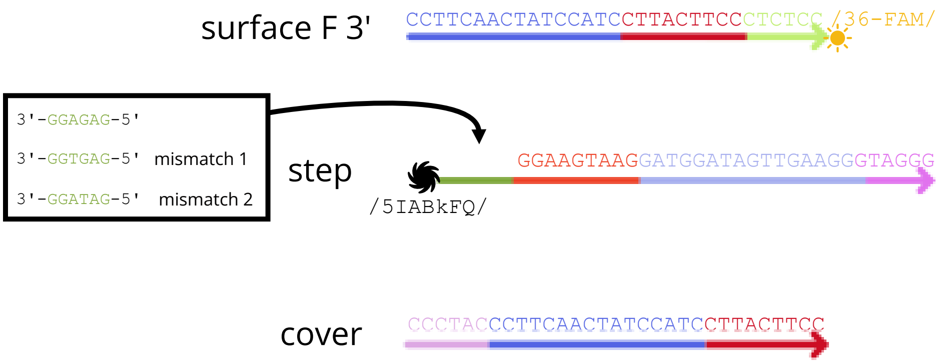


Figure 7.5: DNA Sequences for Toy Model A

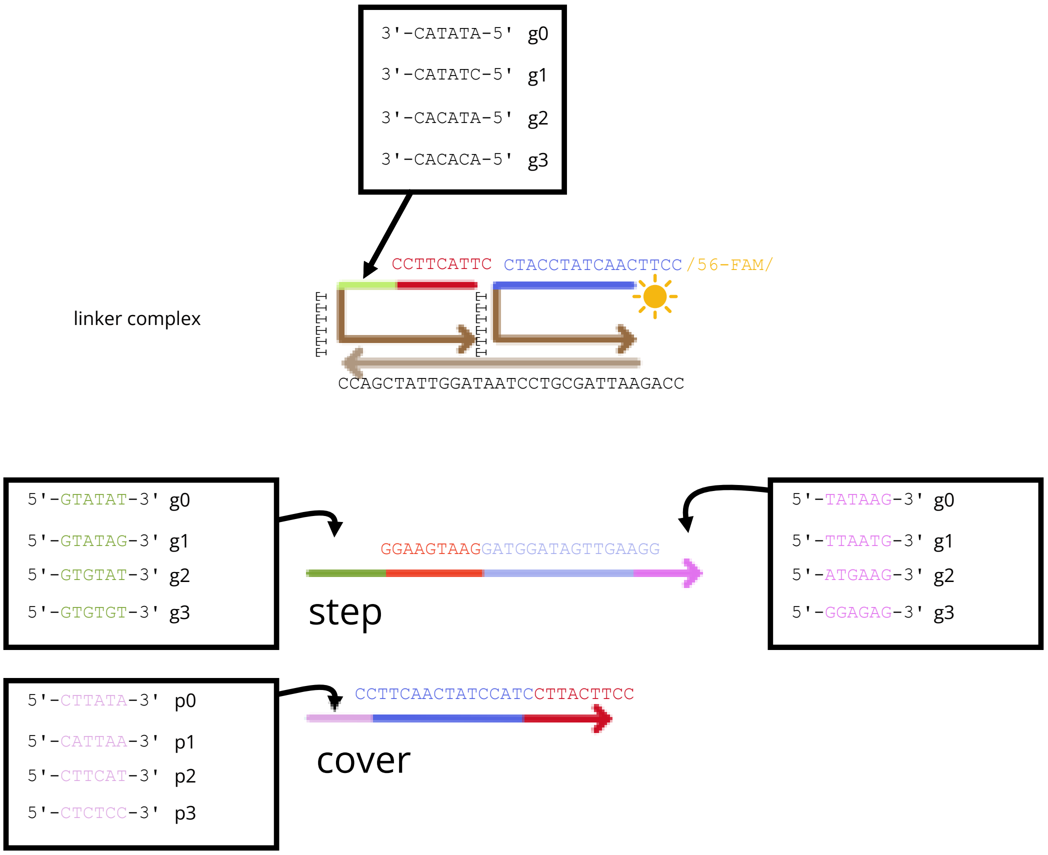


Figure 7.6: DNA Sequences for Toy Model B

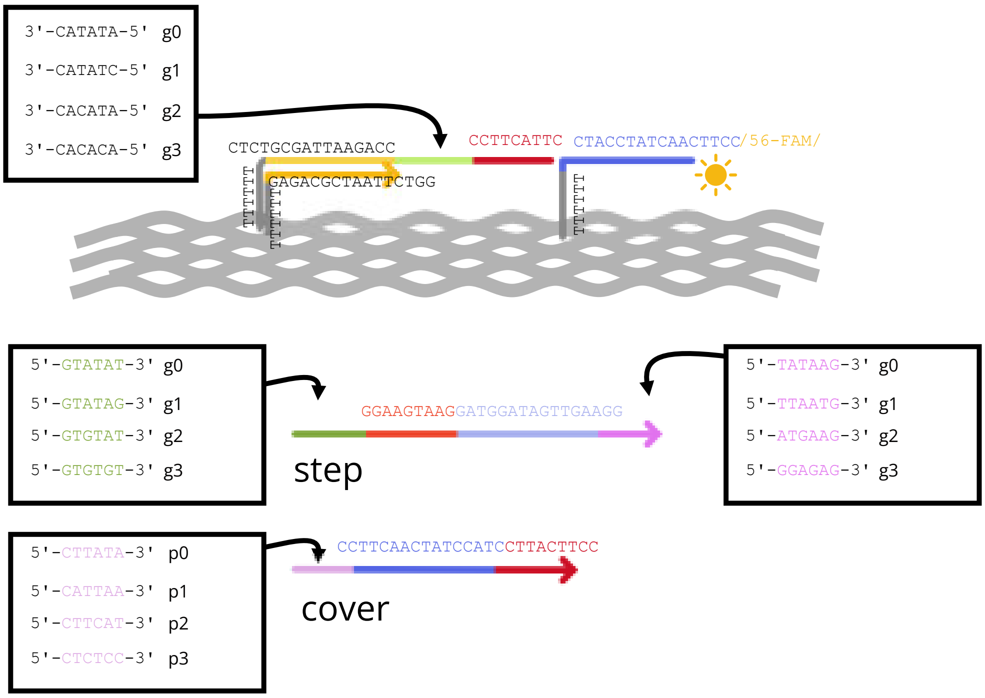


Figure 7.7: DNA Sequences for Toy Model C

Bibliography of the current part

[FPP20] Mark E. Fornace, Nicholas J. Porubsky, and Niles A. Pierce. “A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed”. In: *ACS Synthetic Biology* 9.10 (Oct. 2020), pp. 2665–2678. ISSN: 2161-5063, 2161-5063. DOI: 10.1021/acssynbio.9b00523. (Visited on 05/01/2021).

[Gen+11] Anthony J. Genot et al. “Remote Toehold: A Mechanism for Flexible Control of DNA Hybridization Kinetics”. In: *Journal of the American Chemical Society* 133.7 (Feb. 2011), pp. 2177–2182. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja1073239. (Visited on 09/29/2021).

[KP05] R. T. Koehler and N. Peyret. “Thermodynamic Properties of DNA Sequences: Characteristic Values for the Human Genome”. In: *Bioinformatics* 21.16 (Aug. 2005), pp. 3333–3339. ISSN: 1367-4803, 1460-2059. DOI: 10.1093/bioinformatics/bti530. (Visited on 01/06/2024).

[LAS22] Jory Lietard, Dominik Ameur, and Mark M. Somoza. “Sequence-Dependent Quenching of Fluorescein Fluorescence on Single-Stranded and Double-Stranded DNA”. In: *RSC Advances* 12.9 (2022), pp. 5629–5637. ISSN: 2046-2069. DOI: 10.1039/D2RA00534D. (Visited on 09/06/2023).

[Lor+11] Ronny Lorenz et al. “ViennaRNA Package 2.0”. In: *Algorithms for Molecular Biology* 6.1 (Dec. 2011), p. 26. ISSN: 1748-7188. DOI: 10.1186/1748-7188-6-26. (Visited on 01/16/2024).

[MKT02] Salvatore A E Marras, Fred Russell Kramer, and Sanjay Tyagi. “Effciencies of Fluorescence Resonance Energy Transfer and Contact-Mediated Quenching in Oligonucleotide Probes”. In: *Nucleic Acids Research* 30.21 (2002).

[Sen+21] A. Sengar et al. “A Primer on the oxDNA Model of DNA: When to Use It, How to Simulate It and How to Interpret the Results”. In: *Frontiers in Molecular Biosciences* 8 (June 2021), p. 693710. ISSN: 2296-889X. DOI: `10.3389/fmolb.2021.693710`. (Visited on 09/07/2023).

[Zad+11] Joseph N. Zadeh et al. “NUPACK: Analysis and Design of Nucleic Acid Systems”. In: *Journal of Computational Chemistry* 32.1 (Jan. 2011), pp. 170–173. ISSN: 01928651. DOI: `10.1002/jcc.21596`. (Visited on 12/22/2022).

[Zim+19] Zackary A. Zimmers et al. “Fluorophore-Quencher Interactions Effect on Hybridization Characteristics of Complementary Oligonucleotides”. In: *Analytical Methods* 11.22 (2019), pp. 2862–2867. ISSN: 1759-9660, 1759-9679. DOI: `10.1039/C9AY00584F`. (Visited on 09/06/2023).

[ZW09] David Yu Zhang and Erik Winfree. “Control of DNA Strand Displacement Kinetics Using Toehold Exchange”. In: *Journal of the American Chemical Society* 131.47 (Dec. 2009), pp. 17303–17314. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja906987s`. (Visited on 05/01/2021).

Chapter 8

Results

Outline of the current chapter

8.1 Toy Model A : One-strand Substrate

There was three goals for the experiments with the Toy Model A : 1) To check that we can replicate and analyse toehold exchange strand displacement (TESD). 2) To experimentally find toehold sequences or concentration conditions to balance the system. 3) To work in excess conditions of *cover strand* and *step-cover duplex* .

We will also see that we need to make the *step* attachement toehold really weak in order to reach half fluorescence in conditions with equal concentration of step-cover and cover. For this we tried to first rely on Guanine quenching instead of Dark quencher to reduce the energetic impact of the dark quencher. This was not sufficient, so we introduced a mismatch in the toehold, which, this time, yielded a balanced equilibrium.

8.1.1 Experimental Results

Unidirectionnal Displacement at Low Equimolar Concentration

We conducted two experiments : 1) a forward experiment, starting from *surface* strands (2 nM), with the manual introduction of *step cover* duplexes (2 nM); 2) a backward experiment, starting from of *step-surface* duplexes)2 nM), with the introduction of *cover* monomers (2 nM). In both cases, the two states sytem is expected to go in one direction only, but progressively balanced by either the release of *cover strand* or the duplexation of *step strand* with *cover strand* . Both experiments contain the same concentration of each strand and should theoretically converge to the same equilibrium. This is supported by the fact that the final fluorescence levels are close by. We fit the parameters k_f and k_b at the same time on both curves, using the two-states model. We found $k_f = 2.16 \times 10^6\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ and $3.42 \times 10^5\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. The ratio between the two, which is also the equilibrium constant for the forward reaction is equal to $K = k_f/k_b = 6.3$. The conclusion from this experiment is that the forward reaction is faster than the backward reaction. We can note that the fitted curve for the detachment of *step strand* does not follow precisely the first few minutes of measurements. We discuss the possible explanation in one of the following paragraphs.

