N° d'ordre: 4585

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ BORDEAUX I

ÉCOLE DOCTORAL DES SCIENCES PHYSIQUES ET DE L'INGÉNIEUR

Par M. Jean-Michel ARBONA

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ: LASERS ET MATIÈRE DENSE

Origami d'ADN: étude des propriétés mécaniques et du processus de formation.

Soutenue le 24 Septembre 2012

Après avis de:

F.C. SIMMEL	Professeur, Technische Universität München	Rapporteur
V. CROQUETTE	Directeur de recherche, ABCD Lab	Rapporteur

Devant la commission d'examen formée de :

J.-P AIMÉ	Directeur de Recherche, CBMN	Directeur de thèse
V. CROQUETTE	Directeur de recherche, ABCD Lab	Rapporteur
J. ELEZGARAY	Directeur de Recherche, CBMN	Directeur de thèse
J.F. JOANNY	Professeur des Universités, Institue Cury	Jury
B. LOUNIS	Professeur des Universités, LP2N	Jury
F.C. SIMMEL	Professeur, Technische Universität München	Rapporteur
J. STURGIS	Professeur des Universités, LISM	Jury

Contents

Version résumée **i**
Introduction . . . i
Résumé . . . ii
Conclusion . . . iv

Structure of the document **v**

1 Can we master energy and information at the nanoscale ? **1**
1.1 DNA Origami . . . 3
1.1.1 Principle of formation . . . 4
1.1.2 Design of structures . . . 5
1.1.3 Functionality . . . 6
1.1.4 An active field! . . . 7
1.2 Diving in the nanoworld with DNA . . . 9
1.2.1 Energy scale . . . 9
1.2.2 Length and time scales . . . 10
1.2.3 DNA, a perspective on bioinspired technologies. . . . 11

2 Mechanical properties of DNA constructs **13**
2.1 Introduction: . . . 14
2.2 Analytic model of elasticity. . . . 16
2.2.1 Introduction to an analytic model of DNA. . . . 16
2.2.2 Analytic model of ssDNAs: . . . 18
2.2.2.1 Persistence length . . . 18
2.2.2.2 Force distance . . . 18
2.2.3 Analytical model of a dsDNA: . . . 20
2.2.3.1 Persistence length of a simple dsDNA : . . . 20
2.2.3.2 Energy necessary to bend small dsDNA . . . 21
2.2.4 Simple model to calculate the rigidity of a DNA origami: . . . 22
2.3 Persistence length simulation: . . . 25

2.3.1 Mergel's Energy Model . . . 25
2.3.1.1 DNA electrostatics . . . 26
2.3.2 General description of Monte Carlo Simulations . . . 28
2.3.3 Metropolis algorithm of transition between states in Monte Carlo Simulations . . . 29
2.3.4 Monte Carlo moves . . . 30
2.3.5 Averaging values in MC simulation . . . 31
2.4 Results . . . 34
2.4.1 The twist parameter: . . . 34
2.4.2 The electrostatic parameter: . . . 38
2.4.3 Comparaison with experimental data . . . 40
2.4.3.1 Calculation of the persistence length of DNA construct . . . 40
2.4.3.2 Comparison with indirect measurement ([65]) . . . 43
2.4.3.3 Comparison with direct measurements ([8]) . . . 45
2.4.3.4 Comparison with classical model of elasticity. . . . 46
2.4.3.5 CanDo-DNA website. . . . 47
2.4.4 Study of the persistence length for different constructs . . . 49
2.4.4.1 Effect of the crossover density . . . 49
2.4.4.2 Increasing the number of dsDNA: . . . 50
2.4.5 Force curves measurement of bending on DNA constructs . . . 51
2.4.5.1 Umbrella Sampling. . . . 51
2.4.5.2 Detailed steps of the Umbrella sampling method. . . . 51
2.4.5.3 Result on dsDNA . . . 55
2.4.5.4 Force curves measurement of bending on oDNA. . . . 57
2.5 Conclusion . . . 58

3 Process of formation of DNA origamis: 59
3.1 Introduction. . . . 60
3.2 Experimental techniques to study DNA hybridisation. . . . 63
3.3 Small origami . . . 65
3.3.1 Experimental datas . . . 65
3.3.1.1 Effect of the position: . . . 65
3.3.1.2 Effect of the cooperativity between strands: . . . 67
3.3.1.3 Effect of adding an adaptor. . . . 70
3.4 Thermodynamic model on a lattice for small origami . . . 71
3.4.1 The model . . . 71
3.4.2 Results . . . 74
3.4.2.1 Double stranded DNA . . . 75
3.4.2.2 Holliday junction . . . 75

3.4.2.3 modelling small origami . . . 78
3.4.3 Small origami, comparison with experiment . . . 84
3.4.4 Conclusion on the coarse-grained model . . . 84
3.5 Origami . . . 86
3.5.1 Experimental data. . . . 86
3.5.1.1 Melting curve of different origamis. . . . 86
3.5.1.2 Effect of different experimental parameters on the melting-annealing curves: . . . 88
3.6 Thermodynamic model of the formation of a DNA origami . . . 91
3.6.1 Ideas of the model . . . 91
3.6.2 Folding energy of a given configuration . . . 94
3.7 Comparison with the annealing-melting data of four origamis . . . 98
3.7.1 Kinetic effect. . . . 98
3.7.2 Understanding DNA Origami design . . . 99
3.7.2.1 Circular permutation of the scaffold. . . . 99
3.7.2.2 Decreasing the number of crossovers . . . 100
3.8 Conclusion and perspectives. . . . 102

4 DNA origami as a Raman platform (work done in collaboration with S. Bidault) 103
4.1 Goal . . . 103
4.2 State of the art in using DNA to create SERS structure . . . 104
4.3 Introduction to Raman phenomenon and SERS technique. . . . 109
4.3.1 The Raman phenomenon: . . . 109
4.3.2 The SERS phenomenon: . . . 109
4.4 Experimental results: . . . 112
4.4.1 Protocol for coupling NP with ssDNA . . . 112
4.4.2 Attachment of the NP to the origami. . . . 114
4.4.3 Raman experiment. . . . 116
4.4.3.1 Schema of principle of the experiment . . . 116
4.4.3.2 principle . . . 117
4.4.3.3 Attachment of oDNA to a quartz surface . . . 118
4.4.3.4 Raman experiments . . . 118
4.5 Conclusion . . . 122

5 Conclusion 123

A List of communications 125

B Experimental attempt. 127

C Thermodynamic values for the soDNA **129**

D Computing the probabilities from the law of mass action. **131**
D.1 Remarque on the small origami. 133

E Quenching intermediate states. **135**
E.1 Study of the O1 origami 135
E.2 Study of the O4 Origami 140

F Calculation of the absorbance **145**

G Temperature and salt corrections, calibration and results on small origamis **147**
G.1 Temperature corrections 147
G.2 Salt correction 147

H Manip Supplementary **149**
H.1 Effect of the salt concentration 149
H.2 Effect of the excess on O1 150
H.3 Manip with origami O4 at slower rate. 151
H.4 Manip on one half of the strand. 152
H.5 Manip with a ratio 1 for 1 153

I Treatments of absorbance curves **155**
I.1 Generality 155
I.2 Detail on the treatment of the curves for oDNA experiments. 155

J Strands used in the differents structures **163**
J.1 Small Origami 163
J.2 Origami O1 and O2 163
J.3 Origami O3 165
J.4 Origami O4 166
J.5 Origami O5 170

K Raman spectra of different molecules **175**

Bibliography **179**

Version résumée

Introduction

L' objet d'étude de cette thèse est l' origami d' ADN issue du travail de Rothemund [1]. Le nombre important d'applications utilisant cette technique, et leurs diversités, sont une preuve que ces structures présentées en 2007 sont une avancée importante pour la technologie ADN. Ces applications vont du détecteurs ARN à forte sensibilité [2] à la visualisation par AFM de la variation d'une base d'ADN dans une séquence donnée [3], en passant par une nano-usine fabriquant des assemblages complexes de billes d'or [4]. Les formes des structures, initialement limitée aux structures bi-dimmentionelles, ont été rapidement étendues à de nombreux types de structures tri-dimmentionelles [5, 6].

Au début de cette thèse, du fait de la facilité de création de ces origamis, ainsi que la robustesse du processus de formations, les propriétés mécanique de ces objets, ainsi que leurs processus de formation n'étaient pas étudiés. Cette thèse est donc dédiée à la compréhension de ce processus de formation qui implique 200 brins d'ADN diffèrent ainsi qu'à la modélisation des propriétés mécaniques de ces objets. Finalement une applications de ces structures à la détection SERS est proposée, et les résultats expérimentaux concernant la réalisation de cette structures sont discutés.

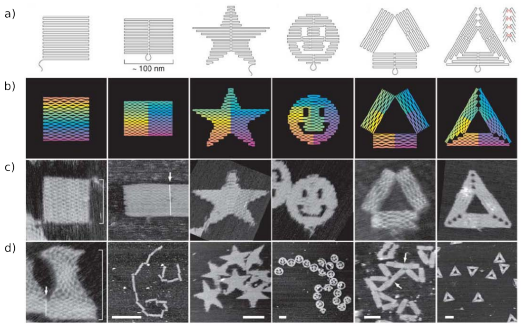


Figure 1: Extrait de Rothemund et al [1]. Différents exemples de schéma d' origamis d'ADN (a,b); et leurs réalisations expérimentales (c,d)

Résumé

La première partie de cette thèse est consacrée à l'étude des propriétés mécaniques des origamis d'ADN. Un modèle couramment utilisé pour décrire les caractéristiques de brins d'ADN est le modèle WLC (Worm-like chain), dans lequel, la longueur de persistance est une mesure de la rigidité de la structure. Pour un double brins d'ADN (dsDNA) cette longueur est de 50 nm. En dessous de cette longueur un brin est considéré comme rigide et au dessus comme un polymère statistique. Un origami d'ADN plan peut être considéré comme un ensemble de dsDNA couplés entre eux par des crossovers (des simples brins d'ADN passant d'un double brin à un autre). Une première approximation pour la rigidité d'une telle structure est de considérer que la longueur de persistance de l' origami est proportionnelle au nombre de dsDNA, le facteur de proportionnalité restant à déterminer (Figure 2).

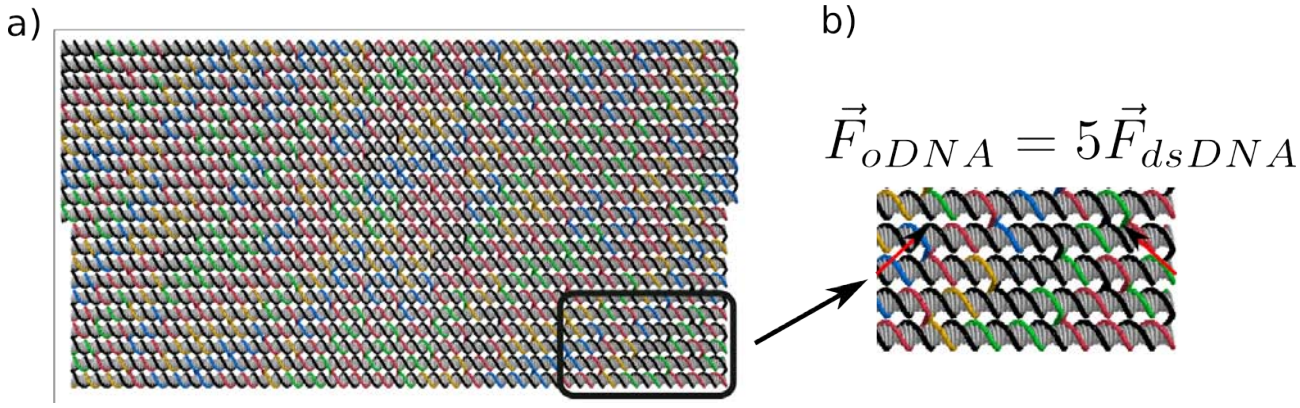


Figure 2: a) Représentation schématique d'un origami d'ADN constitué de 20 brins d'ADN b) Zoom. Est-ce que la force nécessaire pour plier un origami (F_{oDNA} est proportionnelle au nombre de brins qui le constituent?

Des simulations de type monte-carlo, basé sur un modèle [7] publié de dsDNA ont alors été réalisées. Ce modèle a été étendu à plusieurs brins couplés mécaniquement et interagissant entre eux par une interaction de type Debye-Huckel. L'importance de la précision de la connaissance du twist de l' ADN dans un buffer donné a été soulignée car une légère variation de cette valeur peut entraîner une grande déformation de la structure finale. Une comparaison des résultats obtenus par ce modèle avec les différents résultats expérimentaux disponibles a permis la validation du modèle, et le facteur de proportionnalité a été déterminé: 1.7 pour les structures les plus longues (300 bases) et 1.5 pour les structures plus courtes (190 bases). Nous avons également montré que des structures tridimensionnelle pouvait être modélisées par une approche de mécanique classique pour obtenir des résultats en accords avec les expérience menées par Kauert et al [8].

Le processus de formation de ces structures est le sujet de la deuxième partie de cette thèse.

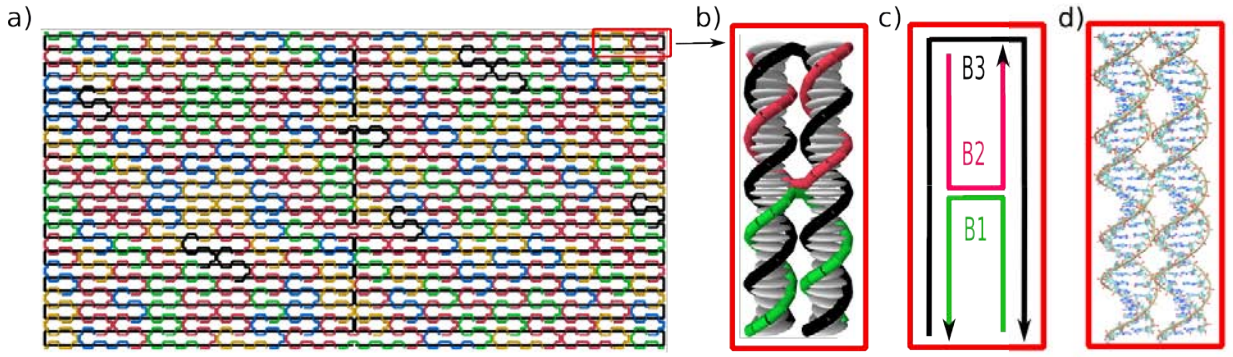


Figure 3: a) Représentation d'un origami composé de 200 brins d'ADN ;b,c) Deux représentation schématique du soDNA ; d) représentation tous atomes du soDNA

Nous nous sommes tout d'abord intéressé à de petites structures d'ADN (soDNA Figure 3) composées de trois brins dont nous avons fait varier la composition. Le processus de leur formation a été étudié par spectroscopie UV, pour obtenir des courbes de melting qui rendent compte de la formation de la structure en fonction de la temperature. En choisissant des brins aux compositions très différentes il a été possible d'étudier les effets coopératifs du processus de formation ainsi que l'effet de la position dans la structure elle même. Un modèle sur réseau à l'échelle de la base a été adapté et validé par rapport à des résultats existant sur les jonctions Hollyday. Il a permis également de rendre compte des principaux résultats observés expérimentalement.

Suites a ces résultats nous avons étudié 4 origamis et établi d'importantes différences dans leurs courbes de melting. Le précèdent modèle ne permettant pas d'étudier des structures aussi grandes, nous avons développé un second modèle à l'échelle du brins d'ADN. Celui-ci est basé sur les résultats obtenus sur les soDNA et à permis une reproduction des courbes de melting des 4 origamis. Il a également permis la création d'un origami plus stable grâce à la compréhension des principes de bases en jeux lors du processus de création.

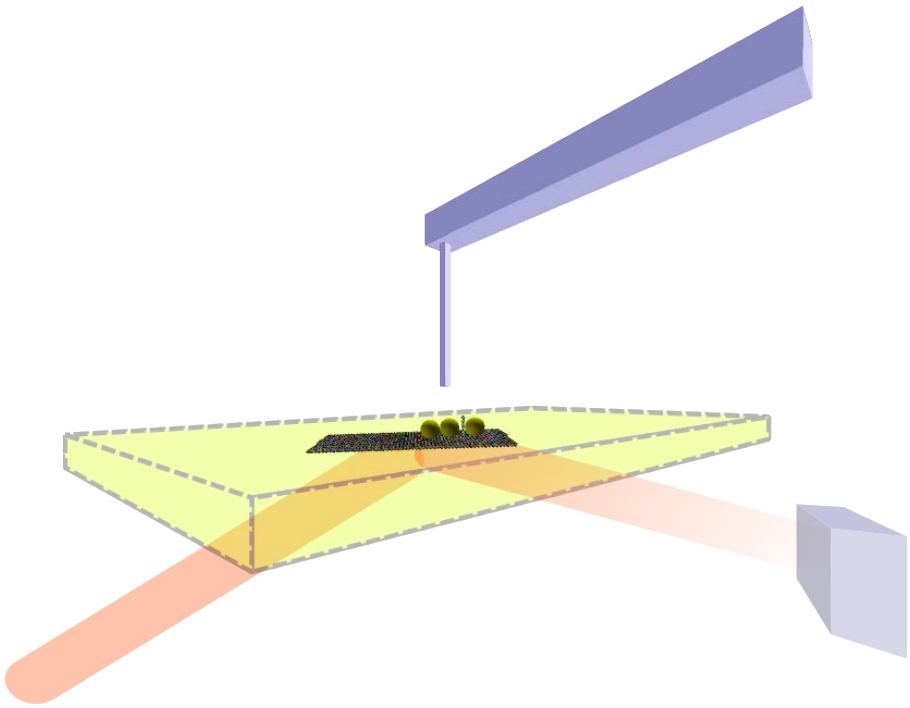


Figure 4: Experimental setup: The oDNA are deposited on a transparent surface. They are then detected by afm, and a Raman spectra is collected by reflection.

Finalement le dernier chapitre de cette thèse est consacré à la réalisation d'une plate-forme SERS (Figure 4) Celle-ci consiste en un origami d'ADN auquel est accroché des nanoparticules d'or de taille variable. Ces particules d'or, à la manière de lentilles permettent de focaliser un champs électromagnétique entres elles. Le positionnement d'un système permettant de piéger la molécule à détecter entre ces particules d'or, permet la détection par spectroscopie RAMAN de cette cible. Il nous a été possible de réaliser de telles structures avec des particules d'une taille de 10 nm. Cependant cette taille de particule ne permet pas l'obtention d'un signal RAMAN d'intensité suffisante pour être séparé du signal des solutions de buffer utilisés. Différentes stratégies ont été utilisées: grossissement des billes de 10 nm par dépôt d'or une fois la structure formée, attachement direct de particules plus grosse, sans succès.

Conclusion

J'espère que ce court résumé vous donnera envie de lire le corps principal de la thèse qui a été rédigé en anglais. La liste des différentes communications orales ainsi que les publications qui ont été acceptées et celles en cours d'écriture est également disponible (cf Appendix A)

Structure of the document

This page is a short overview of the structure of this document and summarises what will be presented in the different chapters:

The introduction is divided in two sections. The first one is a presentation of the DNA origami technique, in which we present the structures one can fabricate, the process of design, and a few examples of new functionalities a DNA origami platform can provide. The second section is a short discussion on the scales of interest in energy, length and time, ending with a focus on the time scale involved in bioinspired technology.

The chapter *Mechanical properties of DNA constructs* deals with the study of the flexibility of DNA origamis. The first approach of this study was through the use of polymer physics, as it is the classical way to study DNA. We then used computer simulations to model the system in a more detailed manner and to extract general rules on the mechanical behaviour of DNA constructs.

The chapter *Process of formation of DNA origamis* first develops an experimental study of the process of formation realised on the simplest origami that we could envision. This study was intended to investigate basic principles on the process of formation of DNA structures. A coarse grain model is then developed to have a first insight onto the formation process. Then an experimental study on large origamis follows with a modeling of the annealing and melting curves based on the principles determined from the study of the simplest origami.

The chapter *DNA origami as a Raman platform* presents the theory of Surface Enhanced Raman Spectroscopy and the progresses made in order to use DNA origamis as a platform for the detection of small concentrations of RNA molecules. The progresses as well as the difficulties encountered when mixing bio-material with metallic colloids are presented.

Then a general conclusion summarises and closes the work done during these four years.

Chapter 1

Can we master energy and information at the nanoscale ?

Contents

1.1 DNA Origami **3**

1.1.1 Principle of formation 4

1.1.2 Design of structures 5

1.1.3 Functionality 6

1.1.4 An active field! 7

1.2 Diving in the nanoworld with DNA **9**

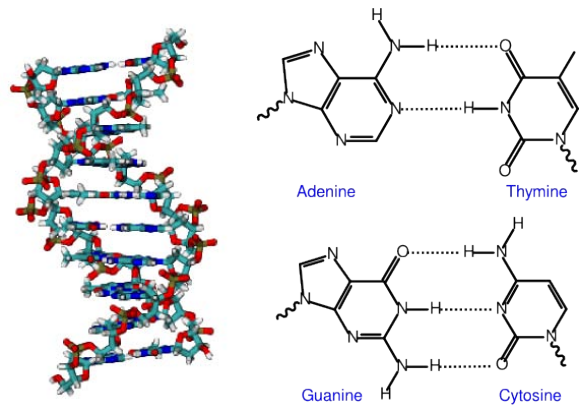
1.2.1 Energy scale 9

1.2.2 Length and time scales 10

1.2.3 DNA, a perspective on bioinspired technologies. 11

DNA nanotechnology is a part of the bioinspired nanotechnologies. The term bioinspired means that we take advantage of what we have learn from life systems to conceive and fabricate new nanomaterials and nanodevices. In that sense, the bioinspired approach is different from the mimicking method, the fundamental first learning process. DNA nanotechnology is the best emblematic example: it uses DNA for its physical characteristics to build structures at the nanoscale. This technology totally disregards the "living" properties of DNA as it uses it as an inert building block. It tries to avoid any contact, at least in the process of formation, with proteins or enzymes, as these ones could interfere with the creation of the structures.

This technology takes advantage of the pairing properties of DNA: the double helical structure of DNA is constituted of two single strands (ssDNA) attached together. These strands composed of a succession of bases (called sequence of the strand) interact thanks to Watson-Crick interactions with the rules that an adenine base (A) attaches with a thymine base (T) from the second strand and a guanine (G) with a cytosine (C). For a strand of 7 bases, 7^4 different sequences are possible. The energy of binding is maximum if every base of the two strands are associated according the Watson-Crick pairing rules. In that case the sequences are called complementary. This high affinity allows a high selectivity between two similar sequences toward the exact complementary sequence. The two strands can be viewed as a key-lock system.



The complementarity and the selectivity allow a precise spatial positioning between two strands. These features are at the origin of the DNA origami technique that will be introduced below.

Experimentally, the use of DNA material is also interesting because:

- The DNA synthesis is relatively affordable.(The cost of one strand of 32 bases is of ≈ 5 euros).
- It is stable and easy to use (Even modelling oriented team like ours can use it).

After introducing the origami DNA which is the main topic of this thesis we will introduce the energy, length and time scales we will encounter in the nanoworld.

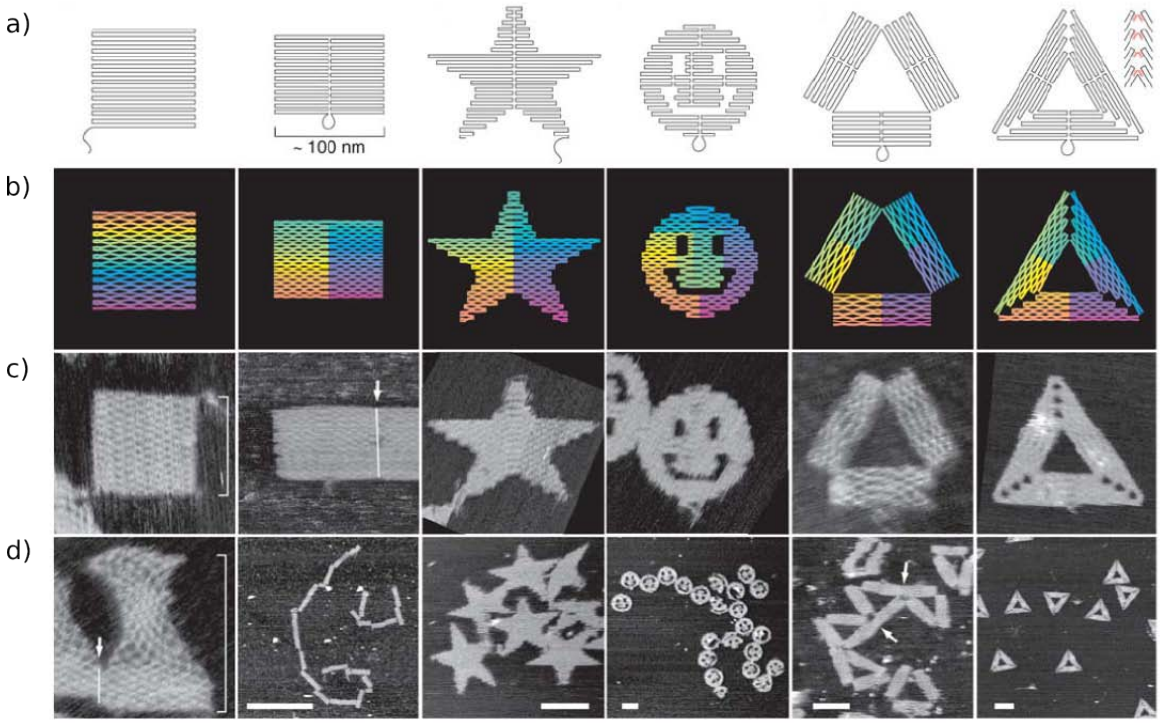


Figure 1.1: From Rothemund work [1]. Several examples of DNA origami, the geometrical shape designed (a,b); the experimental structure obtained (c,d)

1.1 DNA Origami

DNA origamis (oDNA) are geometrical shapes filled with DNA (Fig. 1.1). Observed with a low resolution they appear as a solid shape. Increasing the resolution lets appear a weaving of DNA. The weaving motif is very similar to the one of any wool sweater. The objects are named origami because one long ssDNA (Fig. 1.2 right) typically 7000 bases long is folded into a desired geometrical shape by means of 200 small ssDNA called staples (Fig. 1.2 left). Before being assembled into an origami, the system is composed of a long random coil with small inner parts interacting one with each others, and 200 smaller random coils.

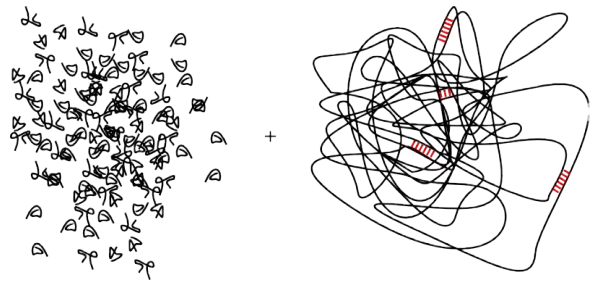


Figure 1.2: Raw material. A long ssDNA and 200 small ssDNA

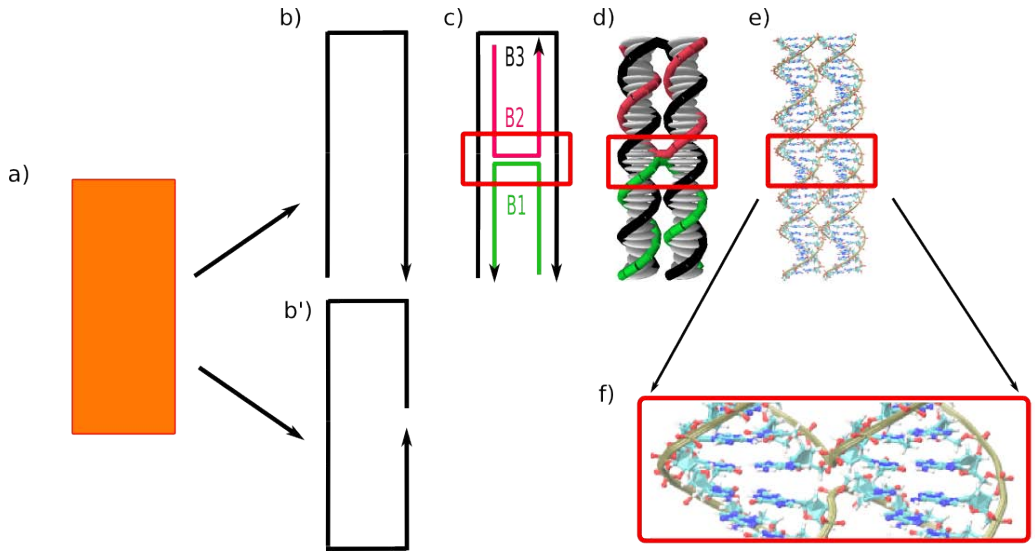


Figure 1.3: a) Shape to be filled. b) and b') two different possible paths for the scaffold. c) d) e) different representation of a small oDNA constituted of one scaffold: B3 and two staple B1 B2. B1 attaches two non contiguous parts of B3, f) Molecular detail at the crossover.

1.1.1 Principle of formation

The process of creation of an origami can be decomposed in different steps. First, one has to choose a geometrical shape. This shape can include holes and angles of any kind. In Fig. 1.3 a) the shape is a small rectangle filled with two lines. Then this shape must be filled by a path that will be the scaffold of the origami. Usually the path is circular, as experimentally the scaffold is a long circular ssDNA of a phage M13mp18 (Fig. 1.3b b'). Different paths can be chosen to fill the same shape. Then staples have to be positioned on all the path, to attach the different lines together (This is why they are called staples) (Fig. 1.3 c). A staple is composed of connected complementary sequences of non contiguous parts of the scaffold. The place where a staple changes of line and attaches to non contiguous parts is called a crossover and its position has to be chosen carefully (Fig. 1.3 d,e,f). Indeed both scaffold and staples are ssDNA. A staple is twisted around the scaffold as it forms with it a dsDNA. The crossovers have to be chosen at the position were the twist of the staple places its phosphate backbone the closest to the line were it must connect (see Fig. 1.3 f) molecular detail of this connection). As the twist of DNA (which is the angle between two consecutive pair of bases) is of 33-34°, connection between two lines are possible approximately every 5 bases (170°). In classical origami "weaving", on a given line a crossover is positioned every 16 bases: one connecting to a lower line, and one and a half turn later, another crossover connecting to an upper line.

Once the weaving done, each staple has a given position. But how could we ensure each staple will attach at the right place along the scaffold? The scaffold which is a ssDNA has a

defined sequence. The sequence of every staple is chosen to be complementary to the parts of the scaffold where it is positioned. All these steps constitute the design of the origami. Then the experimental part takes place. The staples with a given sequence can be ordered to a specialised seller (in our case Sigma - Aldrich). Once the 200 staples are synthesised, they are mixed with the scaffold (at a nM range) in a 10 fold excess of staple in a Magnesium (Mg) buffer. The reason why Mg is necessary will be explained latter. Then an annealing process is used to create the origami. Typically, the solution is heated until 85°C and then slowly allowed to cool to 20 °C. The cooling (annealing) duration depends on the complexity of the structure. For the classical rectangular shape the usual time of cooling is 2 hours.

1.1.2 Design of structures

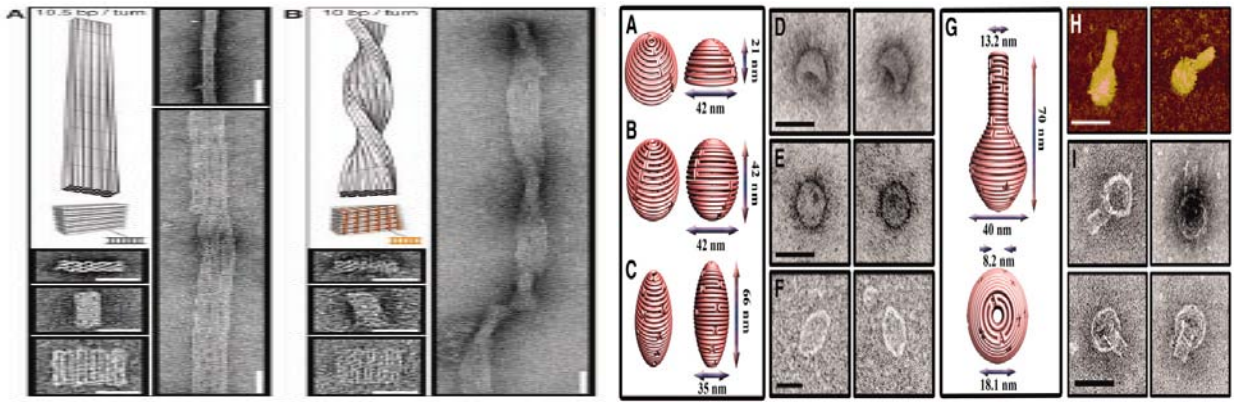


Figure 1.4: left) Dietz et al[5] Three dimensional dense structure, twisted by playing on the space between the crossovers; right) Han et al [6] designed several complex structures in two and three dimensions.

Several kind of structures can be achieved from the initial smiley [1] (Fig 1.1) to the dolphin [9]. A restriction concerns the total surface of the shape that can not be bigger than $\approx 5000nm^2$. For every shape it is possible to fill it with different paths and different sets of staples. This aspect will be studied more in detail in the chapter Process of formation of DNA origami (Chap 3).

To this bestiary of 2D structures can be added complex structures in three dimensions. To create 3D structures two possibilities are available: the first one is to fill the space with DNA [5] (Fig. 1.4 left). Curved and twisted structures can be designed [5] by varying the location of the crossovers (See Chapter 3) so that a wealth of topologies are available. A rectangular or a triangular lattice can be used to fill the space. The second possibility consists in creating shell-like object [6, 10] (Fig. 1.4 right). Andersen et al [10] created a empty cube of DNA with a lock system borrowing ideas from DNA tweezers [11]. At any of these structures can be added

a functionality. As we will see in the next paragraph, since its discovery DNA origamis have boosted researchers creativity to fabricate template biosensing structures or dynamical systems in 2D and 3D.

1.1.3 Functionality

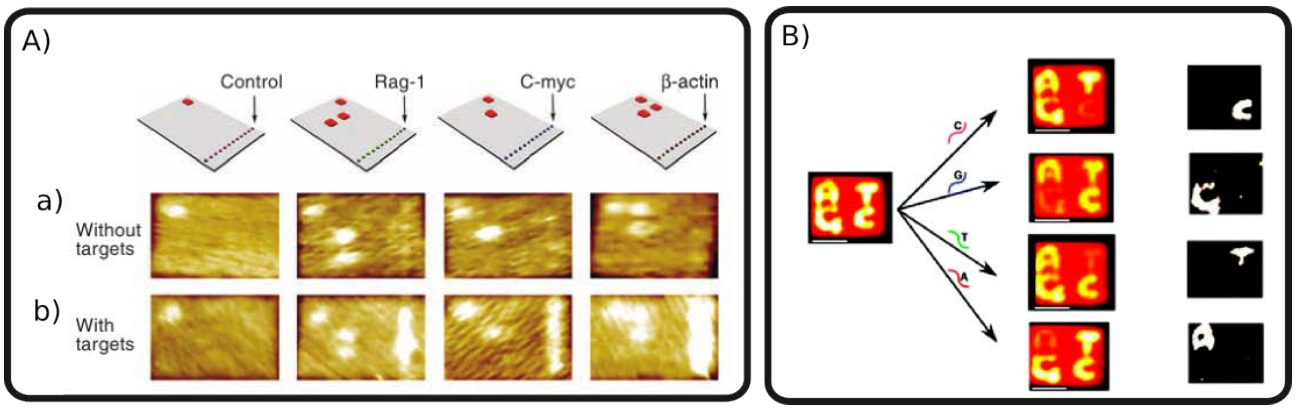


Figure 1.5: A) From KE et al [2]. Comparison of atomic force microscope (AFM) images without a) and with b) the target RNA. The variation of height is clearly visible for the structure detecting Rag-1 C-myc and $\beta-$actin B) From Subramanian et al [3]. oDNA were labelled with apparent letter. The triggering sequence of the four letter variates of only one bases on a 32 bases long sequence. Given the high specificity of the folding process, only the sequence with the perfect complementary sequence triggers the corresponding letter.

oDNA are composed of approximately 200 staples whose positions are exactly defined in the geometrical shape. This gives to oDNA a unique feature in terms of positioning at the nanometer scale. Each staple can be lengthen with a sequence that either becomes a probe or is used to attach a function to the origami. As a consequence, the requirement to attach a function to the origami is to be able to link this object with a ssDNA. At the location were the function has to be attached, the closest staple is lengthen with a small sequence. Then the oDNA is mixed with a strand, complementary to the small sequence, to which the function is attached. Thanks to the pairing properties of DNA the strand will hybridize and the object will be attached to the oDNA.

Functionalization was applied for the detection of RNA [2] (Fig. 1.5 A), proteins [12] or DNA single nucleotide polymorphism [3] (Fig. 1.5 B) which consists in detecting the change of one base on a given sequence. This detection relies on fixing a probe with a high affinity for the target to the origami and then to image the structure with an atomic force microscope (AFM). In the presence of the target, the structure of the origami is modified and the change in height is easily measured.

The effect of the distance between divalent ligands interacting with a target protein [13] was studied in a similar way: the ligands were attached at different distances one to the other, and the binding with the target protein was measured. Chemical reaction such as bond-cleavage reactions were also studied [14] by fixing a strand composed of the sequence staple, the bond to be studied, another small sequence and a streptavidin. A biotin is then added in the solution to make the strands clearly visible with AFM. When the cleavage takes place, the part connected to the streptavidin is detached from the origami and goes in solution. This change is easily visible with AFM. Using aptamer sequences, controlled patterning of protein was also achieved [15, 16].

DNA origami are also considered as possible platforms for electronics. In a first step toward that direction, nanotube were patterned to its surface [17]. Furthermore controlled patterning of origami on different surfaces which is a necessary step was also achieved [18, 19]. DNA origami are also good candidate for SERS platforms and steps toward this direction were reported. Chapter 4 is dedicated to this application.

Besides immediate applications such as biosensors [2], many strategies take use of DNA dynamical behaviours to achieve complex functions or structure reconfiguration. Prescribed tracks have been used for nanomachines and nanorobots [20, 21, 22], reaching a high complexity with an assembly line by Gu et al [4].

Strand displacement technique [23] has been used to reorganise dynamically origamis structures [10, 24].

Strands displacement technique is also the base of the DNA computing field that may also be called Chemical Reaction Network as proposed by Soloveichik et al [25] and which consist in realising logic operations such as the one done by computers but with DNA strands. The complexity reachable by this technique can be seen in the work of Qian et al [26] were a Hopfield associative memory was achieved with DNA strands. Combining oDNA with DNA computing seems to be a promising way to achieve programmable behaviours of high complexity in cellulo.

1.1.4 An active field!

Although the oDNA field is young (<7 ans) the number of groups and publication on the field is rapidly increasing. In 2011 about 40 publications on the origami DNA topic were published on a total of 160 publications related to the keyword DNA technology (study from Sebastien Berger). The connection between the principal researchers working on this field can be illustrated by the graph taken from my personal bibliography and connecting the authors through their published work (Fig. 1.6). Although oDNA is a new tertiary structure of DNA with mechanical properties and a stability still to be determined, the physics of dsDNA structure is a well established domain. Understanding the physic of oDNA seems to us as an important step toward using oDNA as a building block of DNA nanotechnologies. The energies length and time scale at stake in the dsDNA world are the subject of the next section.

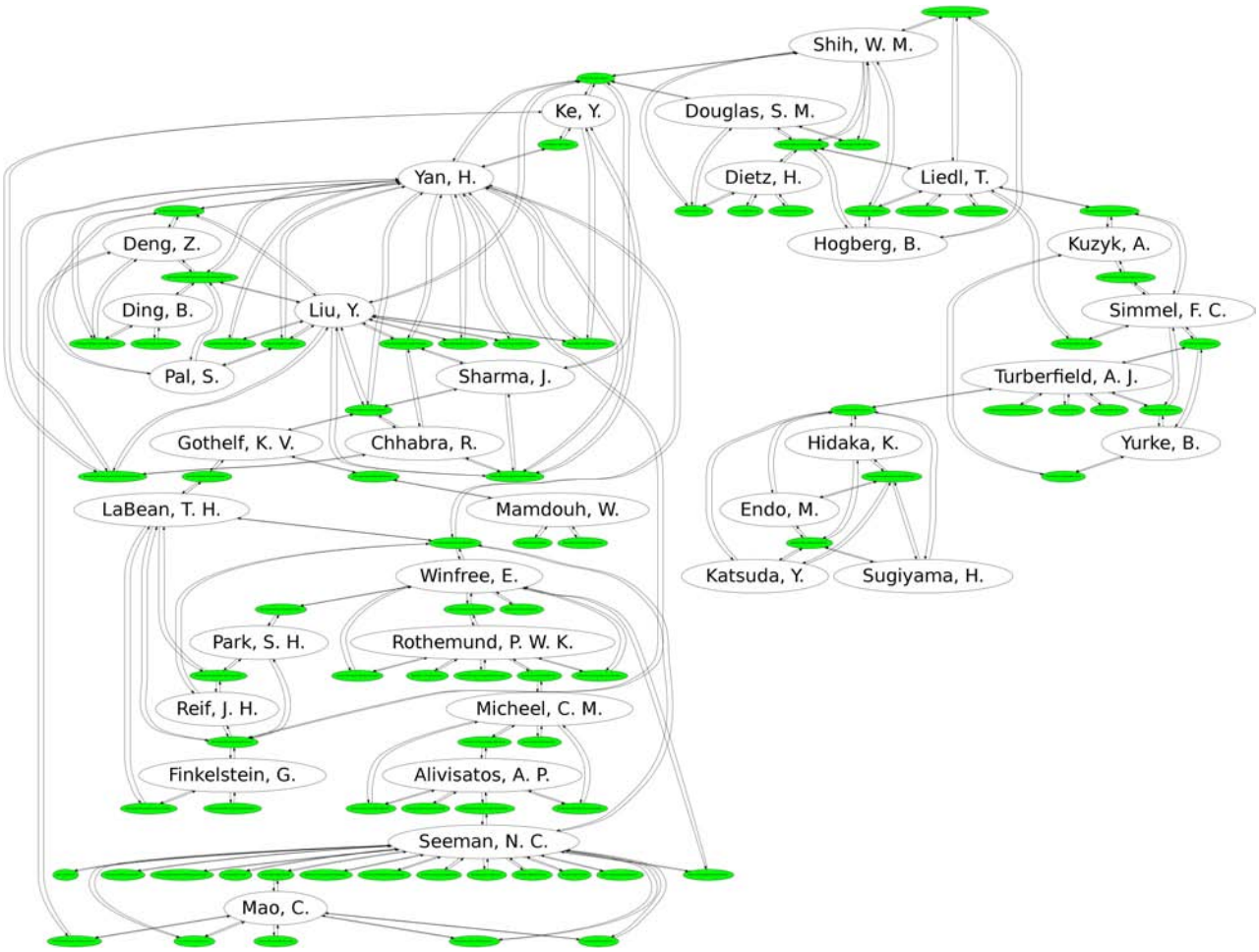


Figure 1.6: In my personal bibliography, authors having at least three publications in the DNA/ origami DNA field. The lines show connection betwen authors through publications

Figure 1.7: Left diving with DNA; Middle DNA in pure water (The tiger represents the electric field pushing away DNA strands from each other) dsDNA-dsDNA interactions as well as ssDNA-ssDNA interactions are repulsive; Right DNA condensation. In solution with ions (fish) the electric field (tiger) diminishes and the dsDNA-dsDNA interaction can even become attractive.

1.2 Diving in the nanoworld with DNA

1.2.1 Energy scale

The energy source the most known and widely used in cell metabolism is the energy coming from the hydrolysis of ATP, producing ADP. This reaction in typical cellular conditions liberates a Gibbs energy Δ G = -14 kcal/mol. This reaction takes place in water solvant were hydrogen bonds structure the water. The energy of these bonds is typically of $\Delta G = -2.4$ kcal/mol. This latest is close from the energy at stake in the pairing of two bases to form the dsDNA secondary structure. For two bases attaching together this energy is of the order of $\Delta G_{bp} \approx -1$ to -2 kcal/mol. The equivalent of the hydrolysis of ATP is approximately the hybridisation of a 7 bases long dsDNA. The exact energy of pairing of two strands depends on the chemical composition and length of the sequence. Indeed the energy of folding comes from two contributions:

- Hydrogen bonds between bases of the two strands. Two liaisons for the A-T pair of bases and three for the G-C pair that is always more stable.
- Stacking energy between two pair of bases, that explains why the energy depends on the sequence of bases and not only on the composition.

Knowing the sequence allows to predict the temperature at which the duplex is more stable than the two single strands, thanks to the Nearest-Neighbour model that also includes the effect of pairing defects [27].

But the chemical sequence is not the unique parameter that governs the dsDNA stability. DNA strands are molecules highly charged, with one negative charge per base. The stability of the complex depends on the salt concentration and on the salt composition. In pure water it would be impossible to bring together two ssDNA to form a dsDNA as the repulsive electrostatic

energy between two bases brought at 2 nm is of 0.21 kcal/ mol/base. (2nm corresponds to the distance between the two strands in the duplex). The electrostatic interaction is a long range one and it scales as the square of the number of charges for charges close with respect to the distance between the strands. The repulsive energy between two strands composed of 10 bases would be of 21 kcal/mol. However in typical cellular condition different salts are presents. These ions screen the electrostatic interactions and allow two ssDNA to get close enough to be able to hybridise. The electrostatic energy between two bases in a magnesium buffer of concentration 25 mM at a distance of two nanometers is of $\Delta G = 0.06$ kcal/mol/bp.