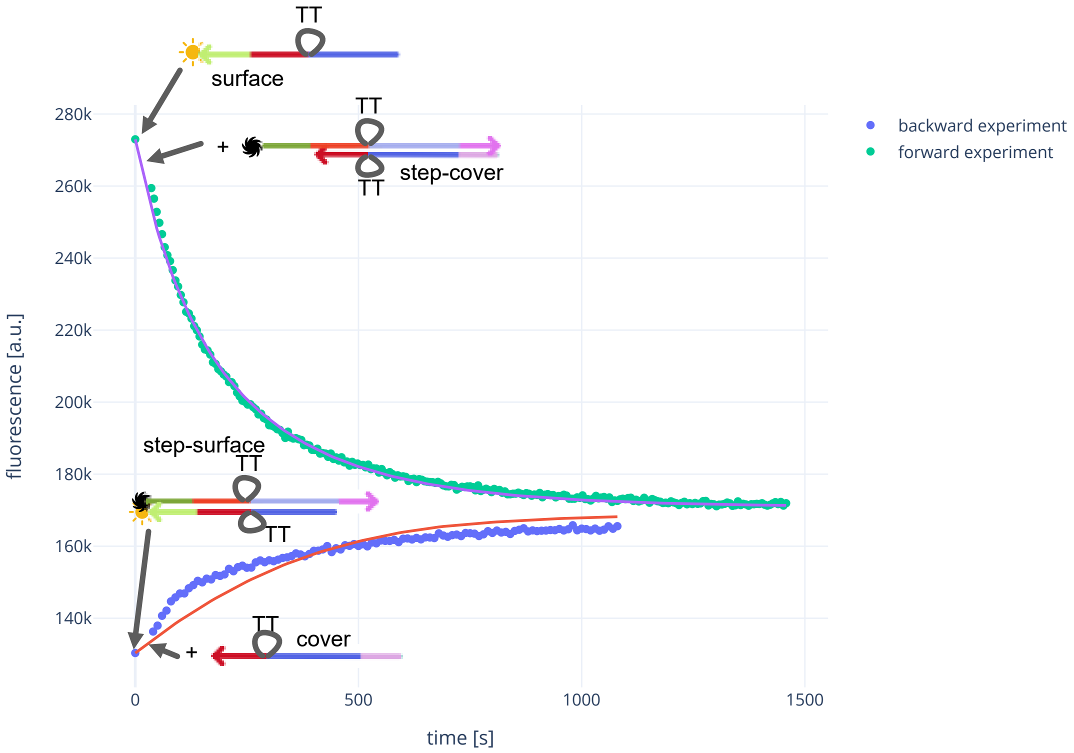


Figure 8.1: Fluorimetric measurements of the displacement of Toy Model A. The top curve (points) is obtained by adding 2 nM *step-cover duplex* to 2 nM of *surface*. The bottom curve (points) is obtained by adding 2 nM *cover strand* to 2 nM of *step-surface duplex*. The lines correspond to the 2-step model fit, with parameters $k_f = 2.16 \times 10^6$ 3.42×10^5.

Displacement on a mix of *step-cover duplex* and *cover strand* We conducted another experiment to validate the model and the rate obtained previously. This time, we started from 2 nM *step-surface duplex* and manually add a mix of 1 nM *cover strand* and 1 nM *step-cover duplex* . In total there is 1 nM *cover strand* less *cover strand* , and 1 nM more *step strand* , so we expect the equilibrium to move towards *step* disassembly and reach a lower fluorescence level. Figure 8.2 shows the experimental curve and the evaluation with the rates from the previous experiments. The fluorescence level at equilibrium is correctly predicted by the model and parameters. However, similarly to the previous experiments, the first few minutes of the curve is not correctly predicted.

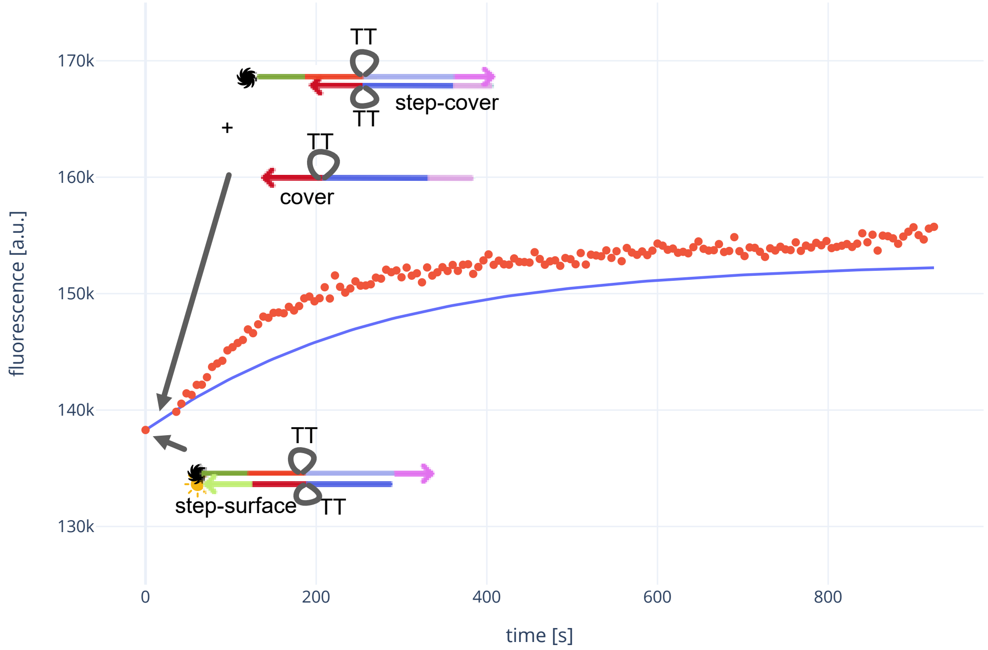


Figure 8.2: Fluorometer measurements of 2 nM *step-surface duplex*, with the manual adding of a mix of 1 nM *cover strand* and 1 nM *step-cover duplex* . Model evaluation from rates parameters $k_f = 2.16 \times 10^6$ $k_b = 3.42 \times 10^5$ ($k_f/k_b = 6.3$)

Reversibility in low concentration conditions. Despite the system not being balanced, we further confirm the strand displacement with an experiment where the system goes in both directions. The experiment startes from 2 nM *step-surface*. Then we manually added 2 nM of *cover*. When the signal stabilized, we manually added 2 nM of *step-cover* . Then we manually add again 2 nM of *cover*. The first two part are exponential behavior, with change of monotony. The third step (when adding 2 nM of cover) have a linear looking behavior. The change of monotony confirms reversible characteristic of toehold exchange strand displacement (TESD).

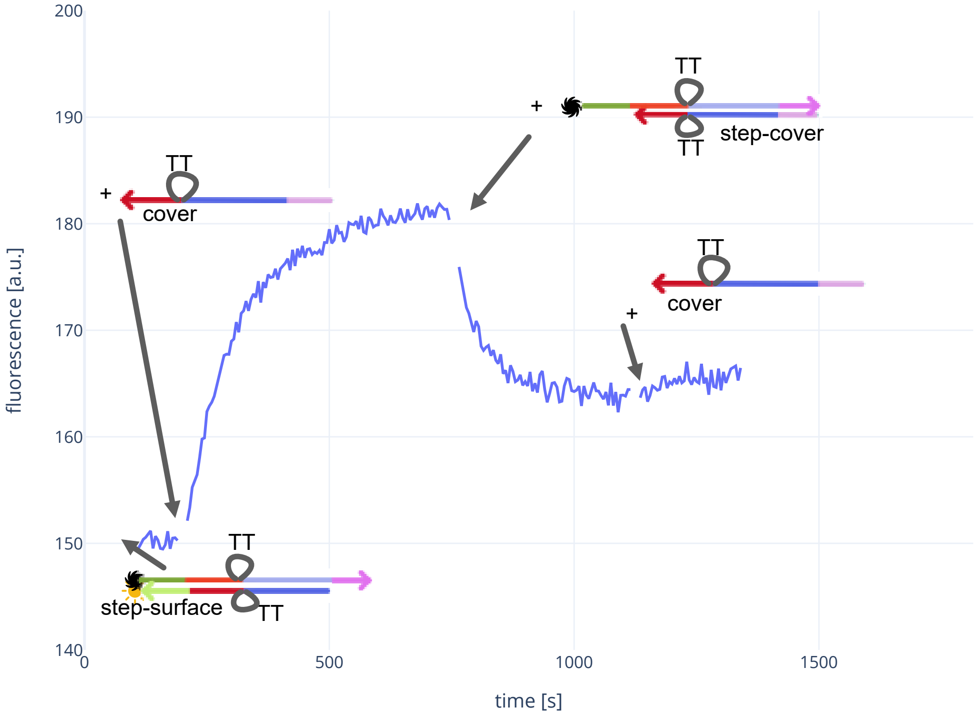


Figure 8.3: Fluorimetric measurements of displacement of 2 nM step-surface by successively 2 nM cover, 2 nM step-cover and 2 nM cover. The calibration could not be set properly due to inconsistencies in fluorescence levels between samples. However, based on previous experiments, a distribution of 0.5 nM *step-surface* and 1.5 nM *surface* is expected at the first peak.

These experiments confirmed the behaviour of TESD. However, we measured a imblance of the system, with the *step* attachment rate being almost 6 times the *step* detachment rate. The modelling using a two-step model is correct, though the *step* detachment curves is not accurately modeled. Further experiments suggest that remaining monomer in the *step-surface* samples could have this kind of impact. However, this most likely does not largely impact the estimation of $K = 6.3$. In the next experiments, we will focus on balancing the forward and backward rates. First we will remove the TT "bulge" parts that add a unecessary complexity. Assuming that the direct contact quenching induces a binding energy bonus, we will also explore how to mitigate it.

Energy Bonus Effect of Direct Contact Quenching

Alternative reporting designs without direct contact quenching Duplexes with direct contact quenching are known to have an associated energy bonus. We wanted here to check that the imblance of the system was not due to this direct contact quenching. For this we made several alternative reporting versions (Figure 8.4). In all these versions, and from now on, we removed the TT part inside the branch migration domains to simplify the study. These versions contains two strategies. The first one is to extend the distance between the quencher and the fluorophore, by adding a 6T sequence. The other strategy is to use Guanine quenching, by adding a GG sequence instead of the regular dark quencher. Indeed, fluorescein inhibits it fluorescence in the presence of Guanine.

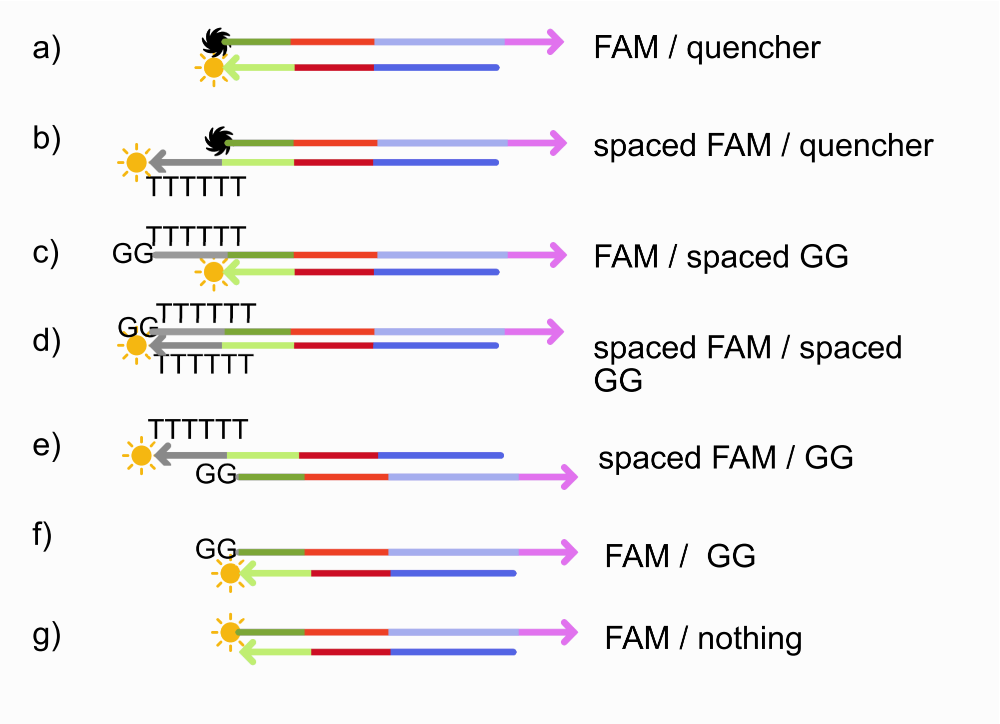


Figure 8.4: Alternative quenching designs. The sun and black hole pictograms represent the modifications with the flurophore and dark quencher.

Melting temperature of the alternative reporting versions. Figure 8.5 shows the melting temperature of the alternative reporting duplexes. The versions without quencher and without GG was not ordered and tested.

First, we discuss the quality of the reporting. Most of the melting curve have an intense peak on the derivative curve. Only the versions with spaced fluorophore / GG (e) and spaced fluorophore / spaced GG (d) have a low peak. It means that the quenching effiency is significantly decreased for these versions. However, the spaced fluorophore/ GG shows a good quality reporting. Because, in all versions, the end of the toehold near the quencher has a GG sequence, we make the hypothesis that the guanine quenching in the FAM / spaced GG (c) and FAM / GG (f) works not only because of the extended GG sequence, but because of the GG sequence inside the toehold. This hypothesis will be later confirmed in Toy Model A, where we completely removed the quencher (g), without replacing it with a GG sequence.

Now we discuss the measured melting temperatures. First we point out that we have to be cautious, since the temperature steps in the qPCR was of 1 °C. However, the results obtained follow our expectations. The direct contact quenching produces the highest melting temperature, which translates into higher stability for the duplex. When preventing the direct contact by spacing the fluorophore, the associated energy bonus decreases. Finally, all the Guanine quenching versions have the same, and lowered melting temperature. We did the same measurement between the *step strand* and the *cover strand*.

We also measured the melting temperature of the *step-surface* duplex. We observed the *step strand* versions with respectively the dark quenchers, quencher, GG-extension, and spaced GG-extension. We measured respecitively 70 °C,71 °C,72 °C. With this temperature being lower than the temperature of all versions of *step-cover*, it again confirms the imblance of our system. However, we cannot explain why the melting temperature of these *step-surface* versions are different. In all further experiments with this Toy Model, we switched to GG-extension reporting (f) to try balancing the forward and backward rates.

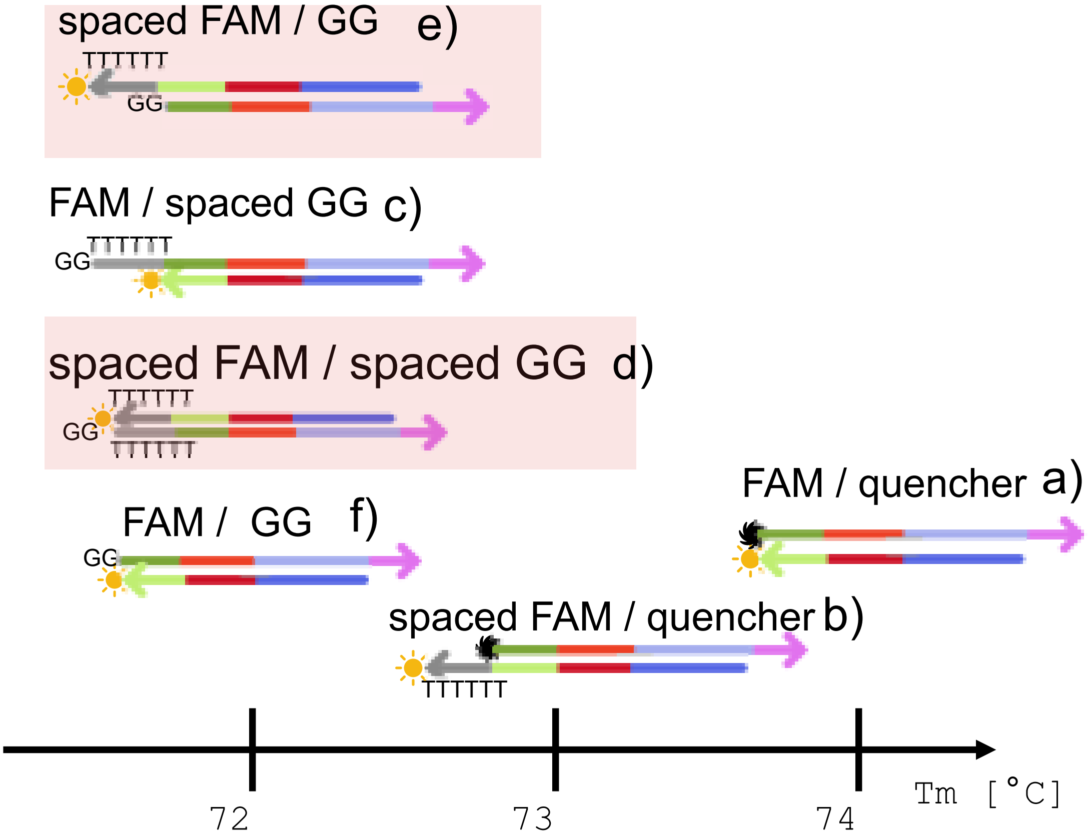


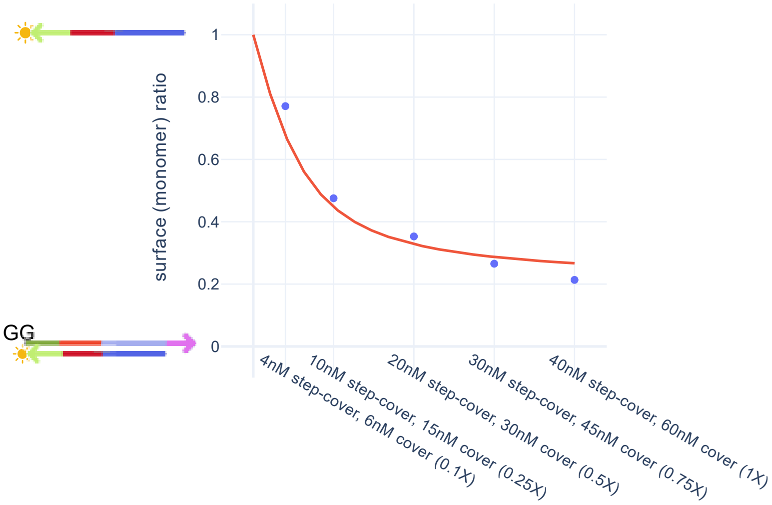
Figure 8.5: Melting temperature of the alternative reporting versions. The melting temperature are measured from hybridization curves, with 1°C steps, at 1 µM of each strand. The red squares indicates samples with a significant decrease of signal level. The letter refers to the Figure 8.4.

Moving the Equilibrium using Medium Range Concentrations Imbalances

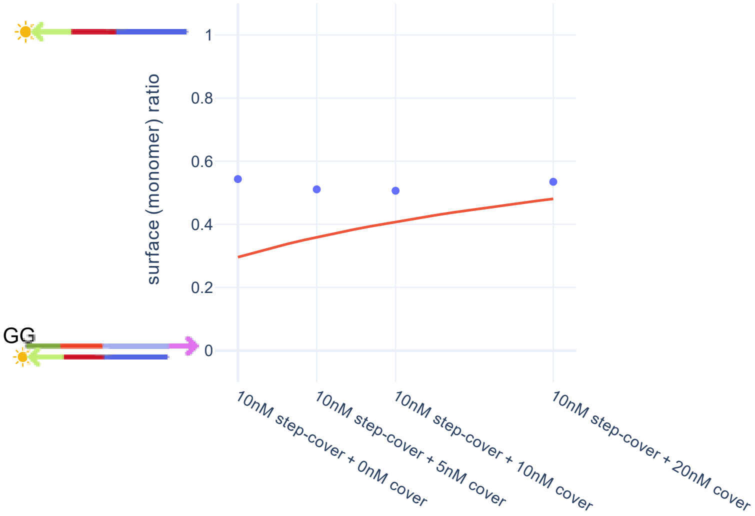
Using the `GG`-extension reporting version, we ran experiments with different *step* and *step-cover* concentrations. Two configurations were evaluated, both on 2.5 nM surface. One with a dilution range of a mix of 40 nM step-cover and 60 nM cover (Figure 8.6b). The other with a fixed 10 nM step-cover plus a concentration range of 0 nM to 20 nM cover (Figure 8.6a). Having replicated successfully the curve of Figure 8.6a, we decided to fit the parameter K only on this curve. We obtain a value $K = 5.66$ which is a bit lower than the value 6.3 estimated based on the low concentration kinetics of the version with the quencher.

This decrease is explained by the removal of the direct quenching energy bonus. The first curve is not accurately described by the model with the parameter value $K = 5.66$. However we are satisfied by the low variation showed in both the measurements and the model.

Figure 8.7 represents the measured conditions of both curves on a phase diagram for the value $K = 5.66$. The phase diagram shows how to reach a balanced system using a concentration imbalance (green part). In the random walk system, a concentration of at least 50 nM would be required, which means about 300 nM of *cover* to balance the system. We think the high concentration of *cover* could cause additional spurious phenomena. This experiment shows that even if we exclude any possible bonus associated with direct contact quenching, there is still a bias towards *step* assembly. The required concentration imbalance to balance the system are not satisfying. In the next paragraph, we explore changing the toehold energies to balance the system.



(a) Equilibrium fluorescences for 10 nM surface with a dillution range of a mix with 40 nM step-cover and 60 nM cover. The line shows the fitting using the two states model, which gives $K = 5.66$.



(b) 10nM surface + (10nM step-cover + 0- 20nM cover)

Figure 8.6: Equilibrium fluorescence for the Toy Model A with guanine quenching, measured with the Qubit.

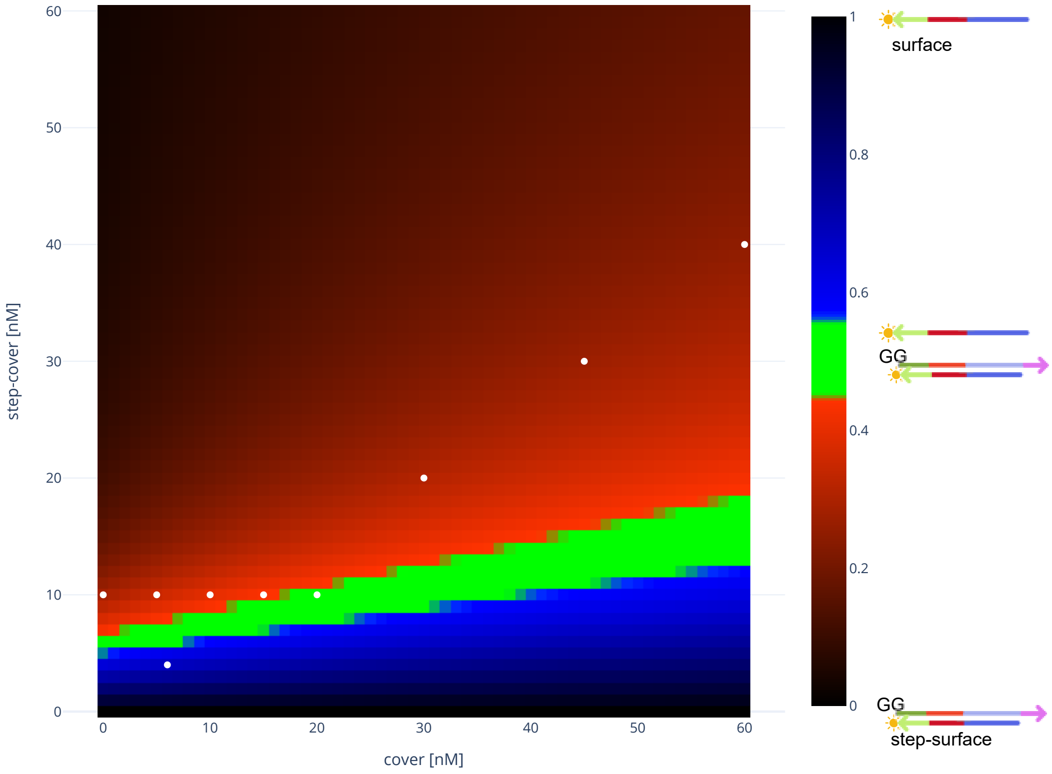


Figure 8.7: Phase diagrams of *step*-assembly yield for the two states models. The ratio $K = k_f/k_b = 5.66$ from the fit of the experiments equilibrium is used. The points represent the experimental conditions that were measured.

Effect of Toehold Mismatch

In the previous section, we saw the limited effect of concentration balancing on the two-state system. In this section, we propose to balance this by drastically changing the toehold energies. To achieve this, we chose two variations of the *step strand* : one with a A turned into a T and one with a G turned into a T. (Figure 8.8a.) When duplexes with the *surface* strand, these variations induce mismatches, which strongly affects the binding energy. For respectively 2.5 nM surface and 5 nM surface we add a mix of 40 nM step-cover and 60 nM cover. As expected, we observe important differences in fluorescence (Figure 8.8b). In Figure 8.8c we represent the evaluation of the reaction constants (K), computed using the minimum and maximum fluorescence, and the concentration conditions. For the non-mismatch version, the observed $K = 12$ for 2.5 nM surface is twice the one evaluated at equilibrium using a range of dilutions of the step-cover mix under 40 nM step-cover 60 nM cover. For the mismatch 1 version, we observed $K = 1.2$ which is promising, as very close to balance. Finally, the mismatch 2 version has a $K = 0.25$, indicating a system biased towards *step* detachment.