This energy is identical in value to the energy of condensation of two dsDNA together. The phenomenon of condensation appears in the presence of mutivalent positively charged ions. dsDNA-dsDNA interactions are no longer repulsive but attractive. This energy has been estimated to be of -0.06 kcal/mol/b [28].

The last source of energy is the thermal energy which is of 0.6 kcal/mol by degree of freedom.

Several models exist to reproduce dynamic characteristics of the transition dsDNA-ssDNA: The Peyrard-Bishop model [29], the Poland-Scheraga model [30, 31], models on lattice [32]. All these models concern the reaction of hybridisation from two ssDNA to a dsDNA. In oDNA structure we are interested in the reaction of 200 staples with one long strand. In this thesis we tried to have a first look inside the mechanisms of formation of the oDNA. Chapter 3 will present experimental and theoretical progress in understanding the process of formation of small origami as the one shown in Fig. 1.3. Then we will try to explain experimental results on different oDNA thanks to a new model based on the use of the Nearest-Neighbour parameters.

1.2.2 Length and time scales

Double stranded DNA is a helical structure with a pitch of 0.34 nm whereas in single stranded DNA the distance between bases is of 0.66 nm. The length of 0.34 nm comes from the strong hydrophoby of the bases that are densely packed in dsDNA with a twist of about 33° between each pair of bases. The force needed to break a pair of bases can be estimated to $f = \frac{\Delta G_{bp}}{l_{hydrogene}} \approx \frac{2kT}{0.2nm} \approx \frac{2\times 4.3pN.nm}{0.2nm} \approx 43pN$. This force has been experimentally measured to be around $10-15$ pN [33].

dsDNA structures are characterised by a long persistence length of 150 bases ($\approx 50nm$). oDNA can be considered to be tertiary structures. Two dimensional oDNA are mostly studied using AFM measurements of the structure deposited on mica surface. As a consequence the structures are planar and usually stuck on the surface. In this thesis we decided to explore mechanical properties of oDNA, using an existing model of the flexibility of dsDNA, and extending it to oDNA structures [7]. Chapter 2 will be dedicated to the results obtained on the study of the dynamic mechanical properties of oDNA.

Another length that is important in the perspective of application for oDNA in the leaving cell is the size of proteins. Proteins have a typical size of a few nanometers. Globular proteins

size can be estimated as a function of the number (N) of amino acid that composes it by the relation $R_g = 0.395N^{3/5} + 7.2$ Å [34]. For a protein of 400 amino acid Rg = 2.2 nm. As typical oDNA are 100 nm long ($\approx$ 300 bases) and 50 nm width a set of several proteins can be attached on its surface [12, 13], combining several functionalities.

Another aspect is the time scale (that will not be considered in this work). The time scale of association and dissociation of dsDNA was measured by Jungmann et al [35]. At 20°C it is respectively of $t_{off} = 0.6s$ and $t_{off} = 5s$ for a 9 and 10 bases long dsDNA. This time is roughly proportional to the exponential of the energy of the complex (see SI [35]). This can be checked as $\frac{t_{off}(10bp)}{t_{off}(9bp)} = 8.3 \approx exp(-\Delta G_{bp}/RT)$ for $\Delta G_{bp} = -1.3$ kcal/mol. In this thesis, the typical length of strands is of 32 bases. By comparison they are very stable at room temperature. On the other hand the association time is independent of the size of the strand [36] as it is a processus of diffusion and nucleation on a few bases. It is dependent on the inverse of the concentration and is for the 9 and 10 bases pair long dsDNA 50s at a 10 nM concentration [35].

1.2.3 DNA, a perspective on bioinspired technologies.

Other timescales at stake with DNA are illustrated in (Fig. 1.8). This figure shows the different steps involved in the emergence of a bioinspired technology. In the case of DNA technology, the molecule itself is implied deeply in the process of emergence as DNA molecules began to code for the human race as we know it 200000 years ago. The transmission and the increase of knowledge to this century leaded to the understanding of the importance of DNA 80 years ago. One breakthrough for this technology was realised 70 years ago, when Watson and Crick discovered the double helical structure. This knowledge allowed Nadrian Seeman 30 years ago to take advantage of DNA characteristics for its own purpose: to create small and self assembled shapes composed of double helical structure of DNA. Finally 6 years ago Rothemund with the technique of the DNA origami amplified the complexity and the possibility reachable with this technique, leading to several applications and probably to a lot more to come. This breakthrough was also possible because of technical progress leading to faster and cheaper synthesis of DNA. We saw that understanding the DNA structure was the first step toward the DNA technology. In order to reach another level of complexity we think that a full comprehension of the mechanisms at stake in the oDNA structures is necessary. Given the complexity of these structures that are the result of the assembly of hundreds of different molecules in a single object, this work is a first step in understanding the processes involved. This understanding could allow to reach higher complexity and functionality in future structures.

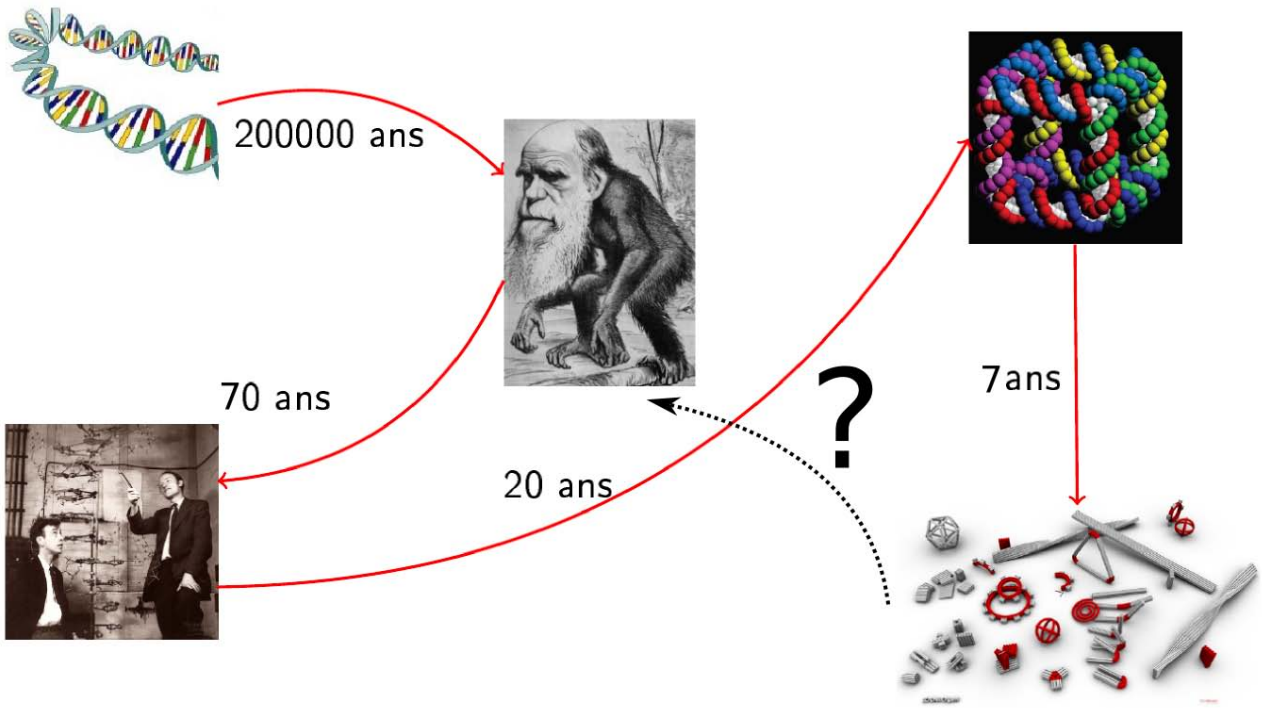


Figure 1.8: Implication of DNA through ages. Bottom left: Watson and Crick;Top right structure. Early DNA structure from the work of Nadrian Seeman; Bottom right: DNA structure realised thanks to the oDNA technique (Work of Shawn Douglas, Picture take from the website cadnano.org)

Chapter 2

Mechanical properties of DNA constructs

Contents

2.1 Introduction: **14**

2.2 Analytic model of elasticity. **16**

2.2.1 Introduction to an analytic model of DNA. 16

2.2.2 Analytic model of ssDNAs: 18

2.2.3 Analytical model of a dsDNA: 20

2.2.4 Simple model to calculate the rigidity of a DNA origami: 22

2.3 Persistence length simulation: **25**

2.3.1 Mergel's Energy Model 25

2.3.2 General description of Monte Carlo Simulations 28

2.3.3 Metropolis algorithm of transition between states in Monte Carlo Simulations 29

2.3.4 Monte Carlo moves 30

2.3.5 Averaging values in MC simulation 31

2.4 Results **34**

2.4.1 The twist parameter: 34

2.4.2 The electrostatic parameter: 38

2.4.3 Comparaison with experimental data 40

2.4.4 Study of the persistence length for different constructs 49

2.4.5 Force curves measurement of bending on DNA constructs 51

2.5 Conclusion 58

2.1 Introduction:

The mechanical properties of a DNA molecule depend on of its secondary structure and may vary significantly depending on with the state of hybridisation. Indeed a ssDNA molecule is one hundred times less rigid than a dsDNA molecule. As a consequence a hundred base ssDNA molecule in solution is a random coil of globular shape with a radius of gyration of approximately 6 nm. This value can be compared to the total length (or contour length L_c) of the strand which is 60 nm. In contrast, the same hundred bases divided into two complementary ssDNA can form a very stiff molecule of 17 nm. The rigidity of dsDNA at this scale is the reason why we are able to image a oDNA as it allows the structure to be planar once attached to a surface. A similar construction with the mechanical properties of the single stranded DNA would look like a blob.

Mixed structures of dsDNA and ssDNA strands [37] give rise to tensegrity structures (Fig. 2.1). A mechanical analogue is a bow in which the string is the ssDNA streched close to its contour length and the flexible limb is the oDNA. The oDNA is curved under the entropic tension of the streched ssDNA. Such structures allow the estimation of the rigidity module of a 6 helix bundle oDNA, rigidity that has been studied more in detail by Kauert et al. [8], and that will be discussed later in this chapter.

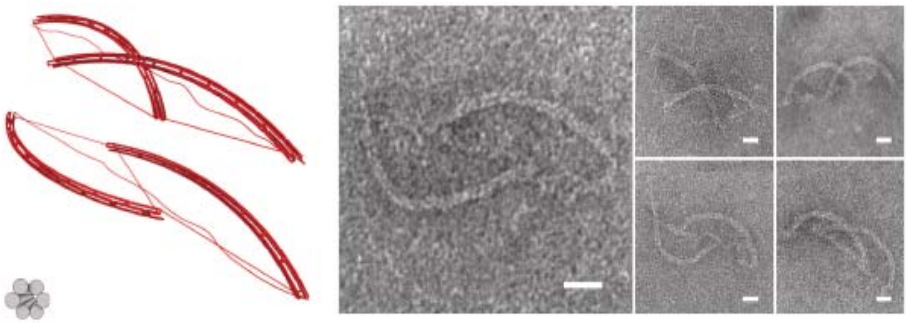


Figure 2.1: Figure from [37], (left) scheme of the tensegrity structure, (right) Transmission electron microscopy(TEM) image of the structure

At the beginning of this project we also wanted to build a planar origami with a long ssDNA strand joining the two extremities of the oDNA. The goal being to achieve a structure with a curvature that would depend on the length of this strand (herafter called a bridge). In order to guide us in the choice of the length of the bridge across the DNA layer, we decided to study a model of this system. The modeling of the system is composed of two parts, the ssDNA

bridge and the planar oDNA. The properties of the ssDNA bridge are modelled by an analytic polymer model [38] (see Sec. 2.2.2). This model can also be used to model long dsDNA (see Sec. 2.2.3.1). However, at the scale of an oDNA ($\approx$ 100 nm) a single dsDNA behaves more like a rigid body than a flexible polymer. Because an oDNA is composed of several dsDNA, a polymer model is no longer applicable and the oDNA must be thought as an elastic body. An existing elastic model for small dsDNA [39] was extended to an oDNA, and gave us a first estimation of the shape of the system that we wanted to build where the shape is a result of an equilibrium between the force exerted by the ssDNA (that behaves like an entropic spring), and the force due to the rigidity of oDNA (cf Sec. 2.2.4).

From our point of view oDNA structures can be considered as several dsDNA coupled together. The previous simple model depends critically on how the strands composing the oDNA are coupled. The interactions coupling the dsDNA are: covalent crossover junctions, and electrostatic interactions (that must be taken into account given the high charge density of dsDNA). This latter interaction (see Sec. 2.3.1.1) can be modelled analytically [40] or numerically [41], and ranges from repulsive forces to attractive ones, leading to the phenomenon called DNA condensation.

We will introduce the numerical model that we used to study the mechanical properties of oDNA in Sec. 2.3. We adapted Mergel's model [7] parameters to be in agreement with recent experimental results and included electrostatic interactions. The results were compared to different experimental data and to another model using a finite-element-based approach [42].

Unfortunately for reasons that became clear to us now, at the end of this work, we were not able to experimentally create bridged oDNA. This will be discussed in Sec. B.

2.2 Analytic model of elasticity.

2.2.1 Introduction to an analytic model of DNA.

From the point of view of polymer physicist ssDNA and dsDNA are polymer molecules constituted of monomer units. Even if constituted of four possible different bases and consequently of sequences that are not necessarily periodic, the monomer unit of ssDNA and dsDNA can be assimilated to one base. In aqueous solution, the mean angle between two bases of a ssDNA molecule is fixed. This is due to the chemical link between the two bases and also to the negative charge carried by every base. For dsDNA molecule, the helical secondary structure makes it very rigid, with a mean angle between bases of approximatively $6°C$. This one is close to zero as the molecule is very rigid. The analytic models used to describe the physic of polymer having a given angle between monomer unit are the freely rotating chain (FRC) model and the worm-like chain (WLC) model [38, 43]. The first is a discrete model used for polymers rather flexible like ssDNA [44], while the second is a continuous model which can be seen as a particular case of the FRC model for rather rigid polymer like dsDNA [38]. As the WLC is a particular case of the FRC model we will first introduce the FRC model.

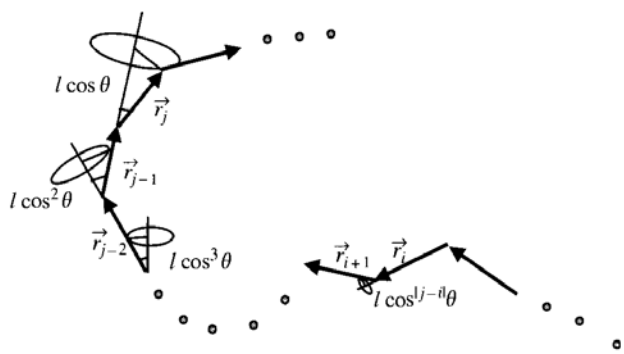


Figure 2.2: Figure from [38], schematic representation of a freely rotating chain.

The freely rotating chain model describes the physics of a polymer represented by N connected segments of length l. The angle between two segments is constant and equal to θ and the torsion angle is free to rotate. As the N segments are free to rotate the angular correlation between two segments is:

$$< \vec{r}(i).\vec{r}(j) >= l^2 cos(\theta)^{|i-j|} \tag{2.1}$$

with $\vec{r}(i)$ the vector between i and $i+1$.

Then we have:

$$cos(\theta)^{|i-j|} = exp(|i-j|ln(|cos(\theta)|)) = exp\left(-\frac{|i-j|}{l_p}\right) \tag{2.2}$$

with $l_p = \frac{-1}{ln(cos(\theta))}$. The later relation defines the persistence length. This length is a parameter of the model that characterises the rigidity of the polymer. A rigid polymer will have a persistence length much larger than the monomer unit length. On the contrary for very flexible polymer the persistence length is close to the distance between monomers. In the case of a rigid polymer, for distances between monomers smaller the persistence length ($|i-j| << l_p$), $cos(\theta)^{|i-j|} \approx 1$. This means that $\theta \approx 0$ and that the segments are aligned. So for distance shorter than the persistence length the polymer chain is often described as a rod. Another interpretation of this length is that the mean angle between two points separated by l_p is 1 radian (about 60 degrees).

The equation 2.2 can be used to evaluate l_p, once given, for example the structure of a set of polymers trapped on a surface [45], or given a numerical model and a good sampling of the system modelled [7].

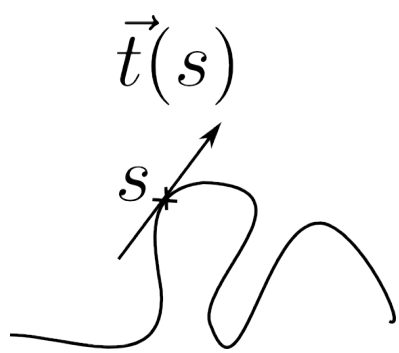


Figure 2.3: Schematic representation of a worm-like chain chain.

This equation (Eq. 2.2) is also true in the case of the WLC model (Fig 2.3). In the WLC model the description of the polymer is continuous and $|i-j|$ becomes simply the curvi-linear distance between i and j. Furthermore as θ is small the relation $cos(\theta) \approx 1 - \frac{\theta^2}{2}$ is used to calculate various quantities.

Given this persistence length, one can evaluate the force distance relation [46] which quantifies the force F needed to constrain the two extremities of a polymer to a given length x. This relation is different for the WLC and the FRC model. The two models agree for small extension and have different asymptotic behaviour for $x \approx L_c$ with $L_c = n \times l$ the contour length of the polymer, and n the number of monomers. For the WLC model:

$$\frac{F(x)l_p}{k_B T} \approx \frac{1}{4(1-x/L_c)^2} - \frac{1}{4} + x/L_c \tag{2.3}$$

This equation is only an approximation as the relation between F and x in the WLC model cannot be solved analytically. It can also be used to determine experimentally the persistence length of dsDNA [47]. As explained previously ssDNA are more flexible and the FRC model should be of better agreement with the experimental data. The FRC model leads to a very

good agreement with the experimental data if one add an extra parameter [44]. However as mentioned in [48] an agreement almost as good can be obtained using the WLC equation.

This is why Eq. 2.3 will be used for both cases. The value of the parameters of the model (the persistence length and the distance between monomers) will be discussed in the following section for both ssDNA and dsDNA.

2.2.2 Analytic model of ssDNAs:

2.2.2.1 Persistence length

A single stranded DNA is a chain of bases linked together by a phosphate sugar backbone. Its persistence length has three contributions:

$$l_p = l_0 + l_{stack} + l_{elec} \tag{2.4}$$

with $l_0 = l/2$ the intrinsic persistence of the chain, $l_{stack} = 2l - 3l$ [49] the persistence length induced by stacking effects and l_{elec} the electronic contribution (each base has one electronic charge) and l the distance between bases.

Sain et al.[49] use a partition function with different proportions of (A,T,C,G) on the ssDNA to estimate l_{stack}. The effect of stacking is important for purine-purine interaction(A-A, G-G), and purine-pyrimidine interaction (A-T) , but it is weak for pyrimidine stack(T-T,T-C, and C-C). They estimate $l_p = l_0 + l_{stack} \approx 11 - 13$ Å by using $l \approx 4.3$ Å. The value of l is a parameter of the model and 5.6 Å is used in [47]. The ionic strength is known to have a strong influence on the persistence length of polyelectrolytes (charged polymers) and lp increases when lowering the ionic strength because of the repulsive phosphate-phosphate interaction ([50] (Tab. 2.1)).

I(mol/L)	10^{-3}	10^{-2}	10^{-1}
l_p(Å)	84.4 ± 21.5	44.1 ± 6.8	13.5 ± 4.5

Table 2.1: Persistence length as a function of the salt concentration (ionic strength). Data from [50] for the model in agreement with [47]

2.2.2.2 Force distance

In the case of a ssDNA Eq. 2.3 gives the force needed to constrain the extremities of the strand to a distance x. The values of the parameters are $L_c = n \times 0.43$ nm with n is the number of bases of the strand and $l_p = f(I)$ (where I is the ionic strength). Weak interaction at the molecular scale can be suitably express in unit of k_BT , where at room temperature $k_BT = 4.3$ pNnm.

For a ssDNA of 260 bases the entropic forces exerted as a function of the distance for different ionic strength are shown in Fig. 2.4. We observe that the force is divided by three when I goes from 10^{-2} to $I = 10^{-1}$ mol.L^{-1}.

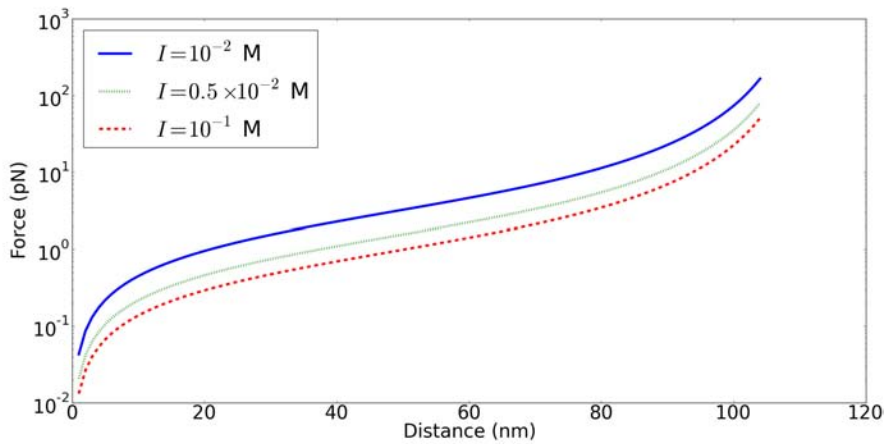


Figure 2.4: Elastic force exerted by a ssDNA of 260 bases (contour length=110 nm) whose extremities are stretched (for various ionic strength).

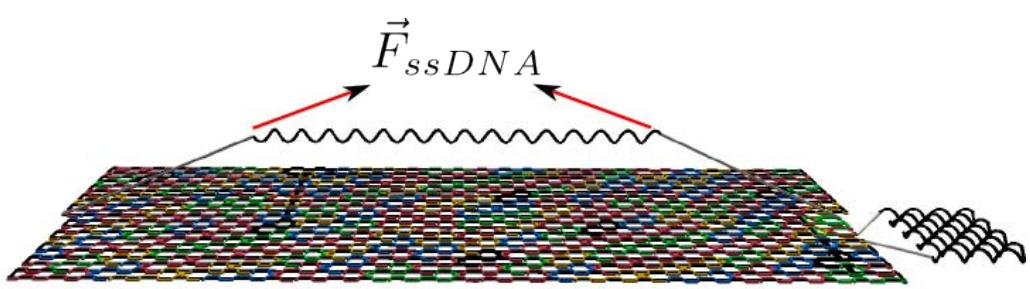


Figure 2.5: Schematic representation of the origami layer and the ssDNA bridge

In our case, the objective was to bend a 110 nm long origami. In this context we wanted to predict

the force a ssDNA bridge (Fig. 2.5) will exert to estimate the relevant size of the bridge. In figure 2.6 is ploted the force generated by such a strand for different lengths between 260 to 800 bases, when it is constrained to a length of 110 nm. The ionic strength chosen of 0.5×10^{-2} M corresponds to that of an usual oDNA buffer (12mM $MgCl_2$).

This force grows very fast when the distance at which the strand is stretched is close to the contour length of the strand. In the following section we compare these forces to the one exerted by a dsDNA , then to a simple model of oDNA.

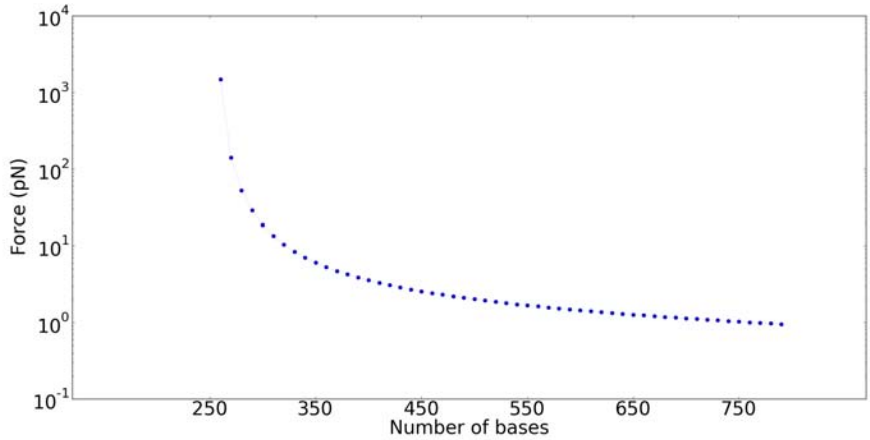


Figure 2.6: Elastic force excerced by ssDNA of various lengths whose extremities are stretched to 110 nm for a ionic strength of 0.5×10^{-2} M (each dot corresponds to a different strand)

2.2.3 Analytical model of a dsDNA:

2.2.3.1 Persistence length of a simple dsDNA :

The persitence length of a dsDNA is around 150 bp ($\approx$ 50 nm) at 25 °C and for a ionic strength of 0.2 M [51]. It is constant between 10 mM and 600 mM ionic strength. This value is the subject of controversy as in [52](p107) it is said that the persistence length of a dsDNA is around 240 base pair, and has always been underestimated because in average one was considering that the dsDNA is straight but in fact there is a deviation of 1 to 3 degrees between each base pair.

As explained in the introduction a dsDNA behaves differently if its length is short or long by comparison to its persistence length ,e.g. like a rod or a random coil. These two behaviours can be illustrated with the experimental formula [53] of the mean end to end distance $< R(L_c) >$:

$$< R(L_c) >= \left(\frac{L_c}{l_p}\right)^{\nu_0} \left(1 + \frac{L_c}{l_p}\right)^{\nu_1 - \nu_0} \tag{2.5}$$

with $\nu_0 = 1.03$ and $\nu_1 = 0.589$ experimentally verified for a range of molecule lengths between 10 and 10000 bp. This formula describes the two behaviours of a dsDNA:

$$< R(L_c) >\approx \frac{L_c}{l_p} \qquad \text{if } L_c << l_p \tag{2.6}$$

$$< R(L_c) >\approx \sqrt{\frac{L_c}{l_p}} \qquad \text{if } L_c >> l_p \tag{2.7}$$

For short contour length (Eq. 2.6) the dsDNA behaves like a rigid body, as its end to end distance is proportional to the number of bases. For longer contour lengths the molecule

behaves like a random coil as its end to end distance grows like the square root of the number of bases. These two equations are equally predicted by the WLC model.

The validity of the WLC model is also demonstrated by Bustamante et al. [48] where the force distance relation of the WLC model is used to fit the experimental force distance relationship. The parameters of best fit are the following $l_p = 53$ nm, $L_c = n \times 0.33$ nm with n the number of bases. For a 48000 bases long dsDNA phage this model is good up to forces of 10 pN [48]. For forces from 10 pN to 50 pN the structure of the strand deviates from the canonical B-DNA structure. For higher forces (from 65 pN to 150 pN) a transition from the B-DNA to the S-DNA form is achieved, and for higher forces the strand melts. It is possible to include all these different behaviours in one model [46] for forces from 0.01pN to 100 pN.

The comparison (Fig. 2.7) of the force distance curves of a 1000 bases long ssDNA and a 1000 bases long dsDNA shows the differences of mechanical properties between these two possible structures of a DNA molecule.

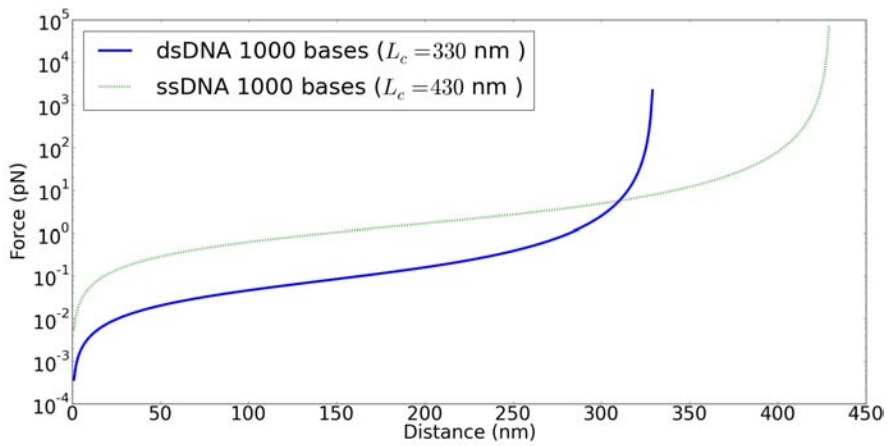


Figure 2.7: Elastic force exerted by a dsDNA and a ssDNA with the same number of bases per strand

Note that the use of a reduced coordinate x/L_c in equation 2.3 gives a universal force-extension curve whatever the contour length of the chain keeping in mind that short dsDNA of about 10 bp behave differently under tension [54]. Indeed the estimated tension before an important deformation is of 0.08 pN. This force is small compared to the force of 10 pN required to induce a change in the canonical B-DNA for longer dsDNA.

In this paragraph we saw that the WLC model can be used to describe dsDNA. We will now use this model to compute the energy needed to bend small dsDNA.

2.2.3.2 Energy necessary to bend small dsDNA

For sizes close to its persistence length, dsDNA can be considered as a rigid rod. For the system we want to model (oDNA) the typical length is a hundred nm, giving the length scale of the

problem. Although a dsDNA in the WLC model is a continuous chain, it is possible to coarse grain several bases together and model it by a succession of segment while keeping the energy dependency of the WLC model. It has been shown [39] that for segment lengths longer than 10-15 nm the link between l_p and the energy of bending of a strand predicted by the WLC model is accurate:

$$\frac{E(\theta)}{k_B T} = \frac{1}{2}\frac{l_p}{l}\theta^2 \tag{2.8}$$

with θ the angle between two consecutive segments and l the coarse grain length. For segments shorter than 15 nm another model should be used [39]. In the following we will consider the dsDNA as if it was divided in two shorter coarse grain monomers and evaluate the force needed to bend these two segments.

The WLC model is still valid but the force needed to bend the strand must be calculated differently. The force predicted by the WLC model tends to zero when the constrained distance tends to zero, that cannot be the case for a rigid rod. If x is the distance between the two extremities of the DNA strand $cos\frac{\theta}{2} = x/L_c$. The force needed to bend this strand is given by

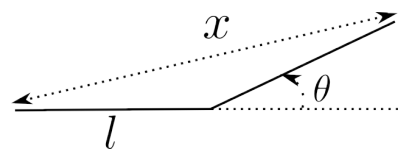


Figure 2.8: Schematic representation of the system divided in two monomers

$$F(x) = -|\overrightarrow{grad}E(x)| = -2k_B T\frac{l_p}{L_c/2}\frac{d\, arcos(x/L_c)^2}{dx}$$

as in our case $L_c = 2l$ (Fig. 2.8)

$$F(x) = 8k_B T\frac{l_p}{L_c^2}\frac{arcos(x/L_c)}{\sqrt{(1-(x/L_c)^2}} \tag{2.9}$$

The force (Fig. 2.9) is evaluated for different length of dsDNA. The force is almost constant for a given size of strand and equal to:

$$F \approx 8k_T\frac{l_p}{L_c^2} \tag{2.10}$$

It increases when the strands are shorter and tends to zero for longer strands.

The generalisation of this model to compute the force needed to bend an origami is discussed in the following section.

2.2.4 Simple model to calculate the rigidity of a DNA origami:

In section 2.2.2 we have evaluated the force needed to stretch a ssDNA. In section 2.2.3.2 we have evaluated the force needed to bend a dsDNA. In this section we want to estimate the effect of attaching a ssDNA bridge over a 110 nm long oDNA. A first approximation for oDNA

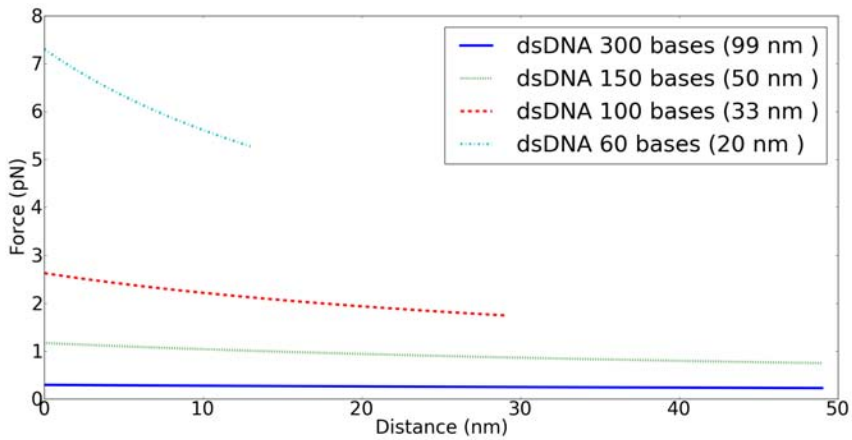


Figure 2.9: Force required to bend short dsDNA

is to consider a linear dependency of the total rigidity with the number N of rows of dsDNA. The total force needed would be N-fold the force needed to bend one dsDNA if we consider the rows as independent. The equivalent persistence length is therefore given by $N \times l_p$ (Fig. 2.10).

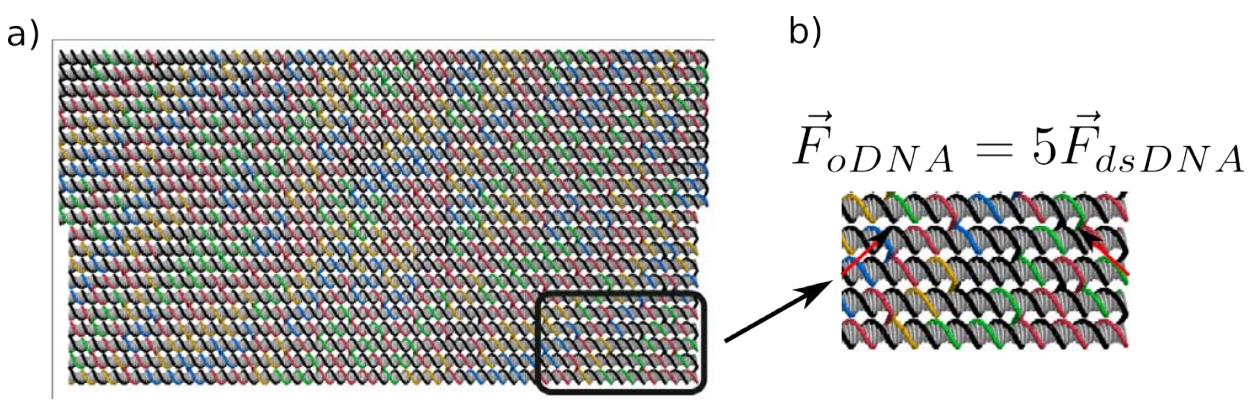


Figure 2.10: a) Top view of an origami constituted of 20 rows of dsDNA of about 300 bases each. b) Zoom on a sub-part and equivalency of the forces

Therefore for an oDNA constituted of 20 row of dsDNA of 110 nm length the bending force is at least 20 times the one of a dsDNA (Eq. 2.9).

In our experimental construct we want to predict what will be the shape of the oDNA under constraint (Fig. 2.11). The final shape will result of an equilibrium between the forces needed to bend the oDNA and the entropic force exerted by the ssDNA bridge.

From figure 2.12 one can calculate the length at which the force of the oDNA and the force exerted by the ssDNA strand cancel. For instance a 300 bases long bridge (Fig. 2.12) will bend the oDNA with a total length of 80 nm whereas the forces exerted by a 750 bases long ssDNA is too weak to curve it.

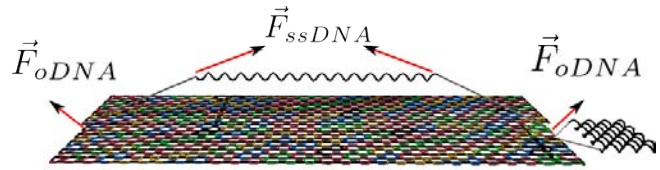


Figure 2.11: Schematic representation of the forces at stake.

The persistence length of a dsDNA corresponds to fluctuations of the molecule in three dimensions. However origamis are anisentropic, and the fluctuation are mostly along the direction perpendicular to the plan of the origami. Trapping a polymer in two dimensions increases its persistence length by a factor 2 [45]. As the force is proportional to l_p, in this hypothesis the force required to bend the origami of figure 2.12 would be multiplied by two (green curve).

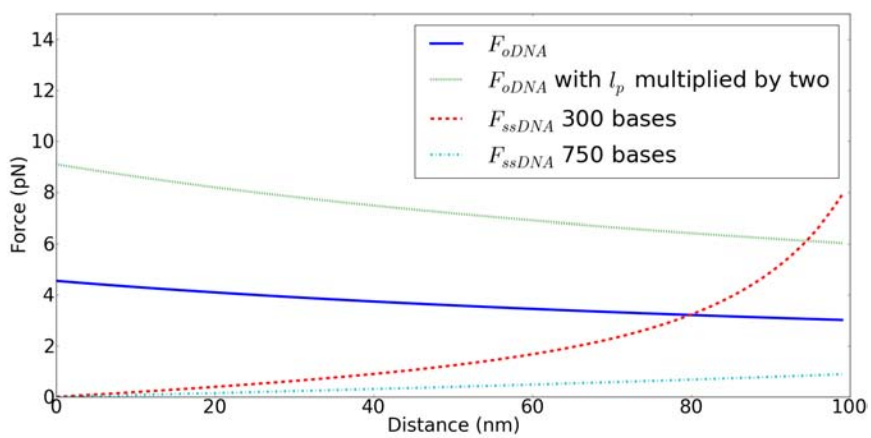


Figure 2.12: Competition between the rigidity of an oDNA and the entropy of a ssDNA.

This analytic model for the two parts of the system gives a first estimation of the forces at stake. However the total persistence length of the oDNA is only roughly estimated. If the total persistence length is underestimated (for example by another 2 factor), the system is not going to be curved. Thus, we decided to use a more detailed model in order to evaluate the total persistence length of the origami.

2.3 Persistence length simulation:

The Mergel's energy model [7] and its adaptation to oDNA is presented in the following, as well as a description of the Monte Carlo method of simulation.

2.3.1 Mergel's Energy Model

In [7] , a dsDNA is modelled by a succession of ellipses, that are connected at their extremities by two springs (Fig. 2.13). One ellipse represents a pair of bases, and a spring represents the phosphate backbone. In an origami, constituted of a succession of dsDNA, the crossovers between two dsDNA are constituted of phosphate backbone, they will also be modelised by a spring.

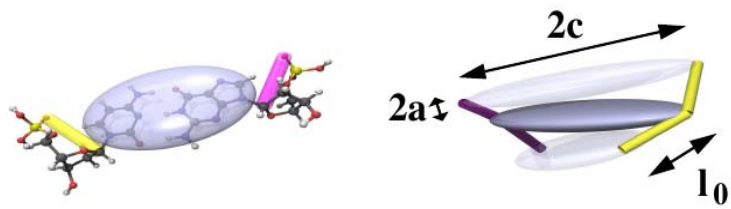


Figure 2.13: Figure from [7], illustration of the model

The model [7] takes into account two energies:

- A stacking interaction between ellipses, that modelises the fact that bases are hydrophobic and attract each other in aqueous solvent.
- A spring energy associated to the phosphate backbone.

The stacking interaction is minimal if the distance between ellipses is 3.3 Å, and the spring energy is minimal if the distance between the extremities of two ellipses is 6 Å. To conciliate these two constraints, there is a rotation of around 36° between two contiguous ellipses.

The model does not incorporate differences between right and left dsDNA. In order to include this difference, in the Monte Carlo trajectories configurations where the twist between two adjacent bases is inferior to -0.2 radian are rejected.

The original model does not include interactions between different strands. To take them into account, we add an interaction of excluded volume between close strands modelised by a simple excluded volume potential between ellipses of two different dsDNA: any configuration with negative distance between ellipse is rejected. The function used to evaluate the distance between ellipses is an approximation of the exact expression and is detailed in [7].

In Mergel et al. model [7] , no charges are explicitly taken into account, as the value of the persistence length is a consequence of interaction between adjacent ellipses. In the case of oDNA constructs, two strands are close one to the other. As a consequence the negative charge of each phosphate of the backbone, interacts with the phosphate of any adjacent strand. The

need to add electrostatic interaction between phosphates to the model will be discussed in the next section.

2.3.1.1 DNA electrostatics

There are three models of electrostatic interaction between dsDNA [55]: attractive electrostatic forces due to counterion correlations, screened Debye-Huckel interaction between helical molecules, and water-structuring or hydratation forces. These three theories have been motivated by the fact that although dsDNA molecules are highly negatively charged they can be compacted tightly in the cell in the presence of counter ions.

This phenomenon of attraction between dsDNA is called DNA condensation [41, 55, 28, 56]. It can take place in presence of divalent charged ion to hexavalent charged ions. However this phenomenon of condensation does not take place for Mg^{2+} ions [40] which is the salt present in oDNA buffers. Consequently only repulsive forces will be considered.

High resolution AFM images (Fig. 2.14) suggest a balance between repulsive forces between strands and the local energy of curvature of the dsDNA, as the strands are not densely packed and can be distinguished individually.

From AFM measurement it is possible to estimate the maximum distance between coupled dsDNA. In the figure 2.14 b) we can observe 9 intervals between strand in a distance of 51.7 nm. The mean interval between crossovers is of 5.7 nm, and given the fact that a dsDNA is 1.8 nm wide, the center to center distance between strands is of 3.9 nm (Fig. 2.14 c))

In this context, a common model for electrostatics is to use a Poisson-Boltzmann solution that describes dsDNA-dsDNA interaction as charged continuous cylinder-cylinder interaction [57]. Mergel's model is less coarse grained as it describes dsDNA at the bases pair level and it is possible to associate to every phosphate (positioned at the extremities of an ellipse) an electric charge. In solution, electrostatic interactions between charges are screened by the ions in solution, and the electrical potential in the Debye-Huckel approximation corresponds to a screened Coulomb (SC). potential given by:

$$E_{SC}(r,T) = E_{Coulomb}(r,T)e^{-r/\lambda_{Debye(T)}} \tag{2.11}$$

where $\lambda_{Debye(T)}$ the Debye length is given by the formula [58]:

$$\lambda_{Debye}(T) = \left(\frac{\varepsilon_0 \varepsilon_r(T) k_B T}{6.10^3 c_0 N_A e^2}\right)^{1/2} \tag{2.12}$$

with ε_0 the permittivity of the vacuum, ε_r the relative permittivity of the solution, c_0 the concentration of divalent ions in solution in mol/L.

The temperature dependence $\varepsilon_r(T)$ [59] should also be taken into account (Tab. 2.2)

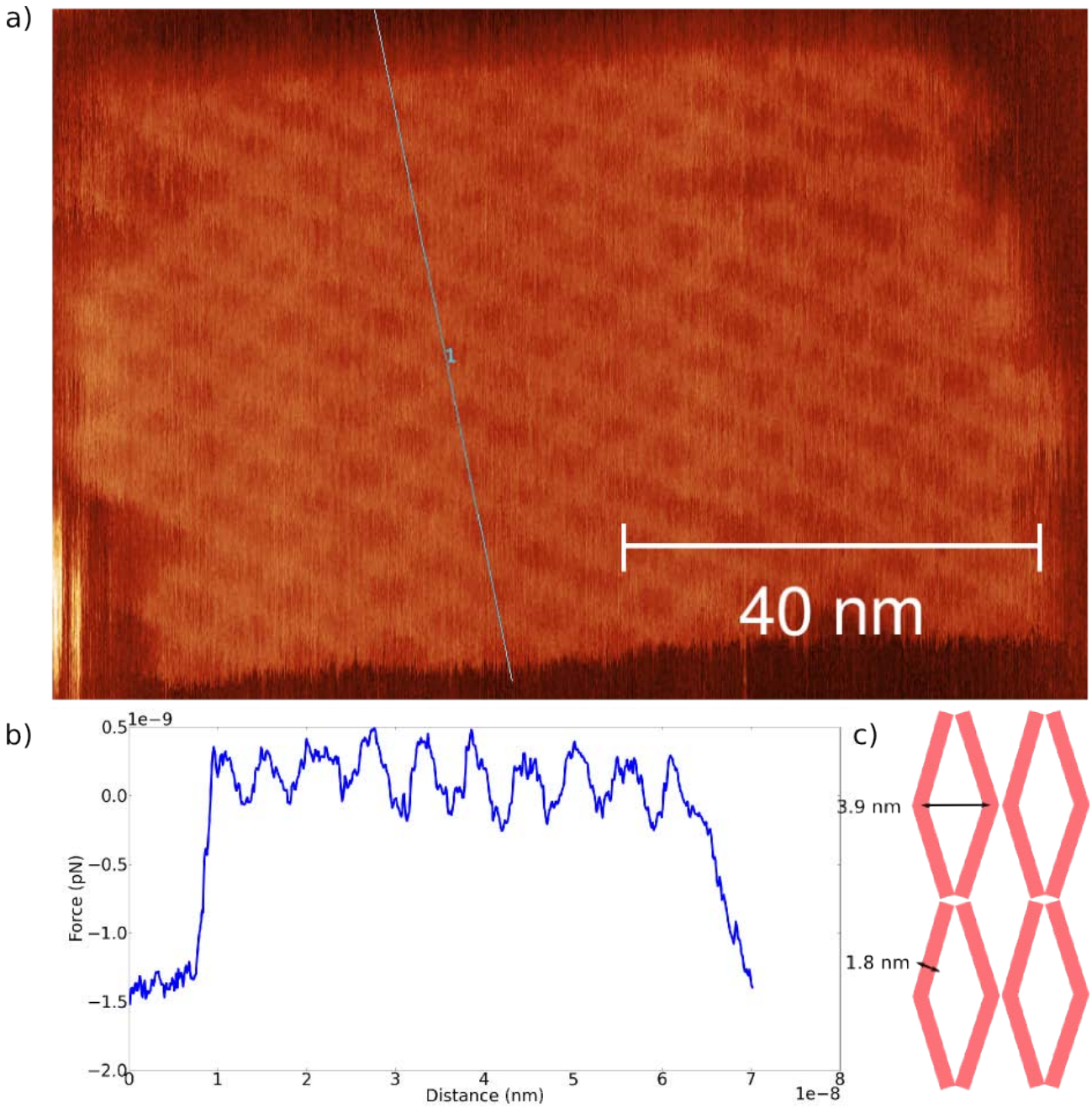


Figure 2.14: a) AFM imaging of an Origami, b) section of the AFM picture, c) schematic representation of the dsDNA as measured by AFM

	T = 0°C	T = 20°C	T = 85°C
$\varepsilon_r(T)$	88	80	59
$\varepsilon_r(T) = 88 - 0.33 \times T$	88	81.3	60.0

Table 2.2: Experimental value of ε_r as a function of the temperature. The second row shows that this dependency is almost linear

$E_{Coulomb}(r, T)$ is given by:

$$E_{Coulomb}(r, T) = \frac{q^2}{4\pi\varepsilon_r(T)\varepsilon_0 r} \tag{2.13}$$

We studied two variations of the model. One where only volume excluded interactions are present, and another one were we add explicit Debye-Huckel interactions between phosphate belonging neighboring dsDNA. Including phosphate-phosphate interactions between phosphates of the same dsDNA would lead to an increase of the persistence length of each dsDNA.