(a) Mismatched path oligos

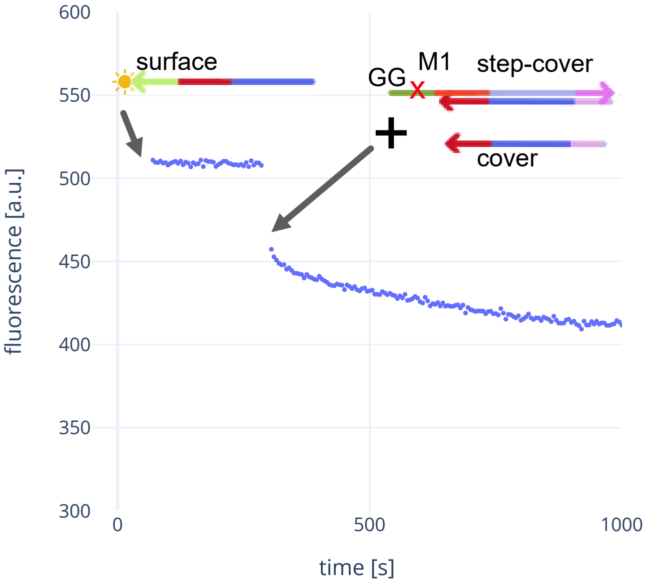
(b) Equilibrium fluorescent measurement of 2.5nM and 5nM surface with a mix of 40nM path-cover + 60nM cover

(c) Equilibrium fluorescent measurement of 2.5nM and 5nM surface with a mix of 40nM path-cover + 60nM cover

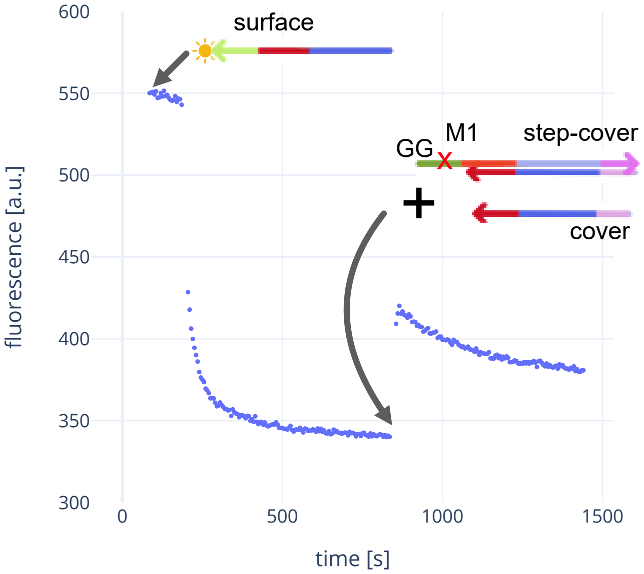
Figure 8.8: Mismatch effect on the equilibrium signal, measured with the Qubit machine.

Effect of excess cover

Finally, we observe the impact of excess conditions of step-cover/cover. These are the intended experimental conditions for the random walk experiment. The motivation was also to check if no abnormal behaviors occurred and if the rate ratio remained constant when scaling the concentrations. Figure 8.9a shows the displacement of 2.5 nM surface with a mix of 40 nM step-cover and 60 nM cover. We observe a fast exponential decrease, followed by an affine decrease in the signal. The two-states model explains that in such conditions, the time constant is very small $\tau = 1/(k_f[\text{step-cover}] + k_b[\text{cover}])$. For the parameters inferred on the 2 nM low concentration, this time constant is equal to about 5s. So the affine decrease confirms the presence of a spurious process. To grasp the misreading impact of this leaking on the evaluation of K parameters, we evaluate the constant K at different points. In Figure 8.9a, the fluorescence after the fast decrease corresponds to $K = 1.05$, and the value after 1000s corresponds to $K = 2.10$. The impact of leaking is stronger in Figure 8.9b. Just after the adding of cover, if the fluoresence signal were to remain constant, then the equilibrium constant would be



(a) Fluoresence signal of displacement for 2.5 nM surface with a mix of 40 nM *step-cover* and 60 nM *cover*.



(b) Fluorescence signal of displacement for 2.5 nM *surface* + 100 nM *step-cover* + 100 nM *cover*. The signal is adjusted to account for signal reduction from dillution. The signal effect of the cover is fast and not visible, however a abnormal decrease of the signal is observed aftewards.

Figure 8.9: Fluorescence curves for the Toy Model A with the *step* toehold having mistmatch 1 sequence.

$K = 1.57$. However, the signal then decreases (which is not expected), and converges to a fluorescence value corresponding to $K = 4.22$.

We have shown that we managed to perform TMSD, but that our original toehold is imblanced, with $K = k_f/k_b \approx 6.3$. Thinking that direct contact quenching could explain this imbalance, we tried several techniques to prevent it, including the complete removal of the quencher. After switching to guanine quenching, we observed how the equilibrium of the system moves with different mixes ratio of step-cover. For this system, we measure a constant $K = 5.6$. This is a bit lower to the first system, which means that switching to Guanine quenching did reduce the imbalance. But this is still important, so to balance the system, we tested two path versions with mismatches on the path-assembly toeholds. The mismatch successfully changed the behaviors, with $K = 1.2$ and $K = 0.15$. However, one caveat is that we observed a spurious process that decreases the fluorescence linearly when assembling path. We think again that *step* strands not duplexed with *cover* strands could explain this phenomenon. These toehold mismatches was conceived to avoid ordering different *surface* strands. The validation of the use of mismatches in the toeholds is a great tool for the future of the project. First, it is an easy technique to drastically change the energy of a toehold. Secondly, we could see an application in the random walk experiments of the project. Indeed, the random walk design uses two versions of the *step* and *cover* strands, sharing two toeholds. Each toehold is used both forthe *step* assembly and the (other) *step* disassembly. In this case, the use of a mismatch on the appropriate *cover* could tune the energy of either the *step* assembly or *step* disassembly independently.

In conclusion, we managed to observe a balanced strand displacement path exchange, even if a strong spurious process occurs. Despite the presence of this spurious phenomenon, we complexified the experiments, moving to Toy Model B.

8.2 Toy Model B : Two-strands Substrate

Toy Model B is a variant of Toy Model A in which the displacement domain is split in two, similar to our final random walk design. We will show that our DNA complexes do not favor path assembly when we simulate their mean structures with molecular dynamics. The three goals of the experimental section are: 1) to make sure we still observe strand displacement on this toy model; 2) to prove that strand displacement is initiated by the toehold and not by leaking. 3) To find a toehold combination that reaches half fluorescence at steady state.

We have conducted experiments on a subset of 16 total toeholds, exploring a vast energetic landscape. We will see that while there is still a toehold effect in this toy model, we failed to find a toehold pair giving equilibrated half fluorescence. We believe that this matches the OxDNA relaxations that suggested a big energy barrier to path assembly.

8.2.1 Molecular Dynamic Simulations

We simulated with oxDNA the structures of the Toy Model B at different steps of the displacement (Figure 8.10). First we see that the linker complex with the *step* strand attached to it is stable and has enough flexibility with the 6T junctions. However, when the *step* strand is not attached, the two parts of the linker complexe are streched far appart. There are two reasons for this. First the choice of length 31 nt for the linker part make both strands "going out" at opposed angle to the helix. Secondly, due to stacking, one of the strands "goes out" on the linker strand orthogonally, while the other tends to stack it single-stranded T nucleotides, which makes it "go out" in the same direction as the linker strand. We imagined an iteration of this design to solve this stacking problem, but it meant adding a fourth strand. The impact of these geometrical remarks, is an energy barrier that could inhibit the strand displacement mechanism from happening. A further analysis of this oxDNA simulation could help predicting the energy barrier, and the resulted rate reduction. However, we choose in this work to focus on the experimental measurement of the rate for this system.

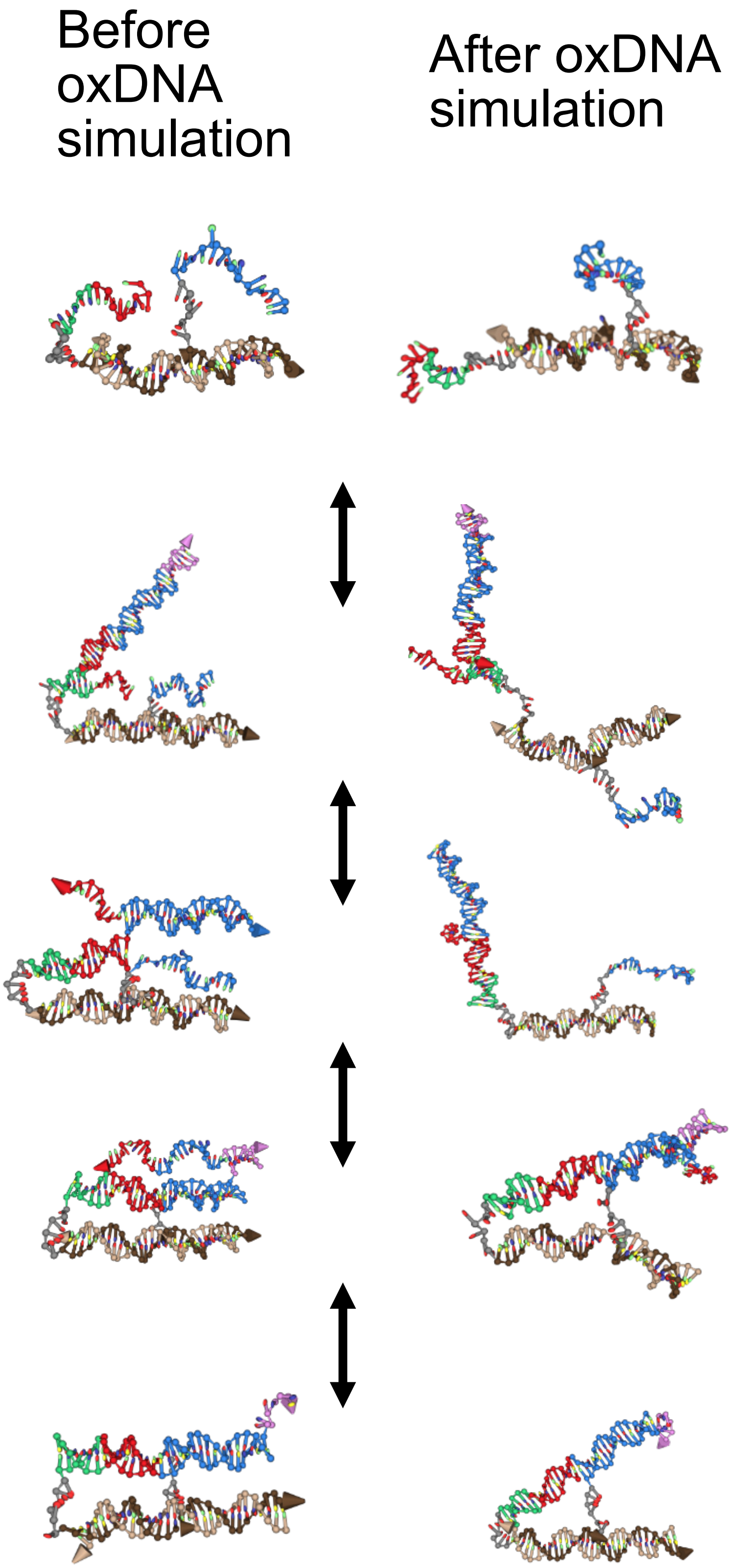


Figure 8.10: Domain steps of *step-cover* displacement on the linker complex. The left part is how we imagined the structures to be, which is also the initial state of the oxDNA simulation. The right part is a representative configuration when computing the trajectory of the complex.

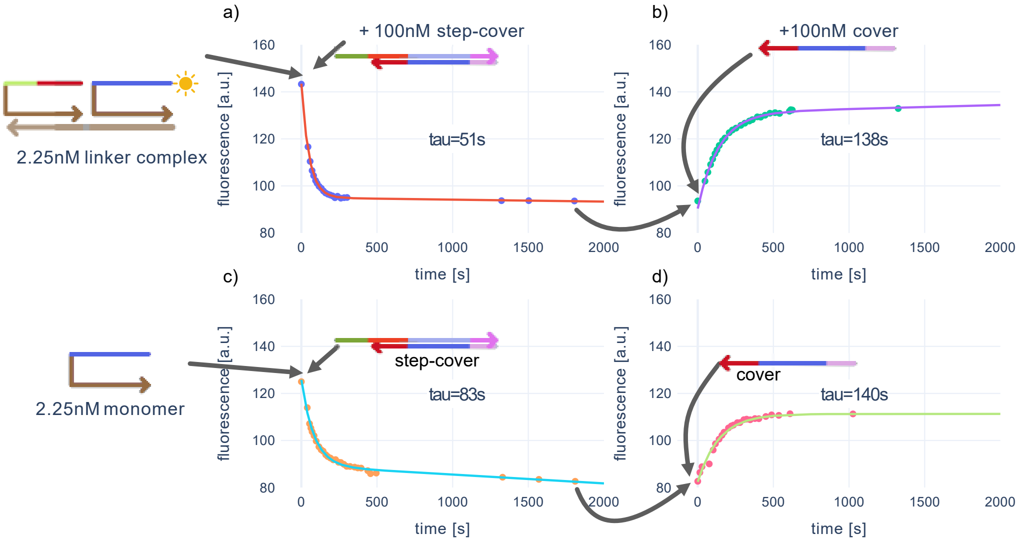
8.2.2 Experimental Results

Our experiments focus first on the particular case of linker complex with toehold versions p0g0 (the weakest toeholds, both for *step* attachment and *step* detachment). We will see that it behaves similarly to the negative control (but that stronger toehold versions behave differently). Also, we observe the path attachment yield by progressive addition of cover strands. In an second part, we study the kinetics of *step* attachment and detachment for different toehold energies. We will observe the impact of toehold energy, validating the cooperative effect of our split toehold exchange strand displacement mechanism.