2.3.2 General description of Monte Carlo Simulations

Monte Carlo simulations are widely used in condensed matter physics. They are mostly used to compute average properties of a physical system. More generally they allow to sample the configuration space. This type of simulation need four different ingredients to run:

- An energy model for the system.
- A set of moves that will modify the system, for example "physical" moves such as the translation of an atom, or of a set of atoms, or "nonphysical" moves such as flipping a whole molecule or cutting the inner part of a chain an reconstructing it with other coordinates.
- An algorithm that constrains the transition between two states according to the distribution that we want to sample. In most cases the transition between the states is such that the degrees of freedom of the system follow a Boltzmann distribution.
- Some computers to run the simulations, or a set of motivated humans as it was done at the beginning [60].

One of the advantages of such techniques compared for example to molecular dynamics is the possibility to move an important set of atoms in a coordinated way and in only one move. The state obtained could have never been reached in molecular dynamics, or reached in a very long computing time. But the advantage of the method is also its drawback, as it requires for

every different system to create a set of intelligent moves that will allow the fast sampling of the phase space.

As the energy model and the set of moves are dependent on the particular system of interest, we will first describe the different algorithms used in classical Monte Carlo simulations.

2.3.3 Metropolis algorithm of transition between states in Monte Carlo Simulations

The objective of a Monte Carlo simulation is to generate a set of configuration according to a probability distribution p, once the system reaches the equilibrium. Mathematically the system is at equilibrium when the probability into and out any state A is identical:

$$\sum_B p_A P_{A->B} = \sum_B p_B P_{B->A} \tag{2.14}$$

with p_A the probability to be in the state A and $P_{A->B}$ the transition probability from A to B

One sufficient condition to achieve this equality is called the condition of detailed balance:

$$p_A P_{A->B} = p_B P_{B->A} \tag{2.15}$$

This equality ensures that given any starting configuration the probability distribution will tend to the distribution p if the simulation runs long enough. Indeed this condition eliminates possible limits cycles in which the system could be trapped, as the detailed balance condition would not be satisfied in this case [61].

In the case of a Boltzmann distribution $p_A = e^{\frac{-E_A}{k_B T}}$. Equation 2.15 requires the condition on the transition probability:

$$\frac{P_{A->B}}{P_{B->A}} = e^{-\frac{E_B - E_A}{k_B T}} \tag{2.16}$$

It is possible to divide the transition probability $P_{A->B}$ into two parts:

$$P_{A->B} = g_{A->B} A_{A->B} \tag{2.17}$$

with $g_{A->B}$ the probability to select a state B from state A and $A_{A->B}$ the acceptance ratio of B given the previous configuration A.

The most used algorithm is the Metropolis Algorithm. In this algorithm the probabilities $g_{A->B} = g_{B->A}$ for every state A and B. The acceptance ratio between the two configurations A and B with energy $E_B < E_A$ have to be chosen in order to satisfy Eq. 2.16. Several choices are possible, but one that ensures a good sampling is to set the acceptance ratio of a new configuration B, given a starting configuration A to $A_{A->B} = 1$. Given Eq. 2.16 the acceptance ratio $A_{B->A}$ is fixed:

$$A_{B->A} = e^{-\frac{E_B - E_A}{k_B T}} \qquad \text{if } E_B > E_A \tag{2.18}$$

$$A_{A->B} = 1 \qquad \text{if } E_B < E_A \tag{2.19}$$

In practice a random number r between 0 and 1 is generated and the new configuration A is accepted if $r < e^{-\frac{E_B - E_A}{k_B T}}$.

The Metropolis algorithm is not the only option. In the case of a spin system other algorithms such as the Wolff algorithm or the Swendsen-Wang algorithm [61] are of interest for temperatures near the critical temperature. These algorithms use the fact that one can choose $g_{A->B} \neq g_{B->A}$.

2.3.4 Monte Carlo moves

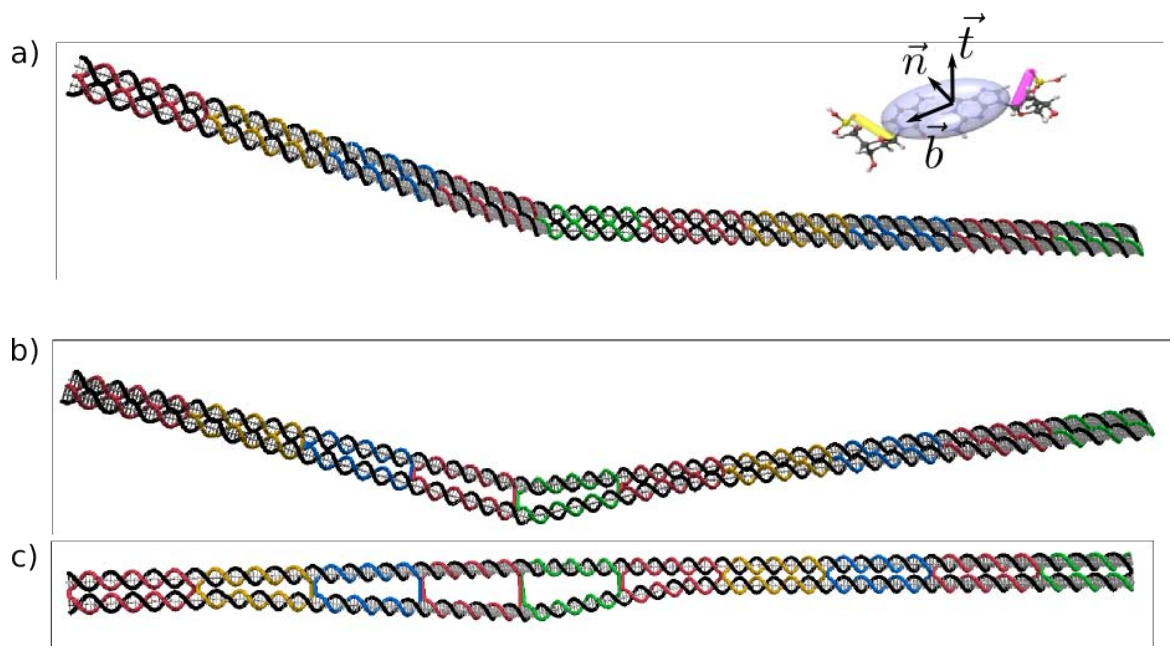


Figure 2.15: a) a pivot move; followed by a cranck-shaft move of a part of the first dsDNA. b) side view, c) top view

Several different moves were proposed for the following simulations besides the translation and rotation of single ellipses. Two different types of pivot moves were tested:

- One base at random B_r is selected then all the bases B_t such as $\vec{t}(B_r).(\vec{CM}(B_t) - \vec{CM}(B_r)) > 0$ ($\vec{CM}$ the center of mass) are rotated. The center of rotation is $\vec{CM}(B_r)$, the axis of the rotation a random vector in the plane $(\vec{b}(B_r), \vec{n}(B_r))$, the angle of rotation is also random. $\vec{b}(B_r)$ and $\vec{n}(B_r)$ are the two principal axis of the ellipse (they are both

perpendicular to the main axis of the strand) and $\vec{t}(B_r)$ is the vector perpendicular to $(\vec{b}(B_r), \vec{n}(B_r))$.

- One row at random is selected and all bases from one side or the other side of the row are rotated in the same way as above

These two moves give an identical result (Fig. 2.15 a)) when the structure is close to the initial (flat) structure, but they are a little bit different for important deformations.

Finally a cranck-shaft move (Fig. 2.15 b,c) is also applied. It consists in a rotation of all the bases between two bases pairs chosen randomly, the axe of rotation being defined by the center of mass of the two pairs, the angle being random.

Usually the parameters of the Monte Carlo moves (for example the size of the maximum translation) are set to have an acceptance ratio of 40 %.

2.3.5 Averaging values in MC simulation

In MC simulation there is no physical time, an analogue is the number of moves tried at a given moment of the simulation. The number of moves needed to reach the equilibrium from an initial state, or to sample the phase space once the equilibrium is reached, are dependent of the size of the system and of the type of moves. If one compares a strand with fifty bases and a strand with a hundred bases, approximately two times more moves are needed to reach the equilibrium in the second case. This is true if only local moves are allowed, because global moves displace several bases at once and are more efficient. To have an analogue of time identical for systems of different size, we introduce the notion of MC step which is the number of moves needed to displace every base at least once in average. This number is given by the formula:

$$MC_{step} = \sum_m p_m * nb(m) \tag{2.20}$$

with m a given MC move, p_m the probability of proposing this move, and $nb(m)$ the number of base displaced by a given move. For local moves like translation of a base, $nb(m) = 1$, and for global moves like the pivot move $nb(m) = 0.5 * Nb$ with Nb the total number of bases. The notion of MC step is important as it ensures that executing the same number of MC step for two systems of different size allow to reach the equilibrium for both systems. But a question remains:

When running a Monte Carlo simulation how long to run a simulation in order to have a good sampling, and enough independent states?

A discussion of this problem can be found in [61]. The simpler answer is to evaluate the equilibrium time τ_e, which is the time needed to go from the initial configuration of the simulation to a configuration at the equilibrium (The configuration is considered at the equilibrium when its energy reaches its mean value in the simulation (Fig. 2.16)) In general this time is longer than the correlation time between two independent configurations.

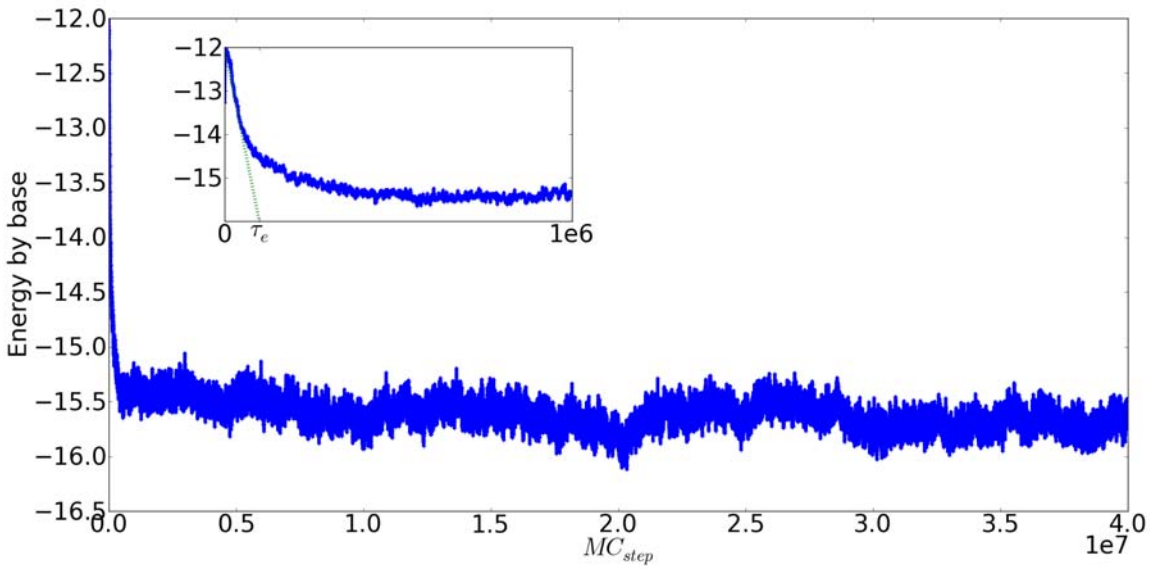


Figure 2.16: Equilibrium time for a DNA construct composed of two 300 bases long dsDNA

In the case shown in Fig. 2.16, the equilibrium time is $\tau_e = 100000$ $\mathrm{MC_{step}}$, so the number of independent samplings is $n = \frac{t_{max}}{2\tau_e} = 400$ as the total length of the simulation is $8 \times 10^7 MC_{step}$. It is also possible to evaluate the correlation time of the energy τ_{ene} by computing its autocorrelation function.

$$\chi(t) = \frac{1}{< E^2 > - < E >^2} \sum_{t'}^{t_{max}} \frac{1}{t_{max} - t'} \left(E(t+t') - < E >\right)\left(E(t) - < E >\right) \tag{2.21}$$

It is expected that $\chi(t) \approx e^{-\frac{t}{\tau_{ene}}}$. In our case (Fig 2.17) the correlation function is not a simple exponential but a sum of exponential functions with two different times $\tau_{ene1} \approx 10^6$ $\mathrm{MC_{step}}$ and $\tau_{ene1} \approx 10^4$ $\mathrm{MC_{step}}$. In a general Markov process the number of times present in the correlation function is equal to the total number of configurations [61]. However these times decrease rapidly and only the longer times are visible on the experimental correlation. To evaluate the number of independent configurations only the longer time τ_{ene1} is relevant. τ_{ene1} gives a number of 40 independent simulations.

Given these results for simulations realised on an oDNA composed of two 300 bases long strands, the number of MC_{step} chosen for computing value on the configuration have been multiplied by almost three compared to this example. The number of MC step will be of 200×10^6 for all the following simulation. For values of particular interest eight simulations were run in parallel in order to estimate the incertitude of the measure.

We will now present the influence of different parameters of the energy model that we explored by Monte Carlo sampling.

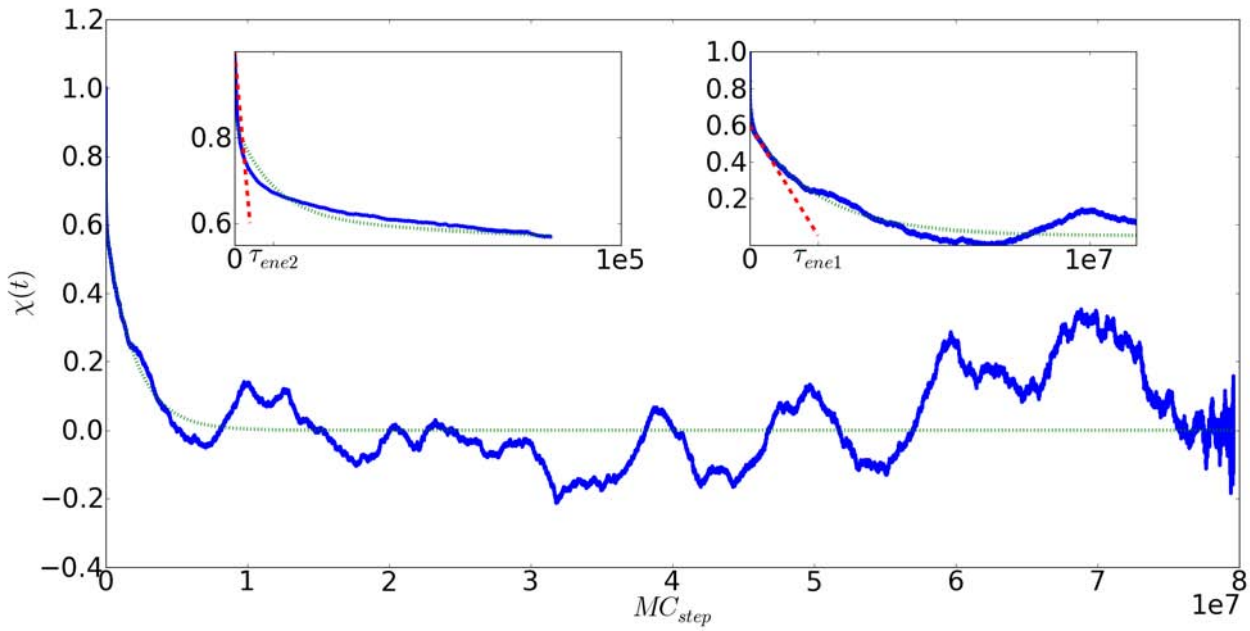


Figure 2.17: The energy autocorrelation function of a DNA construct composed of two 300 bases long dsDNA (blue curves). Fit with a sum of two exponential functions (green curves)

2.4 Results

2.4.1 The twist parameter:

The value of the twist between successive ellipses is imposed by the energy model as a result of an equilibrium between the length of the bond representing the phosphate backbone (6 Å) and the distance of minimal energy between ellipses which imposes a center to center distance of 3.3 Å. This twist corresponds to the twist between successive bases in a dsDNA molecule. In this paragraph we will first show the importance of the experimental twist of the dsDNA in the design of an oDNA structure. We will then compare the theoretical shape of a structure that we obtain in the model to experiments on similar structures and adapt one parameter of the model to improve the agreement.

In natural B-DNA strand the twist is dependent of the sequence of bases (Tab. 2.3) , the mean value being 35.45 °. In the energy model[7], the original parameters give a mean value of the twist (Tw) of $0.617 = 35.35°$, very close to the experimental value.

	GC	AC	AT	TA	CA	CG	GA	AA	GG	AG
N	27	10	34	9	23	10	25	50	3	4
Twist (°)	38.3	35.8	33.4	40.0	36.9	31.1	39.3	35.8	33.4	30.5

Table 2.3: Data from [62]. N represents the number of times that the sequence was present in the 38 different crystals studied

When designing an oDNA structure, the value of the twist is important, because it imposes where to create a crossover between two strands. In order to simplify the process of design of an origami the value of design chosen was initially $Tw = 33,7$ ° [1]. This value corresponds to exactly one turn and an half for 16 bases, this allows to create easily long periodic structures as crossovers are designed exactly every 32 bases (Fig. 2.18).

Experimental evidences [63] show that the oDNA structures are twisted in solution. The twist of the structure is due to a difference between the twist of design and the natural twist of the dsDNA. The consequence of this difference is an internal stress because some liaisons are longer when designed as flat (Fig. 2.18 b)) and the structure of minimal energy is twisted in order to respect the length of liaisons.

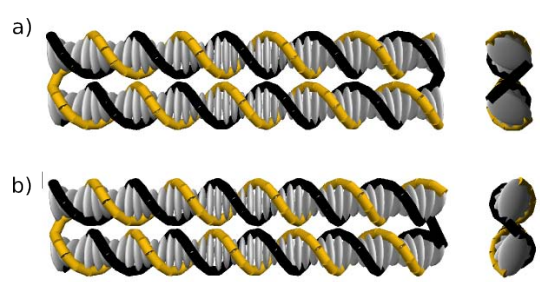


Figure 2.18: a) a structure designed with a twist of 33.7 °. b): If the twist is not 33.7 but 35.45, the position of the crossover has to be changed in order to preserve the length of the phosphate backbone at the crossover

By controlling the number of bp between crossovers, 3D curved structures can be intentionally designed [5]. Also, three dimensional origamis have been fabricated to measure the global twist of a structure, given a twist of design [63]. The datas collected were TEM images. A design of the structure with the hypothesis of a 10.5 bp/turn (34,28 °) for the dsDNA structure gave indeed a twisted oDNA. Planar structures were achieved for a twist of design of 10.44 bp/turn (34.48 °). This experimental observation suggests firstly that the natural twist of dsDNA between two bases in the type of buffer is closer to 34.48 ° than to 35.45 ° and also, that at the crossovers the value of the twist of the B-DNA form is respected.

Therefore, as the structure at the crossover is not a regular B-DNA one could ask wether the natural twist of the DNA is respected or else if the exceeding twist is relaxed in order to minimise the internal stress. Molecular dynamics were performed on small DNA constructs [64]. In these simulations the twist of a B-DNA strand is about $30 - 32$ °. The most stable structures are the ones preserving the natural twist of the B-DNA structure. Otherwise the fluctuations of the twist are larger , specially at the crossovers, although the mean value at the crossover seems to be identical to the twist of the other bases. In conclusion the twist between the two bases at the crossover seems to be preserved, and the value of the twist between two bases is close to 34.48 °.

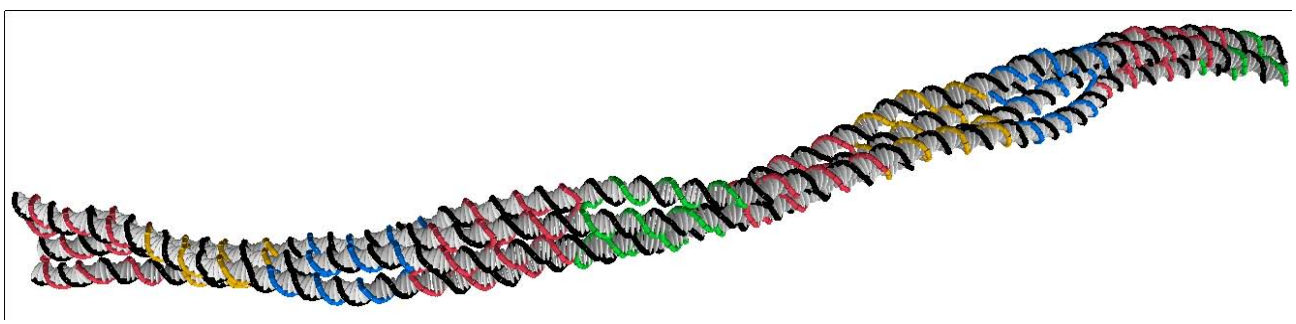


Figure 2.19: Structure of the 4-oDNA(315) after 1 million of Monte Carlo steps. Natural twist: 35.45°, Designed twist 33.7°.

Given these results we now discuss the influence of the twist in our model in order to adjust the parameters to have theoretical results closer to the experimental ones.

Figure 2.19 shows the final configuration of a MC simulation of a 4 dsDNA 315 bases long each coupled every 32 bases by a crossover. We call this structure 4-oDNA(315). If we choose a twist of design of 33.7 °, with the parameter of the model that imposes a twist of 35.45 °, the resulting structure is highly twisted (Fig. 2.19).

In the following we compare the simulations obtained in three different cases for two structures: the 2-oDNA and 3-oDNA (Fig 2.20). Let us denote Tw_N the 'natural' twist, namely the twist naturally adopted by an isolated dsDNA in our model, and Tw_D the twist imposed by the origami design.

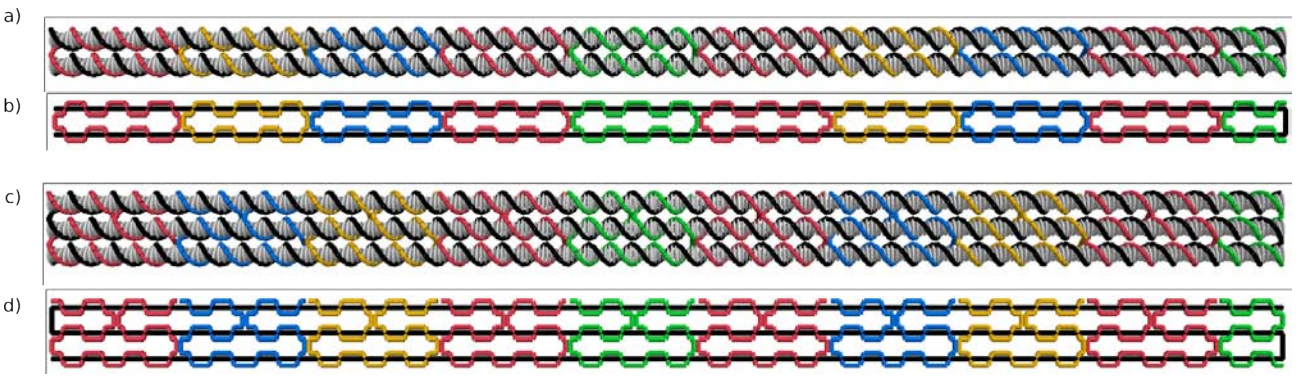


Figure 2.20: Two representations of a) , b) the structure of 2-oDNA(298) , c), d) the structure 3-oDNA(298)

The three different simulations are the following:

- $\mathrm{Tw}_N = 35.35$ ° and $\mathrm{Tw}_D = 33.7$ °(Sim 1)
- $\mathrm{Tw}_N = 34.48$ ° (Obtained by changing the length of the scaffold from 6 Å to 5.9 Å) and $\mathrm{Tw}_D = 33.7$ ° (Sim 2)
- $\mathrm{Tw}_N = 34.48$ ° and $\mathrm{Tw}_D = 34.48$ °(Sim 3)

Given a value Tw_D how to choose the positions of crossovers that respect this value when designing the oDNA. In our structure a crossover takes place every three helical turns. If we choose the first base as a reference, a base that will be a part of the crossover must have the same orientation in order to connect to the upper helix. For the first crossover we must chose the index i that minimise $|\mathrm{Tw}_0 + i \times \mathrm{Tw}_D - 3 \times 360|$ with Tw_0 the twist of the first base chosen equal to 10° for convenience. For the second crossover we will chose the index j that minimize $|\mathrm{Tw}_0 + j \times \mathrm{Tw}_D - 2 \times 3 \times 360|$. To generalise, the index i_n of the crossover n must minimise:

$$|\mathrm{Tw}_0 + i_n \times \mathrm{Tw}_D - n \times 3 \times 360| \tag{2.22}$$

In Sim 1 and Sim 2 as $\mathrm{Tw}_D = 33.7$, the choice of $i_n = 32 \times n$ for the crossover n gives a value close to zero for all n (in Eq 2.22). In Sim 3 the distance between crossovers was chosen to be either 31 or 32: 31,31,32,31,31,32,31,31,32. The crossovers listed here are the ones between the first and second strands. The location of the crossovers between the second and third strands were also modified on the same principle (Sim 3). Another constraint is that the last base must have the opposite orientation of the first base in our construct. The equation to be minimised is $|i_{last} \times \mathrm{Tw}_D = 180\mathrm{mod}(360)|$ with i_{last} the index of the last base. For Sim 1 and Sim2 we choose $i_{last} = 315$ and for Sim 3, $i_{last} = 298$.

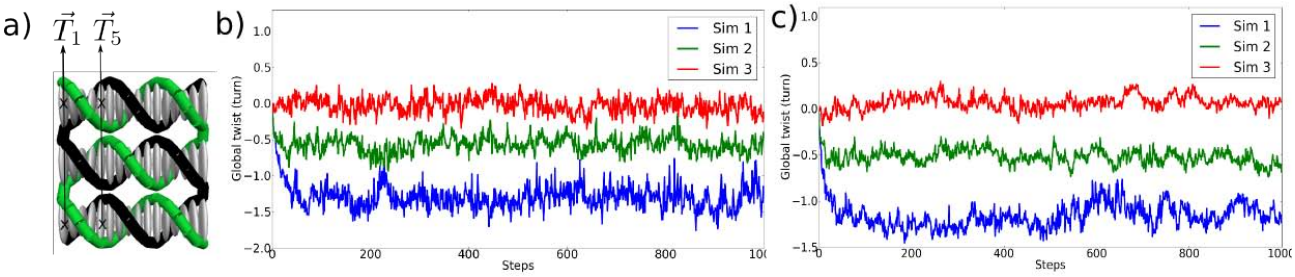


Figure 2.21: a) The vector $\vec{T_i}$ is the vector passing by the center of the ellipse i of the first dsDNA row of the structure and the center of the ellipse i of the last dsDNA row. Evolution of the global twist for b) the 2-oDNA(298) and c) the 3-oDNA(298) structures in the three simulations. For Sim 1 the expected Tw_G is -1.44 turn ($315 \times (33.7 - 35.35)/360$), for Sim 2 it is of -0.68 turn. Each simulation was 200×10^8 MC steps long. The values of the twist was measured every 200×10^5 steps. The values indicated in abscissa have to be multiplied by 200×10^5 to convert them in MC_{step} .

We define a measure of the twist of the global structure Tw_g as the sum of the twist between two columns of bases. The twist between two columns of bases is defined as the angle between the vectors colinear to these columns. (Fig 2.21 a).

$$Tw_g = \sum_{i}^{N_{bases}-1} \alpha_{\vec{T_i},\vec{T_{i+1}}} \tag{2.23}$$

with $\alpha_{\vec{T_i},\vec{T_{i+1}}}$ the angle between the two vectors.

From these experiments it seems that the difference between Tw_N and Tw_D is propagated from the strands to the structure, as the formula $i_{last}(\mathrm{Tw}_N - \mathrm{Tw}_D)$ gives a good estimation of the global twist of the different structures (Fig. 2.21 b,c). The structures designed in Sim 3 (named with a p for planar, eg 2-oDNAp) are flat, and the ones with $\mathrm{Tw}_N \neq \mathrm{Tw}_D$ are twisted. The planar structures were studied more in detail, and the values given on Tab. 2.4 are the average of 8 simulations.

	2-oDNA$_p$(298)	3-oDNA$_p$(298)	4-oDNA$_p$(298)	5-oDNA$_p$(298)
Tw_g (turn)	-3×10^{-3}	3×10^{-2}	3×10^{-3}	2×10^{-2}

Table 2.4: Result for the global twist of the untwisted structures for 8 MC simulation 200×10^8 $\mathrm{MC_{step}}$ long (Sim 3)

The global twist is close to zero for all these structures. Simulations of 4-oDNA$_p$(298) and 5-oDNA$_p$(298) gave similar results. Given the experimental evidences in agreement with this

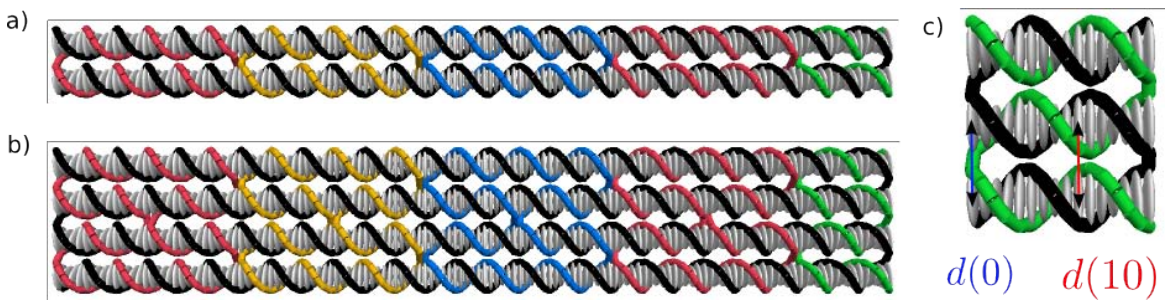


Figure 2.22: Two of the configurations used for the calibration of the electrostatic parameter. a) 2-oDNA$_p$(141) b) 4-oDNA$_p$(141) . c) Distance between bases of same column. The distance between bases of the 0 th and the 10 th column are represented here.

model, this value of the twist has been chosen for the following simulations by setting the length of the bond between ellipses to 5.9 Å.

In the following section we will measure the effect of the electrostatic interaction between strands.

2.4.2 The electrostatic parameter:

For two negative unit charges separated by a distance r, the Coulomb interaction is given by:

$$\frac{E_{Coulomb}(r, T = 300K)}{k_B T} = \frac{e^2}{4\pi\varepsilon_0\varepsilon_r k_B T r} = \frac{(1.6 \times 10^{-19})^2}{4 \times 3.14 \times 8.85 \times 10^{-12} \times 80 \times 1.38 \times 10^{-23} \times 300r} \approx 7.0/r$$

with r in Å.

For the salt concentration of classical oDNA buffer $\lambda_{Debye(T)} = 10.5$ Å. The effective interaction between every phosphate is given by:

$$\frac{E_{SC}(r, T = 300K)}{k_B T} \approx \frac{7.0}{r} e^{-r/\lambda_{Debye(T)}} (2.24)$$

As the electrostatic interaction takes a lot of computing time we worked on a smaller configuration to evaluate the effect of this parameter (Fig .2.22)

On the simulations run with a Debye length of 10.5 Å we notice that at the crossover the stacking between bases was lost (Fig. 2.23). This means that the stacking interaction between pair of bases (Fig. 2.23 b) green arrows) is not attractive enough compared to the repulsive electrostatic interaction (Fig. 2.23 b) red arrows). The energy of stacking is the one of the original model and has been chosen to obtain a persistence length in accordance with the experimental data. However our results indicate that this interaction is not strong enough. In order to prevent this opening that probably is not happening experimentally we added an extra liaison between the bases at the crossovers. (Fig. 2.23 c)

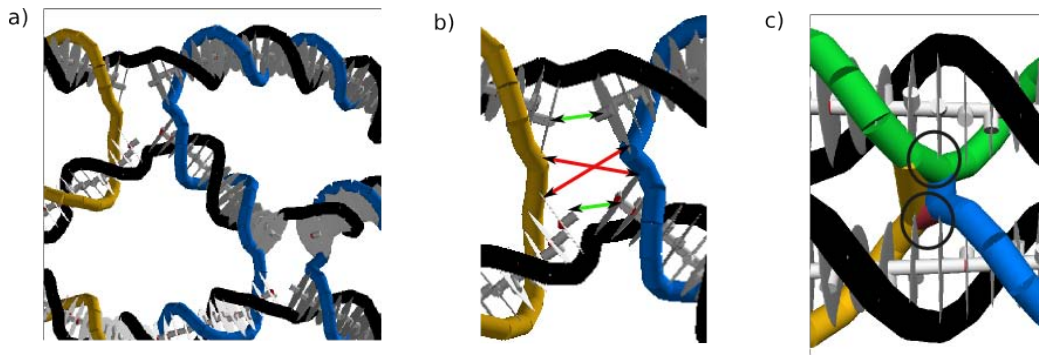


Figure 2.23: a) Opening at the crossover b) Zoom on a crossover. The red arrow represents the repulsive electrostatic interaction, and the green one the attractive stacking interactions c) Extra liaisons added at the crossover.

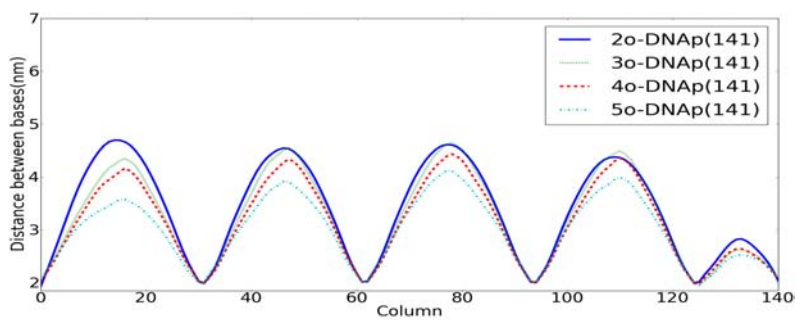


Figure 2.24: Study of the distance between bases of the same column for systems with different number of rows of dsDNA

To evaluate the effect of the electrostatic parameter, an interesting measure is the distance between bases of the same column on two adjacent dsDNA (Fig. 2.22 c)). This distance is of interest as it can be compared to the distance observed by AFM imaging in Fig. 2.14.

We first studied the effect of the number of dsDNA on this distance for a given Debye length of 10.5 Å (Fig. 2.24). In the following, the simulation were 200×10^8 MC steps long, and the values shown are the result of an average on eight simulations.

The curves are not completely periodic, due to the fact that the distance between all crossovers is 31-32 bases, excepted the last one that is 16 bases long. The experimental maximal distance between bases of same column is of 3.9 nm. The maximal distance for the 2-oDNA$_p$(141) construct is of 4.5 nm. The distance is reduced for the 3-oDNA$_p$(141) and 4-oDNA$_p$(141) down to 4.1 nm and for the 5-oDNA(141) down to 3.9 nm. These values are in very good agreement with the experimental value. The global twist of the molecule is also modified from almost 0 to $\approx 0.1 - 0.2$ turns. This means that the electrostatic interaction contributes to the global twist of the molecule, as in the simulations presented in the previous

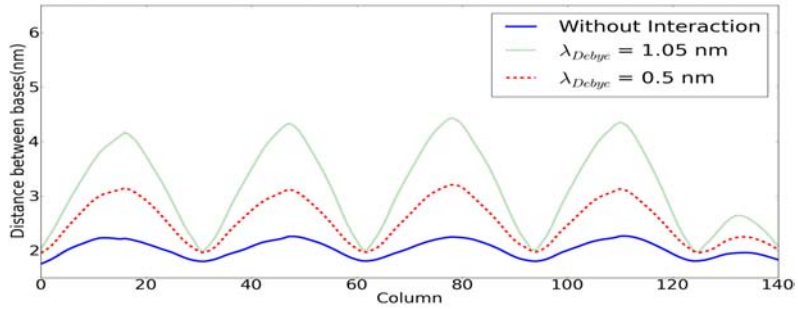


Figure 2.25: Distance between bases of same row in the case of an excluded volume interaction (Without interaction) and for two different Debye length for the 4-oDNAp(141).

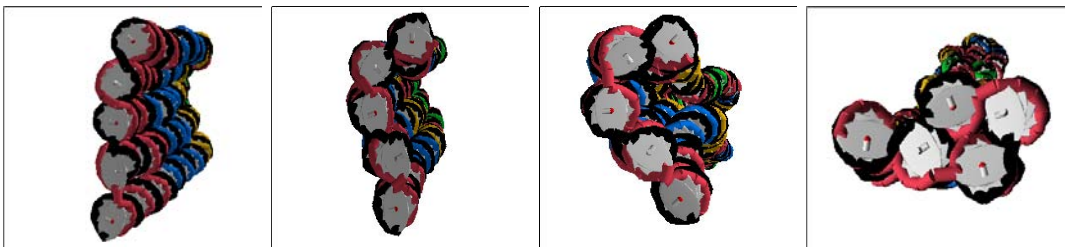


Figure 2.26: Evolution of the shape during the simulation for the 5-oDNAp

paragraph there was no electrostatic interaction, but only steric interaction between adjacent strands.

The effect of the value of the Debye length is illustrated in Fig. 2.25. The distance between bases is almost divided by two when comparing simulations with and without electrostatic interaction.

The absence of electrostatic interaction has a strong impact on the simulation. We noticed that sometimes the configurations are not flat anymore but are trapped in a bundle-like shape (Fig. 2.26)

This is why in the following we will only use the model with electrostatic interactions and a Debye length of 10.5 Å. We will now compare the persistence length of different structures to experimental data available in the literature.

2.4.3 Comparaison with experimental data

2.4.3.1 Calculation of the persistence length of DNA construct

In this paragraph we introduce the calculation of the persistence length used in all the following simulations. As explained in Sec 2.2.1 for a simple polymer modelled by n segments the

tangential vector $\vec{t}(i)$ and the persistence length are linked by the relation:

$$< \vec{t}(j).\vec{t}(k) >= e^{-\frac{|j-k|}{l_p}} \tag{2.25}$$

Practically the correlation function is calculated for a given configuration with the formula:

$$\chi_{\vec{t}}(k) =< \vec{t}(j).\vec{t}(j+k) >_j= \sum_{j=0}^{L-k} \frac{1}{L-k} \left(\vec{t}(j).\vec{t}(j+k)\right) \tag{2.26}$$

Then this function is averaged on 1000 configurations taken regularly on the whole simulation. Finally the function is averaged on all the strands of the construct.

The correlation function for the different DNA constructs (Fig. 2.27) shows an exponential decrease for the simple dsDNA and a more complex behaviour for the other DNA constructs. It is to be noticed that the data shown for the DNA constructs correspond to the correlation function of one of the strand of the construct. The pattern of distance between crossovers (31-32) is visible on the correlation function.

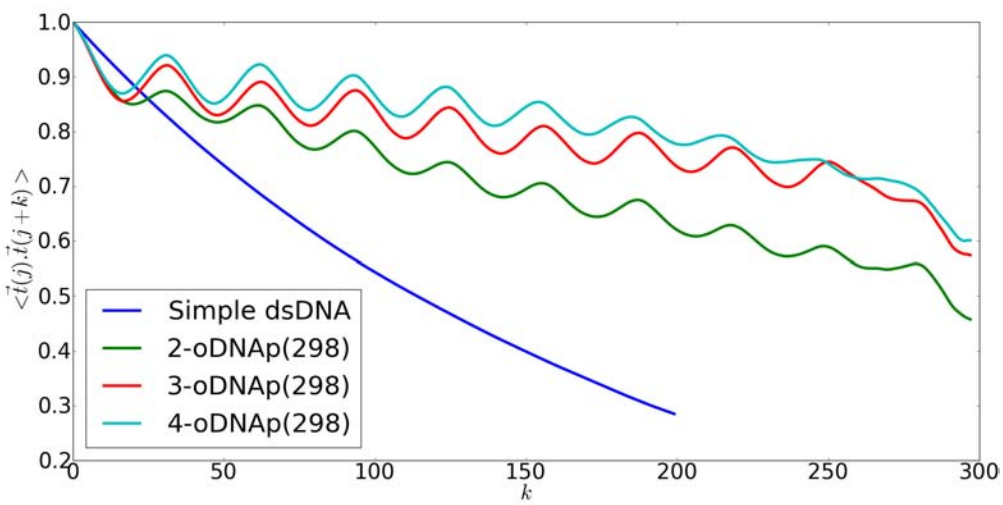


Figure 2.27: Example of correlation function for different constructs. The data shown for the DNA constructs correspond to the correlation function of one strand of the construct

In order to compute l_p, we coarse grain the oDNA structure with reference to the ideal situation where the base-pairs B_{ij} are located on a regular grid where lines (index i corresponding to the dsDNA strands) and columns (index j, in the direction perpendicular to the lines) can be defined. The coarse -graining is done in two steps. First, we average the position of the center of the ellipses belonging to a same column: $B_i^1 = \sum_j B_{i,j}$ (Fig. 2.28). In this way, a family of tangent vectors $\tau_i^1 = B_{i+1}^1 - B_i^1$ is obtained. Second, we average these tangent vectors along the i coordinate: $\tau_i^a = \sum_{l=i-\delta i}^{l=i+\delta i} \tau_l^1$. This averaged tangent vector field is used to compute the persistence length for structures containing more than one dsDNA strand.

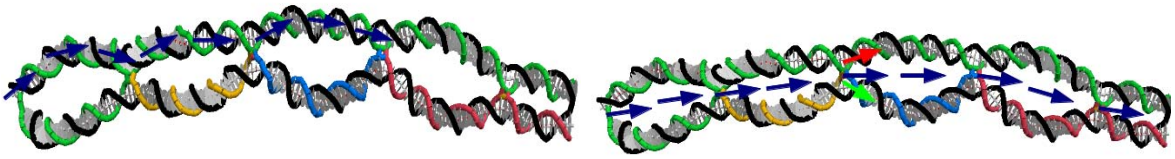


Figure 2.28: (Left) Example of the tangent field of one strand. It shows a the sinusoidal component. (Right) If we add the green and red tangents on one column we obtain the blue tangent field.

In Fig. 2.29 is reported the correlation function of this averaged field for different n-oDNA$_p$(141) structures and its fit by Eq. 2.25. One can notice that the sinusoidal component due to the electrostatic interaction is removed.

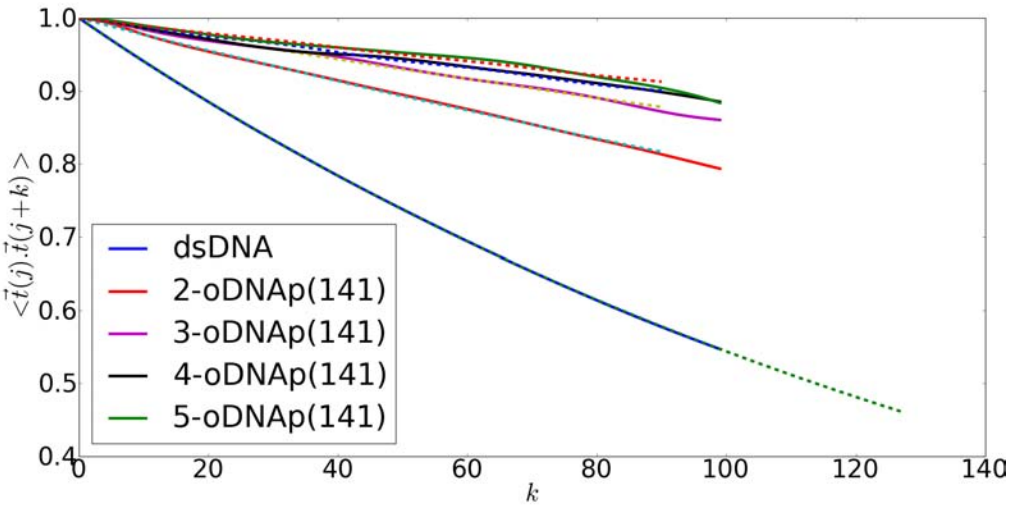


Figure 2.29: The correlation between the tangent vectors of a same column as a function of the distance along the strand. For each structure, the continuos line represents the measured correlation, the dashed line the fitted data

The value obtained for the persistence length of the different systems (Fig. 2.29) will be discussed later.

The correlation function for a given distance is the result of an average on all the points separated by this distance. As a consequence for long distances the function is an average of only few points.

Experimentally it seems that the "regular" behaviour of the curve is observed on 0.9 times the total length of the structure. This is why the fit of the correlation function will be made with a cutoff of 0.9 times the total length.