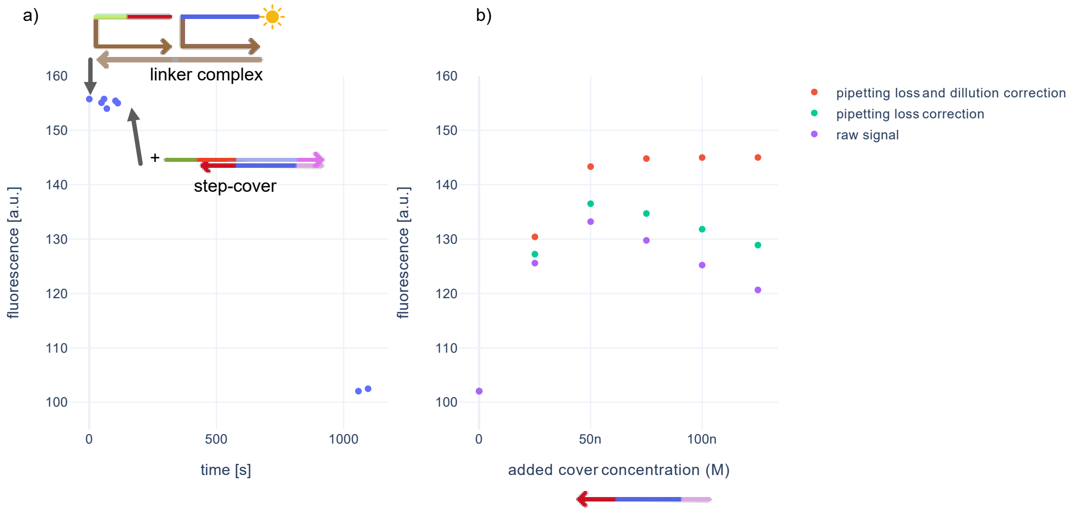
Weak Toeholds

Effectiveness of the displacement on the linker complex. We study the version with the weakest toehold on both sides (p0g0). Figures 8.11a shows the kinetics of step attachment and detachment after adding 100 nM step-cover and 100 nM cover on both dock strand and linker complex. The dock strand does not contain the toehold domain and is used as a negative control. Being in excess conditions, we model each part as an exponential curve plus an affine component to account for spurious phenomena. We can observe that there is a strand displacement effect on the negative control. However the time constant is greater to the one corresponding to the complete linker complex. While the impact of the negative control is greater than expected, this still validate the effect of the toehold for the displacement on this Toy Model B. If we apply the two-states model, we find bimolecular rate of $k_f = 2 \times 10^5\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$, $k_f = 1.2 \times 10^5\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$ for respectively linker complex and negative control.

Interpretation of the results with the two-states model We discuss the validity of the two-step models in this experiments. In both samples, the adding of *cover* produces similar time constant (138 and 140 s). In particular, this detachment time constant is greater than the attachment constants. This is not possible in the two steps model, because $\tau_{\text{detachement}} = 1/(k_f[\text{step-cover}] + k_b[\text{cover}])$ should be lower than $\tau_{\text{attachement}} = 1/k_f[\text{step-cover}]$. We think this is due to the presence of *step* strands not hybridized in the *step-cover* assembly. It would add a rate in the attachment process. However, when adding the excess cover, the cover could bind to the *step* monomer, which make it disapear from the kinetic of the disassembly step. If hypothesis is correct, then the two states model gives $\tau_{\text{NC}} = \frac{1}{k'_f[P\text{mono}]}$ and $\tau_{\text{attachement}} = \frac{1}{k_f[\text{step-cover}]+k'_f[P\text{mono}]}$. Thus, $k_f = (\frac{1}{\tau_{\text{attachement}}} - \frac{1}{\tau_{\text{NC}}})/[\text{step-cover}] = 7.56 \times 10^4\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. The time constant if only this rate k_f applied would be 132.28 s. This is still smaller that 138 and 140 s so something is off. To sum this analysis, the time constants we measured are not compatible with the two-states model evaluated with the correct input concentrations. Using the negative control, and assuming an excess of *step* as a monomer inluences the kinetic of the forward step, we obtain a forward displacement rate of $7.56 \times 10^4\,\mathrm{s}^{-1}\,\mathrm{M}^{-1}$. However, even with this rate, the two states model does not make sense in the *step* detachement phase.



(a) Qubit measurement by displacing 2.25nM of respectively linker complex p0g0 and and dock oligo (negative control) using 100nM path-cover and 100nM cover. Exponential fit are performed, with affine component when necessary.



(b) Qubit measurement path attachement after adding 100nM path-cover (left graph), and the evolution of fluorsence after progressive adding of 25nM to 125nM of cover (right graph). The signal is not monotonic but we show that with both pipetting loss (measured) and volume loss the signal converges.

Figure 8.11: Qubit experiments on p0g0. Every used tube was passivated with BSA. Annealing and preparation of samples for p0g0 version at 1uM. Dillution of the linker complex to 2.5nM were made in 2mL lobind eppedorf first.

Stronger toeholds

Figure 8.12 shows the excess experiment made with four combinations of the toehold versions: from weaker to stronger, p0g0, p1g1,p2g2,p3g3. First, we observe that the fluorescence of each sample varies a lot (from 550 to 750). The fluorescent strand is always the same, but for each version, one strand colocated in the linker complex changes. This difference of colocated sequence could explain this fluorescence variability. Also, as the assembly yield of the linker complex could vary between the sample, because the stock concentrations of the components was not checked. Finally, a more probable explanation is the pipetting error. To keep the linker at 1 µM to prevent loss, we had to pipette 1.25 µL using 2.5 µL pipette. We could measure a maximum 7% error when pipetting this with a precision scale. However being at excess *step-cover* and *cover* conditions, the kinetics is should not dependent on the linker complex concentrations.

To study the kinetics, as previously, we fit a exponential plus affine curve. We observe that the stronger the attaching toehold is (g_i) the smaller the time constant is. For the p1g1 it is not the case, but the strong affine effect could lead to a wrong estimate of the time constant. A similar effect happens for *step* detachment. The stronger the detachment toehold is (p_i), the smaller the time constant is. About the two-step model, the same problem time constant for attachment and detachment happens for the versions p0g0 and p1g1. For the versions p2g2 and p3g3, this is not the case. If our hypothesis is true, it would be because the kinetic term associated to *step* as a monomer becomes insignificant.

Finally, we discuss the final fluorescence level for each curve. This level is interpretable, because remaining monomer *step* should be bound to excess covers at this stage. Because of the difference of fluorescence levels, we are not convinced that every minimum signal level corresponds to fully assembled *step* complexes. However, this still gives a bound for the *step* assembly yield, which is clearly not balanced.

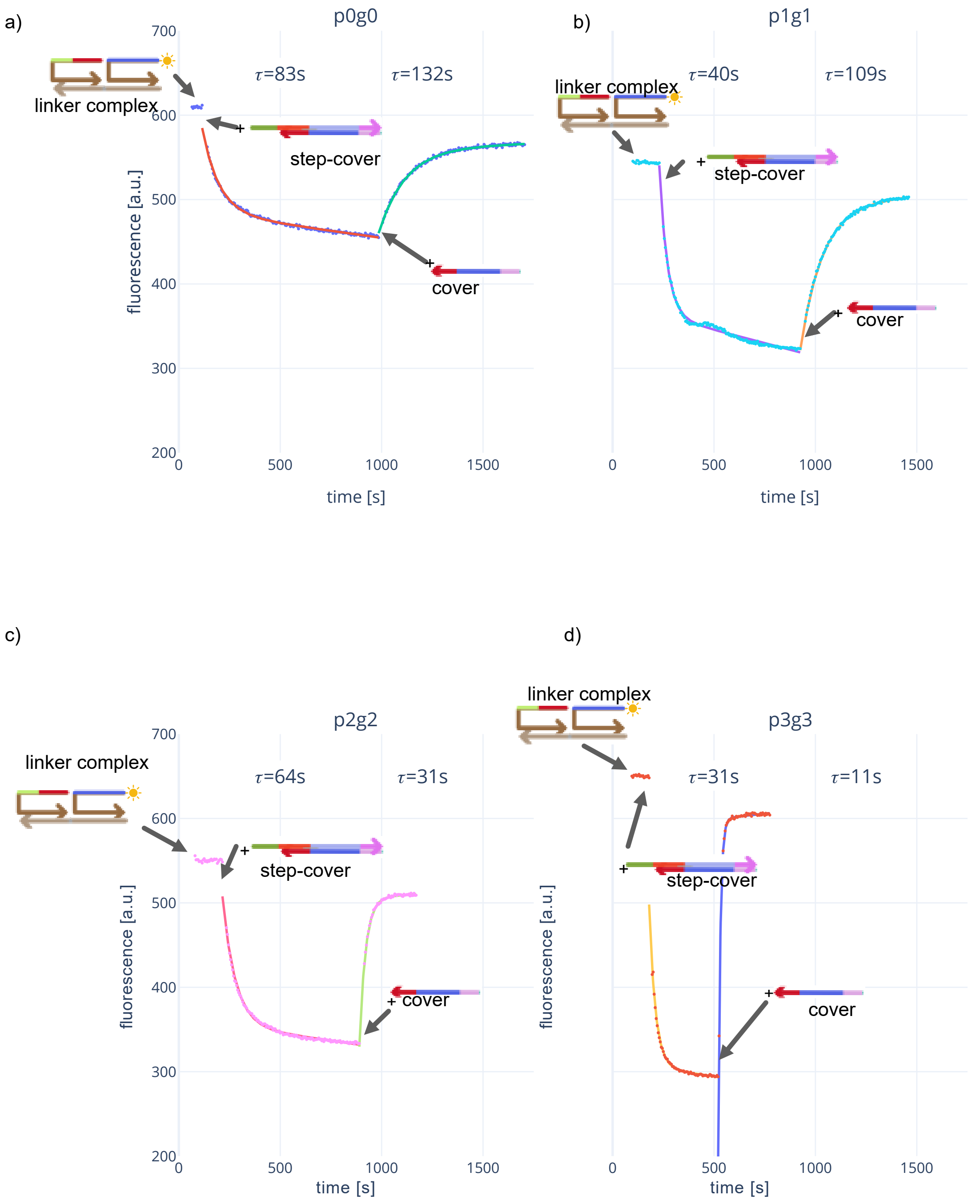


Figure 8.12: We prepared two kind of samples. The path-dock, that we use to know the fluorescence of quenched dock. We used dock only , which is only the dock domain, with 6-FAM on 5'end. For the kinetic experiments, we add the buffer TAE 1X Mg12.5mM, then 1.25uL of 1uM linker complex (2nM in Vtot = 500uL). Then we proceed in two parts: the "path-attachement part" by adding 25uL of path-cover (50nM in Vtot = 500uL), we wait to equilibrium is possible. Secondly the "path-detachment part" by adding 25uL of cover (50nM in Vtot = 500uL).

The Toy Model B validates that our variation of toehold exchange strand displacement works: the energy of the path-attachment toehold has an effect on the path attachment rate, indicating a cooperative effect between the dock and the step-attachment toehold, that are colocalized using the linker complex. We have strong and consistent arguments to explain that the high rate of the negative control is due to the presence of *step* monomers. Even if we could not measure both *step* attachement and detachement rates due to this, we could still analyze the final fluorescence levels. However, we did not find a toehold pair that produced a perfectly balanced system.

8.3 Toy Model C : Origami Substrate

Even if we could not find a balanced system for Toy Model B, we tried to replicate the result of toehold effect on the Toy Model C where the system is now supported on a DNA origami. We designed a new origami for this, in the shape of a line (called origami 1DRW) for future experiments. We describe assembly and purification and then proceed to fluorometer experiments on it.

8.3.1 Results on origami assembly and purification

Assembly We assembled it with a 10X concentration excess of staples. We observed the result both with AFM (Figure 8.14) and on an agarose gel (Figure 8.13). This validates the design as well as the fast annealing because we observe only one conformation on the gel, and it has the correct shape on AFM imaging.