2.4.3.2 Comparison with indirect measurement ([65])

Indirect measurements of the persistence length of structures similar to 2-oDNA (Fig. 2.30) have been performed in [65]. These structures called DAE (for **D**ouble crossovers, **A**ntiparallel ds-DNA, **E**ven number of helical half turns between crossovers) have a distance between crossovers of 21 bases. These structures are the first type of DNA constructs, initially envisioned by Seeman [66].

The measurements were performed using statistical analysis on the fragments (Fig. 2.30 b)) obtained by ligation of the initial DAE molecule [67]. The analysis of the length of the fragments and of the shape dispersion allows to extract the persistence length. Different lengths of the junction between ligations were compared to study the impact of this parameter. The results are gathered in Tab. 2.5.

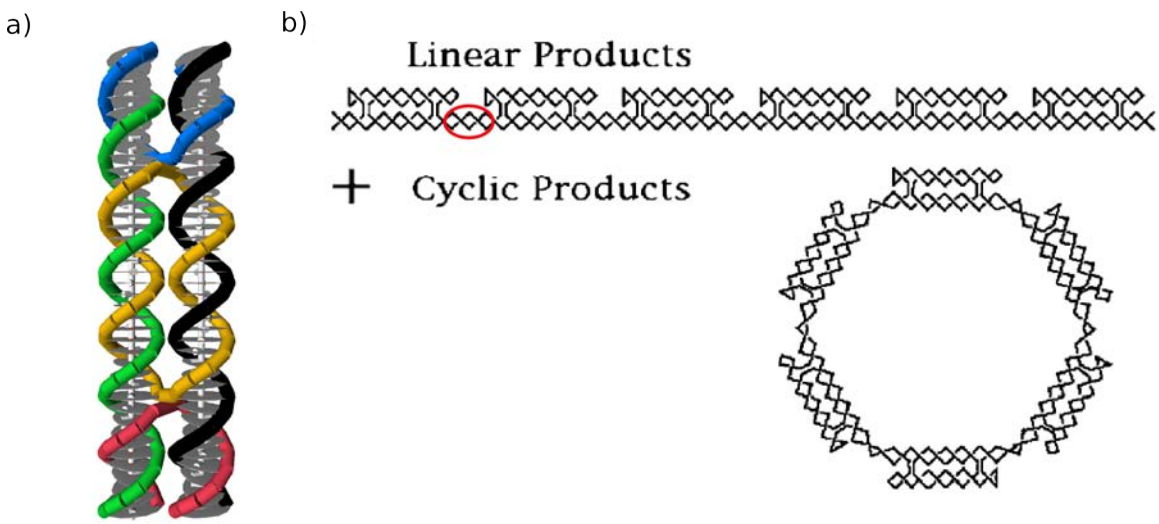


Figure 2.30: a) Schematic representation of DAE-42. b) Products of the ligation [65]. The red circle shows the junction between monomers. In [65] the length of this junction was change from 41 to 43 to evaluate the effect of this parameter

	DAE-42	DAE-41	DAE-43	DAE-nicked	Duplex-42
l_p (bp)	295	350	380	285	130
$l_p/l_{pDuplex-42}$	2.26	2.70	2.92	2.2	1

Table 2.5: Experimental datas from [65] and their normalised value

We studied the monomer DAE (Fig. 2.31 a) and different n-mer (Fig. 2.31 b-f). The values obtained for the persistence length of the different structures are gathered in (Tab. 2.6). These results are qualitatively similar for a simple dae.

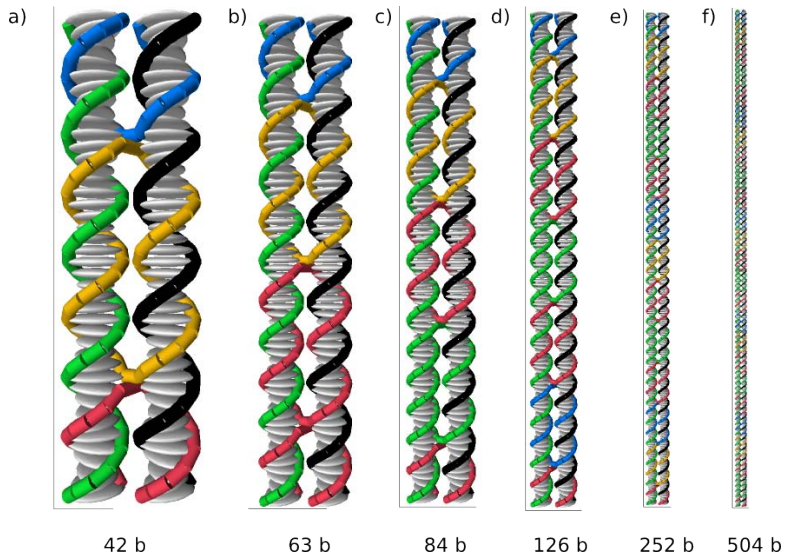


Figure 2.31: The different n-mer studied. The length in bases is indicated At the bottom.

	dae	2-mer	3-mer	5-mer	11-mer	21-mer
l_p/lp_0	3.1± 0.1	3.4± 0.2	3.6± 0.2	3.7± 0.2	3.8± 0.1	3.5± 0.2

Table 2.6: Normalised persistence length for DAE of different sizes. Values obtained by fitting the correlation function of the tangential vector.

A study of the evolution of the persistence length when increasing the structure length leads to a small dependency with the length and the value seems to converge when increasing the number of monomers (a monomer being defined by a segment between two crossovers).

Even if dae molecule is globally stiff it can adopt frequently a reversed bent structure (Fig. 2.32). This structure emerges as the natural twist is not exactly preserved and exerts an internal stress on the molecule and also, because the repulsive interactions between dsDNA tends to move them away.

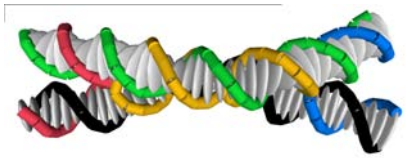


Figure 2.32: Reverse bent structure

For longer molecules such a mode does not contribute to the global flexibility as the next monomer has an opposite curvature resulting in a flat structure.

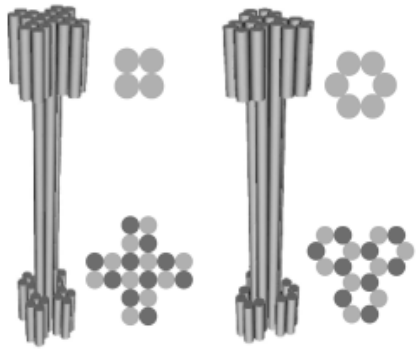


Figure 2.33: Schematic representation of the four (478 nm long) and six helix bundle origami (370 nm long) [8]

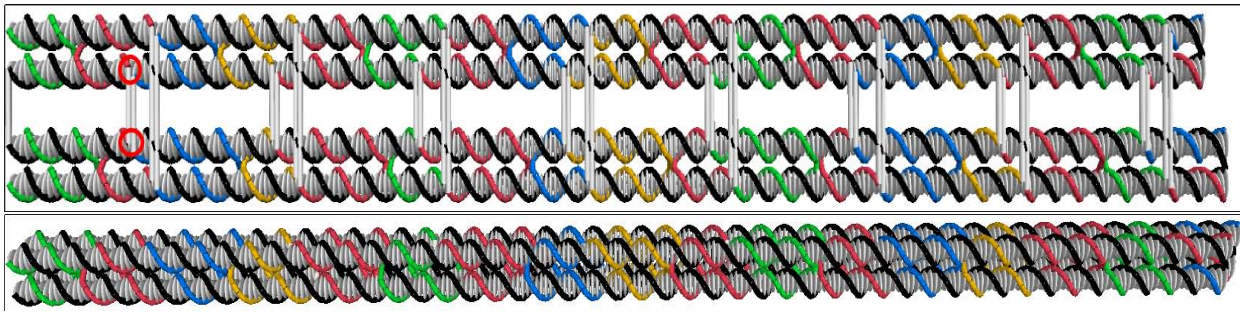
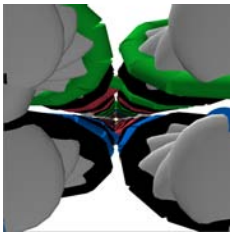


Figure 2.34: Construction of a oDNA similar to a four helix bundle. (Top) shattered view, (Bottom) realistic representation

2.4.3.3 Comparison with direct measurements ([8])

In the work of Kauert at al. [8] direct measurement of the flexibility of three dimensional origami are realised using magnetic tweezers. These origamis are the four and six-helix bundle type origami (Fig. 2.33). Experimental force measurements were realised and compared to a model of dsDNA [57]. The behaviour of the bundle was compared to that of a single dsDNA with higher persistence length. The experimental results are in good agreement with the theoretical model for a total persistence length of 740 ± 140 nm for the four-helix bundle and a total persistence length of 1880 ± 270 nm for the six helix bundle. Compared to a dsDNA with persistence length of 53 nm this gives an increase of 15-fold for the four-helix bundle and of 38-fold for the six helix bundle.



We modelled structures similar to the 4 helix bundle (Fig. 2.34,2.35). The shattered version of the structure is presented in figure 2.34(Top). The junction between the two $2 - oDNA_p$ were designed at the bases where the extremities

Figure 2.35:

4-helix bundle	2-mer	4-mer
l_p/l_0	16.6 ± 2.5	17.5 ± 6.3

Table 2.7: Normalized persistence length for the 4 helix bundles (fig. **??**) as a function of the number of monomers.

of the strand were on the top (Fig. 2.34 red circles).

In order to keep the time required for the simulation reasonable we worked on a 20 nm and 40 nm long structure with respectively two and four monomers.

The results are gathered in Tab. 2.7 Again, we find good agreement between the Monte-Carlo simulations and the experimental results obtained with magnetic tweezers. A value of persistence length for such construct can also be obtained using classical model of elasticity.

2.4.3.4 Comparison with classical model of elasticity.

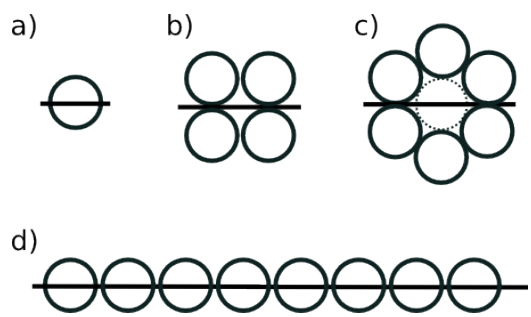


Figure 2.36: Schematic representation of a) a dsDNA b) Four helix bundle c) 6 helix bundle d) oDNA

In this section we want to apply the methods of classical elasticity [68] to study the beam assemblies represented in Fig. 2.36. In a classical model the energy of a beam having a curvature R is given by:

$$F = \frac{1}{2}EI\int_0^L (\frac{1}{R} - \frac{1}{R_0})^2 dz \tag{2.27}$$

with E the module of elasticity, I the second moment area of the beam, L its length and R_0 the curvature at the equilibrium. For a circular rod I is given by $I = \pi r^4/4$ where r is the radius of the rod.

For the continuous model of the worm-like chain the elastic energy is given by Marko et al [69]:

$$\frac{F}{kT} = \frac{1}{2}\int_0^L l_p(\frac{\partial^2 \vec{R}}{\partial s^2})^2 ds \tag{2.28}$$

As a consequence

$$l_p = \frac{EI_{ds}}{kT} \tag{2.29}$$

with $I_{ds} = \frac{\pi r_{ds}^4}{4}$ For a single dsDNA $r_{ds} \approx 1nm$. For the four helix bundle considering that the four strands form a compact road of radius 2 nm,

$$I(4-\text{bundle}) = 2^4 I_{ds} = 16 I_{ds} \tag{2.30}$$

As l_p is proportionnal to I, this gives an increase of 16-fold for the persistence length of the 4 helix bundle (experimentally [8], a factor 15 was found). The result can be derived more precisely with the parallel axis theorem: The second moment area about a new axis parallel to the axis of the original second moment area is given by the relation :

$$I_n = I_i + AD_{ni}^2 \tag{2.31}$$

with D_{ni} the distance between the two parallel axis and A the area of the shape. This gives for the four helix bundle:

$$I(4-\text{bundle}) = 4 \times (I_{ds} + \pi r^2 r^2) = 20 I_{ds} \tag{2.32}$$

And for the six helix bundle:

$$I(6-\text{bundle}) = I(4-\text{bundle}) + 2 \times (I_{ds} + \pi r^2 (2r)^2) = 54 I_{ds} \tag{2.33}$$

These two figures (20 and 54) can be compared with the experimental data (14 and 38): the continuous model exagerates the stiffness. The values obtained for the 4 helix bundle and the 6 helix bundle are 33% and 42% higher than the experimental ones. The distance to the axis $r_e = \alpha r$ that will give a correct estimate of the persistence length for the 4-helix bundle would be the solution of the equation: $4(1 + 4(\alpha)^2) = 15$. This gives $\alpha = 0.88$.

The same effective value gives as a result for the 6 helix bundle: $I(6 - bundle) = 37.5 I_{ds}$.

These calculations even if not predictive, give at least a good order of magnitude, and considering an effective radius of $r_e = 0.88r$ give predictions very close to the experimental value. In the case of planar oDNA the classical model predicts that the persistence length increase linearly with the number of strands.

2.4.3.5 CanDo-DNA website.

Kim et al. [42], developed a framework to predict the shape and the fluctuations of origami structures designed thanks to the caDNAno software [70]. This structures can be given to the website http://cando-dna-origami.org/ and the final shape as well as data on thermal fluctuation can be obtained. The calculations are based on the finite element theory where each dsDNA is represented by a set of two-node beam finite elements whose characteristics stiffness are the one of dsDNA. The 2-oDNA was submitted for calculation and the equilibrium

Figure 2.37: Equilibrium configuration of the 2-oDNA submitted to CanDo

configuration shows a global shape twisted of about -0.5 turn (Fig. 2.37), close to our value of -0.65. Other information such as the lowest normal modes are available. However, it is not easy to extract from these datas the persistence length.

A good agreement is being observed between the model and the different experimental results available:

- The effect of the electrostatic interaction on the distance between strands for the experimental value of the Debye length is well reproduced (Sec. 2.4.2)
- The persistence lengths of the DAE structures (Sec. 2.4.3.2 and the 4-bundle structures (Sec. 2.4.3.3) are in qualitative agreement with the model.
- The twist of the structure is in good agreement with the different observations (Sec. 2.4.3.5 , 2.4.1)

We then decided to study the effect of different parameters such as the density of crossover, and the dependency of the persistence length as a function of the number of dsDNA for oDNA structures.

2.4.4 Study of the persistence length for different constructs

2.4.4.1 Effect of the crossover density

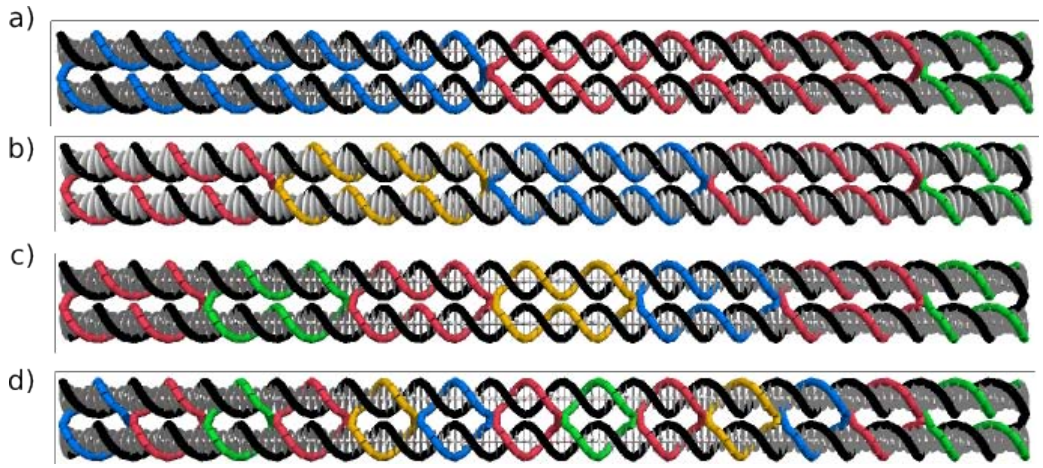


Figure 2.38: The different structures studied in this paragraphs. All are 141 bases long with various crossover density, respectively a) b) c) d)

In this paragraph we study the 2-oDNAp (141) structure and structure identical in length but with different numbers of crossovers (Fig. 2.38). For this oDNA structure the maximum possible number of junctions is 14. There we use the term junction, as connection at the extremities of the structure are also counted.

# junctions	4 j	6 j (2-oDNAp(141))	8 j	14 j
lp/lp_0	0.6± 0.3	3.2± 0.3	3.9± 0.3	3.9± 0.1
Tw_g (turn)	0.0± 0.1	0.1± 0.0	0.1± 0.0	0.1± 0.0

Table 2.8: Evolution of the persistence length with the junction density.

As expected, the structure with two parallel strands is generally stiffer than the double helix (Tab. 2.8), with $l_p/l_0 \sim 3.2$. When only 4 crossovers are present (cf. fig. 2.38(a)), we find that the persistence length of 2-oDNA$_p$(141) is lower than that of one single helix. This result can be understood by noticing that the crossovers are separated by 90 base-pairs. Instead of being parallel, the dsDNA strands are almost free to fluctuate and the electrostatic repulsion between strands increases these fluctuations, leading to contorted structures with a persistence length inferior to that of a single dsDNA strand. It should also be noticed that these structures are difficult to describe by a single coordinate, the coarse-graining procedure reaching here its limit of validity.

2.4.4.2 Increasing the number of dsDNA:

We also wanted to quantify the evolution of the persistence length of oDNA as a function of the number of dsDNA. We saw previously that the persistence length was dependent on the number of crossovers. We decided to study structures 300 bases long which is the classical length for oDNA and structures two times smaller to see the impact of number of monomer. We studied structures with increasing number of dsDNA. For the longer length the number of dsDNA row variates from two to four and for the the shorter from two to five. The results for two length of structures are gathered in Tab. 2.9 and Tab. 2.10

	2-DNA$_p$ (298)	3-DNA$_p$ (298)	4-DNA$_p$ (298)
l_p/l_0	3.5± 0.3	5.7± 0.7	7.4± 1.7
Twist (turn)	0.1± 0.0	0.2± 0.0	0.1± 0.0

Table 2.9: eight different simulations were averaged for each system

	2-DNA$_p$(141)	3-DNA$_p$(141)	4-DNA$_p$(141)	5-DNA$_p$(141)
l_p/l_0	3.2± 0.3	5.1± 0.4	5.8± 1.4	6.6± 3.3
Twist (turn)	0.1± 0.0	0.1± 0.0	0.2± 0.0	0.2± 0.0

Table 2.10: Result for structure with half the length of the previous structures

For the structures with 298 base-pairs per row increasing the number of rows increases monotonously l_p, by roughly 1.8 l_0 added per row. For shorter structures, the situation is very similar, as should be expected for a local measure of elasticity. However one can notice that the incertitude on the measurement is important for the smaller structure. This incertitude come from the fact that the total length of the structures is of about 50 nm or one l_0 which is short by comparison with their persistence lengths, and as a result the fit of the correlation function is less precise. If one only considers the longer structures, a simple rule seems to emerge: the persistence length of a N-oDNA(l) structure with N dsDNA is roughly given by $1.5 \times N$ for l long enough.

The persistence length characterise the mechanical properties of the structure. In the following section, we will use a special Monte Carlo algorithm called Umbrella Sampling, to realise simulated Force-distance experiment, and to access directly to the force needed to bend different constructs.

2.4.5 Force curves measurement of bending on DNA constructs

2.4.5.1 Umbrella Sampling.

In all the previous section we used Metropolis algorithm to sample the phase space. Here we introduce the Umbrella Sampling technique [71, 72] that allows to sample high energy configurations that are not explored by the Metropolis algorithm at room temperature. These configurations can be of interest if we want to explore highly bent configurations of oDNA or dsDNA.

Umbrella Sampling allows the sampling of a restrained part of the phase space by rejecting moves that lead to configurations outside this region. For example only configurations which have a end to end distance (R_{ee}) that belong to a given interval $[l; l + dl]$ are accepted. We obtain in this way the distribution of probability of having a given length in a restrained interval. The free energy can be computed as $\beta A(R_{ee} \in [l; l + dl]) = -ln(p(R_{ee})) + cte(l)$. This additive constant is due to the fact that the probabilities are normalised in a restrained interval $[l; l + dl]$. If the same algorithm is applied for a set of interval constrained in $[l_0; l_1]$, such that each interval slightly overlaps with its neighbours, the total free energy can computed as the constant corresponding to each interval is determined by continuity. So the total free energy is known only up to an additive constant.

In order to have a good sampling in each interval, the size of each interval has to be chosen carefully. If the interval $[l; l + dl]$ presents large variations in energy, a limited sampling will explore only a small part of the interval. This is due to the MC method of sampling that will spend most of the simulation with configuration having 1 k_BT of energy higher than the lowest configuration. So the size of the interval needs to be chosen such as the maximum variation in energy of the interval is not superior to Nk_BT (with N the number of degree of freedom of the system).

2.4.5.2 Detailed steps of the Umbrella sampling method.

By using the umbrella sampling method, it is possible to access to the free energy $A(R_{ee})$ as a function of the distance between the two ends of the structure and "reproduce" force distance experiments. Here we illustrate the different steps in an example for a strand of 40 bases. Then, as this technique allows to obtain force distance curve we will compare the data obtained for the different sizes of strand to the analytic model (Sec. 2.2.3.2)

The first step is to realise simulations on several different slightly overlapping intervals (Fig. 2.39). The figure represents R_{ee} as a function of the MC steps. Every 3300 steps the interval in which the configuration is constrained is changed. For this simulation each window was 1 nm long. For instance, in the first window the strand was constrained to an end to end distance (R_{ee}) between 14 and 15 nm during 3300 steps.

One can notice that the length interval to which it was constrained (symbolised by a vertical line of different color) is not fully explored and that R_{ee} stays close to 14 nm. The strand has

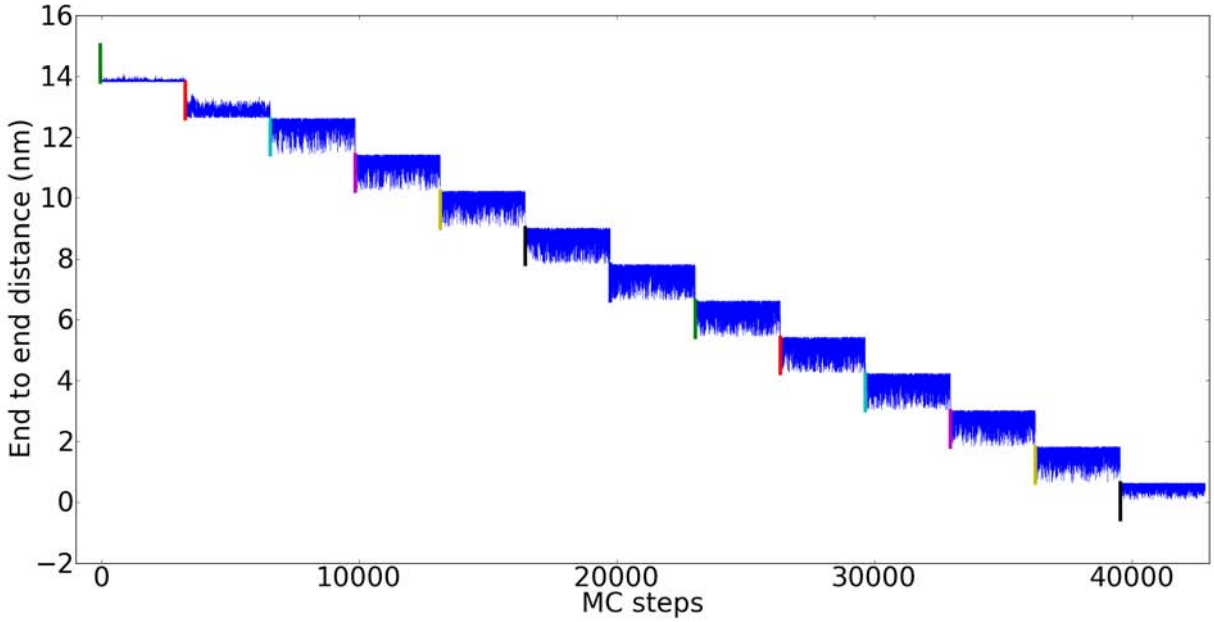


Figure 2.39: Monte carlo trajectory, the windows of constraint are changed during the simulation every 3300 MC steps

a contour length of $13.2nm$, and in this window the strand is constrained to be extended. The configuration of minimal energy is the one having a $R_{ee} \approx 14$ nm and extending it slightly increases rapidly the energy. As MC sampling explores mostly configurations having k_BT by degree of freedom the configurations higly extended are not sampled properly. If we wanted to sample this region properly, we should divide the window in smaller areas.

However, the region of interest is the one where the strand is bent. In the simulation steps it corresponds to the MC steps superior to 80000. As the configurations with minimal energy are the one that are less bent, in each window the more explored configurations are the ones with a longer end to end distance.

The trajectory is separated in individual windows, and the distribution of probability $P(R_{ee})$ is calculated for each window (Fig. 2.40 a). Then, the free energy $A(R_{ee}) = k_BTln(P(R_{ee}))$ is computed for each window (Fig. 2.40 b). However, this free energy is defined up to an additive constant that depends on the window. By continuity each additive constant is adjusted, and the free energy on the total interval is evaluated within an additive constant (Fig. 2.41). The additive constant is set to zero for the minimal energy which correspond to a length slightly smaller than the total length of the strand. This length is smaller because the most likely configuration is a configuration slightly bent. Then the force $F = -\frac{\partial A}{\partial l}$ is computed (Fig. 2.42). The effect of the window size was tested and showed no significant effect on the value of the force (Fig 2.43). However for large window size, the sampling is not accurate and prevents

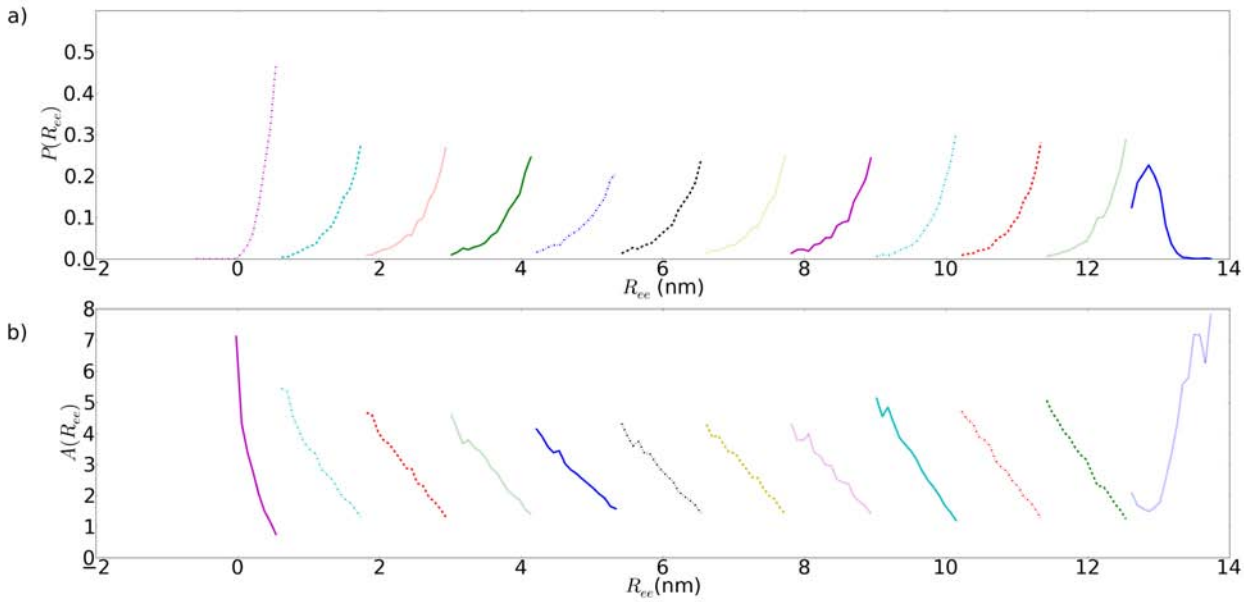


Figure 2.40: a) Distribution $P(R_{ee} \in [l : l + dl])$ of length for each window $[l : l + dl]$, b) Free energy up to an additive constant, different for each window.

some size extension to be measured. We can also notice that the estimated force is very noisy. This is due to the fact that we are interested in a derivative of a function extracted from the simulation.

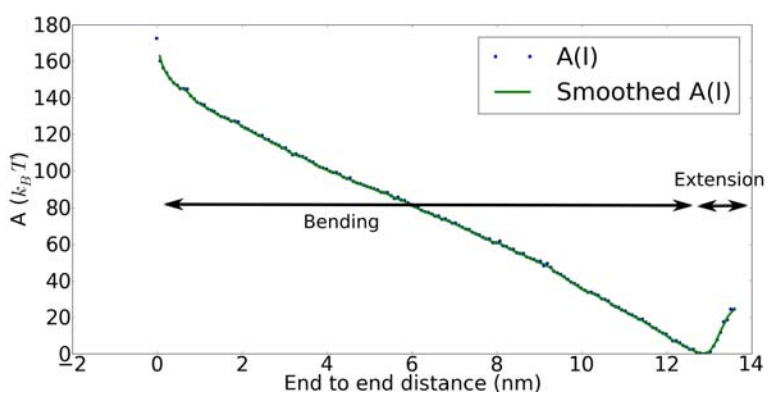


Figure 2.41: Free energy. The different parts bending equilibrium and extension are clearly visible. The slope of the extension part is bigger.

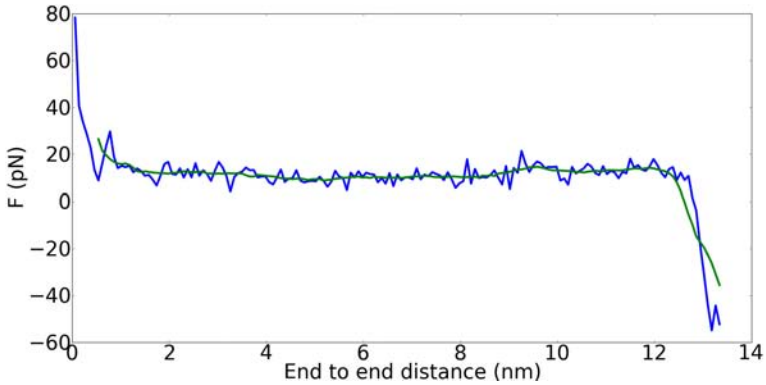


Figure 2.42: Force required to bend the small dsDNA of 40 bases.

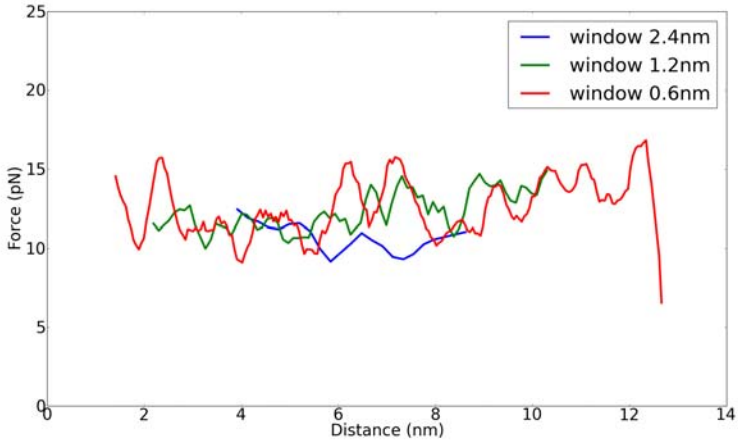


Figure 2.43: Force obtained for different lengths of the sampling window.

2.4.5.3 Result on dsDNA

In Sec. 2.2.3.2 we have shown that for strands shorter than their persistence lengths the force and the energy of structure constrained to a R_{ee} smaller than their contour length L_c are given by:

$$F(R_{ee}) = 8k_BT\frac{l_p}{L_c^2}\frac{arcos(R_{ee}/L_c)}{\sqrt{(1-(R_{ee}/L_c)^2}} \tag{2.34}$$

$$E(R_{ee}) = 4k_BT\frac{l_p}{L_c}arcos(R_{ee}/L_c) \tag{2.35}$$

In (Fig. 2.44) results obtained for different length of strands by umbrella sampling (green curves) are compared to the analytic predictions with $\mathrm{lp}_0 = 160$ bases (blue curve). The agreement between both is good. It is also possible to evaluate the persistence length (lp_f) by fitting the simulation data (dotted red curve) by both theoretical expressions (Eq. 2.34 and Eq. 2.35). One can see that for the shortest strands the agreement is good between the value obtained with these simulations (lp_f) and lp_0. (Tab. 2.11). The limit of validity of the model is clear: for a strands that is two times the persistence length, the strand can no longer be considered as a rigid rod. In that case (340 bases long strand) there is a large minimum and the force to bend is almost null. It is to be noted that the force plots are noisy. A longer sampling would have been required to obtain smoother curves.

length of the strand	40	50	100	340
η (fit force)	-4%	+ 7%	+31%	-189%
η (fit energy)	-5%	+ 7%	+ 14%	-15%

Table 2.11: Relative error $\eta = \frac{lp_f - lp_0}{lp_0}$ for the different lengths of the strands considered in Fig. 2.44

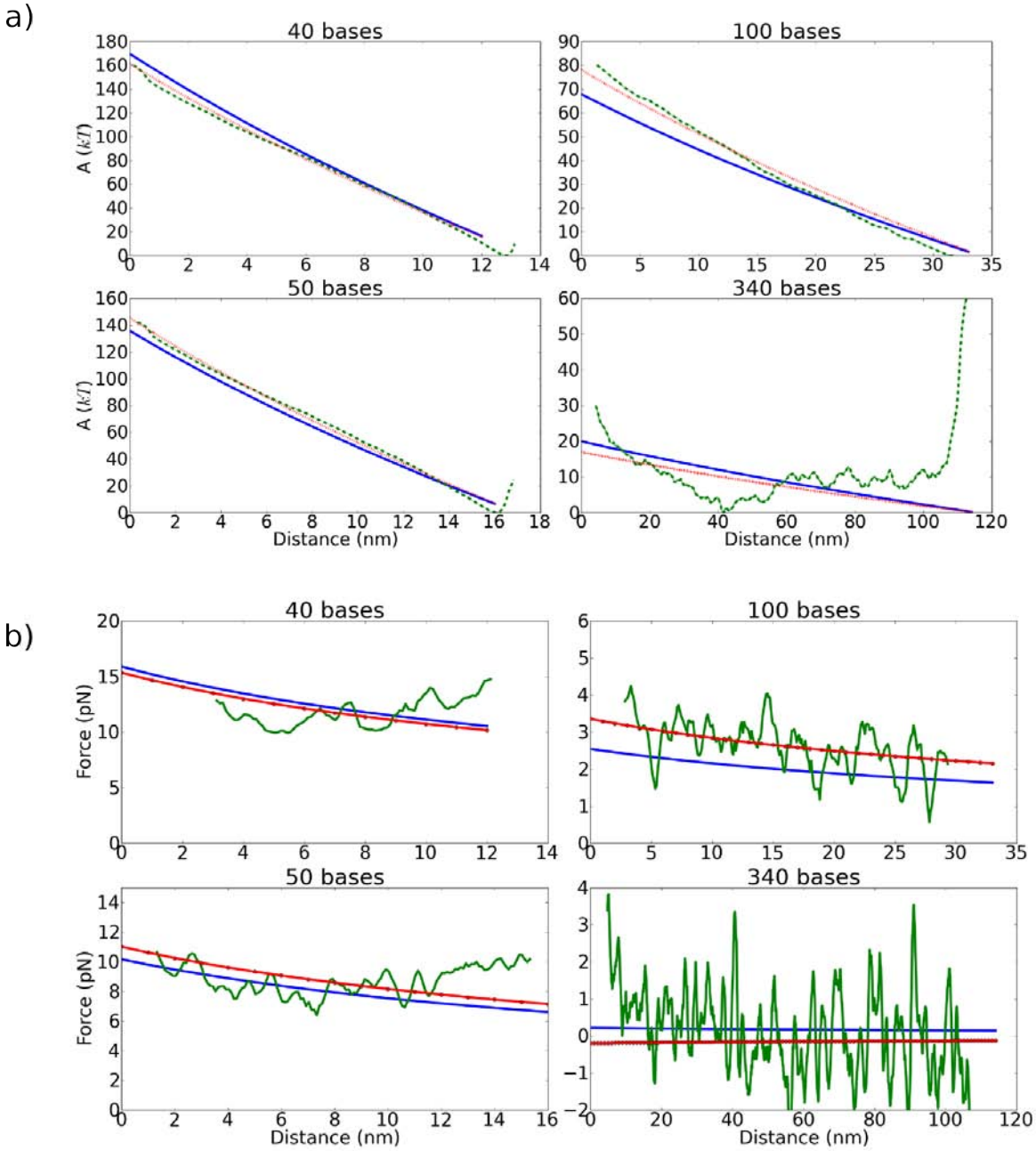


Figure 2.44: a)Comparison of the free energy for the analytic model (blue),Fit of the analytic model (dotted red), MC model (green) for different sizes of strand. b) Comparison of the force for the analytic model (blue),Fit of the analytic model (dotted red), MC model (green) for different sizes of strand

2.4.5.4 Force curves measurement of bending on oDNA.

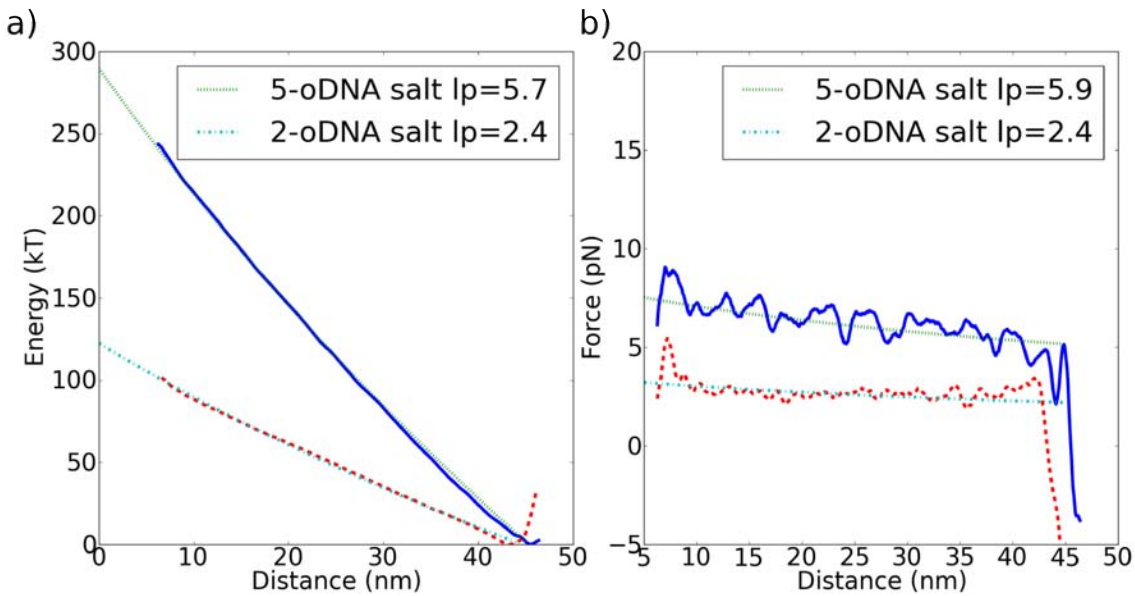


Figure 2.45: a) Fit of the free energy curve with the analytic model for 2-oDNA$_p$(141) and 5-oDNA$_p$(141) . b) Forces curves

In the case of an oDNA construct the constraint is set on the mean end to end distance of the strands. The energies (Fig. 2.45 a) and forces (Fig. 2.45 b) were computed for the 2-oDNA$_p$(141) and 5-oDNA$_p$(141) structures. The fit realised thanks to the analytic WLC model gives a good agreement with the simulation. The analytic model is even good for very bent shapes. The values of persistence length that were obtained with regular MC trajectories were of 3.2 l_0 for the 2-oDNA$_p$(141) and of 6.6 l_0 for the 5-oDNA$_p$(141). The values obtained by fit of the energy of bending are respectively of 2.4 and 5.8. As these values are proportional to the number of strands, they are easier to understand. We think that the value obtained at the equilibrium for these structures were influenced by the fact that the structure were deformed.

It would have been interesting to do the same measurement on the 2-oDNA$_p$(298) and 5-oDNA$_p$(298) structures to confirm this intuition but it was not possible for computational time reasons.

2.5 Conclusion

The analytic model used for dsDNA can be extended to oDNA by adapting only the value of the persistence length of the construct. As discussed in the analytic model in a first analysis this value is proportional to the number of strands n. The extremes values were at our sense nl_0 and $2nl_0$ if we consider that attaching a strand to an oDNA is equivalent to trap it in two dimensions. From the numeric simulation it results that the persistence length by strand is approximately 1.5 l_0. This intermediary value is due to the fact that thanks to the electrostatic repulsion, the structure is not densely packed and the strands can still fluctuate partially in the width of the oDNA. The persistence length also depends on the coupling between strands where the number of crossovers play a more important role than the distance between crossovers.

The formula to compute the force needed to bend an oDNA composed of n strands is fairly constant and is given by:

$$F \approx 8k_T \frac{1.5 \times nl_p}{L_c^2} \tag{2.36}$$

For a classical oDNA composed of 20 dsDNA and 300 nm long the value is 6 pN. The numerical model gives information about the persistence length and also allows to predict the equilibrium shape in term of twist and in term of density of dsDNA as as the distance between dsDNA is in adequacy with the experiments.

To conclude this theoretical work it is possible to graphically extract the lengths of equilibrium of the oDNA composed of 20 dsDNA, each one being 340 bases long, bent by different sizes of bridges (Fig. 2.46), which was the initial purpose of this chapter.

Unfortunately it was not possible to realise the structures corresponding to these predictions. The reasons of this failure are explained in Appendix B.

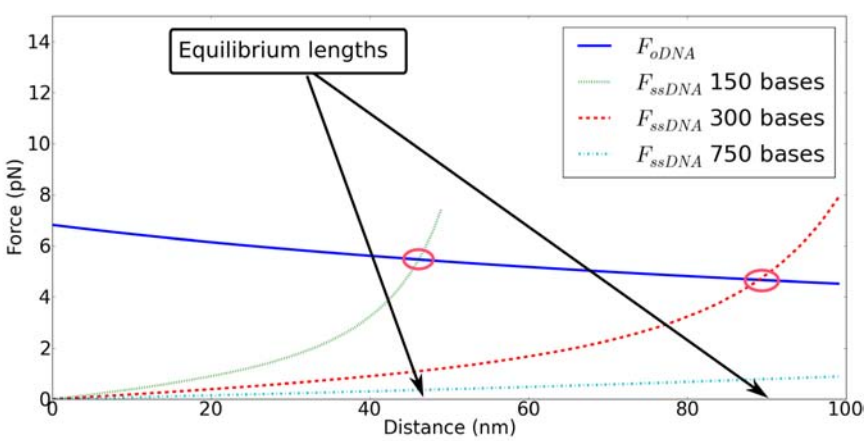


Figure 2.46: Force needed to bend the oDNA (F_{oDNA}) vs entropic force of a ssDNA (F_{ssDNA}) for different lengths.

Chapter 3

Process of formation of DNA origamis:

Contents

3.1 Introduction. **60**

3.2 Experimental techniques to study DNA hybridisation. **63**

3.3 Small origami **65**

3.3.1 Experimental datas 65

3.4 Thermodynamic model on a lattice for small origami **71**

3.4.1 The model 71

3.4.2 Results 74

3.4.3 Small origami, comparison with experiment 84

3.4.4 Conclusion on the coarse-grained model 84

3.5 Origami **86**

3.5.1 Experimental data. 86

3.6 Thermodynamic model of the formation of a DNA origami **91**

3.6.1 Ideas of the model 91

3.6.2 Folding energy of a given configuration 94

3.7 Comparison with the annealing-melting data of four origamis **98**

3.7.1 Kinetic effect. 98

3.7.2 Understanding DNA Origami design 99

3.8 Conclusion and perspectives. **102**

3.1 Introduction.

The process of formation of an oDNA is very robust as the observed yield can be around 90 % [1]. This can be surprising as the formation process involves about 220 staples. Even if all oDNA are not perfectly formed this global yield implies that the process of attachment of every staple has a yield η such as

$$\eta^{220} = 0.9 \tag{3.1}$$

or $\eta \approx 0.9995$. This yield is a result of the high specificity of the position of folding of the staple and of the high energy involved when two strands fold. In this chapter we want to address the following questions:

- For the same shape of oDNA with the same topological scaffold pathway, is the process of formation identical with a different set of staples, as for instance when doing permutation of the strand location along the circular scaffold ?
- The process of folding involves 200 hybridisation reactions. Is it possible to have an insight in the process of formation ?
- Is there any correlation between staples in the process of formation?
- In the case where intermediate states might happen is it possible to predict them?

To obtain the more stable structure the formation process of oDNAs consists of: mixing together the ssDNA in aqueous solution, heating the mix and slowly cooling the solution. It can be analysed and followed in several ways (Sec. 3.2). In this work we used UV measurements at 260 nm that are based on the fact that hybridized bases absorb less than open bases (phenomenon called hypocromism [74]).

The fraction of hybridized bases as a function of the temperature (degree of pairing $\theta(T)$) can be obtained from raw absorbance measurements. The derivative of the melting curve of a short dsDNA displays a maximum at a well defined temperature [73, 75] (Fig. 3.1 a) . This maximum is called the melting temperature (T_m) of the duplex. For longer dsDNA macromolecules, several maxima may appear , as a signature of the existence of contiguous regions that fold (or melt) independently (Fig. 3.1 b). For even longer molecules, all the transitions are superposed, leading to a profile similar to that of a small molecule, but larger in width (Fig. 3.1 c).

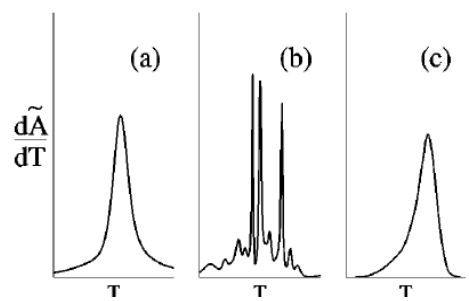


Figure 3.1: From [73]: Schematic view of possible differential melting curves observed for a) short $\approx 10^2$ bp, b) intermediate $\approx 10^3 - 10^4$ bp , and c) long ($\geq 10^6$ bp) DNA sequences.

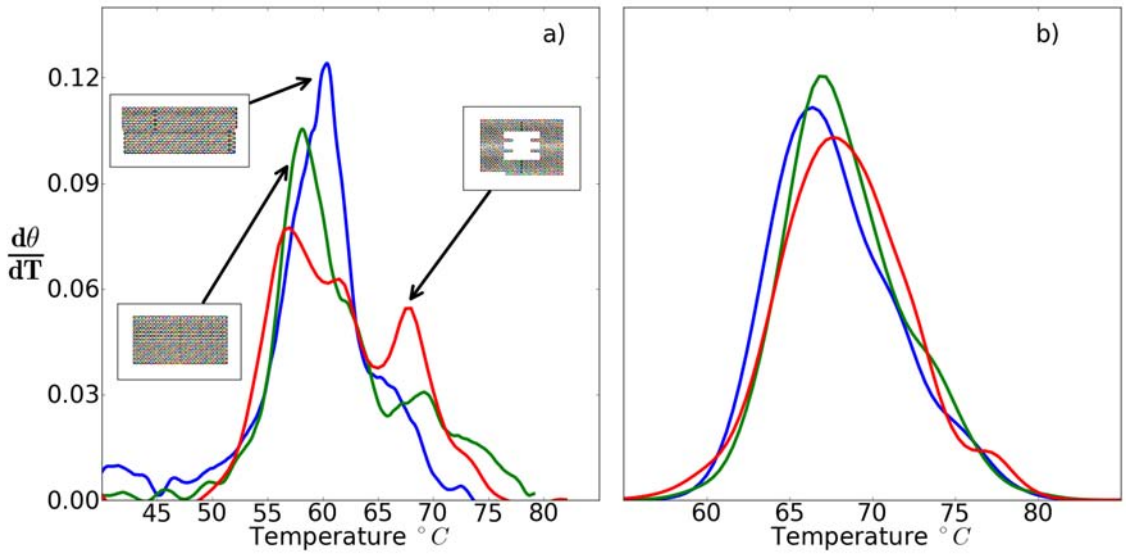


Figure 3.2: (a)Derivative $d\theta/dT$ of the degree of pairing with respect to temperature for the three DNA origamis represented in the insets. (b) $d\theta/dT$ for a model where the staples fold independently.

The set of staples of a oDNA depends on its detailed connectivity. However, the average AT and GC content of two oDNA based on the same scaffold is identical. Fig. 3.2 a) represents the derivative of the melting curves of three different origamis based on the same scaffold. The observed differences (Fig. 3.2 a) point to the existence of mechanisms of folding that are very different from those at work in dsDNA macromolecule. To further stress this difference, (Fig. 3.2 b) represents the melting curve for each oDNA as a linear combination of melting curve of each staple. Notice that the hypothesis of an uncorrelated staple folding process gives a melting temperature 10 $°C$ upper than the one experimentally observed. Furthermore the calculated relative stability of the three oDNA is inverted with respect to the experimental data. This is a second result showing that the hypothesis of an uncorrelated staple folding is a crude, non correct, approximation.

Staples are designed to hold together regions of the scaffold that, otherwise, would be separated by a (possibly) long sequence. The binding of a staple to the scaffold is hindered by an entropic penalty that depends on the length of this region. At high temperature, this region of the scaffold forms a coil. Depending on the T_m at which this staple binds, it may happen that other staples have already folded within the coil, reducing the entropy. Therefore, the binding of any staple depends on the previous binding state of the other staples, leading to a field of interacting loops of various sizes. Stated in such a way, the problem of describing the folding (unfolding) path of oDNA appears untractable.

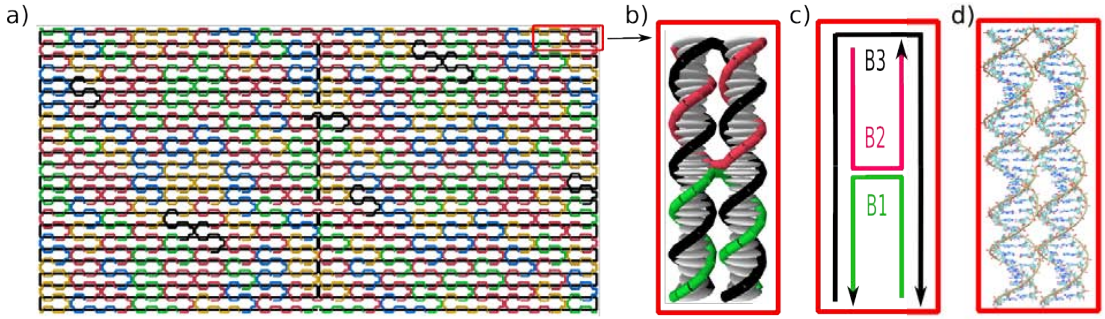


Figure 3.3: a) Representation of a regular origami with ≈ 200 staples;b,c) Two schematic representations of a small origami. This small origami can also be seen as a bent holliday junction; d) All atoms representation of a small origami

In an attempt to have a more quantitative picture, we first reduce the problem down to a simple structure made of three ssDNA . This three strand pattern can be viewed as a building block of the oDNA (Fig. 3.3). This preliminary study intends to shed light on the local process of pairing in a oDNA.

The experimental results on the small oDNA (soDNA) are presented in section 3.3.1. A coarse grained model at the scale of the base was developed to study the interactions between the three strands composing the small oDNA (Sec. 3.4). This model has been published [76]. However this model can not be used at the scale of the oDNA given the complexity of such structure. Then the different experimental results obtained on the classical oDNA structure will be discussed in Sec 3.5. A coarse grain model at the scale of the strand was then developed in a first attempt to understand the process of formation of oDNA (Sec. 3.6).

3.2 Experimental techniques to study DNA hybridisation.

This section details the possible experimental ways to study the process of hybridisation, and what thermodynamic quantities can be extracted from these studies.

Several techniques can be employed to study the process of formation of a dsDNA. The most common are UV measurement [77], Differential Scanning Calorimetry (DSC) [78], collecting the fluorescence intensity of a reporter dye [79], circular dichroism [80], Neutron Scattering [81], attaching a fluorophore and a quencher to two complementary strands (FRET) [82, 83].

We will develop more in detail the DSC method because of the informations that this technique can provide and the UV measurement technique as it is the one used in this work.

A DSC experiment gives access to the specific heat of transition Cp of a complex by measuring the power needed to keep the sample of interest and a reference at the same temperature [78]. Once $Cp(T)$ is known other thermodynamic parameters are accessible, such as, T_m, ΔH and ΔS of transition and the van't Hoff enthalpy ΔH_{vH} . The later is related to the change with temperature of the equilibrium constant for unfolding, and the ratio $\Delta H_{vH}/\Delta H$ is a measure of the cooperative behaviour of the system. If the ratio is greater than one cooperative behaviour is indicated, meaning that several bases unfold together, whereas if it is inferior to one it is likely that there are intermediate states. The partition function can also be directly calculated, and simplifying hypothesis give access to the proportion of different species (such as duplex, simplex and partially unfolded state) at every temperature.

Melting curves through UV measurement can also be used to study the melting-annealing process [77]. These measurements are based on the hypochromism effect: one dsDNA absorbs less than two ssDNA. This effect is attributed to dipole-dipole interactions between stacked bases [74]. DNA bases have a maximum of absorption around 260 nm and no absorption at 310 nm. The absorption recorded at these two wavelengths as a function of the temperature allows to calculate $A(T) = A_{260}(T) - A_{310}(T)$.

The difference is done to remove the effect of absorption of the buffer (this assumes that the buffer has the same behaviour at 260 nm and 310 nm), and the specificity of the cuves used for the experiment. One obtains a curve that needs to be normalised to obtain the fraction of folded elements θ: at low temperature the complex is supposed to be formed and $\theta = 1$ and at high temperature its value is zero. However even if the complex is totally unfolded, the absorption changes with the temperature and this is why the process of normalisation requires the estimation of a baseline at high and low temperature [77] that can be an important source of error. Therefore

$$\theta(T) = \frac{L0(T) - A(T)}{L0(T) - L1(T)} \tag{3.2}$$

with $L0(T)$ and $L1(T)$ the baseline values of the unfolded and folded species, respectively. An example of normalisation of a curve can be found in Appendix I.

Once derived, the fraction of folded bases obtained is very similar to the one given by the DSC technique [84]. Given this curve, it is straightforward to calculate the temperature Tm that yields the maximum of the derivative $\frac{d\theta}{dT}$ This value is roughly a measure of the stability of the complex.

To obtain further information a two states model is supposed. A two states model supposes that only two states are possible for the DNA strands: they are either totally separated , or in the duplex state. No intermediary states are allowed. This hypothesis allows the calculation of the association constant:

$$K_a(T) = \frac{\theta(T)}{C(1-\theta(T))^2}$$

This equation is correct if both initial ssDNA are at the same concentration C. K_a is linked to Gibbs enthapy by:

$$\Delta G = -RTln(K_a) = \Delta H_{vH} - T.\Delta S$$

And by plotting :

$$ln(K_a) = -\frac{\Delta H_{vH}}{RT} + \frac{\Delta S}{R}$$

as a function of $1/T$ one can determine ΔH_{vH} and ΔS. All these computations are justified if one assumes that the system is close to the equilibrium. In the case where the equilibrium is not reached (if there is an hysteresis between the melting and the annealing curve) one can access the kinetic constants $k_{off}(T)$ and $k_{on}(T)$ (in the hypothesis of a bi-molecular reaction [85, 86]):

$$\frac{d\theta}{dt} = k_{off}\theta - k_{on}C(1-\theta)(E-\theta) \tag{3.3}$$

where C is the concentration of the ssDNA , E an excess of one of the specie and $\frac{d\theta}{dt} = \frac{d\theta}{dT}\frac{dT}{dt}$. This equation 3.3 is true at every temperature, and given the fact that an hysteresis is present between the cooling and the melting process it is possible to solve at every temperature the system:

$$\frac{d\theta_{heating}}{dt} = k_{off}\theta_{heating} - Ck_{on}(1-\theta_{heating})(E-\theta_{heating})$$
$$\frac{d\theta_{cooling}}{dt} = k_{off}\theta_{cooling} - Ck_{on}(1-\theta_{cooling})(E-\theta_{cooling})$$

and obtain $k_{on}(T)$ and $k_{off}(T)$. Then by plotting $ln(k_{off}) = f(1/T)$ and $ln((E - 1/2)Ck_{on}/2) = f(1/T)$, the intersection gives the thermodynamic T_m and the linear regression gives E_{on} and E_{off}. And finally one can access to an equilibrium property: $\Delta G = E_{on} - E_{off}$ the Gibbs energy of the equilibrium.

In conclusion the UV technique allows the calculation of several thermodynamics parameters at the equilibrium and out of equilibrium, given the hypothesis of a two state reaction. In the following section we present the experimental results obtained in the study of soDNA structures.

3.3 Small origami

The small origami DNA (soDNA) is made of two ssDNA 32b long (staples) and a 64b long ssDNA (scaffold) (Fig 3.4 a). This structure is similar to DAO structures [87] and comparable in size and shape to JX and PX structures which have already been studied experimentally [88], and theoretically [89, 64]. Such structures being the building block of the DNA technology have been the subject of different studies. The assembly of a structure similar to the Holliday junction was studied using FRET technique [83], allowing to evaluate the thermodynamic of formation of the structure. FRET technique was also used to study more in detail the thermodynamic of attachment of preformed monomer structure composed of several strands into a dimer. It allowed to evaluate the effect of the flexibility of the monomer structure on the dimer formation [90].

To study our structure, three different sets of staples based on the same structure were chosen to quantitatively evidence cooperative effects during the binding of the staples. In the first two sets, the two staples (B1 and B2) have very different compositions: the sequence of B1 only contains A or T nucleotide whereas B2 only contains G or C nucleotide. Accordingly, their melting temperatures are far apart, respectively 57 $°C$ and 91 $°C$. This allows to differentiate the two staples in the melting curve. The third set has two staples, B1m and B2m, designed with chemical sequences different enough to avoid mispairing with the 64b template B0 and with close melting temperatures (respectively 77 $°C$ and 80 $°C$) as their AT/GC ratio are similar.

In this section, for all the experiments the concentration of each strand is 1μM. The buffer used is the one of classical oDNA: a TAE buffer with magnesium ([Mg] = 12.5 mM). The experiments have been realised in 200 µL UV cuvets and to prevent evaporation during the annealing-melting experiments, a paraffin oil invisible to the UV spectra has been dropped on the top of the solutions.

3.3.1 Experimental datas

3.3.1.1 Effect of the position:

The topology of the binding is illustrated in (Fig. 3.4): each staple contains two contiguous parts, 16b long, that bind to the scaffold. In (Fig. 3.4 a), B1 is in the 'outer' position, B2 is in the 'inner' position. The difference between these two ways of binding can be further stressed by considering what happens when only half of the staple is hybridized. In the 'outer' position, the unbound parts of the staple and the scaffold are located on the same side of the bound moities (Fig. 3.4 b). In the 'inner' position, the unbound parts are on the opposite sides (Fig. 3.4 c). Besides the existence of this entropic hindrance, the inner position requires that double-helical domains stay in close contact, which could result in additional instability. Motivated by these considerations, we have investigated two different cases:

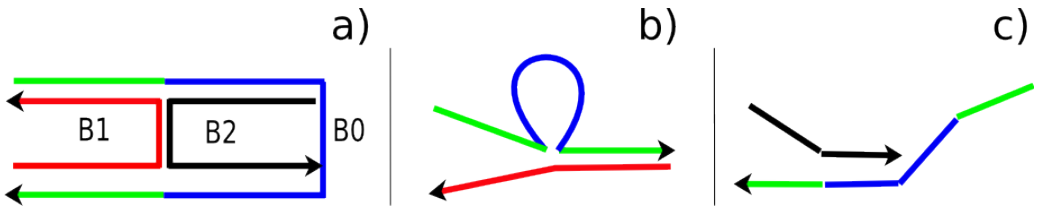


Figure 3.4: (a) Schematic representation of the connectivity of the small origami.(b) B1 staple is in the 'outer' position,(c) B2 staple in the 'inner' position. (b) and (c) show that the binding of staples in the outer (b) or inner (c) positions are very different.

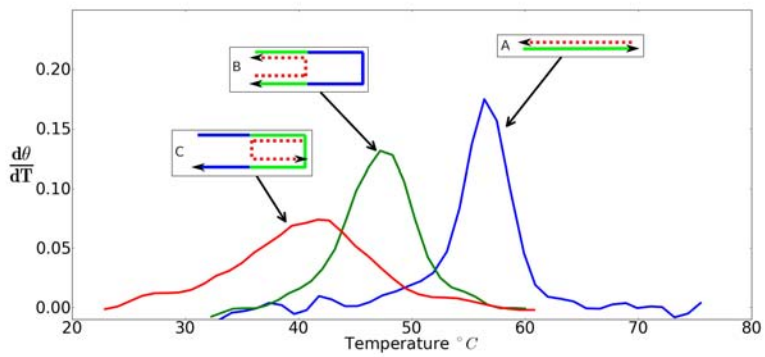


Figure 3.5: Experimental data on the folding of B1(AT) in the absence of B2(GC) in the different positions. A) dsDNA. B) outer position. C) inner position

- B2 is located in the inner part, B1 in the outer part.
- B2 is located in the outer part, B1 in the inner part.

In (Fig. 3.5) we report the derivative of the melting curves that show the behaviour of the B1 strand for different configurations. Fig. 3.5 A corresponds to the melting curve of B1 with its complementary $\overline{B1}$: it shows a maximum peak at 57 $°C$. We analyse first the case where B1 is outer. Without the staple B2 (Fig. 3.5 B), the structure produces a loop or bulge, as described in [27] that shifts the free energy towards a lower value and decreases the melting temperature to 48.5 $°C$. In the inner position the staple hardly folds with the B0 template. The maximum peak is located at 41.5 $°C$, 15 K lower than the value of the dsDNA and with a half width as large as 12 K.

Similar trends can be observed for the B2 staple. B2$\overline{B2}$ (Fig. 3.6 A) has a maximum peak at 91$°C$ while when B2 is in the inner position (Fig. 3.6 B) the peak is located at 83 $°C$. When B2 is in the outer position (Fig. 3.6 C) the peak is located at 86.5$°C$ and is slightly narrower.

Therefore, the same differences between the inner and outer positions are observed whatever the chemical sequence involved.

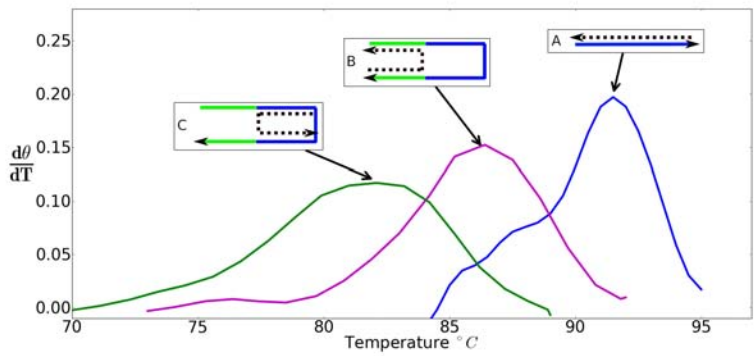


Figure 3.6: Experimental data on B2(GC) without B1(AT) in the different positions A) dsDNA. B) outer position. C) inner position

3.3.1.2 Effect of the cooperativity between strands:

When B1 and B2 are both present two separated events appear on the melting curve (Fig. 3.7). The events are clearly separated and this allows the identification of the different staples. In the following we report the events individually in order to highlight the cooperative effects.

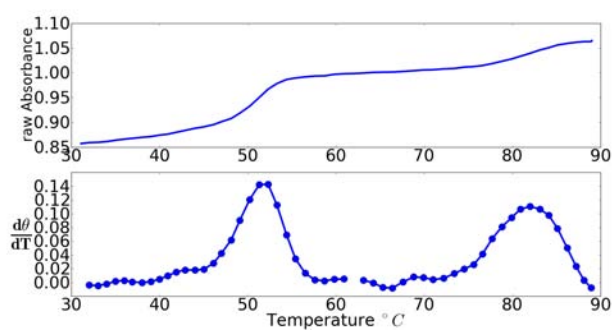


Figure 3.7: Top: Absorbance curve. Bottom: derivative of the absorbance curve showing two events

We first study the case where B2 is in the inner position. The folding of B2 at 83 $°C$ is not modified by the presence of B1. By contrast, B1 folds at a temperature higher (Fig. 3.8 C) than when it is alone (Fig. 3.8 B), with a maximum peak at 51.5 $°C$.

Therefore, the inner staple B2 helps the pairing of the staple B1 located in the outer part by suppressing part of the entropic penalty related to the bulge formed by the scaffold. When B1 is in the inner part and when the B2 staple is added (Fig. 3.9 E), the pairing of B1 is also stabilized with a maximum peak much higher located at 50 $°C$. These experimental results support the evidence of a strong correlation between the two strands that appears when B2 is added: the presence of B2 helps the folding of B1 whatever its location. Moreover, these results also show that the location of the strand is of importance, the inner location being much less favourable. The origin of this difference is not obvious. It may in part be the result of an entropic penalty larger than the one a bulge induces, as it occurs when B1 is located in the

outer domain, but it may also be the consequence of the energy cost of a local curvature the inner strand imposes to the B0 template.

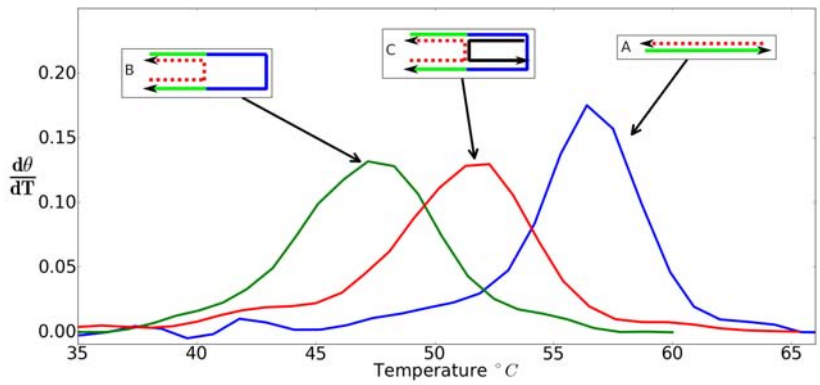


Figure 3.8: The derivative $d\theta/dT$ reported in the three cases corresponds to the folding of the dotted staple. Experimental data on the folding of B1(AT) in the **outer** position. A) dsDNA. B) B1 outer alone. C) B1 outer and B2 inner

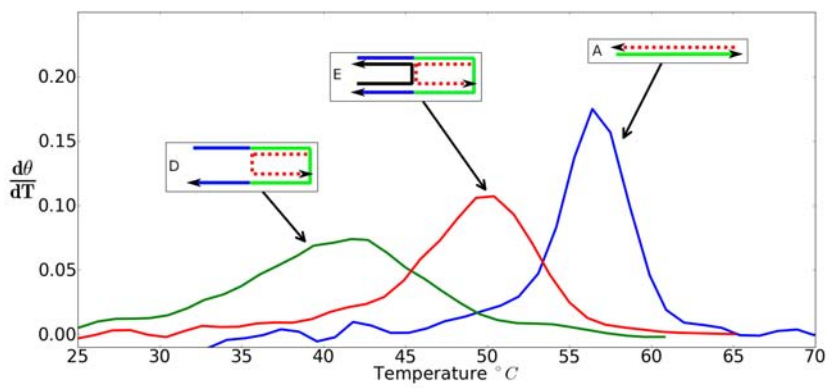


Figure 3.9: The derivative $d\theta/dT$ reported in the three cases corresponds to the folding of the dotted staple. Experimental data on the folding of B1(AT) in the **inner** position. A) dsDNA. B) B1 inner alone. C) B1 inner and B2 outer

The experimental results obtained with the B1m-B2m ($T_m = 77\ °C$ and $T_m = 80\ °C$) set of staples are shown in (Fig. 3.10 and 3.11). As the T_m of the two staples are close, when the two staples are present, only one global event is observable on the melting curve. Again, the same trends as before are observed but with less pronounced effects. When compared with the results obtained with the previous structure, the temperature shifts and the increase of the half width are smaller with respect to the melting curves of the double strands B1m$\overline{B1m}$ and B2m$\overline{B2m}$.

A correlation effect is also noticeable, and is now observed for both strands B1m and B2m. When the staple B1m is in the solution, the staple B2m, which folds at a higher temperature, shows a narrower peak at a maximum 1.5 $°C$ higher than when it is alone. Therefore partial folding of the staple B1m helps the folding of the staple B2m. Similarly to what was observed in the previous case, the correlation effect is even more effective when we consider the influence of the B2m staple on the folding process of the B1m staple (Fig. 3.10 D) shifting the T_m from 67 $°C$ to 72 $°C$. Thermodynamic values were extracted from the melting curves and are listed in Appendix C

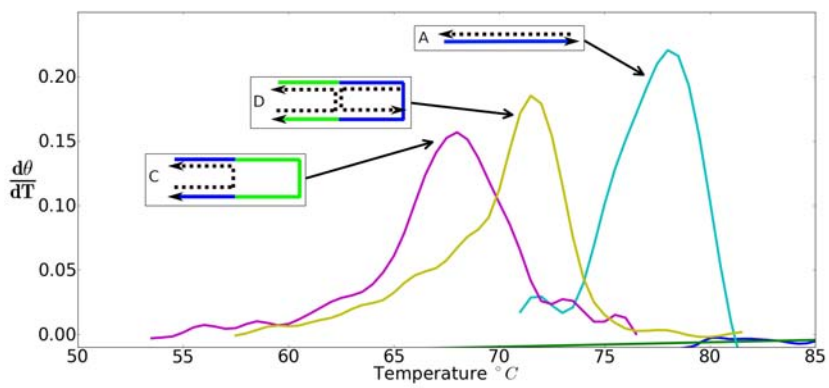


Figure 3.10: Experimental data on the folding of B1m with and without B2m. A) dsDNA C) B1m outer alone. D) B1m outer B2m inner folding in a single event.

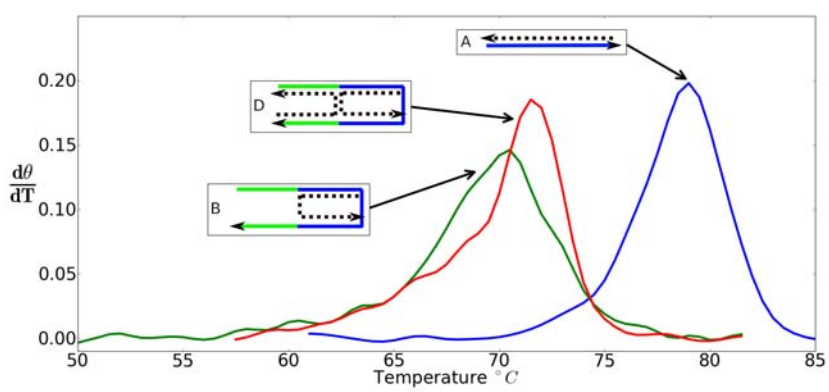


Figure 3.11: Experimental data on B2m with and without B1m. A) dsDNA B) B2m inner alone. D) B1m outer B2m inner folding in a single event.

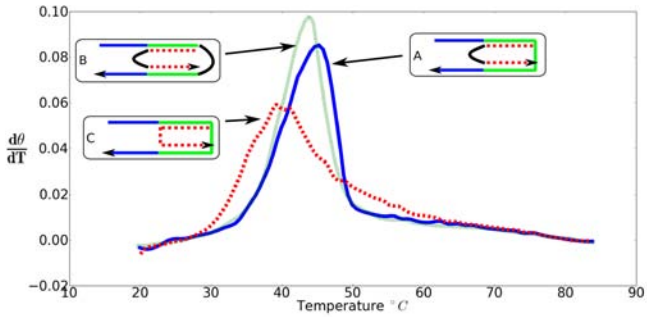


Figure 3.12: A) B1 inner with adaptator . B) B1 inner with adaptator and scaffold with adaptator. C) B1 inner.

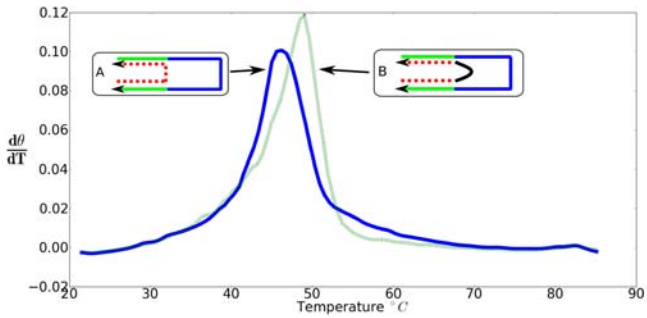


Figure 3.13: A) B1 outer. B) B1 outer with adaptator.

3.3.1.3 Effect of adding an adaptor.

As explained in section 2.3.1.1 repulsive interactions take place between adjacent dsDNA in the structure. We decided to modify crossover junctions by adding four non-complementary bases (called adaptor) to increase the distance between the dsDNA parts. The effects on the melting temperature are gathered in (Fig. 3.12 , 3.13). When inserted, the adaptor is drawn in black in the middle of the staple.

The presence of the four bases (GGGG) added to the inner AT strand seems to stabilise the complex (Fig. 3.12 A). The regular inner position (Fig. 3.12 C) is less stable by ≈ 5 °C. Adding an adaptor to both strands (Fig. 3.12 B) seems to have little effect. In the outer position the adaptor stabilises the complex by 3°C (Fig. 3.13 B). Adding an adaptor seems to stabilise the structure in all cases. However it is possible that the free G bases of the adaptor interact with the part of the scaffold that contains GC bases (Drawn in blue in all figures)).

The experimental data on soDNA lead to two results. First the position of the staple with respect to the scaffold is important. Second, the strands exhibit a strong cooperative behaviour. We decided to develop a model based on these experimental results, in order to see if they are predictable , and also to explore other positions and parameters.

3.4 Thermodynamic model on a lattice for small origami

3.4.1 The model

The Poland-Scheraga (PS) model describes double stranded DNA as a sequence of either helical structures or coiled loops [31]. The free energy $\Delta F_{NN} = \Delta H_{NN} - T\Delta S_{NN}$ to form a helical structure of m stacked pairs of neighbouring base pairs is given by the nearest-neighbour model [91]

$$\begin{aligned} \Delta H_{NN} &= \sum_{i=1}^{m} \Delta h_{NN,i} \\ \Delta S_{NN} &= \sum_{i=1}^{m} \Delta s_{NN,i}, \end{aligned} \tag{3.4}$$

where $\Delta h_{NN,i}$ and $\Delta s_{NN,i}$ are respectively enthalpy and entropy terms associated with one of the 10 possible pairs of bases. To this sequence dependent terms, it is customary to add initiation free energies ΔG_{ini} for each of the two ends of the helical segment.

The PS model describes the free energy change involved in a loop formation using the concepts from polymer physics. The corresponding term in the partition function for a loop of length N is written as

$$Z_{loop}(N) = \sigma N^{-c} \exp \sum_{i=1}^{N} \beta(\Delta h_{NN,i} - T\Delta s_{NN,i}). \tag{3.5}$$

The exponent c depends on the interaction of the loop with the rest of the chain, it is equal to ~ 1.75 for non-interacting self-avoiding loops [32]. The prefactor σ is the so-called *cooperativity factor* for bubble opening.

The DNA structures we wish to consider involve more than two complementary strands. For instance, in the case of small origamis, three strands s1,s2 and s3 hybridize partially. Bubbles can form between s1 and s2 as well as between s2 and s3, and the steric interactions between these two loops need to be taken into account explicitly. Furthermore, thermodynamic properties of structures such as helical dsDNA can be easily computed because recurrence relations exist for sequences of increasing length. Such relations become more complex for branched structures such as those considered here.

Recent theoretical works have adressed the process of formation of similarly branched structures such as Holliday junctions [92] and Yurke tweezers [93]. However, as demonstrated by these studies, the numerical modelling of the dynamics and/or thermodynamics of such objects remains challenging. Here, we introduce a coarse-grained model that is able to simulate cycles of annealing-melting of non standard structures, but also the self-assembly of DNA duplexes.

In this model, each base is represented by an occupied site of a 3D triangular lattice. A single strand is modelled as a self-avoiding walks (SAW) in this lattice. Restricting the possible

locations to the sites of a triangular lattice, together with an appropriate free-energy model, has been shown [94] to provide with accurate RNA secondary structure predictions, outperforming other leading RNA pseudoknot prediction methods. It can be concluded that the restriction to a grid is not incompatible with, at least, an approximate description of nucleic-acid statistical properties.

The energy model is that of [32]. More specifically, four independent terms need to be computed for each configuration:

- The sequence of stacked base pairs determines the ΔG_{NN} contribution (eq. 3.4). Watson-Crick pairs are only computed if two complementary sites are neighbours and belong to different chains. Two consecutive Watson-Crick pairs are stacked if the bases belong to two antiparallel chains (the generalization to three chains leads to unphysical globular structures). Note also that stacking interactions between bases that belong to the same loop are not considered.

- an interfacial free energy is associated with each interface between single and double stranded segments. This term can be related to the cooperativity factor σ.

- the number of free ends of double stranded segments determines the ΔG_{ini} contribution.

- to each kink in a double stranded segment it is associated a constant enthalpic term. This term ensures a correct bending rigidity.

The link between the PS model and the present lattice model is made through the entropy attributed to each base pair. This parameter was determined [32] by making equal the contribution of a loop in the lattice model and in the PS model (which is given by eq. 3.5). This equality will ensure that the thermodynamic description of both models is the same. One possible way to ensure this equality is to assimilate a loop to a SAW that starts and returns to the origin. In the triangular 3D grid we consider, the number of such walks of length N has the form $\exp(-1.8668 + 1.8774N)$ as $N \to \infty$. Hence, the requirement that the partitions functions of the lattice and the Poland-Scheraga models are equal implies that $e^{-\beta T S_{PS}} = e^{-\beta T S_{Latt}}\mu$, with $\log \mu = 2 \times 1.8774 = 3.7548$, so that $S_{latt} = S_{PS} + \log(\mu)k$ (k is the Boltzman constant). However, this method is not accurate enough to take into account the contributions of short loops. Instead, we will consider μ as a parameter of the model that can be adjusted in order to fit the melting curves obtained for simple dsDNA with the methods of [95] (cf. Fig. 3.16 below). In the following, the value $\log \mu = 4.3$ will be used.

To sample the different configurations, we consider a Monte-Carlo method with two types of moves: pull moves [94] applied to a single strand and simultaneous pull moves applied to two strands. A pull move proceeds as follows (Fig. 3.14): randomly, a site is moved to a nearest-neighbour empty site. In doing so, it is possible that some neighbour sites of the chain become disconnected. To fix this, recursively each of these disconnected sites is moved to the position

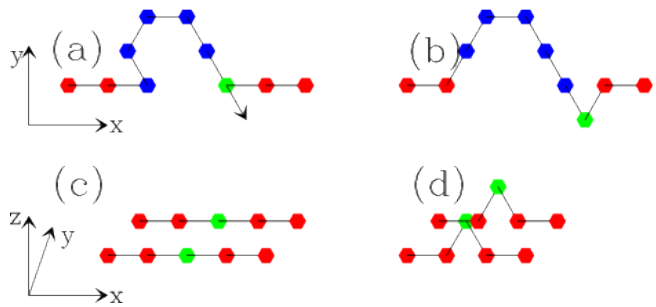


Figure 3.14: Schematic representation of a pull movement. (a) before the pull. (b) after a pull on a single strand. Notice that, in this case, the motion is planar only for the sake of simplicity. (c) and (d) illustrate a simultaneous pull on two strands. (c) before the pull. (d) after the pull.

of the previous (along the chain) site, until all the sites in the chain become nearest-neighbours once again. It can be shown that pull moves are both reversible and complete (any configuration can be transformed to another configuration by a sequence of pull moves). They generalize flip moves [94] and have a semi-local property: the average number of relocated elements is only a small constant that depends on the length of the chain to be 'pulled'. To the set of pull moves we added the possibility to perform simultaneous pull moves on two single strands that are at least partially hybridized. More precisely, if a pull move is done on a site i of the strand S, and site i is hybridized with site i' of strand S', a pull is tried also on site i'. This pull can be either in the same direction as the pull on i (in which case a kink is introduced simultaneously on strands S and S'), or in the opposite direction, in which case a bubble forms.

From a numerical point of view, the Monte-Carlo sampling is enhanced by a replica exchange algorithm: a fixed number N_{MC} (typically $N_{MC} = 10$) of MC simulations is run simultaneously, with temperatures $T_i, i = 1, ..N_{MC}$, separated by $\Delta T = T_{i+1} - T_i = 2K$. The probability to exchange configurations between contiguous simulations is such that the Boltzman distribution with corresponding temperature T_i is the asymptotic probability distribution for each simulation [96].

In these simulations, we disregard concentration effects. More precisely, in the simulation of a structure formed by n_s ssDNA, only *connected* structures will be considered. Notice that this restriction does not imply that any ssDNA is connected to any other ssDNA. Rather, for each strand there is at least one WC pair with some other strand.

The usual way to characterize the formation of a DNA double helix is to compute the number of stacked WC pairs. For instance, the measurement of UV absorption gives access to this number, because UV absorption is lower for hybridized basis (hypochromic effect). Thus, for each configuration the *degree of pairing (d.o.p.)* $\theta(T)$ will be computed: it is the number of stacked WC pairs (a WC pair is not stacked if it is flanked by two unpaired bases). This definition of $\theta(T)$ is different of the one introduced in the beginning of this chapter. Indeed in the experimental analysis we supposed that all the bases fold and $\theta(T)$ is normalised between

0 and 1. In all the subsections of Sec 3.4, $\theta(T)$ is not normalised and varies between 0 and n, the maximum number of bases folded.

3.4.2 Results

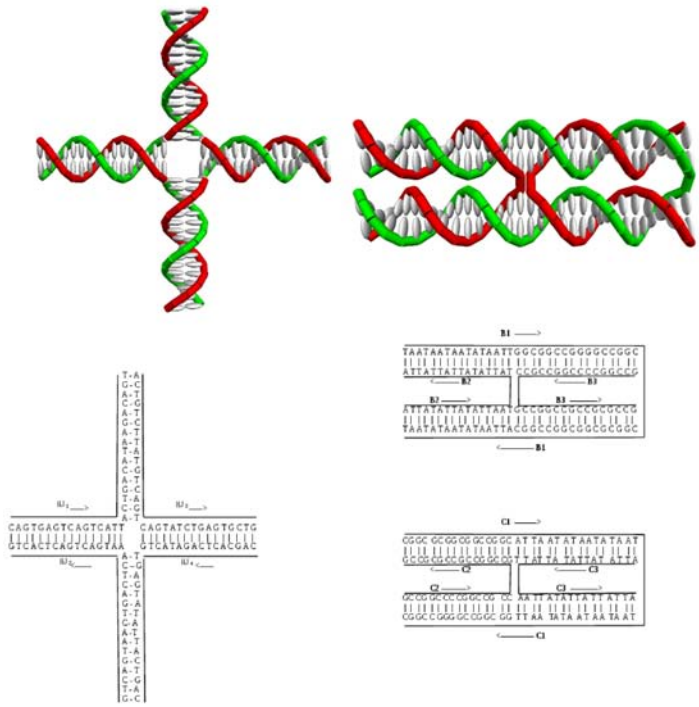


Figure 3.15: Schematic representation of the DNA structures considered in this section. Up: (left) Holliday junction (right) small origami. These two schemes show the helicity of an ideal B-DNA structure. Helicity is not taken into account in the lattice model. Down: (left) Sequence and connectivity of the small origami (see text) (right) Sequences and connectivity of the two small origamis.

We have considered three sets of topologically different structures (cf. Fig. 3.15). *Double stranded* (DS) chains consist of two antiparallel chains that are Watson-Crick complementary, with exactly the same number of sites. *Holliday junctions* (HJ) are formed of four chains of equal length that, in a first approximation, form a cross-shaped structure. *Small origami* (soDNA) that corresponds to a modification of the usual HJ where two single stranded chains are merged in a single chain that is necessarily bent.

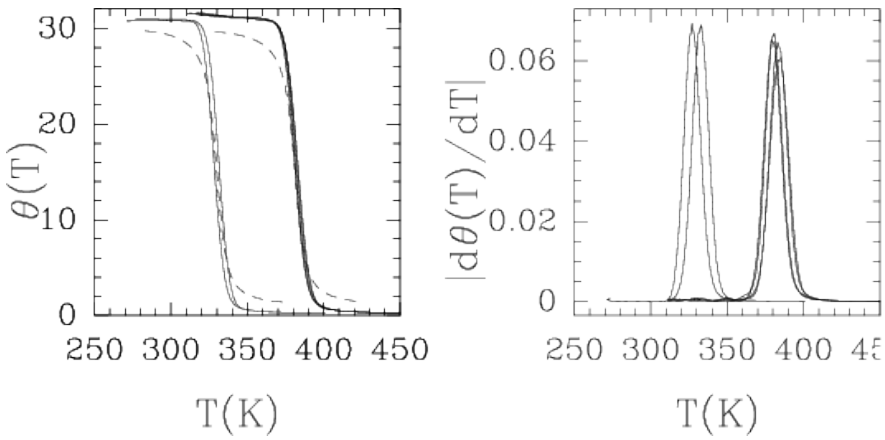


Figure 3.16: (a) $\theta(T)$ versus temperature for the two short strands (AT and GC rich) that compose the soDNA of type B (see text). The dashed lines are computed with the partition function algorithm [95]. (b) Derivative of the d.o.p. $|d\theta(T)/dT|$ for the annealing and melting curves. The slight hysteresis observed in $\theta(T)$ explains the existence of two peaks for each $|d\theta(T)/dT|$. The position of the maximum determines the melting temperature: T_m(AT rich) = 330K, T_m(GC rich) = 380K.

3.4.2.1 Double stranded DNA

Here we consider the simplest DNA structures: double stranded DNA. In Fig. 3.16 are represented the d.o.p. $\theta(T)$ for the two short strands (respectively rich in AT and GC) that compose the soDNA of type B (see below), each strand folding with its complementary strand. These two sequences are 32 bp long and have very different melting temperatures. This is illustrated in Fig. 3.16(b), where the absolute value of the derivative $|d\theta(T)/dT|$ is shown. The data of Fig. 3.16 have been computed with 10^7 MC steps and still show some hysteresis (around 2K difference between the heating and cooling cycles for a given value of $\theta(T)$). As a comparison, we also report the melting curves obtained with the partition fonction method [95]. Both results agree reasonably in the transition regions and deviate somewhat outside. This difference could be due to the way the loop entropy is computed in the partition function method [95], wherein the asymptotic behaviour also describes the contribution of short loops.

3.4.2.2 Holliday junction

Holliday junctions [97] are mobile junctions between four strands of DNA. Here, somewhat abusively, we also call Holliday junctions stable nonmigrating analogs of the Holliday recombinatorial intermediate. Analog structures have been studied experimentally [66]. In this work, the authors performed calorimetric and optical measurements to characterize the melting transitions of the junction and of each of its arms. They reached the conclusion that the sum of

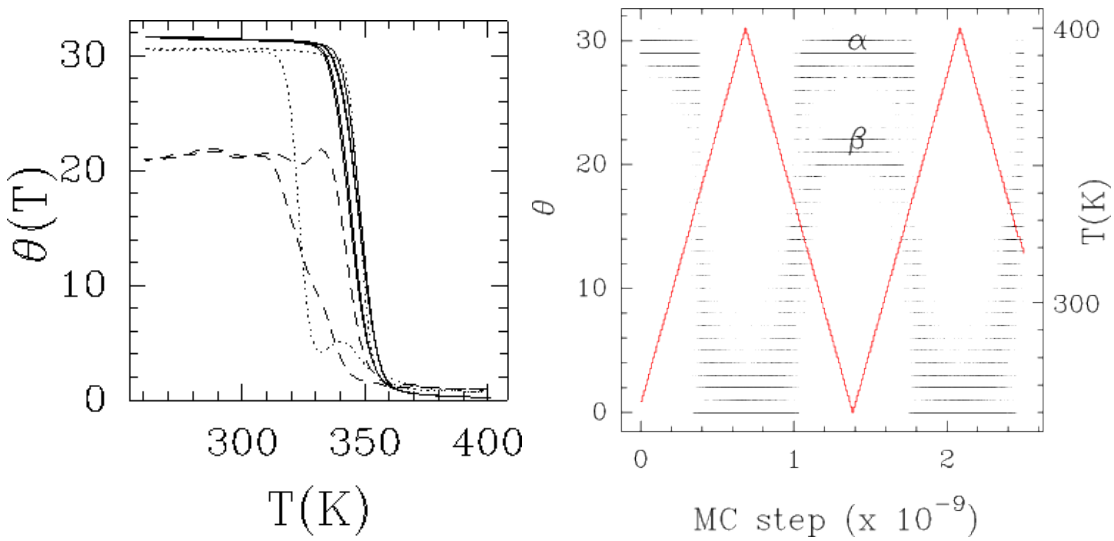


Figure 3.17: (Left) Comparison of the computed melting curves obtained for HJ_1 in presence of its complementary strand (continuous lines) or in the presence of the other strands that form the Holliday junction HJ (dashed line). Similar results are obtained for HJ_2, HJ_3 and HJ_4 (data not shown). The dotted line is obtained from the HJ simulation by filtering the structures to keep only those with the highest d.o.p.;(right)d.o.p. for the HJ_1 as a function of the MC step. The starting point is the ideal HJ structure. After a cycle of melting and quenching, a second state appears (around 1.4 10^9 MC step), showing a lower value of d.o.p. The red line indicates the temperature associated with each MC step (vertical scale on the right y axis).

the transition enthalpies of each arm is very similar to the transition enthalpy of the junction. This suggests that the formation of the junction does not significantly perturb the B-helical conformation of the individual arms.

Our HJ is formed of four different branches, named HJ_1, HJ_2, HJ_3 and HJ_4, each 32b long, with sequences generated randomly:

HJ_1: CAGTGAGTCAGTCATTACTGACATAAGACAGT
HJ_2: GTCAGTAACTGACTCAAATGACTGACTCACTG
HJ_3: ACTGTCTTATGTCAGTCAGTATCTGAGTGCTG
HJ_4: CAGCACTCAGATACTGTGAGTCAGTTACTGAC

Successive cycles of melting-annealing starting from the ideal cross-shaped form of this junction show that 80% of the native contacts are recovered in this simulation. Previous simulations [92] also considered a similar structure, half of the size of the present one. Monte-Carlo simulations starting from fully disorganized (eventually disconnected) structures succeeded to show how the cross-shaped structure is formed, although with a rather low yield. The same authors also used umbrella sampling in order to evaluate the free energy of formation. The critical difference with the work in [92] seems to be the connectivity constraint that we used.

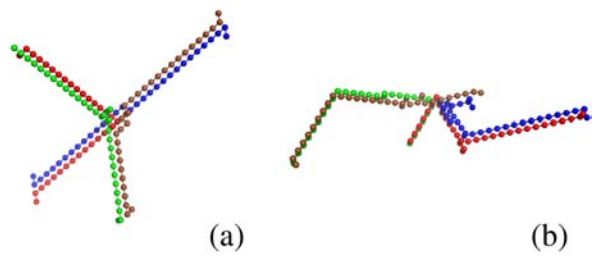


Figure 3.18: Typical low temperature configurations of HJ.(a) and (b) correspond respectively to the configurations of HJ noted α and β in Fig. 3.17(right).