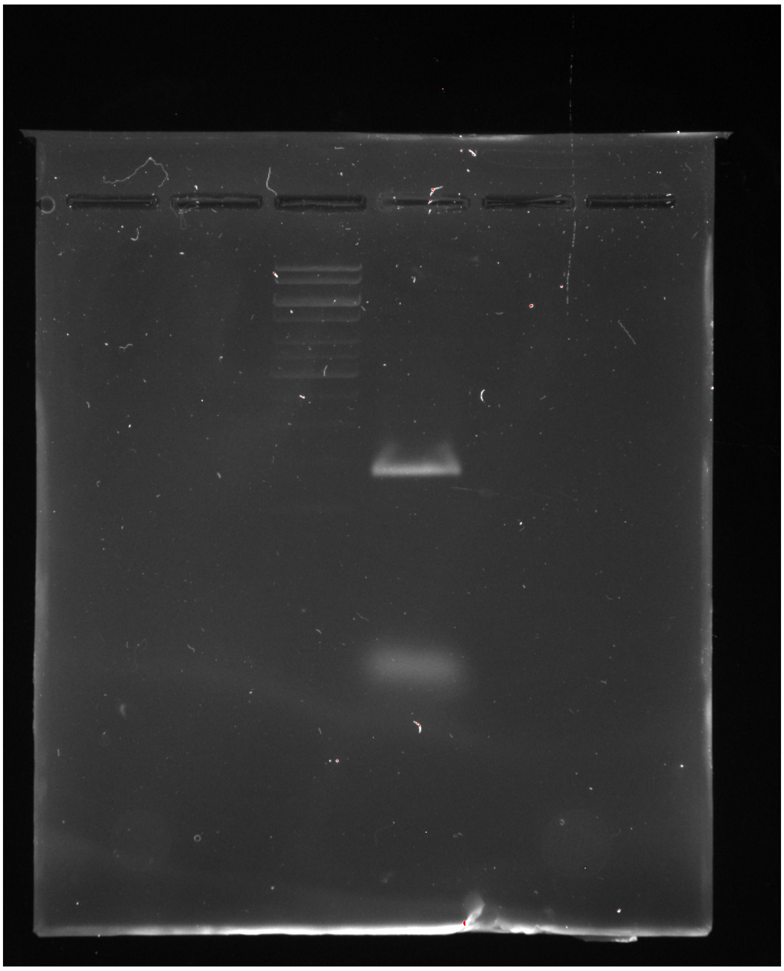


Figure 8.13: Gel of 1DRW annealed dock on lane 4. Lane 3 contains 1kb generuler ladder.

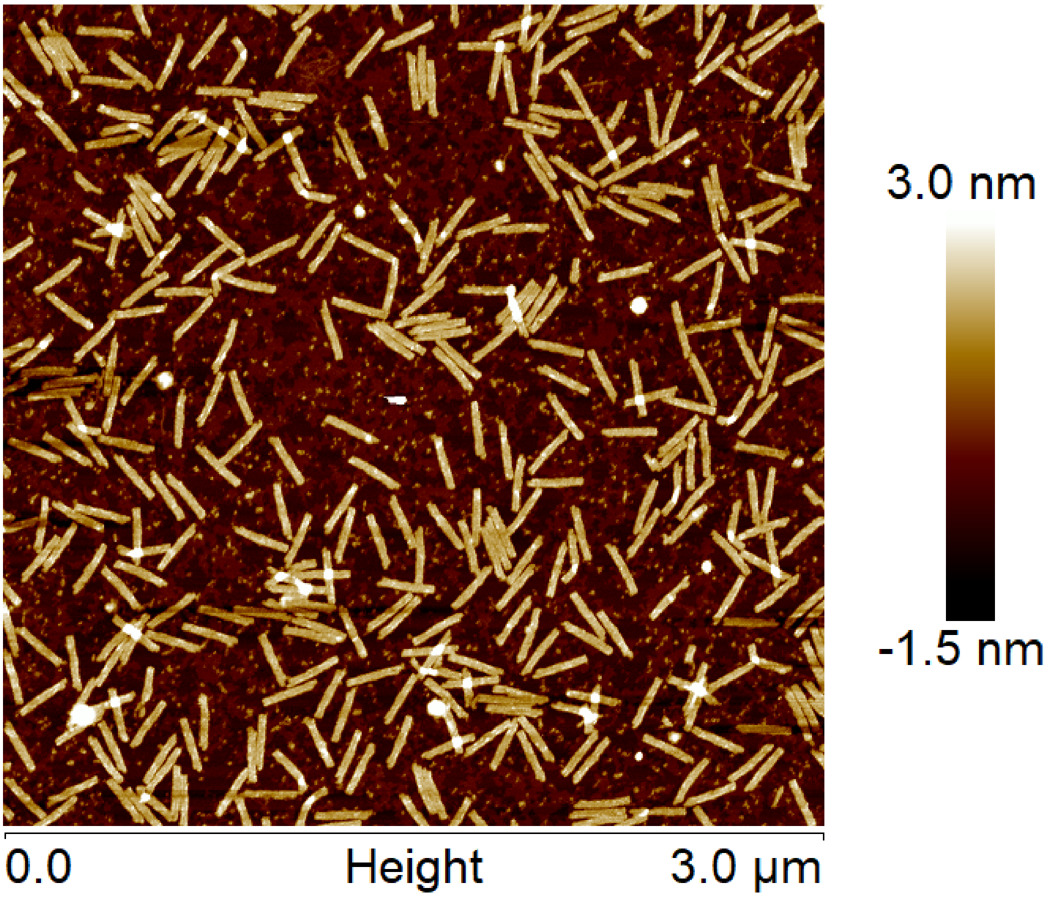
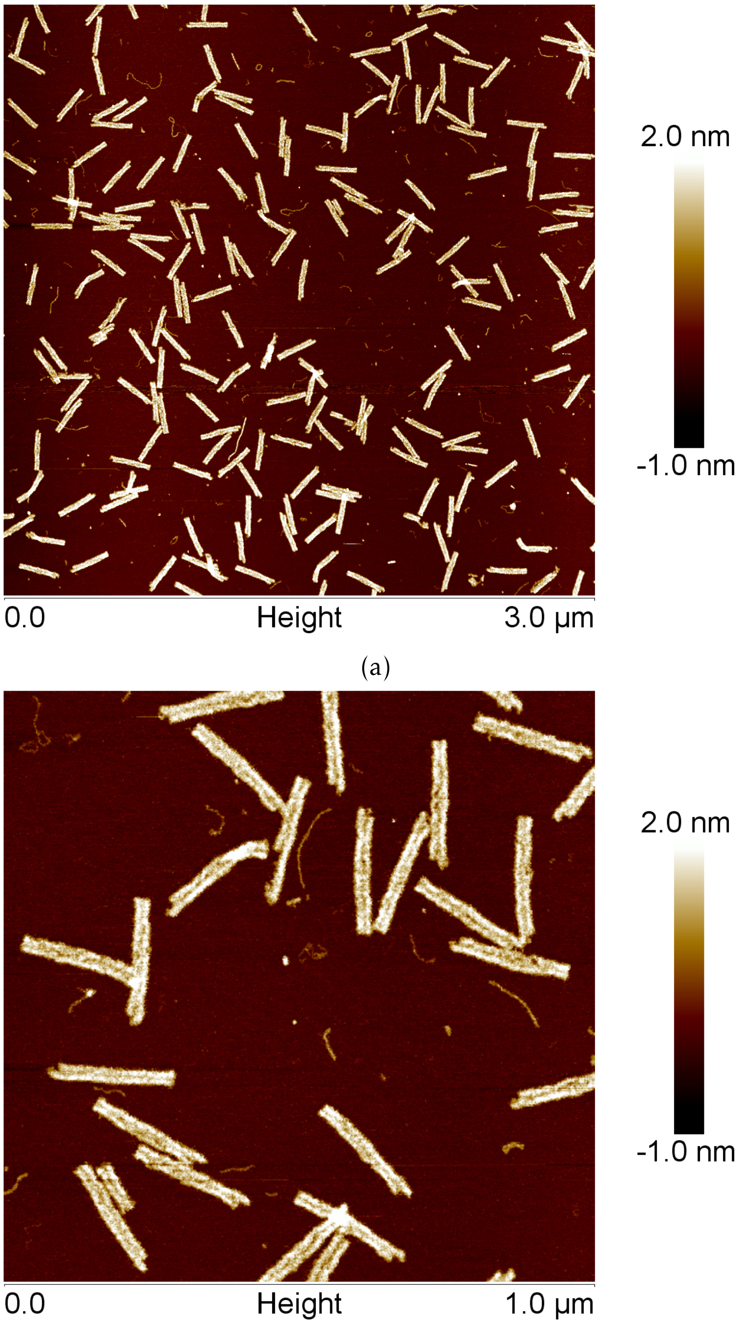


Figure 8.14: AFM image in air of the 1DRW origami. The origami is not purified, many staples are visible on the mica.

Purification We validated the purification on agarose gel 1%, with intercalating dye only in the sample. This makes sure that only a minimal amount of Intercalating dye remains in the buffer after Montage extraction kit which only extract the buffer and its content from the extracted agarose parts. This is important for us because we want no signal from the intercalating dye, which is furthermore more prone to photobleaching than our fluorophore.



(a)

(b)

Figure 8.15: AFM image in air of the purified DNA origami for Toy Model C. In a) or b), very few staples are visible.

8.3.2 Fluorometer experiments

Preliminary results of the cooperative effect of path-attachement

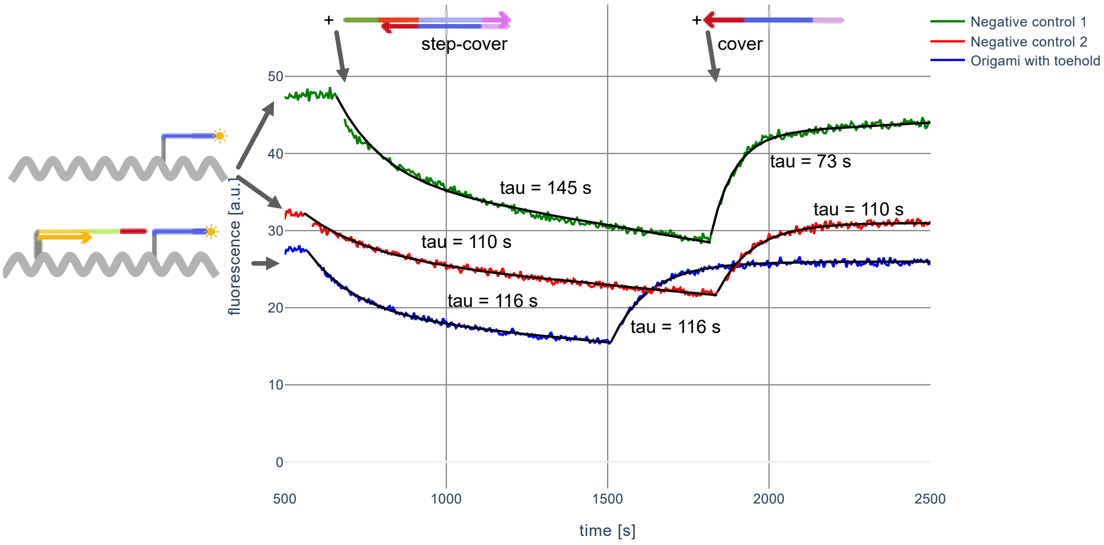


Figure 8.16: Flurometer experiment on the line origami. The experiment is conducted in two steps : 1) the assembly of the step strand with the introduction of the step-cover complex. 2) the diassembly of the step strand with the introduction of the cover strand.

We prepared two versions of the origami. The first one is the negative control, with only the dock staple to observe toehold-less step attachment. The second one has the dock staple, and the previous step staple, showing the g3 toehold (stronger version).

We observe in a fluorometer the signal emitted by the dock staple. First a sample of a 6-FAM-modified oligo purified in HPLC, and the same buffer gave us a signal reference for 2 nM. This let us estimate the origami concentration based on the initial fluorescence. The fluorescence of the origami could be slightly decreased due to the presence of the surface of the origami. Similarly as in the Toy Model B, because the experiments are done in excess *step-cover* and *cover* conditions, the results should not be affected by the variations of initial concentrations of origami.