In Fig. 3.17(left) is shown the average $\theta(T)$ of component HJ_1 compared to the same quantity obtained for the separate strand in presence of its complementary (similar results are obtained for HJ_2, HJ_3 and HJ_4). Due to the limitations of our sampling procedure, hysteresis is apparent in the simulation of the annealing of the HJ structure. The values of $\theta(T)$ at low temperatures show that 80% of the native contacts are recovered. It is interesting to note that the simulation of the HJ actually yields the ideal HJ structure for a significant number of MC steps. The fact that, on the average, only 80% of the native contacts are recovered, is related to the existence of an alternative configuration where HJ_1 does not hybridize with HJ_3 nor HJ_2 with HJ_4.

This is apparent in Fig. 3.17(right), where the non averaged values of the d.o.p. are reported as a function of the MC step. The data obtained for the low temperatures clearly display two different clusters, noted α and β. Configurations illustrating each of these clusters are shown in Fig. 3.18.

The comparison between different speeds of cooling shows that the probability to observe these non ideal structures diminishes with the speed of cooling. The parameters we used here (20×10^6 Monte-Carlo steps for each temperature, with $\Delta T = 2$ K between two points of the annealing-melting simulations) are the best combination we obtained, taking into account the computing limitations. Thus, the cluster β should be considered as a spurious configuration only related to the limitations of the algorithm we have used to generate the structures. Filtering these spurious structures is possible by clustering the conformations and rejecting those that belong to the group with the lowest value of the d.o.p. (dotted line in Fig. 3.17(left)). The resulting melting-cooling curve is similar to the one of the DS (more than 90% of the natif contacts are recovered), albeit shifted by almost 10K. Overall, the effect of the presence of the other strands in the HJ structure is a destabilization.

These results can be compared with the experiments of [66], where the difference between the thermodynamic properties of each separate strand and the whole HJ structure could be measured. In [66], the authors showed by calorimetric methods that $\Delta H(HJ) \sim$

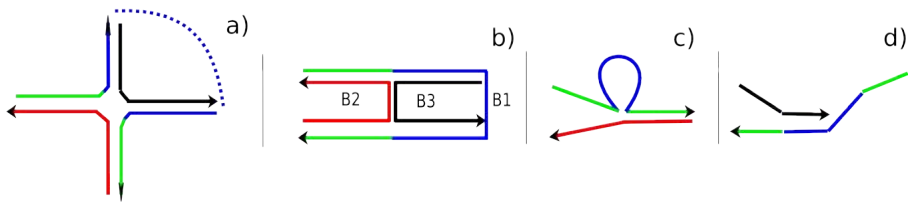


Figure 3.19: (a) Schematic representation of the connectivity of the Holliday junction the dotted line show the connection with the bent Holliday junction. (b) Schematic representation of the connectivity of the bent Holliday junction. The template strand B1 is colored in green and blue to stress the fact that different parts hybridize with different staples. (c) Intermediate situation where B2 staple is in the 'outer' position and only half of it is hybridized with B1. (d) B3 staple in the 'inner' position. (c) and (d) show that the binding of staples in the outer (c) or inner (d) positions are very different.

$0.9(\Delta H(HJ_1)+\Delta H(HJ_1)+\Delta H(HJ_3)+\Delta H(HJ_4))$, indicating also a slightly weaker stability of the Holliday junction compared to the separate strands. Assuming a base-independent ΔH for base opening, the experimental 0.9 factor is compatible with the 90% recovery of the native contacts in the simulations. In our case, the existence of hysteresis makes difficult to push further the analysis in order to get a more accurate estimate of $\Delta H(HJ)$.

3.4.2.3 modelling small origami

Using the same algorithm, let us now address the formation of the small origami that can be seen as bent Holliday junctions (Fig. 3.19). When two of the strands of a HJ structure merge, a small origami is obtained. This introduces some asymmetry on the system, as the two short strands are not expected to behave in the same way. The strands used in this section are similar to the ones presented in the experimental section, however, in the AT staple a G and a C base were included in order to diminish the potential mispairing.

If the inner staple binds first, the hybridization of the whole inner is expected to take place in two steps: when half of the staple attaches the scaffold, the other half is going to be kept away from the region of the long strand where it is expected to bind (Fig 3.19 d). Besides, this second step of pairing could be hindered by curvature effects or steric hindrances. If the outer staple binds first, there is a bulge in the structure which implies an entropic penalty. In both cases (inner or outer staple), the binding is expected to happen at a temperature lower than the melting temperature of the isolated strand. Let us now consider the binding of the second strand. In the outer position, the situation is not very different from the formation of a double stranded linear DNA, because the bulge is constrained by the inner strand. In the inner position, curvature effects are expected again.

Let us now be more specific and consider the following two small origamis B and C composed respectively of three strands (B_1, B_2, B_3) and (C_1,C_2,C_3) with sequences:

B_1: TAATAGTAAAATATTTGGCGGACGGGGCCGGCCGGCGCGGCGGTCGGCATTTATATACTATAAT

B_2: ATTATAGTATATAAATAAATATTTTACTATTA = C_3

B_3: GCCGGCCCCGTCCGCCGCCGACCGCCGCGCCG = C_2

C_1: CGGCGCGGCGGTCGGCATTTATATACTATAATTAATAGTAAAATATTTGGCGGACGGGGCCGGC

In the preceding terminology, B_2 and C_2 are outer, B_3 and C_3 are inner. Owing to the large difference in melting temperature between the GC and AT-rich strands (respectively 380K and 330K), both B and C fit well in the above simple description.

In Fig. 3.20 we show the non averaged d.o.p. as a function of the MC step: for each MC step, a point corresponding to the value of $\theta(T)$ is drawn. The more populated structures appear therefore as thicker lines. Starting from completely disorganized structures (very low values of θ, high temperature), the soDNA recovers its initial structure upon annealing: both strands show configurations where d.o.p. is close to their length. However, even for the lower temperature values, the ideal structure where $\theta(T)$ reaches its maximum is in equilibrium with less folded structures. Thanks to the large difference in melting temperature between the two strands, the location of the GC strand in the soDNA structure is found to be of primary importance and determines the whole double stranded structure of the soDNA final state. When the GC-rich strand is located at the inner position (soDNA B), two main types of configurations (noted as α and β in Fig. 3.20) are present, with average values $\theta^{\alpha}_{inner} \sim 27$ and $\theta^{\beta}_{inner} \sim 20$. The corresponding values of the associated AT-rich strand are very similar: $\theta^{\alpha}_{outer} \sim \theta^{\beta}_{outer} \sim 30$. Representative configurations of α and β are illustrated in Fig. 3.21(a) and (b), respectively: α configurations show a very uniform distribution of hybridization, β configurations show more deficient patterns of hybridization in one of the arms of the GC-rich strands.

When the GC strand-rich strand is in the outer position (soDNA C), two main types of configurations (γ and δ, illustrated in Fig. 3.21(c) and (d)) can also be distinguished, with very similar values for the d.o.p. The main difference between the inner and outer positions of GC-rich strands comes from the less frequent conformations, with values of the d.o.p. much more spread out for the inner than for the outer positions (compare frames (b) and (d) of Fig. 3.20). This is reflected in the temperature variation of the average d.o.p. $\theta(T)$. In the annealing process, the outer position for the GC-rich strand appears to be more stable than the inner position: the equilibrium value of $\theta(T)$ is already reached for $T = 350K$ for the outer, and only at $T \sim 320K$ for the inner. Notice that the same conclusions can also be reached for the AT-rich strand (see the variations of the thin blue lines in Fig. 3.22).

From these two simulations, it can be concluded that the inner position is a destabilizing factor in the hybridization of DNA strands. This conclusion holds both in the presence or in the absence of another strand. However, these simulations also reveal a cooperative effect. More precisely, the hierarchical hybridization that takes place in the formation of B and C (GC-rich attaches before AT-rich) is at the origin of a stabilizing effect. To stress this point, we have

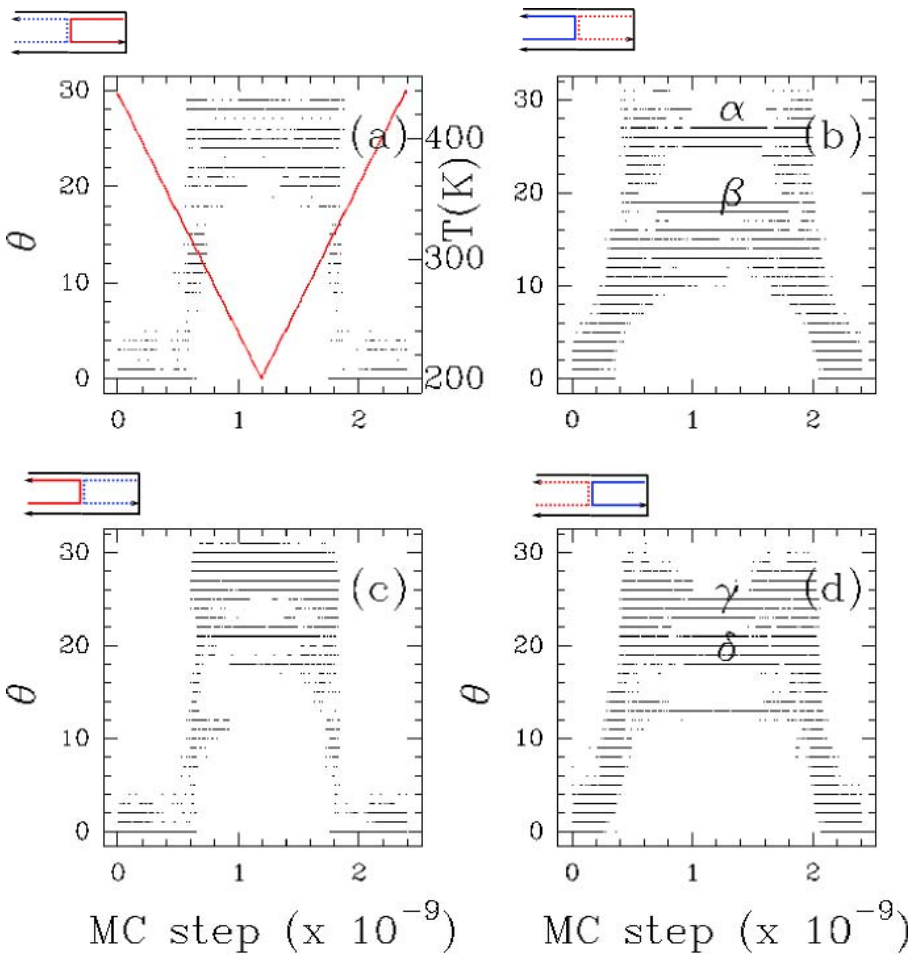


Figure 3.20: In all the plot the staple of interest is represented with dot (blue for AT and red for GC). Non averaged d.o.p. as a function of the MC step for the two short strands that compose B and C. (a) AT-rich strand of soDNA B. (b) GC-rich strand of soDNA B. (c) AT-rich strand of soDNA C. (d) GC-rich strand of soDNA C. The temperature associated to each MC step is represented by the red line in (a), the corresponding scale being given by the right coordinate axis. In the soDNA schema (b) and d)) the AT staple has been represented, however when GC folds the AT staple is not attached to the scaffold.

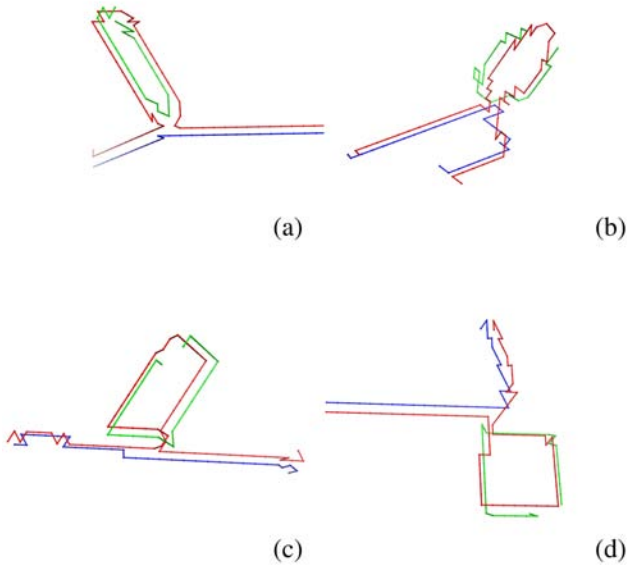


Figure 3.21: Typical low temperature configurations of the soDNA.(a) and (b) correspond respectively to the configurations of soDNA B noted α and β in Fig. 3.20. (c) and (d) correspond respectively to the configurations of soDNA C noted γ and δ in Fig. 3.20. In all these figures, the inner strand is in green and the outer strand is in blue, the template strand is in red. Notice that these are 2D projections of structures in 3D. This is why the arms of (a) seem to be of different length.

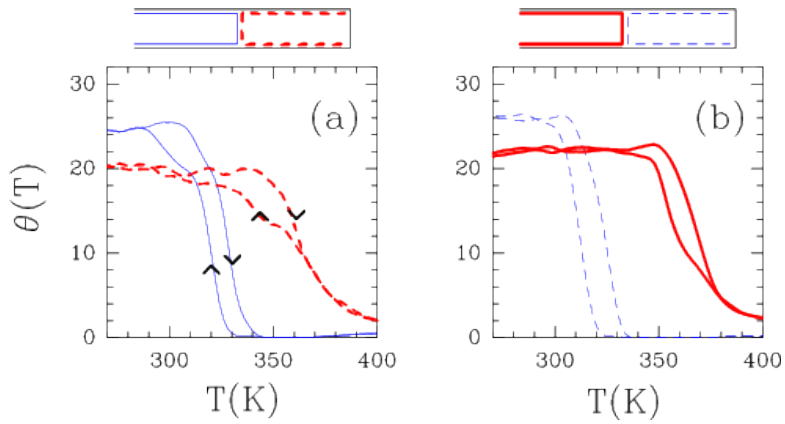


Figure 3.22: (a) and (b): Average d.o.p. as a function of the temperature for the two short strands that compose B and C. Thick (red) lines represent the GC-rich strand, thin (blue) lines the AT-rich. The strands in the inner position (resp. outer position) are represented by dashed (resp. continuous) lines. (a) B soDNA (b) C soDNA. The two schemes above the graphs show a schematic representation of the position of each strand in the soDNA. The arrows in (a) indicate how the temperature varies along the annealing-melting curves.

considered the formation of structures without the presence of the GC-rich strand, e.g. the B and C structure with only the AT-rich strand in the inner or outer place: B_1 and B_2 or C_1 and C_3.

Fig. 3.23(a) shows that the outer AT rich strand is only weakly stabilized by the presence of the GC-rich strand in the inner position (average shift of $\sim$ 2K when the GC-rich strand is present with no significant modification of the annealed values of the d.o.p.). The opposite situation is illustrated in (Fig. 3.23(b)) where in the absence of the outer GC-rich strand, the pairing is mostly incomplete with only half of the strand (in the average) attached. Therefore, with cooperativity assimilated to an increase in $\theta(T)$, the supression of the loop entropy (soDNA of type B) has only a mild cooperative effect on the hybridization in the outer position (Fig. 3.23(a)), while the cooperativity becomes very pronounced when a staple has to bind in an inner position with the outer part being frozen by another staple (Fig. 3.23(b)).

In order to check the generality of these conclusions, let us now consider a situation where the melting temperatures of the two staples are similar which correspond to the B1m-B2m case (The strands are similar in GC content but not identical to the experimental case). The soDNA structure D is also formed by one long (D1) and two short (B1m and B2m) strands:

D1: CAGTGAGTCAGTCATTACTGACATAAGACAGTCAGCACTCAGATACTGTGAGTCAGT-TACTGAC

B1m: GTCAGTAACTGACTCAAATGACTGACTCACTG

B2m: ACTGTCTTATGTCAGTCAGTATCTGAGTGCTG

Figure 3.24 compares, for the outer strand Fig. 3.24(a) and inner strand fig 3.24(b), the

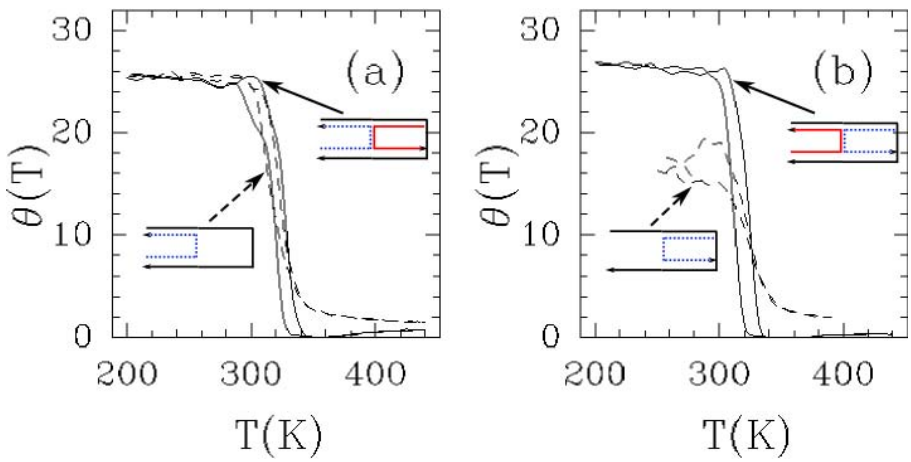


Figure 3.23: d.o.p. $\theta(T)$ as a function of the temperature for the AT-rich strand of B, B_2, hybridized with B_1 in the outer position (a) and with C_1 in the inner position (b). The continuous line refers to a simulation where both B_1, B_2 and B_3 (a) (resp. C_1, C_2 and C_3 (b)) are present. Dashed lines refer to simulations where only two strands are present: B_1 and B_2 (a), C_1 and B_2 (b).

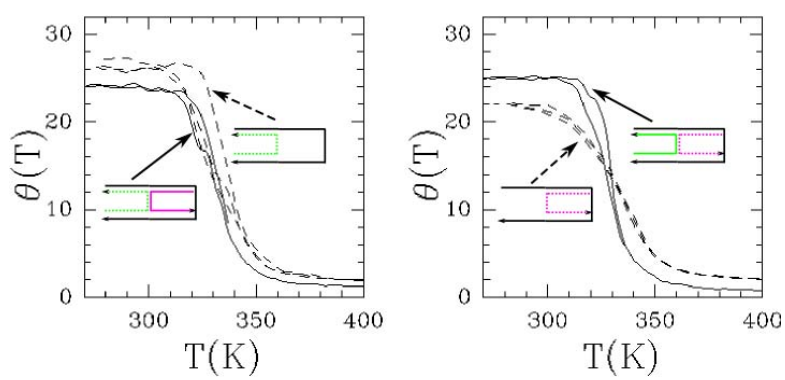


Figure 3.24: Melting of the soDNA D structure: average d.o.p. $\theta(T)$ as a function of the temperature for the two short strands (a) outer, B1m (b) inner B2m. The continuous lines refers to simulations where D1, B1m and B2m are present. The dashed line refers to simulations where only two of them are present: (a) D1 and B1m (b) D1 and B2m.

average d.o.p. obtained in the presence (or not) of the other strand. The outer strand B1m is slightly destabilized by the presence of the inner strand B2m, the maximum d.o.p. decreases by 2 units and the melting temperature is mostly unchanged. A similar effect can be observed in Fig. 3.20(d), where the non averaged d.o.p. of the CG-rich strand decreases when the inner strand hybridizes. On the other hand, the inner strand B2m takes benefit from the presence of B1m and folds better (increase of almost 4 units in the d.o.p.), with no significant change in the temperature behaviour. Thus, as in the study of soDNA B and C, we observe a more pronounced cooperativity for staples in the inner position.

3.4.3 Small origami, comparison with experiment

The model showed clearly the importance of the position of the staple in the final structure. The AT rich staple in the inner position is hardly folded compared to the one in the outer position. These results are in good agreement with the experiment. The absolute variation of d.o.p being very different in these two simulated experiments, we studied the experimental variation of absorbance for AT in both positions (inner and outer), expecting a lower decrease for the inner staple. However to the precision of our experiment the two variations are identical and this variation has the same value than the one observed for the annealing experiment of the staple with its complementary strand.

The cooperative effect is also well observed for the second staple folding in the inner position (Fig. 3.23 b) and 3.24 right). This is also in good agreement with the experimental data. However , for the folding in the outer position, the cooperative effect was not well described by the model showing only very small differences with and without the inner staple already folded. In the case of B1m B2m the effect even seems to be reversed, the presence of the outer staple having a destabilising effect.

3.4.4 Conclusion on the coarse-grained model

To summarize, we have shown that our coarse-grained model is able to simulate melting-cooling cycles for different types of DNA constructions. We have shown that different types of staples behave differently according to their position in a soDNA. Inner staples are strongly destabilized compared to the usual DS structure, similarly to the experimental data. Outer staples bind much in the same way as in the dsDNA structure with a small loop entropy penalty when no inner counterpart is preformed. Preformed soDNA, either prepared with outer or inner staples, show good examples of cooperative effects. When topology constraints are relaxed to recover the unbent Holliday junctions, these differences are erased. From these simulations we conclude that the chemical composition of the sequences plays an important role not only in the determination of the optimal temperature for the formation but also in the existence of hierarchical folding.

In general, the staple strategy, used in the formation of more complex structures such as Rothemund origamis, leads to a destabilization (decrease of the T_m melting temperature) compared to the analogous DS structure. It also creates competing, less folded structures which tend to decrease the observed (average) degree of pairing. So far, we cannot exclude that these are artifacts due to a poor sampling, as in the case of the simulated Holliday junctions.

Helicity plays an important role in the design of more complex structures. Taking into account helicity in the context of a grid model is not easy. However, we do not expect that more elaborate DNA models will change significantly the conclusions of the present study. The main reason is that the lenght of the oligonucleotide sequences have been chosen so that crossovers in the soDNA are possible with no major distorsion of the underlying double helix structure. The situation would be very different if this length was not conmensurate with the period of the double helix structure.

Experimentally, the annealing of soDNA does not show significant hysteresis and we expect the same for the structures considered here. Therefore, in order to be able to compare our results with experimental data, further progress is needed to minimize the hysteresis phenomenon inherent to these simulations. Besides increasing the number of MC steps at each temperature, it could be useful to modelize the sequence of bindings by an ensemble of coupled reactions. Once all the kinetic constants have been computed, the steady state can be derived ([77]). However, this approach assumes that each binding proceeds in a two-state manner with an Arrhenius kinetics, an hypothesis that needs to be checked for each case. Finally, the application of the parallel tempering to a more accurate coarse-grained model [93], where both helicity and full flexibility of the chains are taken into account, needs also to be envisaged.

In the following section we will develop the results obtained with structures having a more respectable size.

3.5 Origami

3.5.1 Experimental data.

3.5.1.1 Melting curve of different origamis.

Melting curves of oDNA structures give a global information on the total number of bases folded. The process of folding of individual staples can not be extracted. However as pointed out in the introduction the melting curves of different origamis are different. We considered four different oDNA (Fig. 3.25) with the same scaffold (M13mp18 virus) and about 200 staples. O1 is the rectangle in the original Rothemund work [1], the staples are mostly 32b long, with sequences divided in 8-16-8 patterns. O2 is another rectangular oDNA that includes a hole [98] and presents the same 8-16-8 pattern. O3 has the same connectivity pattern as O1, but some staples have been merged two by two in four areas (black area in 3.35c), so that the typical staple pattern is 8-16-16-16-8. Finally, O4 is another rectangular oDNA where a 700b long subset of the scaffold goes from one side to the other of the rectangle, forming a ssDNA 'bridge'.

For each oDNA, a series of annealing-melting cycles was performed coupled to UV-absorption measurements. The concentration in M13mp18 was set as close as possible to 1 nM for each experiment and in all the experiments an excess of 10 staples by M13mp18 strand was used (unless specified otherwise). For all UV-experiments we did in parallel the annealing-melting curves of the staple strands alone, of the M13mp18 alone and of the mix staples plus M13mp18. This was done with the purpose of removing the signal due to the excess of staples by subtracting the curve of the staple alone to the curve of the mix staples plus M13mp18. The protocol to obtain the normalised degree of pairing $\theta(T)$ is detailed in Appendix I. The temperature ramp (0.4 $°C$ min^{-1}) is typical for this one-layer origamis. The annealing-melting process is not symmetrical, the hysteresis between the two phases of a cycle is such that the melting takes place at temperatures higher than the annealing.

Several thermodynamic parameters can be extracted from these curves in the very simple hypothesis of a bi-molecular equilibrium (Sec. 3.2, [77]) between the virus and all the staples (Tab. 3.1). The fits of $ln(k_{on}(1/T))$ and $ln(k_{off}(1/T))$ show that these quantities do not follow a simple Arrhenius equation (Fig. 3.26) as the curves are not linear. The values obtained for ΔG correspond to the energy expected for the annealing of a 20 bases long dsDNA (oDNA 01 and 02) and for the annealing of a 36 bases long strands (oDNA 04). The energy obtained is an energy by strand as $\theta(T)$ was normalised between 0 and 1. We were not able to understand the differences between the values obtained for the different origamis and won't comment them much, as the poor fit of $ln(k_{on})$ and $ln(k_{off})$ is an indication that this simple model is not adapted to the description of the melting and annealing process of an oDNA.

However such a simple model can still help to understand the effect of the concentration on the melting curve (Sec. 3.5.1.2).

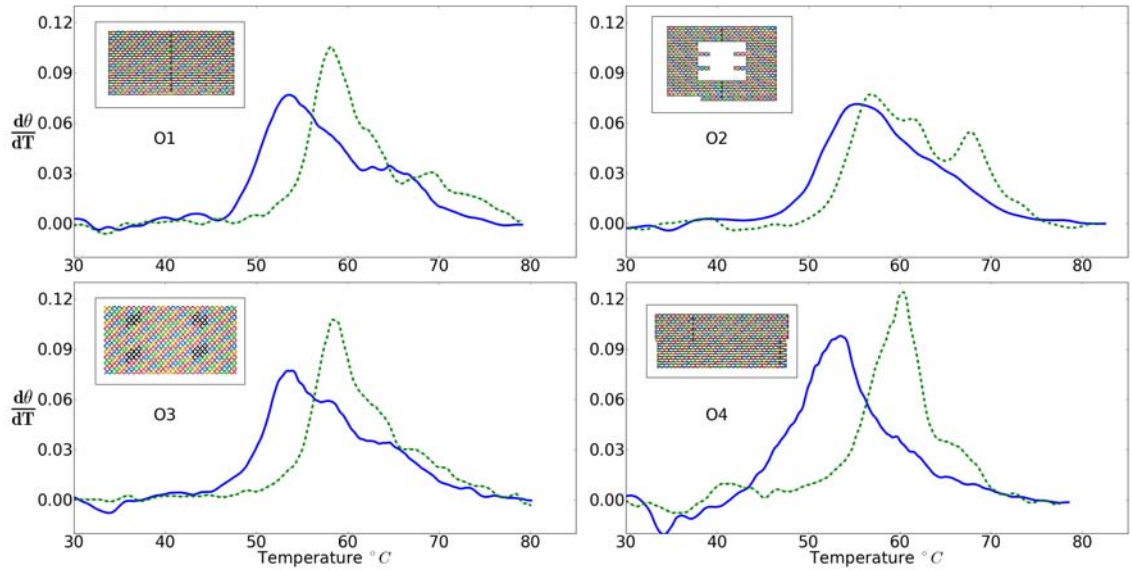


Figure 3.25: Derivative of the pairing degree versus temperature. The data corresponding to annealing are in blue, melting in green.

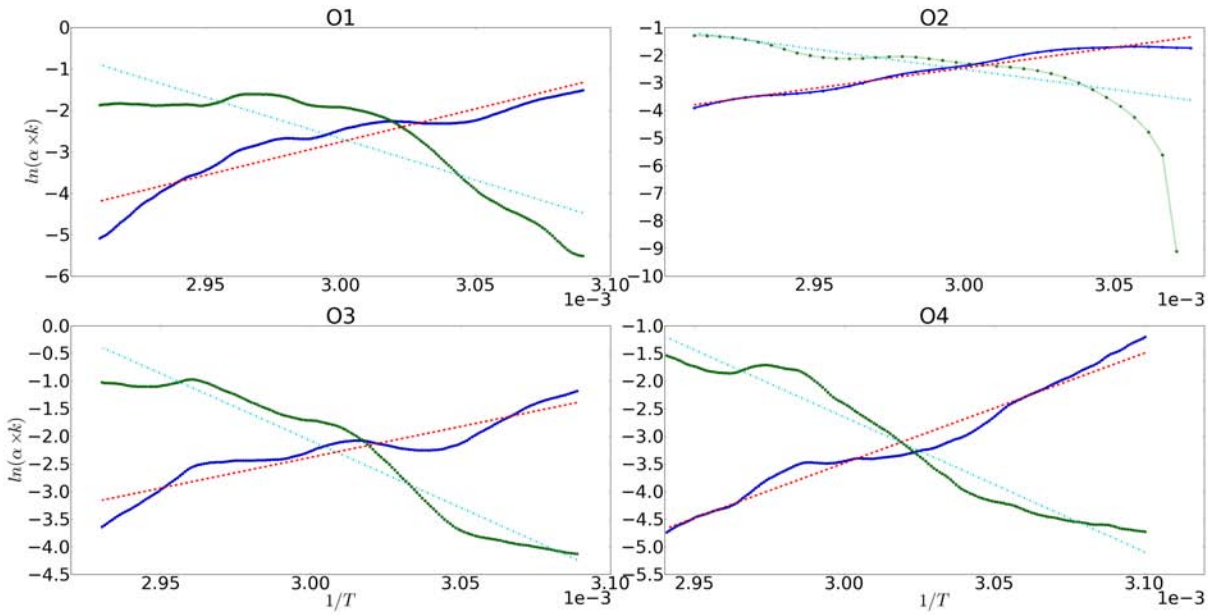


Figure 3.26: Plot of $ln(k_{off})$ (green = experimental, cyan = fit) and $ln((E-1/2)Ck_{on}/2)$ (blue = experimental, red=fit) and their fits for $\theta \in [0.1, 0.9]$ (With E the excess of staple and C the M13mp18 concentration)

Origami	E_{on}	E_{off}	ΔG	T_m
O1	-21.07	31.03	-52.10	48.7
O2	-35.59	17.89	-53.47	50.7
O3	-15.12	44.28	-59.40	50.1
O4	-14.09	77.18	-91.26	52.1

Table 3.1: Thermodynamic value extracted from the non-equilibrium curves E_{on}, E_{off} and ΔH are in kcal/mol and T_m in °C. O1,O2,O3,O4 are the value obtained for $\alpha_{cooling} \in [0.2; 0.9]$ and $\alpha_{heating} \in [0.2; 0.9]$.

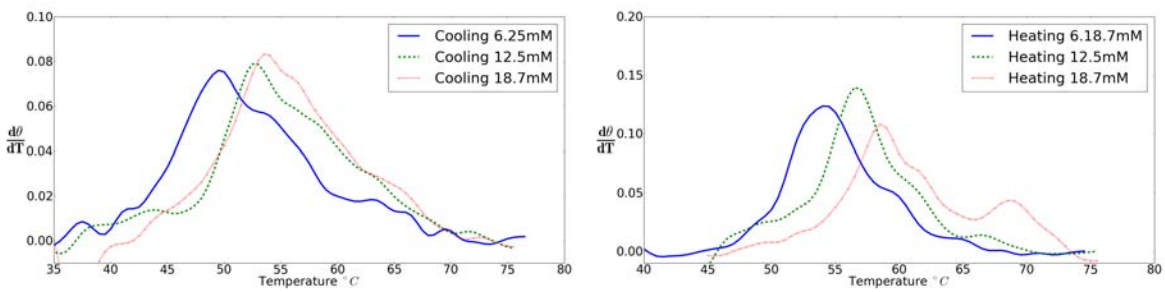


Figure 3.27: Annealing (left) and melting (right) curves of the O1 origami for different concentrations in magnesium.

3.5.1.2 Effect of different experimental parameters on the melting-annealing curves:

The reproducibility of the experiments is good when repeated with a given buffer. However we noticed that the use of a new buffer had no effect of the annealing curve but the derivative of the melting curve exhibited sometimes another secondary peak at high temperature and high Mg concentration. This led us to study the effect of the concentration in magnesium in the buffer. We compared three concentrations in magnesium: 6.25 mM (x0.5), 12.5mM(x1) and 18.7mM(x1.5) (Fig. 3.27).

The absorption curves for the different constituents (Virus, Staple, Origami) and for the different salt concentration can be found in appendix H.1. We can see that the peak at high temperature increases with the salt concentration for the heating experiment. We also noted that for other oDNA the same peak appears. Furthermore, melting experiments on the virus strand alone show a small transition at $65°C$ for a magnesium concentration of 18.7 mM. For these reasons it is possible that this transition is related to the virus strand alone.

To further comfort this idea, an experiment on an oDNA with an incomplete set of staples was realised. We compared the melting curves of a complete O4 oDNA (Fig 3.28 a.A) to the same oDNA where 37 staples were omitted (Fig 3.28 a.B), and to another where 47 staples were

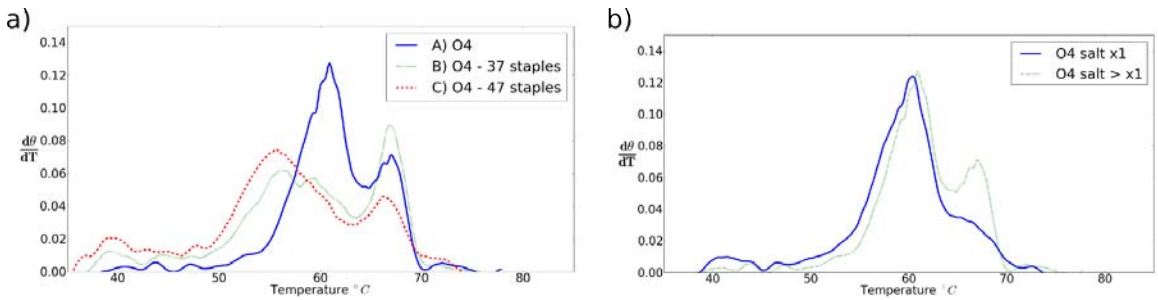


Figure 3.28: a) Comparison of the melting curve of O4 A) and the same oDNA with several staples removed B) and C) (Mg > 1); b) Comparison of an origami in the regular buffer (salt x1) and an origami with a salt buffer slightly superior to one.

omitted (Fig 3.28 a.C). It should be noticed that the 37 staples were not chosen randomly. All of them have a melting temperature between 61 and 63 °C. The 47 staples are the 37 previous staples plus the staples having a melting temperature between 66 and 68 °C. One can see the presence of the peak at high temperature for all the oDNA, even if they are probably only partially folded (because of the missing staples). Also, the comparison of this experiment with one realised with a 12.5mM Mg buffer (Fig. 3.28 b) (salt x1)) indicates an effect similar to the one observed for the origami O1.

	0.5x	1x	1.5x
T_m Cooling	49.6	52.7	53.6
T_m Heating	54.2	56.6	58.6

Table 3.2: Melting temperatures as a function of the salt concentration

The results of the dependence of T_m with Mg concentration for O1 are gathered in (Tab.~3.2) The salt is important in the melting process as it screens the negative charges of the ssDNA and allows the two ssDNA to get close-by and to fold. In the case of oDNA the presence of the salt is even more important as discussed in section 2.3.1.1: the dsDNA that constitute the oDNA are in interaction via repulsive charges. Once the oDNA formed, the screening effect of the salt diminishes the repulsive interaction between dsDNA. This could explain the increase of the melting temperature (both for melting and annealing) when increasing the salt concentration.

The dependency of the melting temperature with the heating and cooling rate has been studied on the origami O4. (See Appendix H.3 for the curves). The annealing curve is one degree higher when changing the rate form 0.4°/min to 0.2°/min and the melting curve is hardly shifted. This means that the annealing curve is farther from the equilibrium than the melting curve.

The excess of staples has been studied by comparing a 5 and 10-fold excess (Appendix H.2).

Once again the annealing curve is more affected than the melting curve. The annealing curve is stabilised by 1.5 degree when increasing the excess, whereas the melting curve is stabilised only by 0.5°C. This can be understood in terms of a simple second-order kinetic reaction model.

3.6 Thermodynamic model of the formation of a DNA origami

The model developed in this section proposes a mechanism of formation for oDNA. This model is based on the cooperativity observed on the small origamis. However, given the complexity of an origami we need to use simplifying hypothesis to allow the calculation of the folding process (Sec. 3.6.1) The melting curves of the experiment on the small origami allowed us to extract the different possible cases of folding for a staple in the origami and help us to calibrate the model (Sec. 3.6.2). The results of the model for the origamis are compared with the experimental data in section 3.7. A correct agreement between the model and the data being achieved, we will use the model to play with the design of origamis to increase the thermal stability. (Sec 3.7.2).

3.6.1 Ideas of the model

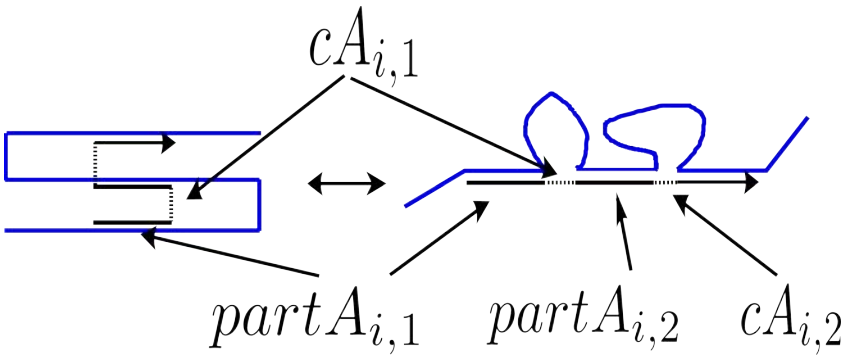


Figure 3.29: Notations

The small origamis show cooperative and topological effects that help to start a study on folding process of oDNA that bear hundreds of staples. For this, we need to make a few hypothesis in order to make the problem tractable. The difficulty is due to the huge number of possible configurations that need to be handled to compute average properties such as the number of open base pairs. Long linear structures of double stranded DNA can be computed rigorously because recurrence relations can be established in such cases [75, 30]. DNA origamis are highly connected structures that prevent the use of linear recurrences.

Let us enumerate the working hypothesis of the model. Each staple S_i of length $|S_i|$ can be divided in parts that hybridize to non-contiguous regions of the scaffold (Fig. 3.29). Let us note $S_i = \text{part}S_{i,1} + \text{part}S_{i,2} + \ldots$ such a division of the strand sequence (typically, each 32b staple is divided in three parts, see Fig. 3.29 and Fig. 3.30, but other partitions are possible). Two consecutive parts ($\text{part}S_{i,j}$, $\text{part}S_{i,j+1}$) are connected by a crossover noted $cS_{i,j}$

Hypothesis 1: we will focus on configurations where each part $S_{i,k}$ is either completely hybridized or completely free. Moreover, we also disregard misfolded configurations, that is, staples that partially hybridize to the 'wrong' part of the scaffold. Notice that this assumption is plausible for the one layer origamis we consider here. For more complex, multi-layered structures, the staples are divided in smaller parts so that the probability to bind to the wrong part of the scaffold is considerably increased.

We will call *crossover* the connection between two contiguous parts of any staple. A crossover is not associated with a particular DNA base, it is only a convenient notation to describe the connectivity of the origami. In the examples of Fig. 3.30, typical staples are 32 bases long and composed of three parts (8,16,8 bases long respectively) linked by two crossovers $cS_{i,1}$ and $cS_{i,2}$. On the scaffold side, a crossover is associated with a loop, a subset of the scaffold that is hybridized (or not) depending on the presence of other staples.

Hypothesis 2: configurations with non contiguous hybridized parts are not considered. This hypothesis is verified when the central part of the staple is much longer than the other parts. In the following, we will note $S_i(k,l)$ the configuration where part$S_{i,k}$, part$S_{i,k+1}, \ldots$, part$S_{i,l}$ are hybridized, the other parts being unpaired.

The model aims to compute the probability $p(S_i(k,l),T)$ of having a particular folded state of the staple S_i at temperature T. We will assume that at very high temperature $T = T_h$, all the staples are unfolded: $p(S_i(k,l),T_h) = 0$ (in practice, $T_h = 90\ ^\circ C$). The model is *recursive*: $p(S_i(k,l),T+dT)$ is computed based on the knowledge of $p(S_i(k,l),T)$. The increment dT can be both positive or negative: the algorithm starts from T_h, the temperature decreases down to a value T_l, then increases again. At any temperature T, the probability to observe a given configuration $S_i(k,l)$ will depend upon the presence (or not) of neighbour staples. Therefore, for each staple a set of neighbour staples $\{N_\alpha(S_i)\}$ needs to be defined. How many staples one has to consider in this set is a parameter of the model. In the following, the set of neighbour staples will be limited to the staples that are in the same row of the origami scaffold, and separated by less than 75b. With these notations, the probability to observe the staple S_i in a given configuration $S_i(k,l)$ and for a given neighbourhood $N_\alpha(S_i)$ is modelized by an equilibrium reaction: $S_i(k,l) + N_\alpha(S_i) \rightleftharpoons N_\alpha(S_i)S_i(k,l)$.

This modelling therefore does not consider any real kinetic effect. However the value of the step dT can play such a role (Sec 3.7.1).

Hypothesis 3: because the model only keeps track of the single probabilities $p(S_i(k,l),T)$ and not of the joint probabilities $p(S_1(k,l),S_2(k',l'),\ldots T)$, it is necessary to make an additional approximation to determine $p(N_\alpha(S_i),T)$. Based on the data from the small origamis, we assume that there is a strong correlation between the different staples. As the processes of annealing and melting are monotonous, for two staples S_{i^1} and S_{i^2} we venture the hypothesis that if $p(S_{i^1}(k,l),T) < p(S_{i^2}(k',l'),T)$, then the S_{i^2} staple was present in the structure when S_{i^1} staple began to fold. In order to compute $p(N_\alpha(S_i),T)$, let us generalize this idea and order the staples in $N_\alpha(S_i) = \{S_{i^1}, S_{i^2}, \ldots\}$ in such a way that $p(S_{i^1},T) \leq p(S_{i^2},T) \leq \ldots < p(S) = 1$

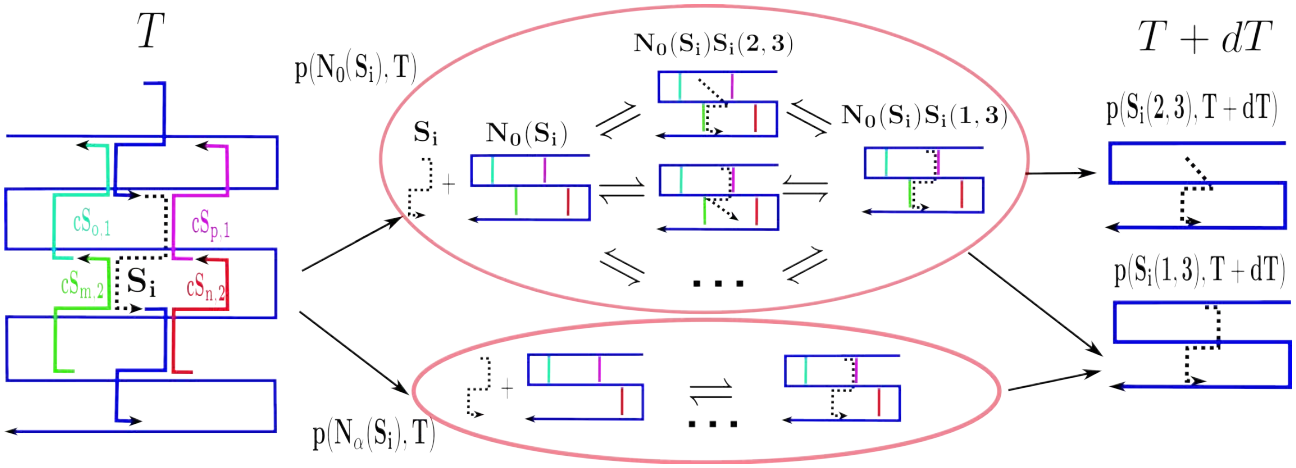


Figure 3.30: To evaluate the fraction of the staple S_i folded at the temperature $T + dT$, one considers the nearby staples of the staple i at T and calculates the probability of the different neighbouring crossovers configurations ($cS_{m,2}, cS_{p,1}$,etc) around S_i. The origami is then subdivided in different partially folded state(eg $N_\alpha(S_i)$) with a given probability (eg $p(N_\alpha(S_i))$..). For each of these partial states the equilibrium constant for a partial folded configuration ($N_\alpha(S_i)S_i(m, n)$) of the staple within this restricted local state is calculated as explained in the energy model. The law of mass action for each partial configuration folded gives a set of coupled equations. Once solved they allow to determine the fraction of partial configuration folded in this environment $p(N_\alpha(S_i), T + dT)$. Then we can calculate the total fraction of each configuration folded $p(S_i(m, n), (T + dT))$, as the sum of the fraction of those configurations in the different local states, weighted by the probability of each state. This probability will allow to calculate the probabilities of the neighbouring configurations of every staple at the next step.

where S stands for the scaffold. According to this *high correlation hypothesis*, we approximate the joint probabilities in the following way:

$$\begin{aligned} p(S_{i^1}, T) &= p(S_{i^1}, S_{i^2}, \ldots) \\ p(S_{i^2}, T) - p(S_{i^1}, T) &= p(S_{i^2}, S_{i^3}, \ldots), \ldots \end{aligned}$$

For instance, in the case where only two crossovers influence S_i, with probability $p(S_{i^1})$ both cS_{i^1} and cS_{i^2} are present, the probability of only having cS_{i^2} is $p(S_{i^2}) - p(S_{i^1})$ and the probability of only having the scaffold is $1 - p(S_{i^2})$. We show (Appendix D) that the set of equilibrium reactions determines $p(S_i(k, l))$ provided the equilibrium constants of these reactions are known. This amounts to define an energy model which is detailed in the next section.