We did three measurements :

- **Negative control 1** : origami with the dock staple. The origami is estimated to 0.39 nM. Then 10 nM of *step-cover* are added, followed by 10 nM of *cover*.
- **Negative control** 2: origami with the dock staple. The origami is estimated to 0.26 nM. Then 10 nM of *step-cover* are added, followed by 10 nM of *cover*.
- **toehold with p3g3** : origami with the dock staple, and the previous step staple. The origami concentration is estimated to 0.22 nM. Then 7.2 nM of *step-cover* are added. followed by 7.2 nM of *cover* are added. The idea was to keep the same ratio [step-cover]/[cover] as in the negative control 1.

The negative controls are not made exactly in the same conditions of excess of step-cover and cover. However due to the similar concentrations (10 nM and 7.2 nM), as a first analysis, we will consider that this concentration was identical. First, the two negative controls signals gives different time constants. According to the two-state model, it should not be the case, because they have with different origami concentrations, but the same step-cover and cover concentrations. The linear effect on the signal could affect the time constants, and explain this difference. Then by comparing the negative control 2, and the sample with the toehold (p3g3), the origami concentrations are similar, as well as the time constants. This seems to indicate that we don't see a cooperative effect between dock staple, and the previous step staple on the strand displacement.

However, due to a mistake, the previous step does not have a single-stranded T^6 sequence between the origami and the toehold domain. First, this error could result in the impossibility to assemble the step strand by mediation of the toehold. This could either be because of an infeasible distance, or a too high energetic barrier to fully assemble the path without leaking. Two things should be done: 1) perform oxDNA simulations to observe the behavior of the oxDNA DNA model, with and without the removal of T^6 sequence on the staple, when the *step* strand is attached. 2) redo the same experiment with this T^6 sequence.

The same experiment with the single-stranded T^6 on the staples should be done, in order to valid

> We successfully designed and assemble a DNA origami for supporting the Toy Model C. On this Toy Model C, we tried to replicate the effect of Toy Model B, using the strongest toehold we had. In the fluorometer curves, the time constants with the negative control are too similar to confirm the presence of cooperative effect with the toehold. We do not know if no cooperative effect occurs, or if the high rate of leaking due to a possible effect of *step* monomers hide the cooperative effect.

Conclusion and Perspectives

Summary of progress. In this work, we explored ways of implementing of DNA random walk, supported by a DNA origami.

In the first part, we focused on the assembly of paths on DNA origami. We experimented with two new designs of single-layered origami to prevent transient folding. It gave us the opportunity to perfect our design and assembly techniques. In particular, we pointed out difficulties such as the control of landing face on mica. On this DNA origami, we tried annealing paths made of multiple strands. We could validate that this kind of path can be visualized with our equipment. Because of the successful assembly of a hard-coded path, we challenged our design by relaxing the constraints. With this, we could assemble paths using a single set of 4 DNA strands. We tried to carefully design the energies to be able to control the kinetics of assembly similarly to kTAM technique of assembly.

In the second part, we designed a random walk system based on the strand displacement technique. We approached it with a bottom-up methods using three Toy Models or increasing complexity. Toy Model A let us acquire knowledge about strand displacement experiments. We observed an imbalance in the system, and tried several methods, such as the use of guanine quenching or toehold mismatch, to try to solve it. It let us explore different ways of measuring the fluorescence, which in turn sensitized us to the possiblity of strand loss on the tube surfaces. The experiments in excess *step-cover* and *cover* cover conditions possess spurious behaviors. This is possibly due to the presence of non-duplexed *step*. In the future, the purification of such duplexes prior to experiments could give some insights. Then, the experiments on Toy Model B proved that our variation of toehold exchange strand displacement works, despite the energetic barrier created by the separation into two colocated strands. The two-state model, that was effective for Toy Model A, did not make sense for all the toeholds versions. We could measure that a high contribution of the signal evolution was not due to the presence of the toehold. A hypothesis is again the presence of non-duplexed *step*, which has to be taken into account in the kinetics. The fact that the stronger toehold versions are compatible with the two-states model supports this idea. Unfortunately, based on the final fluorescence levels, we did not observe a toehold pair leading to a balanced two-states system. In the future, the careful study of this data could help predict the appropriate sequence to reach a balanced system. Finally, as for Model C which operates on a DNA origami, we cannot conclude yet. We observe similar rates with the negative control, again because of non-duplexed *step* in the solution. There are two possibilities: 1) a cooperative effect could happen. But due to the effects with the origami surface effect, the rate would be low, and the effect would be hidden by the displacement without the toehold due to non-duplexed *step* in the solution; 2) the sequence error could spatially spread the two staples too much, which makes the presence of the toehold useless. The effectiveness of the Toy Model B suggests that the latter possibility is more likely.

Perspectives. This work opens many perspectives for the future of the project. About the methodology, we see several area of improvement. First many doubts would be solved by accurately measuring and adjusting the concentration of each strand. Similarly, one should use purified strands, in order to be sure that the measurement corresponds to the desired strand concentation. In our experiments we could measure and control fluoresence (and so strand) loss when pipetting inside the cuvette, and when measuring the signal in Qubit plastic tubes. We had experiments where some loss happended in Eppendorf tubes. Our solution was to use only 1 µM concentrated samples, which is not always practical. Some work could be done to control the concentration from the preparation to the use in experiments. When assembling duplexes or linker complexes, we had to be in stoechiometry conditions, which means that there are some monomers left.These monomers are likely to explain many leaky observations. In the future, the use of acrylamide gel could solve this problem.

Finally for the experiment with a set of different toehold energies, it would be more advised to use one long toehold sequence on one strand, and prefixed of the complemery toehold in variations, similarly to the litterature. This way, the relationship between the binding energy and the energy sequence would be more predictable.

Now we discuss potential strategies for the completion of the project. First, there is work to be done on oxDNA simulations to obtain an estimation of the energy barriers and rates in Toy Model B and Toy Model C. Also, the reduction of displacement rate on the negative control will be crucial. First, it affects the correct measurement of the rate we are interested in. Then, if this is due to leaking and not due to the presence of *step* monomers, it will strongly affect the random walk. Finally, replication of the result of Toy Model B, on the origami, and with the correct sequence, will be a significant milestone for the implementation of the random walk on the origami.

Bibliography

[23] *Optimizing the Specificity of Nucleic Acid Hybridization - PMC*. https://www.ncbi.nlm.nih.gov/pmc/a Jan. 2023. (Visited on 01/26/2023).

[Bai+12] Xiao-chen Bai et al. “Cryo-EM Structure of a 3D DNA-origami Object”. In: *Proceedings of the National Academy of Sciences* 109.49 (Dec. 2012), pp. 20012–20017. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1215713109. (Visited on 11/01/2023).

[BK16] D. W. Bo Broadwater and Harold D. Kim. “The Effect of Basepair Mismatch on DNA Strand Displacement”. In: *Biophysical Journal* 110.7 (Apr. 2016), pp. 1476–1484. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.02.027. (Visited on 02/06/2023).

[Cha+19] Jie Chao et al. “Solving Mazes with Single-Molecule DNA Navigators”. In: *Nature Materials* 18.3 (Mar. 2019), pp. 273–279. ISSN: 1476-1122, 1476-4660. DOI: 10.1038/s41563-018-0205-3. (Visited on 05/01/2021).

[Che12] Xi Chen. “Expanding the Rule Set of DNA Circuitry with Associative Toehold Activation”. In: *Journal of the American Chemical Society* 134.1 (Jan. 2012), pp. 263–271. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/ja206690a. (Visited on 01/08/2024).

[Dey+21] Swarup Dey et al. “DNA Origami”. In: *Nature Reviews Methods Primers* 1.1 (Jan. 2021), p. 13. ISSN: 2662-8449. DOI: 10.1038/s43586-020-00009-8. (Visited on 12/24/2023).

[Dir+07] Robert M. Dirks et al. “Thermodynamic Analysis of Interacting Nucleic Acid Strands”. In: *SIAM Review* 49.1 (Jan. 2007), pp. 65–88. ISSN: 0036-1445, 1095-7200. DOI: 10.1137/060651100. (Visited on 05/24/2021).

[DP03] Robert M. Dirks and Niles A. Pierce. “A Partition Function Algorithm for Nucleic Acid Secondary Structure Including Pseudoknots”. In: *Journal of Computational Chemistry* 24.13 (Oct. 2003), pp. 1664–1677. ISSN: 01928651, 1096987X. DOI: 10.1002/jcc.10296. (Visited on 11/03/2021).

[DZ04] Roumen A. Dimitrov and Michael Zuker. “Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids”. In: *Biophysical Journal* 87.1 (July 2004), pp. 215–226. ISSN: 00063495. DOI: 10.1529/biophysj.103.020743. (Visited on 05/25/2021).

[Eva] Constantine Evans. “Crystals That Count!” In: ().

[EW17] Constantine G. Evans and Erik Winfree. “Physical Principles for DNA Tile Self-Assembly”. In: *Chemical Society Reviews* 46.12 (2017), pp. 3808–3829. ISSN: 0306-0012, 1460-4744. DOI: 10.1039/C6CS00745G. (Visited on 10/09/2023).

[FG53] Rosalind E. Franklin and R. G. Gosling. “Molecular Configuration in Sodium Thymonucleate”. In: *Nature* 171.4356 (Apr. 1953), pp. 740–741. ISSN: 1476-4687. DOI: `10.1038/171740a0`. (Visited on 01/06/2024).

[FPP20] Mark E. Fornace, Nicholas J. Porubsky, and Niles A. Pierce. “A Unified Dynamic Programming Framework for the Analysis of Interacting Nucleic Acid Strands: Enhanced Models, Scalability, and Speed”. In: *ACS Synthetic Biology* 9.10 (Oct. 2020), pp. 2665–2678. ISSN: 2161-5063, 2161-5063. DOI: `10.1021/acssynbio.9b00523`. (Visited on 05/01/2021).

[Gen+11] Anthony J. Genot et al. “Remote Toehold: A Mechanism for Flexible Control of DNA Hybridization Kinetics”. In: *Journal of the American Chemical Society* 133.7 (Feb. 2011), pp. 2177–2182. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja1073239`. (Visited on 09/29/2021).

[KMS83] Neville R. Kallenbach, Rong-Ine Ma, and Nadrian C. Seeman. “An Immobile Nucleic Acid Junction Constructed from Oligonucleotides”. In: *Nature* 305.5937 (Oct. 1983), pp. 829–831. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/305829a0`. (Visited on 12/25/2023).

[KP05] R. T. Koehler and N. Peyret. “Thermodynamic Properties of DNA Sequences: Characteristic Values for the Human Genome”. In: *Bioinformatics* 21.16 (Aug. 2005), pp. 3333–3339. ISSN: 1367-4803, 1460-2059. DOI: `10.1093/bioinformatics/bti530`. (Visited on 01/06/2024).

[LAS22] Jory Lietard, Dominik Ameur, and Mark M. Somoza. “Sequence-Dependent Quenching of Fluorescein Fluorescence on Single-Stranded and Double-Stranded DNA”. In: *RSC Advances* 12.9 (2022), pp. 5629–5637. ISSN: 2046-2069. DOI: `10.1039/D2RA00534D`. (Visited on 09/06/2023).

[Lor+11] Ronny Lorenz et al. “ViennaRNA Package 2.0”. In: *Algorithms for Molecular Biology* 6.1 (Dec. 2011), p. 26. ISSN: 1748-7188. DOI: `10.1186/1748-7188-6-26`. (Visited on 01/16/2024).

[LP00a] Rune B. Lyngsø and Christian N. S. Pedersen. “RNA Pseudoknot Prediction in Energy-Based Models”. In: *Journal of Computational Biology* 7.3-4 (Aug. 2000), pp. 409–427. ISSN: 1066-5277, 1557-8666. DOI: `10.1089/106652700750050862`. (Visited on 01/06/2024).

[LP00b] Rune B. Lyngsø and Christian N. S. Pedersen. “RNA Pseudoknot Prediction in Energy-Based Models”. In: *Journal of Computational Biology* 7.3-4 (Aug. 2000), pp. 409–427. ISSN: 1066-5277, 1557-8666. DOI: `10.1089/106652700750050862`. (Visited on 01/08/2024).

[LS] Nicolas Levy and Nicolas Schabanel. “ENSnano: A 3D Modeling Software for DNA Nanostructures”. In: ().

[MA20] Christopher Maffeo and Aleksei Aksimentiev. “MrDNA: A Multi-Resolution Model for Predicting the Structure and Dynamics of DNA Systems”. In: *Nucleic Acids Research* 48.9 (May 2020), pp. 5135–5146. ISSN: 0305-1048, 1362-4962. DOI: `10.1093/nar/gkaa200`. (Visited on 01/19/2024).

[Mac+14] Robert R. F. Machinek et al. “Programmable Energy Landscapes for Kinetic Control of DNA Strand Displacement”. In: *Nature Communications* 5.1 (Nov. 2014), p. 5324. ISSN: 2041-1723. DOI: `10.1038/ncomms6324`. (Visited on 01/10/2024).