3.6.2 Folding energy of a given configuration

The gain in Gibbs free energy for hybridizing $S_i(k,l)$, $\Delta G(S_i(k,l),T) = \Delta H(S_i(k,l),T) - T\Delta S(S_i(k,l),T)$ contains two contributions: $\Delta G = \Delta G_{NN} + \Delta G_{top}$. The local contribution ΔG_{NN} only depends on the sequence of $S_i(k,l)$. It quantifies the gain in free energy associated with the local formation of double helices. We use the parameters of the nearest-neighbour model [27] with a temperature correction given in [99] (see Appendix G).

ΔG_{top} gathers several contributions that depend on the connectivity of the origami (ΔG_{NN} depends mostly on the sequence of the scaffold and, to a less extent, on the density of crossovers, but not on the connectivity). With any crossover, we associate an entropic penalty. This penalty reflects the difficulty for a staple to hybridize non-contiguous parts of the scaffold. In a first approximation, the longer the region of the scaffold that connects the two parts to be hybridized, the larger the penalty. Our previous results obtained with the small origami show that this needs to be refined. Based on these data, we consider three situations characterized by transient arrangements of staples that we call local intermediate states (LIS).

In the first one (LIS outer), the staple hybridizes to the scaffold, forming an internal asymmetric loop [27] (Fig. 3.31a). The length of this loop corresponds to the number of unpaired bases of the scaffold linked by the crossover. This is a generalization of the 'outer' position found for the three strands origami. Before the crossover formation, when only part of the staple is folded, the scaffold and the non hybridized part of the staple are on the same side of the hybridized part of the staple (Fig. 3.4b). In this case, the staple is not involved in the path that connects the two extremities of the crossover.

A particular case of LIS outer, which we call LIS crossover, is the situation where the length of the loop is zero: the crossover forms locally a Holliday junction, the staple hybridizes in the close vicinity of an already hybridized staple (Fig. 3.31b).

The third LIS, LIS inner, corresponds to the inner position in the small origami: the shortest path that connects the two ends of the crossover involves the staple itself (Fig. 3.31c). Because of this, before the crossover forms, the non hybridized parts of the strand and the scaffold are located on opposite sides of the hybridized parts (Fig. 3.4c). Therefore, LIS inner implies a larger penalty than LIS outer or LIS crossover. In the small origami, the shift in T_m was less than 10 $°C$ for LIS outer, between 5 $°C$ and 7 $°C$ for LIS crossover and up to 15 $°C$ for LIS inner.

The ΔG_{top} contribution for the insertion of a staple in LIS outer can be written:

$$\Delta G_{top}(\text{outer}) = -T\Delta S_{bulge}(n_T - 0.8 < nb_{folded} >) \tag{3.6}$$

(Fig. 3.31). The function $\Delta S_{bulge}(n_T)$ is that of ref. [27]. n_T corresponds to the number of bases along the scaffold and $< nb_{folded} >$ is the average value of bases folded along the scaffold (this average takes into account the probabilities of all the possible neighbouring configurations). For example the value for a bulge of one base is $\Delta G_{top} = 2.7$ kcal/mol and for a bulge of 30 bases of

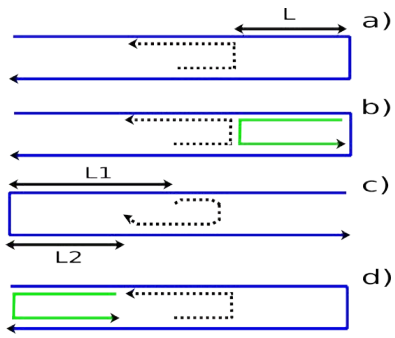


Figure 3.31: Computing the entropic penalty for the three different local intermediate states (LIS). The staple to be inserted is represented by the dotted line, the scaffold by the continuous line. (a) LIS outer (b) LIS crossover (c) LIS inner. Here, we assume that, because of the curvature constraints imposed by this configuration, the staple remains partly unfolded. (d) A typical situation where two types of LIS (LIS outer at the right side of the staple, LIS inner at the left side) can be attributed to a given crossover.

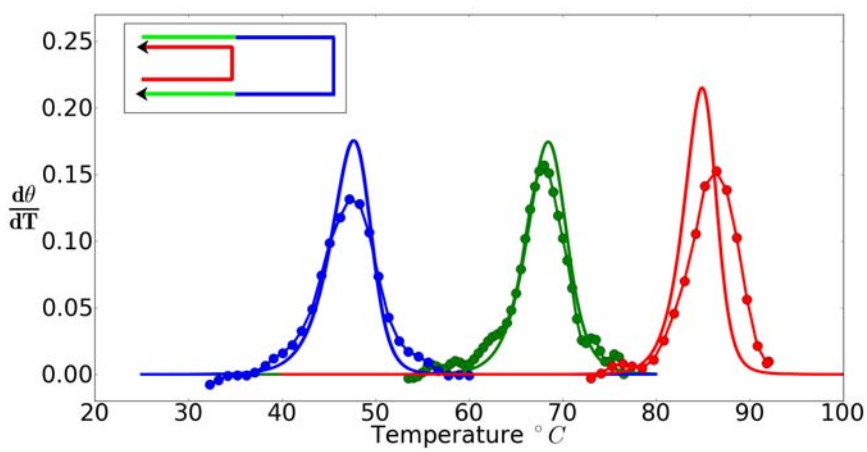


Figure 3.32: (dotted line = experiment,continuous line = theoretical) Derivative $d\theta/dT$ for the insertion of a staple in LIS outer. blue=B1,green=B1m,red=B2

$\Delta G_{top} = 5.9$ kcal/mol. The comparison between this model and the experimental results from the small origami structure is illustrated in Fig. 3.32

For the LIS crossover, we use the following expression:

$$\Delta G_{top}(\text{crossover}) = \Delta H_{cross} - T\Delta S_{cross} \tag{3.7}$$

with $\Delta H_{cross} = 25.3$ kcal/mol and $\Delta S_{cross} = 65.0$ cal/mol/K. This constant contribution has been derived so as to fit as well as possible the B1-B2 experimental data (left of Fig. 3.33) and then applied to the B1m-B2m data (right part of Fig. 3.33). At the crossover, two bases

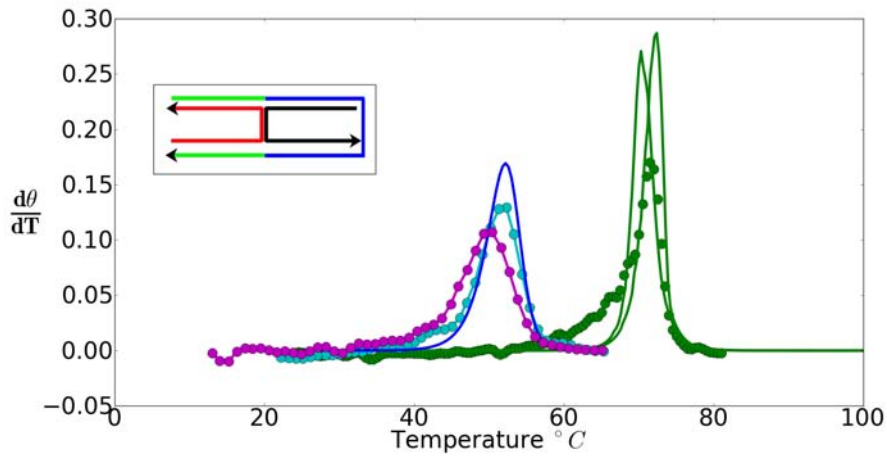


Figure 3.33: (dotted line = experiment,continuous line = theoretical) Derivative $d\theta/dT$ for the insertion of B1 in LIS crossover(cyan=B1 in position outer,magenta=B1 in position inner,blue=theoretical identical for both experiements), in green simultaneous binding of B1m and B2m.

that belong to S_i face the bases constituting the crossover made by the other strand. The initial enthalpic and entropic contribution of these two bases is subtracted from ΔG_{NN} as they are not nearest-neighbours anymore. One half of the contribution (nearest-neighbour model) of the two new pairs of bases is added. Wherever the staple is located, in the inner or outer position, the correlation effect is roughly the same showing nearly identical melting curves with a difference of about 3 K. As a consequence, in our calculation, the same set of parameters is used to describe the correlation effect, whatever the location of the strand, in the inner or outer location. The two melting curves corresponding to AT folding in the inner position and AT folding in the outer position are displayed in Fig. 3.33 (cyan and magenta curve). This approximation is also motivated by the fact that in the oDNA structure the separation between crossovers is 32 bases so that the effect of being in the inner position is less penalising than the similar situation in the small origami where the distance is twice smaller.

For LIS inner, we use the following expression:

$$\Delta G_{top}(\text{inner}) = \Delta H_{unbind} - T\Delta S_{unbind} + \Delta G_{top}(\text{outer}) \tag{3.8}$$

ΔH_{unbind} (resp ΔS_{unbind}) quantifies the loss of enthalpy (resp gain in entropy) associated with the partial unfolding of the ends of the staple involved in this type of LIS. The number of bases that unfold is a parameter of the model. The data in (Fig. 3.34) correspond to the unfolding of a total of 8 bases (two bases for each of the four extremities of the staple, see Fig. 3.31c). This is a generalisation of the inner position in the small origami, where a free loop of length $L_1 + L_2$ can be present (Fig. 3.31c) adding a contribution $\Delta G_{top}(\text{outer})$.

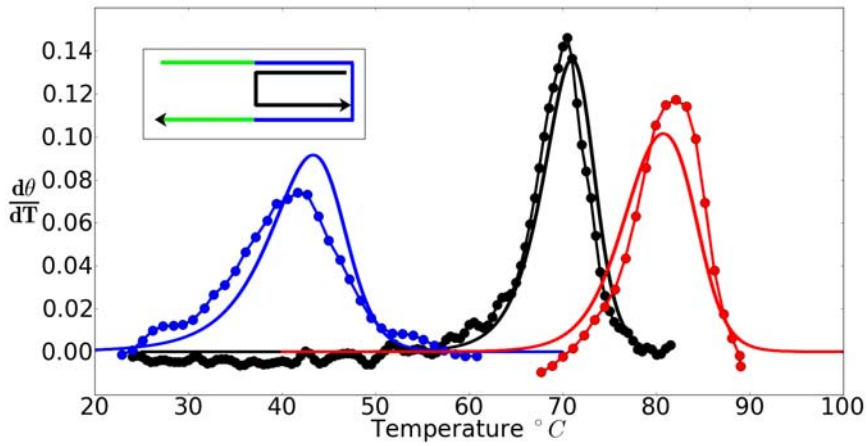


Figure 3.34: (doted line = experiment,continuous line = theoretical) Derivative $d\theta/dT$ for the insertion of a staple in LIS inner. blue=B1,black=B2m,red=B2

Under some circumstances (Fig. 3.31d), two types of LIS can be attributed to a given crossover. In such cases, the LIS with the smaller ΔG_{top} is taken into account.

The modelling obtained with the contributions ΔG_{NN} and ΔG_{top} is quite satisfactory except for a constant negative shift (~ -2 K) of the melting temperatures. This shift indicates that another stabilizing mechanism that is not present in small constructions such as the small origamis has to be invoked. We first considered mechanisms such as correlations between counter-ions and hydration forces that are present when DNA condensates [56]. Indeed adding an energy per base $-0.132 kcal/mol/base$ (similar to the one needed for DNA condensation $10^{-1} k_B T/base$ to $10^{-2} k_B T/base$ [28]) correct the shift of 2 degres. However as we saw in section 2.3.1.1, it seems at least that at short distance the interactions are repulsive. Another hypothesis to explain this additional stability could be a smaller entropy of the dsDNA in the origami structure, as only two dimensional fluctuations are possible.

The stabilising term added is proportional to the number of neighbouring bases n_n that are close to the staple S_i, and to the length $|S_i(k,l)|$ of the partial configuration $S_i(k,l)$ considered.

$$\Delta H_n(S_i(k,l)) = -0.132 n_n |S_i(k,l)|/|S_i| (kcal/mol) \tag{3.9}$$

The salinity of the buffer is taken into account via the correcting terms in [100] in the case where Mg is dominant (see Appendix G.2). Finally the number of bases folded is converted to a theoretical absorbance [101] (Appendix F).

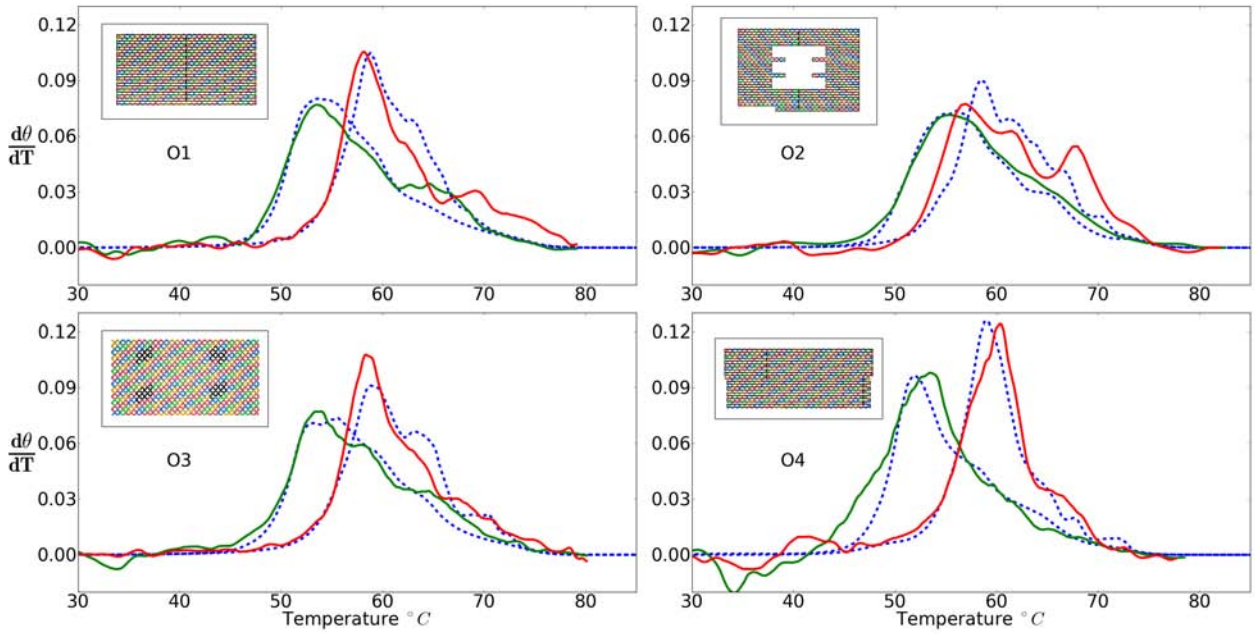


Figure 3.35: Derivative of the pairing degree versus temperature for the oDNA O1,O2,O3 and O4. The data corresponding to annealing are in red, melting in green, the model is in blue for both processes.

3.7 Comparison with the annealing-melting data of four origamis

The overall agreement (Fig. 3.35) between the melting-annealing curves observed experimentally and computed with the model is satisfactory. The model captures the hysteresis between the annealing and melting processes, as well as the relative strength of this hysteresis between different origamis (O2 has only 4K shift between annealing-melting, whereas O4 has 10K shift). The maximum value of the derivative, which can be linked to the overall enthalpy of the transition in a two-state model, is also reproduced. This feature is robust against small variations of the parameters of the model.

3.7.1 Kinetic effect.

As presented before, the model does not take into account any kinetic of reaction. Indeed at every temperature step we consider that each staple is at the equilibrium in a state defined by the neighbouring crossovers. However, in this model two coupled parameters have an influence on the hysteresis of the annealing-melting process. The first one is dT, the size of the step in temperature between two successive updates. The second parameter is the number of times

that the equations to compute the probability of each staple $p(S_i(j,k),T)$ are iterated at a given temperature. In fact at each step, the probability for a staple to be folded is computed according to its neighbouring crossovers. This allows to compute the probability for each crossover. At that point of the algorithm one could recompute at this temperature the new probability for the staple to be folded given this new set of crossover probabilities. This is equivalent to set $dT = 0$.

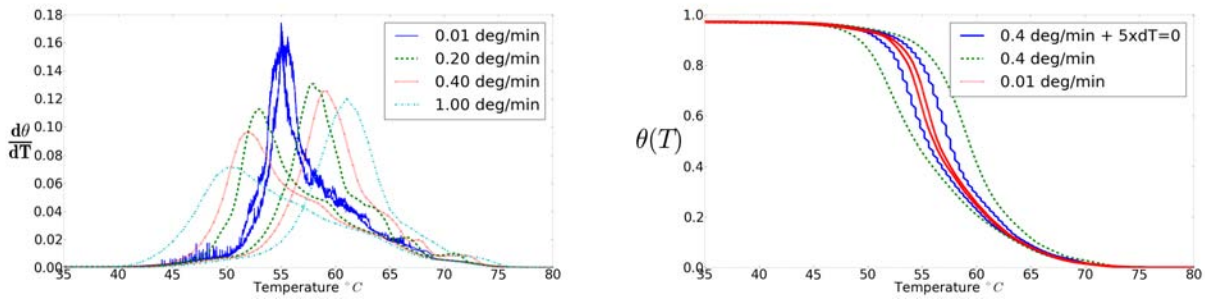


Figure 3.36: (Left) Comparison of different melting rate for O4;(Right) Comparison of the 0.4 and 0.01 rate with the 0.4 rate with five dT=0 steps between each temperature steps

Diminishing the temperature step leads to a convergence of the annealing-melting curves to an equilibrium curve (Fig 3.36). The same phenomenon can be observed when including between every step in temperature a given number of steps with $dT = 0$. The comparison between the theory and the experiments was realised with $dT = 0.4\ °C$.

3.7.2 Understanding DNA Origami design

In this section, we use the model to explore how the melting temperature of the rectangular origamis depends on the specific connectivity. In many instances, this type of considerations are relevant to improve the stability of the template against temperature when origamis are sought as platforms controlling chemical reactions or other applications including grafting inorganic species.

3.7.2.1 Circular permutation of the scaffold.

The scaffold used for the design of rectangular origamis is a circular phage. Thus, it is possible to choose the beginning of the scaffold sequence anywhere so that 7248 different sets of staples (with same length and position, but with a different composition) are possible. We compared the melting curves given by those permutations on the O4 origami, permuting the sequence in steps of 16 bases as this shifts the middle of one staple to the middle of the nearby staple. The distribution of temperatures corresponding to the maximum of the derivative for the annealing

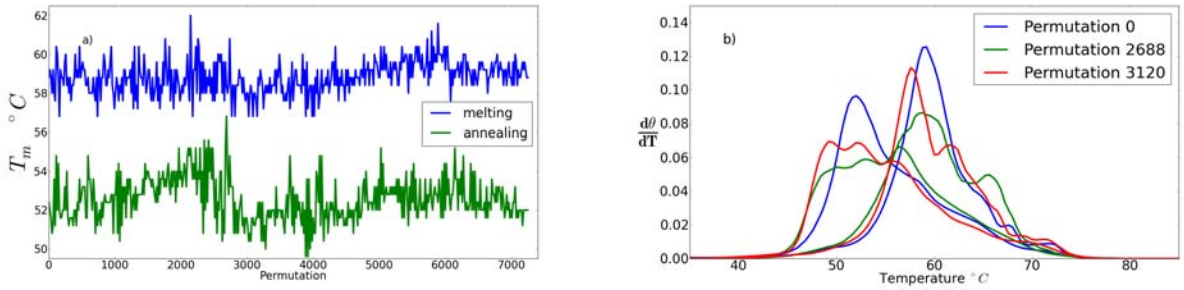


Figure 3.37: (a) Distribution of melting temperatures (annealing and melting) as a function of the order of the circular permutation of the scaffold strand. (b) two different melting curves corresponding to two permutations with the lowest and highest annealing temperatures.

and melting curves (Fig. 3.37a) shows an amplitude of variation of about 4 $°C$. Also, correlations exist between consecutive permutations. Depending on the permutation, the melting curves can be very different in shape (Fig. 3.37b).

3.7.2.2 Decreasing the number of crossovers

In our model, a penalty is associated to each crossover. Reducing the number of crossovers should in principle increase the stability in the annealing process. We use the most regular design O1 to investigate the effect of the reduction of the number of crossovers. In the initial origami there is a length of 32 bases between two crossovers, which corresponds to three periods in the double helix. Increasing the distance between crossovers leads to consider 54 bases (5 double-helix periods). We considered two possibilities, illustrated in Fig. 3.38: staples 27b long, split in 13-14 (O5 origami), and staples 54b long, split in 13-27-14 (O6 origami). Indeed, there is a trade-off between the gain in enthalpy when increasing the length of the staple, and the additional penalty of having two crossovers/ staple instead of only one. Our model shows that the net gain in stability, compared to the initial 8-16-8 staple strategy, is almost 20 $°C$ (Fig. 3.38 O6). Again, the comparison with the experiments is excellent. Notice that decreasing the number of crossovers could have an impact on the flexibility of the origami.

Experimentally the hysteresis between the annealing and melting curve of O6 is small ($≈ 2$~$°C$). This means that the system is closer to the equilibrium compared to the other oDNA. This could be explained by a smaller number of staples for the O6 oDNA: it only has 144 staples compared to the 210 staples of the other oDNA. Having less staples leads to a smaller number of possible configurations and probably to less hysteresis. Furthermore, the length of the smaller sub-part of the staples is 13 in the oDNA O6 compared to 8 in the other oDNA. This should lead to less miss-paring in the annealing process and a faster equilibrium.

Theoretically we predict a melting curve shifted from 6 degrees compared to experimental one: the maximum of the predicted melting curve is at 73°C whereas the experimental maximum

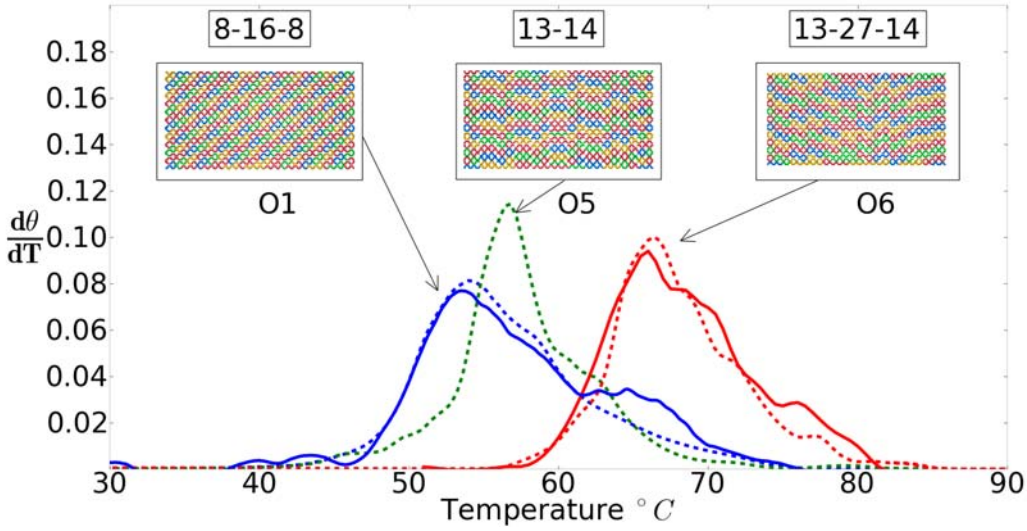


Figure 3.38: Annealing curves of the O1 and O6 origamis. The two origamis correspond to the same scaffold pattern, but different staple pattern (solid line = experimental data, dashed line = theoretical curves)

is at 67°C. This lower stability when melting could come from a higher flexibility of the structure compared to the other oDNA, and could require a different value for some of the parameters.

3.8 Conclusion and perspectives.

Simplifying the complex origami to its smallest component : the small origami , helped us to to understand the cooperative process at stake in the folding and the importance of the staple position with respect to the scaffold.

Numerically most of the properties of folding of the small origami were reproduced with a first coarse-grained model on a grid. However the complexity of real sized origami was not reachable in a the reasonable amount of computational time.

This is why a new coarse grain at the level of the strands was introduced. This algorithm provides a reasonable account of the observed melting and annealing behaviour of DNA origamis. The model reproduces hysteresis and melting temperatures, as well as the width of the melting curve. It emphasizes the role of cooperativity in the folding process by introducing correlations between the probability of presence of neighbour staples. Finally, it allows to improve the thermal stability by quantifying the effect of different construction factors such as staple length and density of crossovers. Extensions to 3D [5, 102] and structures other than origami [103] are envisioned, as well as tests at the single molecule level (FRET). AFM measurement of Origami as function of the temperature [104] could be envisaged to compare the observed structure with the model. Preliminary results of AFM imaging gave an agreement on several points but showed that the model still needs to be improved (Appendix E). and a collaboration Jie Song in Mingdong Dong teams has partially confirmed the result of the model [105]. Neutron scattering experiments could also be envisioned .

Chapter 4

DNA origami as a Raman platform (work done in collaboration with S. Bidault)

Contents

4.1 Goal **103**

4.2 State of the art in using DNA to create SERS structure **104**

4.3 Introduction to Raman phenomenon and SERS technique. **109**

4.3.1 The Raman phenomenon: 109

4.3.2 The SERS phenomenon: 109

4.4 Experimental results: **112**

4.4.1 Protocol for coupling NP with ssDNA 112

4.4.2 Attachment of the NP to the origami. 114

4.4.3 Raman experiment. 116

4.5 Conclusion **122**

4.1 Goal

The general context of this chapter is the detection of low concentrations of molecules such as DNA, RNA or, more generally, biomolecules involved in any disease. We would like to develop a methodology to map the distribution of a molecular agent diluted in a nanoscale structured or a randomly distributed medium without altering the medium (i.e. introducing labels or external

agents which modify the structure of the medium) and to capture in a nonintrusive manner the interaction of this agent with a target molecule. More precisely, for the analysis of enzymatic DNA repair activities, we would like to develop miniaturized DNA- based biosensors that would provide new powerful tools to further substantiate the role of excision repair capacity in the development of cancer and human diseases, and to develop screening platforms to search for new DNA repair activity modulators. The difficulty in the direct spectroscopical detection of a single molecule is related to the discrimination of its own signal to the surrounding response. In this respect, the introduction of fluorescent labels was an important achievement since it allowed the in vivo observation of the intimate structure of cells (cytoskeleton fiber, nuclei compartments ...). However, this method can be considered as non perturbative if the amount of label remains negligible with respect to the molecule to be recognized. When its concentration drops down to a few femto-moles the ratio of (number of label)/(number of molecules to be probed) tends to one. In that case the detection is perturbative, and the response of the molecule can be changed by the presence of the labelling agent.

The label-free methods, such as Raman spectroscopy, rely on the fact that some molecules can be detected due to their optical signatures, such as the Raman spectrum. Both approaches (label or label-free) present advantages and difficulties. In this chapter, we want to explore new ways to circumvent specific problems related to Surface Enhanced Raman Spectroscopy (SERS) to conceive devices that overcome the lock of a lack of methods for producing of reproducible SERS substrates which have a reliable enhancement factor. Roughly speaking, SERS detection requires two ingredients: (i) the creation in a confined region of a high electric field and (ii) the closeness of the target molecule to this hot spot. In the standard use of SERS, both (i) and (ii) take place randomly. The basic approach we want to use is based on origami DNA to locate the hot spot and the target molecule with a few nanometers spatial resolution.

4.2 State of the art in using DNA to create SERS structure

Understanding the interactions of DNA, RNA and proteins is of enormous complexity and requires analytical tools for screening and sequencing DNA and RNA. These tools usually rely on the molecular recognition of two strands of oligonucleotides which are complementary to each other: the key issue is to have a probe that allows molecular recognition and that converts this response to a kind of output that can be amplified and analysed. In recent years, new technologies have appeared which allow massively parallel analysis on a single device. In the case of so-called DNA array, large numbers of DNA molecules or oligonucleotides are immobilised onto various spots. The array is treated with a sample solution containing single-stranded DNA fragments having a specific label: if the sample DNA hybridizes completely or partially to the immobilised DNA fragment, it stays on the array whereas if not, it stays in

solution and it is washed away in the next reaction step. Depending on the nature of the label, various strategies can be designed: if the label is fluorescent, the array may be scanned and the spots interrogated by a laser fluorescence confocal microscope. It was also suggested to monitor hybridisation through fluorescence resonance energy transfer (FRET) ([106]). When two complementary DNA oligonucleotides, labelled with a different fluorescence dye at the 5'-or 3'-end, hybridize, the donor and the acceptor fluorophores become very close to each other (less than 8 nm.) so that FRET takes place: there is an enhancement of the acceptor fluorescence and a quenching of the donor fluorescence.

Although all these approaches are sensitive and powerful, they all need the attachment of a specific fluorescence dye to the probed molecule. The approach we propose relies on the use of vibrational spectroscopy (Raman spectroscopy), which is very selective and requires no specific label. Richard C. Lord and co-workers at MIT in the late 1960s [107] first demonstrated that Raman spectroscopy is an effective experimental probe of nucleic acid constituents. Then, Raman methods were implemented in many laboratories to investigate native, model nucleic acid structures and their biological complexes [108]. A vibrational Raman spectrum is obtained by excitation of a sample in a transparent region of its absorption spectrum. All of the Raman allowed vibrational modes of the sample appear, albeit with variable intensity. Each base (A, C,

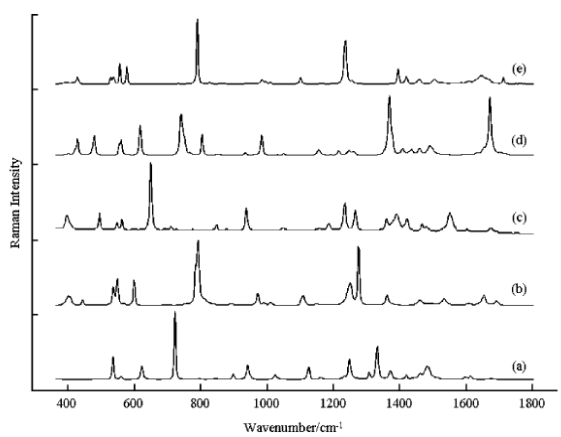


Figure 4.1: Raman spectra of DNA and RNA bases, ex=785 nm, X 100, P=45 mW, acquisition time 60s, from De Gelder et al [109]) a) adenine; b) cytosine; c) guanine; d) thymine; e) uracyle

G, T, U) has a specific Raman spectra, as presented Fig. 4.1, exhibiting specific "marker bands". However, Raman signals are naturally weak, and there are two basic approaches to overcome this weakness: one is based on the use of an appropriate incoming exciting laser wavelength, near or within the energy of an electronic transition of the molecule (260 nm for DNA or RNA), causing an enhancement of several orders of magnitude (10^6) of the Raman intensity (this is Resonant Raman Scattering RRS). The other is based on the adsorption of the probed molecules on a rough metal surface (typically silver, gold, copper...), with adapted roughness

in the nanometer size range (e.g. colloids): the resultant Raman intensity of the adsorbed molecules, obtained under proper illumination, may exceed by several orders of magnitude the corresponding bulk Raman spectrum. This effect is known as Surface Enhanced Raman Scattering (SERS) ([110, 111]).

Our challenge is to detect this specific Raman marker band for each base, at low concentration, using SERS spectroscopy, in order to be able to quantify the amount (percentage) of each base present in single strand, and to reveal the base not involved in a canonical interaction during the process of hybridisation. Our objective is to detect directly the marker bands of each base. In the literature, the SERS detection of the process of hybridisation is always done using a probe [112] . In some cases, the SERS effect becomes stronger because the frequency of the excitation light is in resonance with an absorption band of the probe (Surface Enhanced Resonance Raman Scattering: SERRS) ([113, 114]).This probe has a high absorption in the excitation wavelength providing the resonance Raman effect. The sensitivity with SERRS methods is high but the characterisation of DNA or RNA is poor. The common situation in SERS experiments is that single molecular responses are typically obtained by chance with an emitter aggregated in the spacings of silver or gold nanoparticles ([115, 116]). There are only two published demonstrations in 2009 and 2010 of single molecule SERS by design in gold nanostructures ([117, 118]). However, these examples do not work as label-free sensors since the target molecules are built directly in the nanostructure. The measured SERS enhancement is of the order of 10^{10} which is what is expected in a 1 nm spacing between gold particles of 80 nm of diameter excited at plasmon resonance. There is, in practice, no reported demonstration of single- molecule SERS measurements following the specific binding of a target molecule on a recognition site as it would potentially be the case using gold functionalized DNA origamis. In fact, as shown in recent careful experiments ([119]), eventhough the number of Raman active molecules per metallic nanoparticle in a typical SERS experiment can be relatively high (10^3 – 10^4), the number of molecules at the origin of the observed signal is believed to be much lower (probably of the order of one).

Synthetic strategies to produce gold nanostructures around with DNA ligands have been developed over the last decade after the pioneering works of C. A. Mirkin and coworkers ([120]) and A. P. Alivisatos and coworkers ([121]). In particular, it is possible to obtain well-defined NP grouping geometries by controlling precisely the number of DNA single strands grafted on each gold particle through electrophoretic purification ([122, 123]). This allows combining gold particles with other nano-objects (such as quantum dots, [124]) on the same template or producing symmetric 1D ([125]) or 2D ([126]) NP lattices. However there are three major difficulties when producing DNA templated NP groupings for SERS applications : (i) obtaining short particle spacings to enable large field enhancements (ii) using a rigid DNA scaffold to control precisely the grouping geometry (iii) introducing specific binding sites on electromagnetic hot- spots. S. Bidault and coll. recently demonstrated that reproducible $\approx$ 1 nm NP spacings could be obtained in groupings of 5, 8 or 18 nm diameter particles using two DNA double strands (see

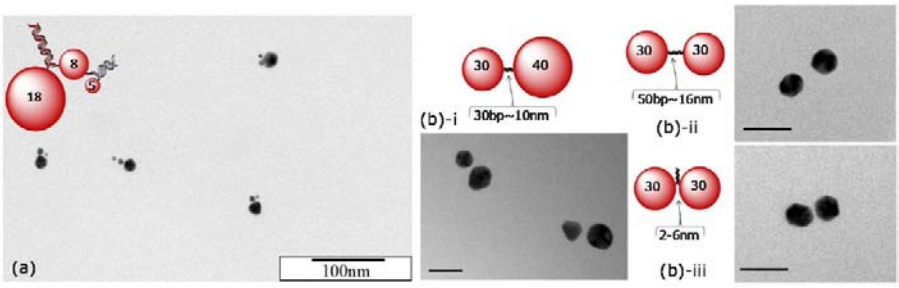


Figure 4.2: From S. Bidault unpublished work. TEM images of gold NP groupings (a) 5 / 8 / 18 nm groupings assembled on two independent 50 base and 100 base long DNA double strands in a "head to tail" geometry. (b) 30 / 40 nm groupings linked with 30 base or 50 base long DNA strands in "head to head" (i / ii) or "head to tail" (iii) geometries (scale bar is 50 nm).

Fig. 4.2 a) ([127]). Using several particles of different sizes allows optimizing field enhancement effects at one specific position of the structure (at the surface of the smallest NP) ([128]). This geometry still presents two drawbacks : it is not possible to tune the spacing lengths and the three particles cannot be properly aligned to optimize field enhancements ([128]).

In order to tune NP spacings, Bidault recently managed to graft a controlled number of DNA single strands as short as 7 nm on particles as large as 40 nm in diameter [129]. After hybridisation of these building blocks, it is possible to obtain the symmetric or asymmetric NP dimers shown on Fig. 4.2 b with spacings ranging from $\sim$ 2 nm to $\sim$ 15 nm (unpublished result). DNA origamis are good candidates to overcome the last two issues : controlling precisely the nanostructure geometry and introducing specific binding sites. The experimental conditions for the observation of single molecule SERS spectra have been carefully studied in several recent publications ([130, 119]). From these and other studies, it becomes clear that the existence of a narrow gap of the order of 1nm between metallic NPs is a necessary condition for the observation of single molecule SERS, at least for Raman dyes such as Rhodamine G that exhibit strong Raman cross-section. However, in these studies, the metallic NPs are uncoated. In the present project, we need to stabilize the NPs with a coating of approximate width 1nm. This implies that the minimum gap achievable in our approach is around 2nm. It has been shown ([131]), using numerical simulations based on multipolar expansions, that single molecule SERS is still possible provided the radius of the NPs is larger than 10nm. Moreover, the Raman effect can be significantly improved by forming branched arrangements of metallic NPs.

Several works involving oDNA and metallic NP to metallize a whole oDNA are already reported. The principle is to use NP as seeds and then to use metallizing techniques to cover the whole oDNA with different metals. The objective is to incorporate oDNA to electronic circuits. In these works NP of different composition and of small size are first attached to the oDNA, then the metallisation of all the structure by different metals is achieved by using the NP as seeds for growing the metal. Gold NP were used to metallize oDNA with gold [132] [133],

or to metallise it with silver [134]. Other works involved silver seed for the silver metallization [135] and gold metallization [136]. Finally metallisation of oDNA with Pb was also achieved [137]

In a work closer to our objectives, oDNA were used to structure NP. In [138] Ding et al observed a shift of the plasmon resonance with NP of different size (5,10 15 nm of diameter) in a specific design on an oDNA. Helical metallic structures were also achieved with two different approaches in Kuzyk et al [139] and Shen et al[140] works. Both observed through circular dichroism measurements a spectra having the characteristic of the metallic helical structure formed. In Kuzyk et al [139] metallisation of the structure also increased the observed effect. These three works proves that oDNA are good templates to achieve complex NP structures. Plasmonic structures of higher efficiency can be achieved by using particles of different sizes and shape. A bestiary of all the NP available as well different method to attach them in ordered DNA array can be found in the review by Tan et al[141]. Asymmetric structures are also reachable through the use of asymmetric nanoclusters, leading to complex Janus structures [142].

Once a NP structure has been created it is possible to predict the enhancement in the signal that it will create. We will now introduce more in detail the Raman phenomenon as well as the SERS principle to evaluate the gain in signal obtained with a given structure.

4.3 Introduction to Raman phenomenon and SERS technique.

4.3.1 The Raman phenomenon:

The Raman phenomenon can be explained by classical electromagnetic theory. In the following discussion, all the quantities are simplified and considered as scalar quantities. Consider an emitted electromagnetic wave $E = E_0 cos(\omega_0 t)$ on a diatomic molecule $q = q_0 cos(\omega_v t)$ with q_0 the equilibrium distance between the two atoms. The electrical dipole of the molecule can have two contributions: a permanent dipole (Eq. 4.1) and a dipole induced by the incident electromagnetic wave (Eq. 4.2):

$$\mu = \mu_0 + \left(\frac{\partial \mu}{\partial q}\right)_0 q \tag{4.1}$$

$$p = \alpha E = \left(\alpha_0 + \left(\frac{\partial \alpha}{\partial q}\right)_0 q\right) E \tag{4.2}$$

The infrared spectrum of the molecule is related to the permanent dipole whereas Raman phenomenon is related to the induced dipole. We will only consider the induced dipole and develop Eq. 4.2:

$$p = \alpha E = \alpha_0 E_0 cos(\omega_0 t) + q_0 E_0 \left(\frac{\partial \alpha}{\partial q}\right)_0 cost(\omega_0 t) cos(\omega_v t) \tag{4.3}$$

and by using a classical trigonometric relation

$$p = \alpha E = \alpha_0 E_0 cos(\omega_0 t) + \frac{1}{2} q_0 E_0 \left(\frac{\partial \alpha}{\partial q}\right)_0 (cos\left[(\omega_0 - \omega_v t)t\right] + cos\left[(\omega_0 + \omega_v t)t\right]) \tag{4.4}$$

The three terms are respectively at the origin of the elastic scattering, the Raman Stokes scattering and the Raman Anti-Stokes scattering. Classical Raman experiments consist in a laser of excitation with a fixed wavelength, a sample and a detector for the emitted photon.

Typical spectra represent the number of photons emitted as a function of their frequency. The frequency reported is the shift in frequency with respect to the frequency of the laser source. The usual recorded frequencies are detailed in Fig. 4.3. As observed in Fig. 4.1 each base has a unique spectra composed of several maxima which make the identification of every component possible in a global spectra. However typical Raman signals are very weak leading to a need of amplification for the detection of small quantity of molecules.

4.3.2 The SERS phenomenon:

The Surface Enhanced Raman Spectroscopy (SERS) is based on the same physical effect that Raman spectroscopy. The higher intensity of the signal is explained by the fact that the molecule

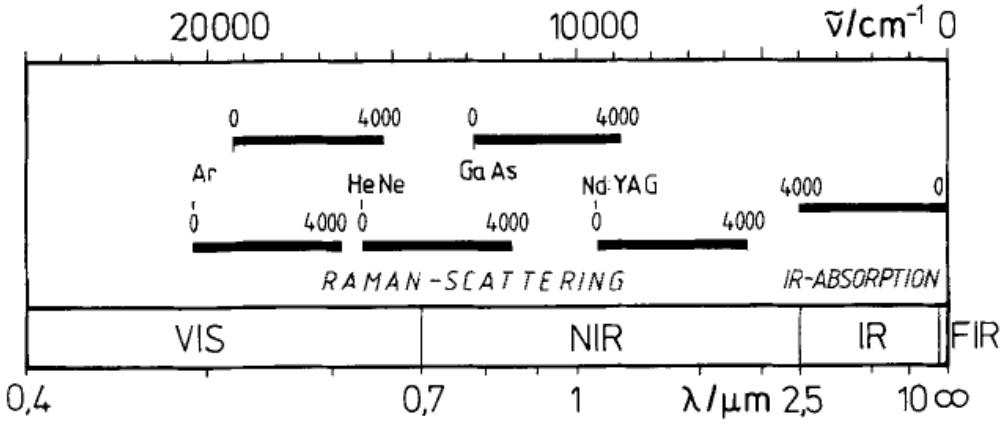


Figure 4.3: from [143]: The visible, near, middle and far infrared region of the spectrum drawn in a scale linear in wavenumbers. The infrared (IR) and far-infrared (FIR) spectrum is recorded by absorption of light from a continuous spectrum in the range of $\lambda = 2.5 \ldots 100\mu\text{m} \equiv \tilde{\nu} = 4000 \ldots 100\text{cm}^{-1}$ and $\lambda = 100 \ldots 1000 \equiv \tilde{\nu} = 100 \ldots 10\text{cm}^{-1}$. Raman spectra can be excited by monochromatic radiation, emitted by different lasers in the visible (VIS) or near-infrared range (NIR). Molecules emit Raman lines with a frequency difference $\Delta\tilde{\nu}$ to that of the exciting frequency $\tilde{\nu_0}$ between 0 und + 4000 or - 4000 cm^{-1}. Usually only the Raman spectrum which is shifted to smaller wavenumbers, the 'Stokes' Raman spectrum, is recorded. Its range is indicated by bars for different exciting lines: Ar$^+$ laser at 488 and 515 nm, HeNe laser at 623 nm, GaAs laser at 780 nm, and Nd:YAG laser at 1064 nm.

of interest is no more excited directly by the electromagnetic field. Instead a metallic structure is excited and will create locally a higher electric field that will be the incoming electromagnetic wave for the molecule of interest [144]. The initial electromagnetic wave has to be close to the plasmon excitation frequency of the metallic structure for the effect to be maximal.

For simple circular metallic particle the polarizability [110] of a bead of radius R is given by:

$$\alpha = \frac{R^3(\varepsilon_b\omega^2 - \omega_p^2) + i\omega\gamma\varepsilon_b}{((\varepsilon_b + 3)\omega^2 - \omega_p^2) + i\omega\gamma(\varepsilon_b + 3)} \tag{4.5}$$

with ε_b the contribution of interband transitions to the dielectric function, ω_p the metal plasmon resonance and γ the electronic scattering rate. γ is large for poorly conductive metal reducing the amplitude of the resonance. Also, considerations on ε_b allow the classification of silver > gold > cupper for SERS experiments. In SERS experiment, the cross-section for such NP is given by Gosh et al [145]. It is proportional to the cube of the radius and the resonance frequency will depend on the dielectric constant of the medium surrounding it. The intensity of the resonance being driven by the same consideration than for the polarisation. As the phenomenon is dependent of the radius to the power of three one would wonder why not

to increase the size of the NP indefinitely. The reason is that for this phenomenon to occur the size of the particle has to be inferior to the initial wavelength. As a consequence the maximum size for the particle is around 200nm. Though important enhancement can be achieved by a single particle, the enhancement achieved with two close-by particles can be of the order of 10^{11} for the Raman signal of a molecule in the middle of the two NP. Coupling two beads also shifts the plasmon resonance of the coupled system. This shift decreases exponentially with the separation to diameter ratio [141].

Detailled calculation of the enhancement factor for different size of particles and different spacings can be found in Xu et al work [146]. The total photon flux in the SERS experiment being related to this enhancement factor by

$$\Phi = \frac{2\pi}{\omega_I}\sigma_{RAM} I_I \sum_i^N M_i^{EM} M_i^{Ch} (4.6)$$

N being the total number of particles, σ_{RAM} the cross-section of the molecule, M_i^{EM} and M_i^{Ch} respectively the electromagnetic and chemical enhancement. By using boundary charge method [141] the maximal theoretical enhancement is calculated for gold NP of radius 10 nm ($M^{EM} = 10^6$).For a radius of 50 nm the enhancement achieved is 10^{11} for a spacing of 1 nm and for a wavelength of $\approx$ 600 nm. The origin of the chemical enhancement is not clear and is specific of the metal-molecule complex. This factor is added because electromagnetic effect alone can not explain enhancement of the order of 10^{14} reported in the literature.

Once the different parameters that play a role in the enhancement are known it is possible to understand and create adequate structures to detect DNA. The following section reports the experimental progress made in this direction.