[McK+03] Sean A. McKinney et al. "Structural Dynamics of Individual Holliday Junctions". In: *Nature Structural Biology* 10.2 (Feb. 2003), pp. 93–97. ISSN: 10728368. DOI: `10.1038/nsb883`. (Visited on 09/12/2023).

[MKT02] Salvatore A E Marras, Fred Russell Kramer, and Sanjay Tyagi. "Effciencies of Fluorescence Resonance Energy Transfer and Contact-Mediated Quenching in Oligonucleotide Probes". In: *Nucleic Acids Research* 30.21 (2002).

[Mon] Fabien Montel. "Dynamique à l'équilibre et hors d'équilibre de la chromatine visualisée par microscopie de force atomique: effet des variants d'histones et des facteurs de remodelage". In: ().

[Oul+13] Thomas E. Ouldridge et al. "DNA Hybridization Kinetics: Zippering, Internal Displacement and Sequence Dependence". In: *Nucleic Acids Research* 41.19 (Oct. 2013), pp. 8886–8895. ISSN: 1362-4962, 0305-1048. DOI: `10.1093/nar/gkt687`. (Visited on 06/24/2021).

[Por91] Dietmar Porschke. "Persistence Length and Bending Dynamics of DNA from Electrooptical Measurements at High Salt Concentrations". In: *Biophysical Chemistry* 40.2 (May 1991), pp. 169–179. ISSN: 03014622. DOI: `10.1016/0301-4622(91)87006-Q`. (Visited on 01/16/2024).

[QW11] Lulu Qian and Erik Winfree. "Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades". In: *Science* 332.6034 (June 2011), pp. 1196–1201. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.1200520`. (Visited on 10/20/2022).

[Rey+00] Luis P Reynaldo et al. "The Kinetics of Oligonucleotide Replacements". In: *Journal of Molecular Biology* 297.2 (Mar. 2000), pp. 511–520. ISSN: 00222836. DOI: `10.1006/jmbi.2000.3573`. (Visited on 01/11/2024).

[Rot] Paul W K Rothemund. "Folding DNA to Create Nanoscale Shapes and Patterns Supplementary Notes 1–1". In: (), p. 82.

[RPW04] Paul W. K Rothemund, Nick Papadakis, and Erik Winfree. "Algorithmic Self-Assembly of DNA Sierpinski Triangles". In: *PLoS Biology* 2.12 (Dec. 2004). Ed. by Anne Condon, e424. ISSN: 1545-7885. DOI: `10.1371/journal.pbio.0020424`. (Visited on 10/09/2023).

[San98] J. SantaLucia. "A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics". In: *Proceedings of the National Academy of Sciences* 95.4 (Feb. 1998), pp. 1460–1465. ISSN: 0027-8424, 1091-6490. DOI: `10.1073/pnas.95.4.1460`. (Visited on 05/23/2021).

[SAS96] John SantaLucia, Hatim T. Allawi, and P. Ananda Seneviratne. "Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability". In: *Biochemistry* 35.11 (Jan. 1996), pp. 3555–3562. ISSN: 0006-2960, 1520-4995. DOI: `10.1021/bi951907q`. (Visited on 01/05/2024).

[Sch+17] Joerg Schnitzbauer et al. "Super-Resolution Microscopy with DNA-PAINT". In: *Nature Protocols* 12.6 (June 2017), pp. 1198–1228. ISSN: 1754-2189, 1750-2799. DOI: `10.1038/nprot.2017.024`. (Visited on 01/17/2024).

[See82] Nadrian C. Seeman. "Nucleic Acid Junctions and Lattices". In: *Journal of Theoretical Biology* 99.2 (Nov. 1982), pp. 237–247. ISSN: 00225193. DOI: `10.1016/0022-5193(82)90002-9`. (Visited on 02/23/2024).

[Sen+21] A. Sengar et al. “A Primer on the oxDNA Model of DNA: When to Use It, How to Simulate It and How to Interpret the Results”. In: *Frontiers in Molecular Biosciences* 8 (June 2021), p. 693710. ISSN: 2296-889X. DOI: `10.3389/fmolb.2021.693710`. (Visited on 09/07/2023).

[SH04] John SantaLucia and Donald Hicks. “The Thermodynamics of DNA Structural Motifs”. In: *Annual Review of Biophysics and Biomolecular Structure* 33.1 (June 2004), pp. 415–440. ISSN: 1056-8700, 1545-4266. DOI: `10.1146/annurev.biophys.32.110601.141800`. (Visited on 05/23/2021).

[Sri+13] Niranjan Srinivas et al. “On the Biophysics and Kinetics of Toehold-Mediated DNA Strand Displacement”. In: *Nucleic Acids Research* 41.22 (Dec. 2013), pp. 10641–10658. ISSN: 1362-4962, 0305-1048. DOI: `10.1093/nar/gkt801`. (Visited on 09/29/2021).

[Thu+17] Anupama J. Thubagere et al. “A Cargo-Sorting DNA Robot”. In: *Science* 357.6356 (Sept. 2017), eaan6558. ISSN: 0036-8075, 1095-9203. DOI: `10.1126/science.aan6558`. (Visited on 05/01/2021).

[Tin+97] Bernard Tinland et al. “Persistence Length of Single-Stranded DNA”. In: *Macromolecules* 30.19 (Sept. 1997), pp. 5763–5765. ISSN: 0024-9297, 1520-5835. DOI: `10.1021/ma970381+`. (Visited on 01/16/2024).

[TPQ17] Grigory Tikhomirov, Philip Petersen, and Lulu Qian. “Fractal Assembly of Micrometre-Scale DNA Origami Arrays with Arbitrary Patterns”. In: *Nature* 552.7683 (Dec. 2017), pp. 67–71. ISSN: 1476-4687. DOI: `10.1038/nature24655`. (Visited on 01/18/2024).

[Wag+17] Klaus F. Wagenbauer et al. “How We Make DNA Origami”. In: *ChemBioChem* 18.19 (Oct. 2017), pp. 1873–1885. ISSN: 14394227. DOI: `10.1002/cbic.201700377`. (Visited on 05/01/2021).

[WB04] Erik Winfree and Renat Bekbolatov. “Proofreading Tile Sets: Error Correction for Algorithmic Self-Assembly”. In: *DNA Computing*. Ed. by Gerhard Goos et al. Vol. 2943. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 126–144. ISBN: 978-3-540-20930-0 978-3-540-24628-2. DOI: `10.1007/978-3-540-24628-2_13`. (Visited on 01/17/2024).

[Win+23] Christopher M. Wintersinger et al. “Multi-Micron Crisscross Structures Grown from DNA-origami Slats”. In: *Nature Nanotechnology* 18.3 (Mar. 2023), pp. 281–289. ISSN: 1748-3395. DOI: `10.1038/s41565-022-01283-1`. (Visited on 04/03/2024).

[Win+98] Erik Winfree et al. “Design and Self-Assembly of Two-Dimensional DNA Crystals”. In: *Nature* 394.6693 (Aug. 1998), pp. 539–544. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/28998`. (Visited on 12/25/2023).

[Win06] E. Winfree. “Algorithmic Self-Assembly of DNA”. In: *2006 International Conference on Microtechnologies in Medicine and Biology*. Okinawa: IEEE, May 2006, pp. 4–4. ISBN: 978-1-4244-0337-0 978-1-4244-0338-7. DOI: `10.1109/MMB.2006.251471`. (Visited on 10/09/2023).

[Woo+19] Damien Woods et al. “Diverse and Robust Molecular Algorithms Using Reprogrammable DNA Self-Assembly”. In: *Nature* 567.7748 (Mar. 2019), pp. 366–372. ISSN: 0028-0836, 1476-4687. DOI: `10.1038/s41586-019-1014-9`. (Visited on 02/06/2022).

[WSD17] Klaus F. Wagenbauer, Christian Sigl, and Hendrik Dietz. "Gigadalton-Scale Shape-Programmable DNA Assemblies". In: *Nature* 552.7683 (Dec. 2017), pp. 78–83. ISSN: 1476-4687. DOI: `10.1038/nature24651`. (Visited on 01/08/2024).

[YM03] Bernard Yurke and Allen P. Mills. "Using DNA to Power Nanostructures". In: *Genetic Programming and Evolvable Machines* 4.2 (June 2003), pp. 111–122. ISSN: 1573-7632. DOI: `10.1023/A:1023928811651`. (Visited on 09/15/2023).

[Zad+11] Joseph N. Zadeh et al. "NUPACK: Analysis and Design of Nucleic Acid Systems". In: *Journal of Computational Chemistry* 32.1 (Jan. 2011), pp. 170–173. ISSN: 01928651. DOI: `10.1002/jcc.21596`. (Visited on 12/22/2022).

[Zha+] Jinny X Zhang et al. "Predicting DNA Hybridization Kinetics from Sequence". In: (), p. 10.

[Zim+19] Zackary A. Zimmers et al. "Fluorophore-Quencher Interactions Effect on Hybridization Characteristics of Complementary Oligonucleotides". In: *Analytical Methods* 11.22 (2019), pp. 2862–2867. ISSN: 1759-9660, 1759-9679. DOI: `10.1039/C9AY00584F`. (Visited on 09/06/2023).

[ZW09] David Yu Zhang and Erik Winfree. "Control of DNA Strand Displacement Kinetics Using Toehold Exchange". In: *Journal of the American Chemical Society* 131.47 (Dec. 2009), pp. 17303–17314. ISSN: 0002-7863, 1520-5126. DOI: `10.1021/ja906987s`. (Visited on 05/01/2021).

Acronyms

A | B | D | E | I | K | M | N | P | Q | R | S | T

A

AFM Atomic force microscopy. 22, 35, 37, 53, 56, 60, 66–68, 74, 76, 81
aTAM abstract tile assembly model. 23–25

B

bp base pair. 6
BSA Bovine Serum Albumine. 107

D

DNA Deoxyribonucleic acid. 6, 7
dsDNA double-stranded DNA. 6, 8, 9, 11, 106, 108

E

EDTA Ethylenediaminetetraacetic acid. 6, 29

I

IEL intuitive energy landscape model. 28

K

kTAM kinetic tile assembly model. 23–25

M

MFE Minimum Free Energy. 12

N

NN Nearest-neighbor. 10, 11, 15

P

PCR Polymerase Chain Reaction. 9, 10, 56, 68
PEG Polyethylene glycol. 22, 32

Q

qPCR Quantitative Polymerase Chain Reaction. 9, 56, 78, 119

R

RFU Raw Fluorescence Unit. 71

S

ssDNA single-stranded DNA. 4, 7, 9, 11, 32, 86, 96, 106, 108

T

TAE Tris, Acetic Acid, EDTA buffer. 55, 107
TBE Tris, Borate, EDTA buffer. 108
TE Tris-EDTA buffer. 6, 29
TEM Transmission electron microscopy. 22
TESD toehold exchange strand displacement. 35, 86, 87, 96–98, 115, 117, 118
TMSD toehold mediated strand displacement. 27–29, 32, 35, 42, 86, 125

Index

Contents

SCALABLE DNA ALGORITHMS

Abstract

The *DNA computing* field consists in using DNA as dynamic building blocks. By interacting together, they can implement small algorithms and effectively compute.
Many successful approaches were made. For instance, by implementing logical circuits where reconfigurations of DNA complexes progressively evaluate the network. Another approach is to attach DNA strands according to defined rules to a substrate made of large DNA objects called *DNA origami*. However, all the current approaches face the challenge of scalability. In most designs, the size of the input is linked to either the DNA origami or the number of strands. The number of strands, is limited not only technically but also theoretically, as there is an inherent chance of hybridization error between two strands that are not fully complementary.
In this thesis, we want to solve this scalability issue on the particular problem of *maze solving*. This problem was already solved in both in a non-reversible and non-scalable fashion. We propose to implement a reversible random walk walker on a DNA origami. Our point is twofold. First, we can make a design with only four different strands, no matter the size of the maze. Most importantly, using reversibility is a key factor, as it can harness randomness to reverse hybridization errors.
In the first part, we conducted experiments where we attached static paths made of DNA strands on a DNA origami. We will validate our ability to both conduct, observe and process these experiments. In the second part, we propose an implementation of a reversible random walk using a variation of the *toehold mediated strand displacement* technique. We have conducted and developed experiments on this variation using a bottom-up approach. Our experiments led to preliminary results of the technique on a DNA origami.

Keywords: dna computing, dna origami, strand displacemement, scalable, random walk

---

Laboratoire de l'Informatique du Parallélisme
ENS de Lyon, – 46, allée d'Italie, – 69364 LYON CEDEX 07, FRANCE