4.4 Experimental results:

On the works coupling NP and oDNA [139, 138], the classical approch to attach NP in a controlled manner is to select the different places where to attach the NP on the oDNA and to extend the oDNA staple at this point with a DNA sequence (Seq_1) at least 15 bases long. Several close staples can be modified to have more than one point of attachment. Then thiol modified strands complementary to the sequence Seq_1 are attached to the NP. Usually the NP are covered by the thiol modified strands. Then the oDNA and the NP with the strands are mixed together.

In our case the target of interest is ssDNA. As a consequence it is not possible to have several ssDNA strands attached to the NP, as the signal of these strands would interfere with the detection of the strand of interest. In consequence we applied a protocol established by Sebastian Bidault to obtain NP with only a single ssDNA attached to it. The ssDNA is attached to a trithiol group (Fig .4.4) at its 5' extremity. The advantage to use a trithiol compared to a thiol is a higher strength of interaction.

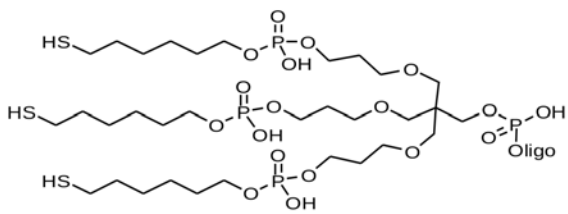


Figure 4.4: 5' extremity

The protocol to obtain NP with only a single ssDNA attached at the surface is detailed in the following section.

4.4.1 Protocol for coupling NP with ssDNA

From available commercial solution Au NP of 10 nm radius are obtained at a concentration of 0.01 nM. We need to concentrate the solution first in order to be able to use it with oDNA whose concentration is $\approx$ 1 nM. The solution (50mL) is centrifugated at 7500rpm during 30 minute (2 or 3 times) leading to higher concentrations in the bottom of the solution. The more concentrated part is preempted and the solution is once more concentrated by centrifugating in smaller tubes. Once the volume has been reduced to a hundred of µL a solution of BSPP (5mg BSPP for 20 mL mQ) is added $\approx$ 100 µL and the solution is concentrated once again. This operation has to be repeated 2 times to remove the initial buffer. The BSPP solution is used in order to stabilise the NP. Indeed at low concentration the particles are stable but when increasing the concentration there is a high risk of aggregation. These steps of concentration allow to reach a NP concentration of about $\approx$ 100 nM.

The trithiolated strand whose sequence is complementary to the free part of the staple on the oDNA is then hybridised on 10 bases with a 50 bases long DNA strand. The purpose of this strand is to allow the separation of the NP with one strand from the NP with two or more strands when running a gel. For NP of higher radius, more than one 50 bases long strands have to be added.

Then the NP are mixed with the strands during one night. For size of NP superior to 10 nm the NP have to be shacked in order to prevent the sedimentation. Thirty minutes before running the gel the NP are mixed with a PEG solution (M=356g/mol) that will protect the NP from aggregation in the oDNA buffer. Then a 2% agarose gel preparated with a 0.5 TBE buffer is run to separate NP with no strands to NP with one and two strands. (Fig. 4.5)

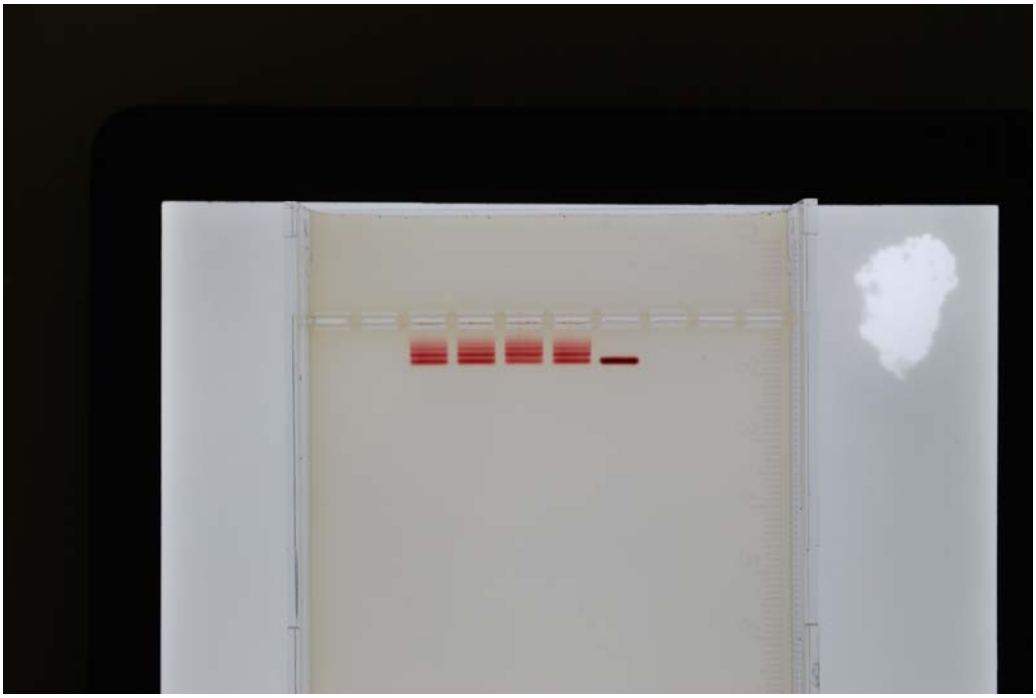


Figure 4.5: Agarose gel after separation. Several bands can be observed. The right track contains NP alone. All the other tracks contain NP with zero, one, two and more strands

Finally a slide of gel is removed in front of the NP with one strand, replaced by the buffer, and the current is turned on. The NP continue to migrate and go in the buffer. The buffer with the NP is then removed and the operation is repeated until no NP are left in the gel. The solution of NP obtained is then concentrated by centrifugating. In our case when binding to the oDNA a small annealing protocol was carried out to remove the 50 bases and to improve the pairing with the oDNA.

The next step of the experiment is to build the SERS structures by attaching the NP to the oDNA.

4.4.2 Attachment of the NP to the origami.

As already explained, the classical method to attach any "object" on a oDNA is to extend one staple with a small sequence, and then to mix the oDNA with the object of interest. Since this object is attached to the complementary sequence, a reaction of hybridisation takes place. The object is positioned with a high precision but can still fluctuate in solution because of the flexibility of single dsDNA.

First we studied the attachment of a single NP to an oDNA where only on site of attachment is possible. We compared different values of the ratio origami/Au NP to evaluate the proportion of well formed structures by AFM. These experiments are gathered in Tab 4.1. A 10 fold excess gave a yield of 60 percent.

Name	Yield (%)	C_{oDNA}	Excess Au	Annealing procedure
PTA1G1	8	0.5 nM	1	20 min 40°C; 40 to 24 (40 min)
PTA1G2	60	0.5 nM	10	20 min 40°C; 40 to 24 (40 min)

Table 4.1: Yield of structure having one bead with different NP excess.

We then worked on structure with three separated binding positions. These positions were chosen far enough on the oDNA in order to be able to evaluate rapidly the number of beads by structure. The results are gathered in Tab. 4.2. Differents annealing procedures were followed showing the importance of a long enough annealing reaction.

Name	Yield[1] (%)	C_{oDNA}	Excess Au	Annealing procedure
PTD2G1	0.1	0.34 nM	3.9	20 min 40°C; 40 to 24 (40 min)
PTD2G2	0.5 to 0.9	0.34 nM	3.9	then 20 min 40°C; 40 to 24 (80 min)

Table 4.2: [1] Yield of the structures having at least one bead. This experiment shows that a slow cooling process improves the result

A study of the number of beads by origami gave the following percents: 30 % without beads 30 % with one bead 30 % with 2 and 10 % with three beads. This statistic was established on a hundred of oDNA structure.

We finally concluded this set of experiments by deciding to work with a 10-fold excess of NP per binding site and an annealing procedure from 40°C to 20 °C in 120 minutes (Fig. 4.6). The AFM images show that almost each oDNAs bear at least one bead, and several oDNAs have two beads or more. The result seems to be better when the binding sites are closer.

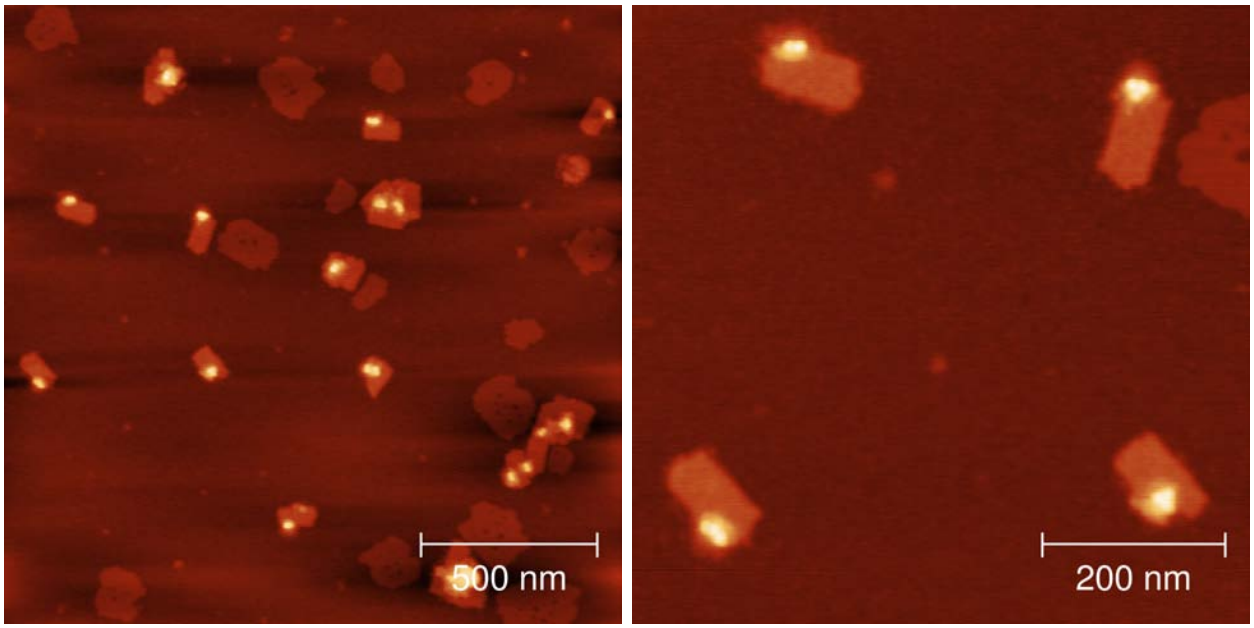


Figure 4.6: AFM images. On the left 3 oDNA have the same orientation with the expected structure if the beads are on the top of the structure. One has the wrong orientation meaning that the beads are in contact with the surface. Surprisingly the oDNA have no clear deformation.

4.4.3 Raman experiment.

4.4.3.1 Schema of principle of the experiment

Given these encouraging results we decided to realise Raman experiments coupled with AFM (Fig. 4.9) to further characterize these constructions.

The schema of SERS oDNA structure is represented in Fig. 4.7. The structure include a single strand having the complementary sequence to the RNA to detect. In the shema Fig. 4.7 the probe is attached to the sequence to detect and is the dsDNA formed by the probe and the target is drawn vertically. The different states with and without the RNA strand to detect are illustrated in Fig. 4.8. In the initial state the probe is partially hybridised with a strand attached to a Rhodamine molecule. This molecule is known for having an important Raman cross section. Thanks to the NP the initial state has a strong Raman signal. In the presence of the target RNA molecule, the strands with the Rhodamine is freed because the RNA has a stronger affinity with the probe, leading to a decrease in the Raman intensity.

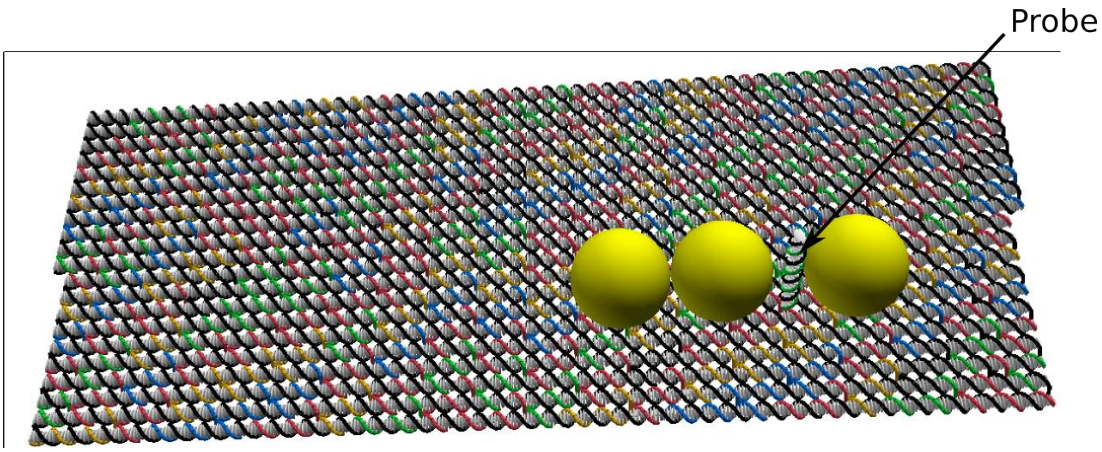


Figure 4.7: Schematic representation of the SERS active structure. The structure is composed of the oDNA, three gold NP of 10 nm and a probe

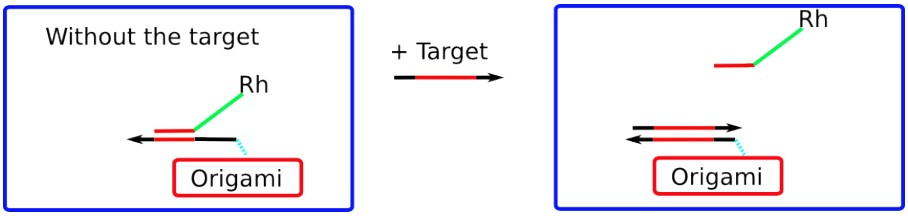


Figure 4.8: State of the origami without the probe (left) and with the probe (right)

Two wavelengths were available to realise the SERS experiments: 514 nm and 633 nm. We choose the 633 nm wavelength to excite the plasmons as single Au particles absorb at 540 nm

and the plasmon wavelength of excitation for two close particles is bigger than the one of the particle alone (Cf [129] for plasmon wavelength of coupled Au particle of 36 nm of diameter).

4.4.3.2 principle

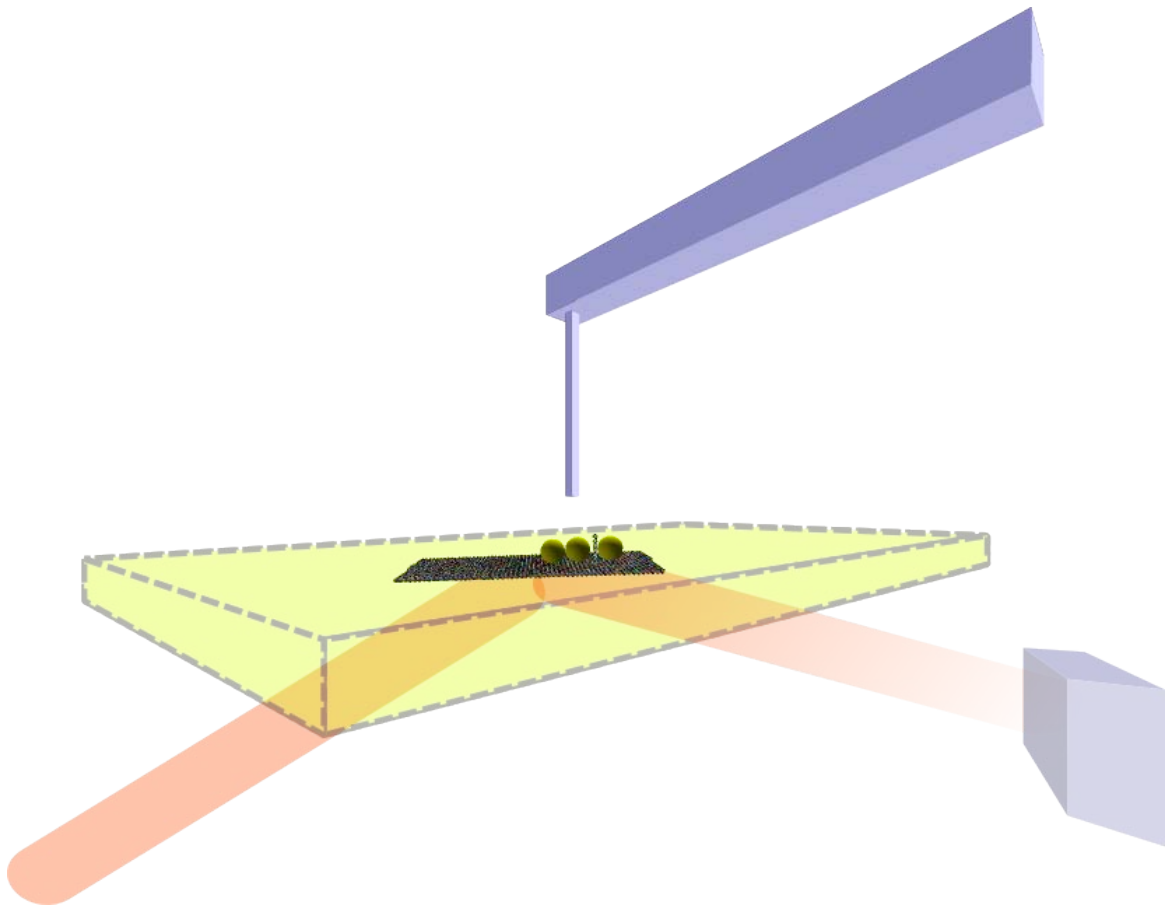


Figure 4.9: Experimental setup: The oDNA are deposited on a transparent surface. They are then detected by afm, and a Raman spectra is collected by reflection.

Two type of experiments were envisioned: a classical Raman experiment in solution, and a Raman experiment coupled with an AFM apparatus.

The principle of the second is to deposit the sample on a transparent surface. The oDNAs are detected by classical AFM technique. Once an oDNA with NP is found, the laser of excitation is focused on the tip of the AFM. The tip is mainly composed of silicium (Si), the intensity of the Raman spectra of the Si help to focus the laser on the bottom of the tip. Once this step done, the emitted light is collected by a detector.

One difficulty of this technique was to deposit the oDNA on a transparent surface. We

choose quartz surfaces that are flat up to $\lambda/10$ because the surface has to be plan enough in order to be able to detect oDNA with AFM. The protocol followed to deposit oDNA on a quartz surface is detailed as follows.

4.4.3.3 Attachment of oDNA to a quartz surface

To attach DNA to a surface this one has to be charged. Mica surfaces that are the classical support for imaging DNA are negatively charged. Mg ions or positively charged ions in solution form a bridge between the mica and the oDNA. Unlike Mica sheets, the quartz surface has to be treated to charge it before depositing oDNA on it.

The following procedure was chosen to negatively charge the surface:

- The surface was immersed in a detergent for 2h at 50 °C
- immersed for 12h in miliQ water
- immersed for 3h in ethanol.
- dried with a nitrogen flux.

3µL of the solution containing oDNA and gold NP is dropped for 1 minute on the quartz then washed with the buffer solution, and then dried with a kimsweep. The surfaces obtained were very in-homogeneous on origami concentration, some parts having no oDNA and others being densely covered (Fig. 4.10). Once the structure and the deposit were obtained the Raman experiment was carried on.

4.4.3.4 Raman experiments

We first realised coupled AFM SERS experiments on quartz surface, but no correlation between the oDNA and Raman spectra of the PEG molecules was observed. In fact no Raman spectra excepted the one of the quartz surface was detected. As this experiments are more complex than simple Raman spectra, we switched to bulk experiments , where the oDNA are in a drop and the spectra is gathered by reflection on the surface where the sample is dropped. Three beads were positioned closely on the oDNA.

From the AFM observation of the dry sample, the beads seem to be touching each other. The expected Raman spectra correspond to the Raman signal of the PEG molecules as they are positioned between the NP and to Raman signal of the Rhodamine molecule. In solution the water molecule dominates the signal. Once the sample dried, the signal observed were coming from the different compounds constituting the buffer (Fig. 4.11). None of the observedpeaks can be related to the spectra of the PEG molecule (Fig 4.12 a) or to the Rhodamine (Fig 4.12 b). (The Raman spectra of all the individual chemical compounds used in all the experiments are gather in Appendix K).

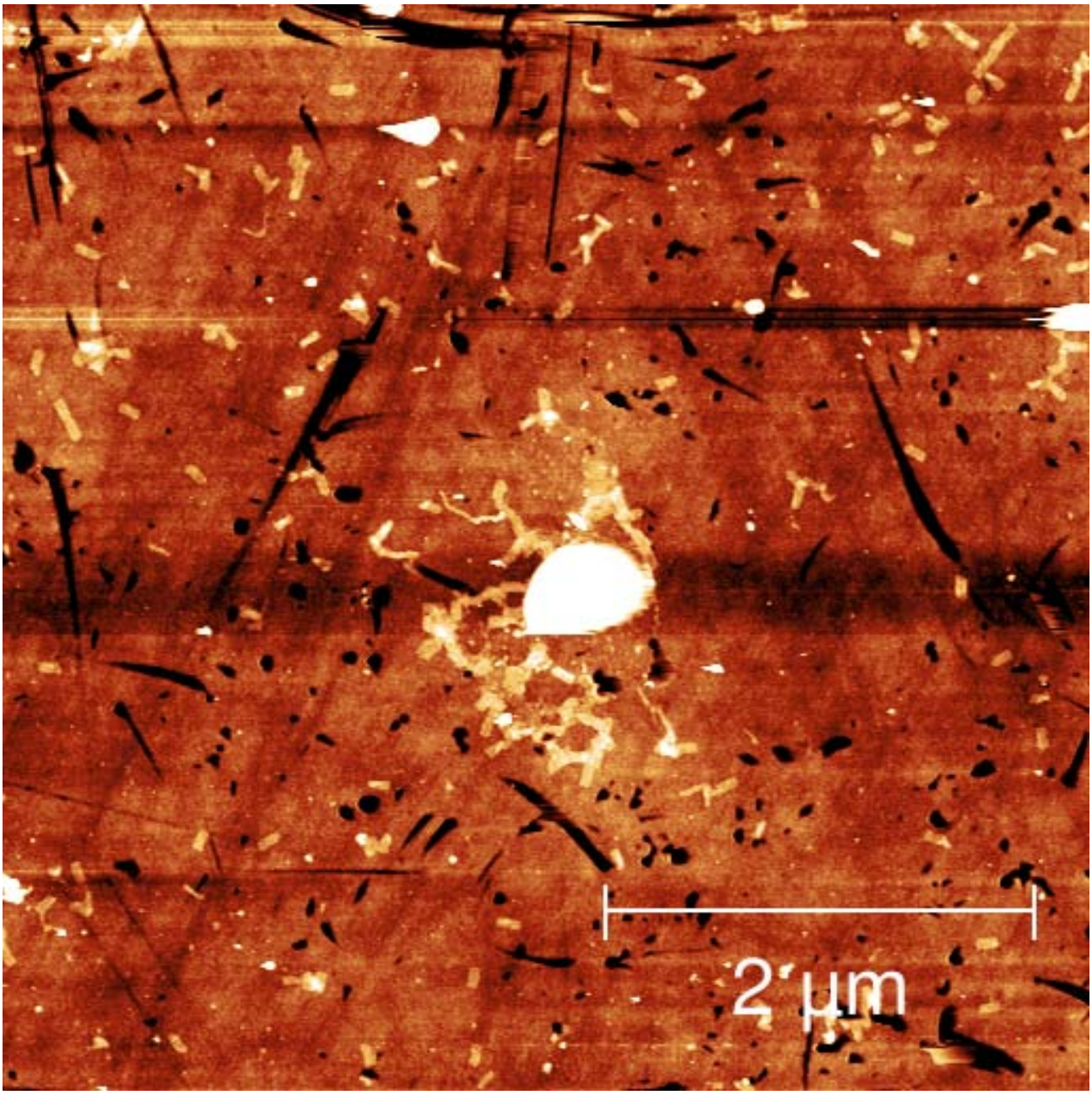


Figure 4.10: AFM images of oDNA attached to quartz surface

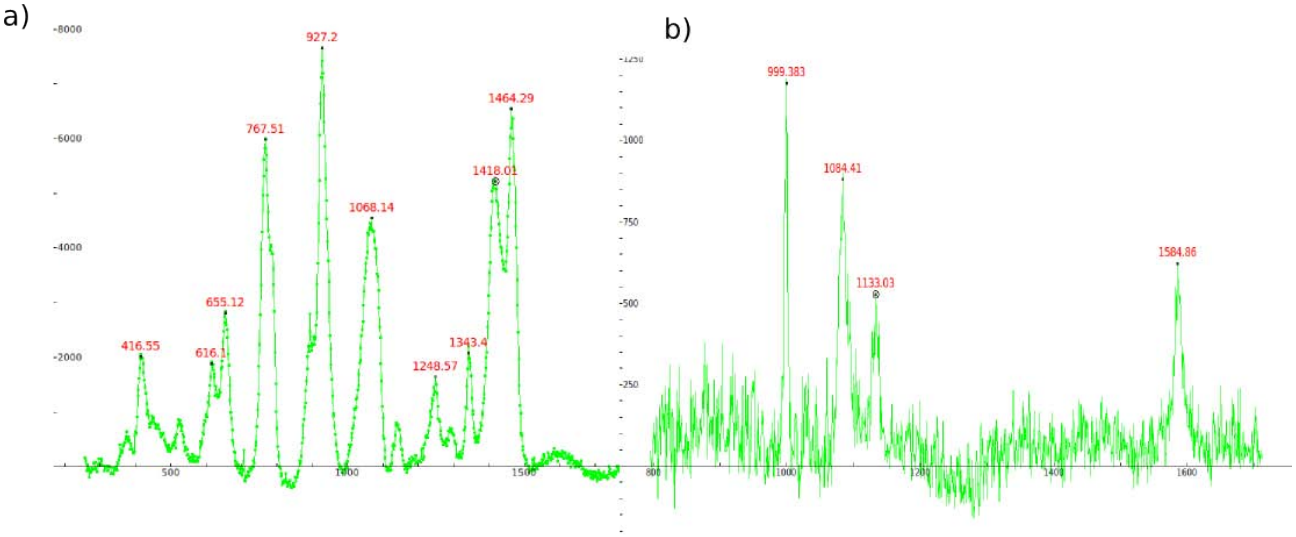


Figure 4.11: Spectra observed for in our Raman experiment. The unit of the axis are cm^{-1} ;a) TAE buffer Bottom right BSPP buffer

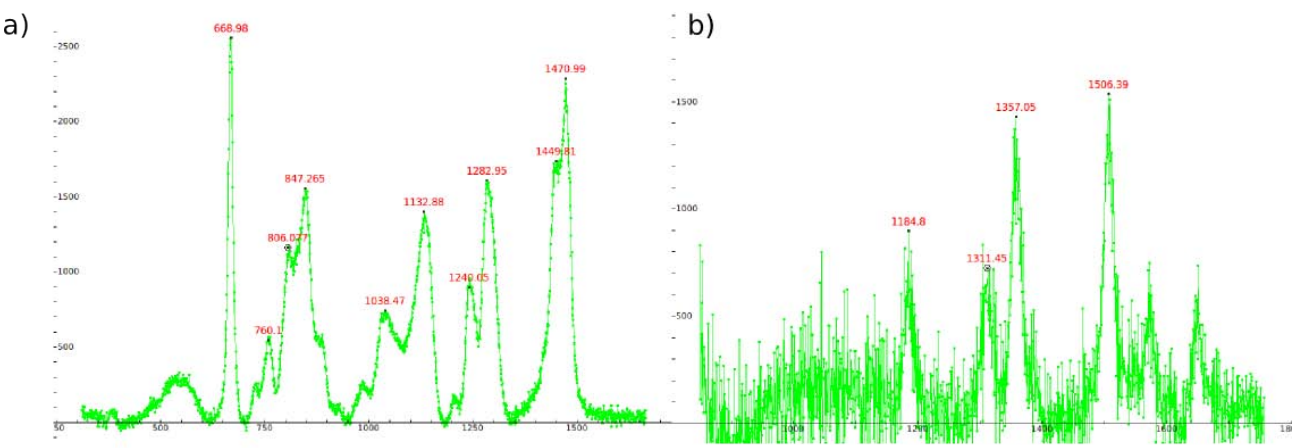


Figure 4.12: a) Raman spectra of the PEG molecule alone; b) of the rhodamine 6G (The unit of the axis are cm^{-1})

The fact that the spectra of the PEG molecule or the Rhodamine was not observed means that its signal was too small compared to the signals of the other compounds in solution and to the noise of the experiment. The signal coming from the origami is proportional to its concentration (1nM) , to the cross-section of the molecule and to the enhancement factor. This last can be estimated to be of the order of 10^2 - 10^3 (Fig. 4.13) because the PEG molecules are preventing the NP from being too close ($\approx$ 2 nm of distance) and the beads are of radius 5nm. So for the PEG or the Rhodamine the signal intensity is of the order of $\sigma 10^{-7}$ (with

σ the Raman cross-section) while for the compounds in the the buffer it is of the order of $\sigma_{compound}10^{-3}$.

The enhancement factor that would be needed to overcome the signal from the buffer for experiment in solution is at least of the order of $10^6 - 10^8$. From Xu at al [146] on can notice (Fig. 4.13) that 10^6 is the maximum theoretical factor of enhancement for NP of radius 10 nm at a distance of 1 nm. Given the fact that the distance between the NP is 2nm we considered working with NP of 15 nm and 20 nm radius. Unfortunatly the NP are no longer stable in the oDNA buffer because of the presence of 25mM of magnesium. Tests on the stability of the NP in a salt buffer revealed that they are not stable in the presence of a mM range of Mg. They were stable for a concentration of 30 mM of Na and aggregated for a concentration of 50 mM. However, at this concentration in salt the oDNA are no more stable. Indeed the electrostatic repulsion between strands become too strong and the oDNA "explode". The stability of the gold particle is related to the size of the PEG that covers it. Increasing the size of the PEG molecule increases the stability of the NP but also increases the distance between them, leading to no real improvement in the enhancement factor.

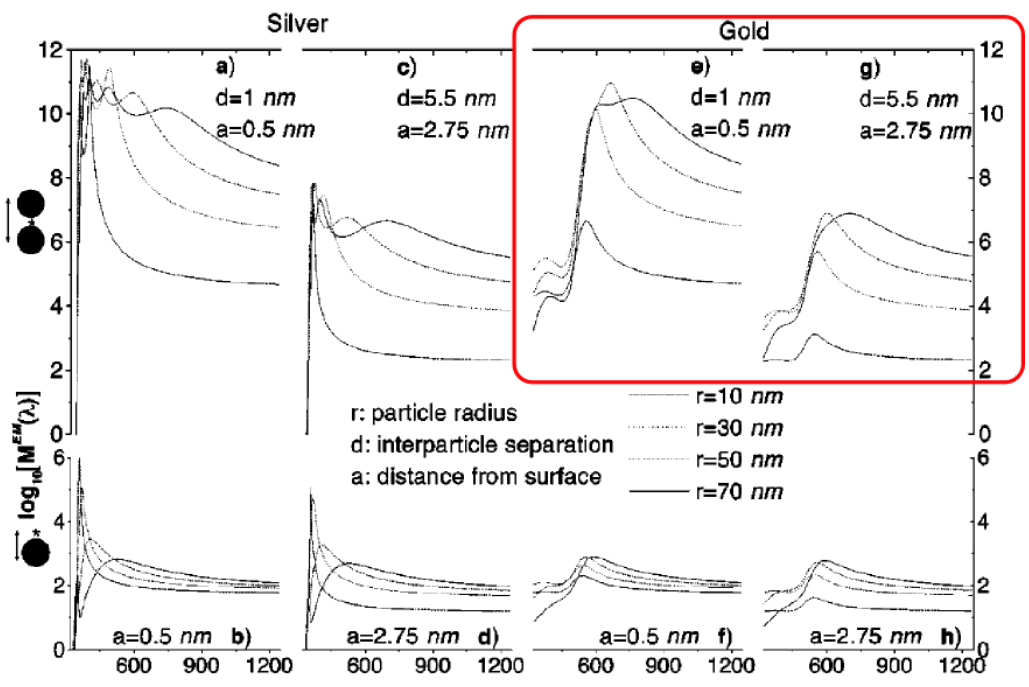


Figure 4.13: From Xu et al [146]: Enhancement factor for silver (left) and gold (rigth)particle. Different parameter are compared. The top plot are the enhancement factor for two beads, and the bottom ones only for one bead.

We then considered another direction which is to increase the size of the NP once they are attached to the oDNA, using a process of metallisation. After attaching the NP to the oDNA

we followed the protocol from Kuzyk at al [139] which consist in adding to the solution three different solutions that will allow the beads to grow. Unfortunately no noticeable increase in size was observed in our experiments. We think that the main difference between our work and the work from Kuzyk is the way NP are stabilised.

In their work NP are covered with ssDNA whereas in our case the NP are protected with PEG molecules. The latter are likely to be more densely packed on the NP as they are smaller. We tried to decrease the quantity of PEG at the surface of the NP by varying the concentration and the time for mixing both compounds. When the concentration of PEG is too low, the NP are no more stable in the oDNA buffer. At the minimun concentration in PEG where the NP are stable we did not observe a growing of the NP.

4.5 Conclusion

Despite good progress in building an active SERS platform we were confronted to the difficulty of the instability of the gold particle in "biological" buffer. A direction that was envisioned but not explored was to protect the gold beads with other compound whose Raman spectra would not interfere with the Rhodamine and that would still allow the gold particles to grow. For example protecting the NP with DNA could be envisioned as long as the strands used to cover the particle would not interfere with the binding of the strand of interest.

Chapter 5

Conclusion

DNA origamis are very flexible tools that virtually allow to couple any nano-constituents together. These DNA platforms with pixel resolution of about 5 nm are already widely used to create several applications that ranges from RNA probes [2] to small assembly line [4]. When realising a SERS platform, we saw that mixing DNA with inorganic material (such as Gold nano-particles) was not that easy as the conditions of common stability of the two constituents were narrow and even do not exist for the biggest NP we have used.

In this work we mainly focused on understanding the mechanical properties and the process of formation of oDNA structures. Regarding the mechanical properties, the main result was that a dsDNA in a flat oDNA structure has a persistence length of 1.5 its natural persistence length. Knowing the persistence length allows to characterise the mechanical properties of these structures and more particularly to estimate the energy and the forces at stake. The fluctuation of the structures in solution could also be derived. In the case of complex 3-D structures a continuous mechanical approach is enough to give a good estimation of their persistence length. The importance of the electrostatic interactions was also highlighted as we showed that without them oDNA structures could be totally huddled up.

The information obtained on the process of formation showed that the correlations with nearby surrounding stapples previously formed is of primary importance. This coupling was highlighted by working on the simplest oDNA structure possible, a template of 64 bp and two staples of 32 bp each, that we named small origami. This is an important general rule that I will try to remember: when trying to understand something, simplify the most. The stability of the oDNA with the temperature could be summarised by: every crossovers decrease the stability of the staple compared to its stability on the duplex form. By keeping the number of crossovers constant and increasing the length of the staple we were able to construct a stabler structure. Stabler structures could be of interest if one wants to work with oDNA at higher temperature

than 20 $°C$. Recent progress have been made in the direction of using oDNA to target specific cells with an aptamer system, and to deliver a cargo once these cells are reached [147] . As the temperature is of 37 °C in human cells, having platforms that remain stable up to 40°C is a necessary requirement to conceive devices for biological applications such as investigating a cell behavior.

The last project of this thesis that was not presented in this manuscript because it was only started lately concerns DNA logic gates. In this project we try to attach DNA logic gate onto oDNA structures. The goals are to increase the rate of reactions by confining them on a small area, and also to increase the complexity of the calculations possible by assembling independent logic modules on oDNA structures, that would lately be combined to build more elaborate combination of logical operation.

Appendix A

List of communications

This work gave rise to two publications:

- Arbona, J. M. Aime, J.-P. & Elezgaray, J. (2012)Folding of small origamis . The Journal of Chemical Physics 136, 065102.
- Song, J. Arbona, J.-M. Zhang, Z. Liu, L. Xie, E. Elezgaray, J. Aimé, J.-P. Gothelf, K. V. Besenbacher, F. & Dong, M. (2012)Direct Visualization of Transient Thermal Response of a DNA Origami . Journal of the American Chemical Society

One publication submitted to Physical Review Letters : Arbona, J. M. Aime, J.-P. & Elezgaray, J. Modelling the folding and melting of DNA-based nanostructures. And two publications are in preparation:

- Liliya A. Yatsunyk , Olivier Peitrement , Delphine Albrecht, Phong Lan Thao Tran, Daniel Renciuk, Hiroshi Sugiyama, Jean-Michel Arbona, Jean-Pierre Aimé, Jean-Louis Mergny. Guided assembly of tetramolecular G-quadruplexes
- Arbona, J. M. Aime, J.-P. & Elezgaray, Modelling the mechanical properties of DNA nanostructures.

In the course of this thesis three poster presentations were done at:

- FNANO 2011 (Salt Lake City)
- Statistical Mechanics and Computation of DNA Self-Assembly 2011 (Finland)
- 2011 DNA Workshop (Munich)

And one oral communication was realised at the NanoSWEC 2010 conference (Bordeaux)

Appendix B

Experimental attempt.

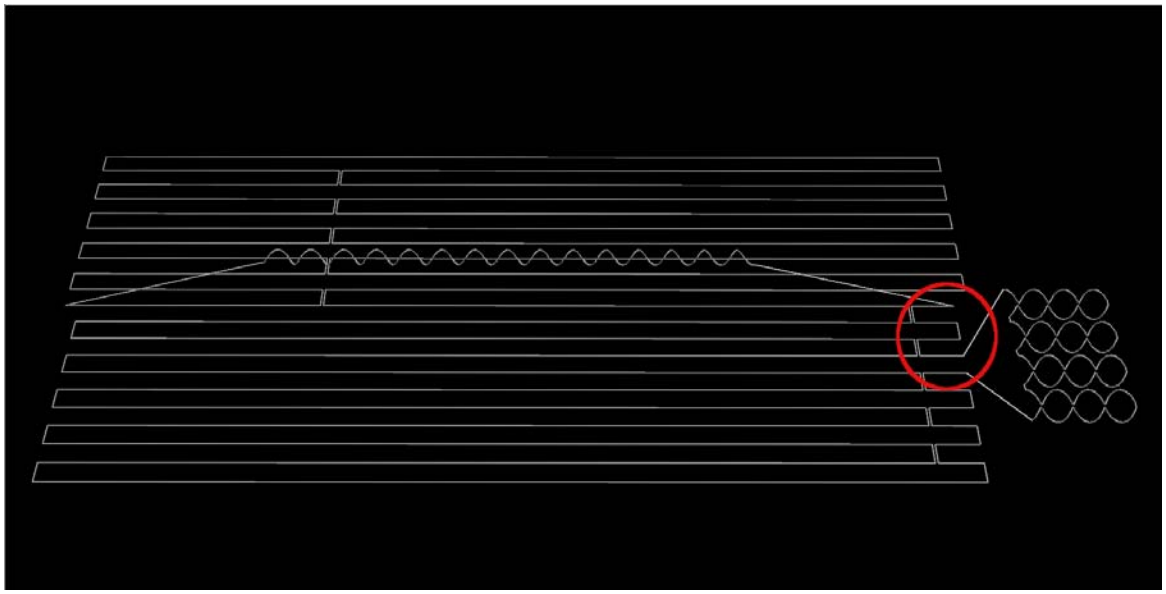


Figure B.1: Schematic representation of the path followed by the scaffold with the presence of a reservoir

The structure that we wanted to create is illustrated in : Fig. B.1. This structure is composed of three parts: an origami layer, a small reservoir, and a bridge. The reservoir is here to modulate the size of the bridge. Indeed the choice of four strands in the red circle, imposes a given part of the virus at this position and so a given size of the bridge. The maximum possible size for the bridge is 700 bases. As calculated in 2.2.2 this ssDNA bridge should exert a neglectable effect on the origami layer. However shorter bridges of size 240 bases (L_c=81nm) should exert a force and bend the origami. Unfortunatly this did not work as expected. The bridge seems to bind to nearby origamis (Fig. B.2). This is probably due to the fact that the

staples which determine the size of the bridge are short (20 bases) and probably fold at lower temperature than the other staples, so that the origamis are already formed when these staples are able to bind. The persistence length of these structures being important the fluctuations are probably small. So the probability to have the origami bent with a high curvature and at the same time the staples coding for the size of the bridge in the vicinity is too small. Furthermore the energy to bend the structure is of the order of 200 kT. This is higher than the energy of folding of the four staples.

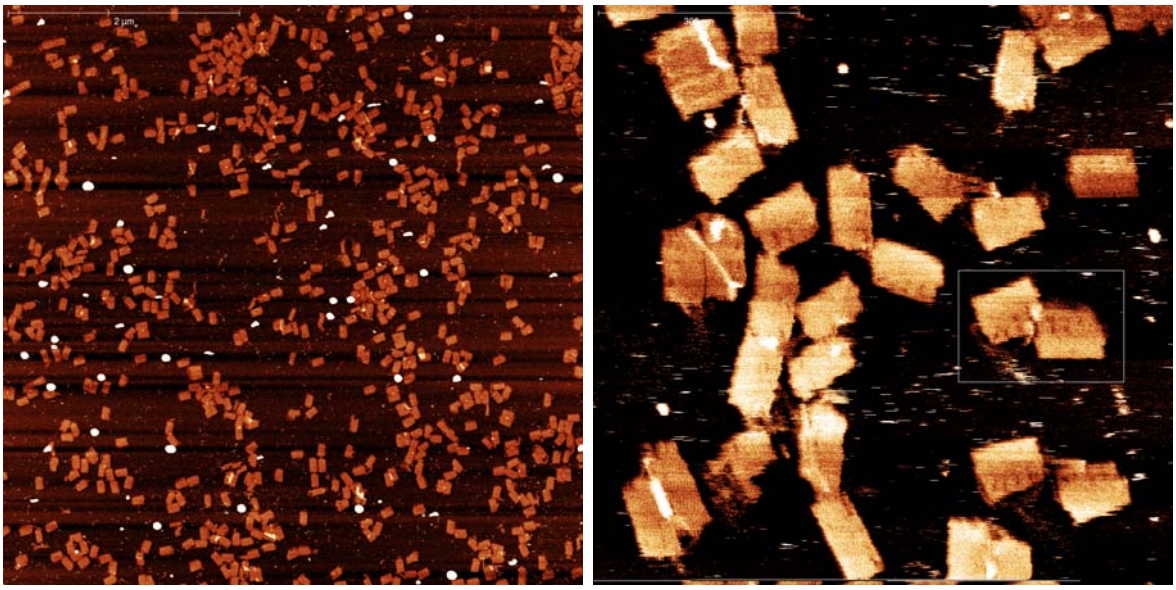


Figure B.2: AFM images of an origami structure with a supposed bridge of 240 bases. (Rigth) Zoom on the structures. On some structures one can notice that some staples had been remove of the origami.

Appendix C

Thermodynamic values for the soDNA

The thermodynamic values were calculated by the fit of the curve $\frac{d\theta}{dT}$ in the hypothesis of a two states hybridisation. The incertitude on the values are given by the measure of different concentration from $0.5\mu M$ to $2\mu M$. Strands with their complementary (c), in the outer position (o), in the inner position (i), when both strands are present (co) and $_a$ means with adaptor. For all the strands δH and δS are experimental values. For the strands with their complementary Δ_0 refers to the experimental value minus the theoretical. For the other case Δ_0 refers to the experimental one minus the theoretical with the complementary strand. The experiement of the adaptator were realised only once and have no errror bar. The error indicated in the last column corresponds to the mean square deviation between the fit of the log of the equilibrium constant as a function of the temperature and the experiemental data, on the range were the constant goes from 0.01 to 0.99.

Mix	δH	δS	ΔH_0	ΔS_0	ΔG_0	ΔT_0	Err
B1 AT (c)	-225.1 ± 7.4	-656.1 ± 23.2	-2.9 ± 7.4	-8.2 ± 23.2	-0.4	0.2 ± 0.6	
B2 GC (c)	-289.3 ± 0.0	-762.5 ± 0.0	-1.9 ± 0.0	2.3 ± 0.0	-2.6	3.5 ± 0.0	
B1m (c)	-238.5 ± 0.0	-651.6 ± 0.0	9.1 ± 0.0	27.0 ± 0.0	1.0	0.6 ± 0.0	
B1 AT (o)	-159.4 ± 11.1	-469.1 ± 35.2	62.8 ± 11.1	178.9 ± 35.2	9.1	-8.8 ± 0.2	
B2 GC (o)	-271.6 ± 2.8	-727.7 ± 9.6	15.9 ± 2.8	37.1 ± 9.6	4.7	-3.7 ± 0.4	
B1m (o)	-250.9 ± 14.0	-703.8 ± 40.4	-3.3 ± 14.0	-25.2 ± 40.4	4.3	-7.8 ± 0.1	
B1 AT (i)	-82.5 ± 4.0	-233.3 ± 11.8	139.6 ± 4.0	414.7 ± 11.8	15.3	-15.8 ± 0.8	
B2 GC (i)	-212.1 ± 35.0	-568.0 ± 100.3	75.4 ± 35.0	196.8 ± 100.3	16.3	-7.9 ± 0.3	
B2m (i)	-218.1 ± 18.5	-603.6 ± 53.7	69.3 ± 18.5	161.2 ± 53.7	20.9	-17.4 ± 0.4	
B1 AT (co)	-176.7 ± 2.6	-515.8 ± 7.8	45.5 ± 2.6	132.2 ± 7.8	5.9	-4.4 ± 0.0	
B1 AT (co)	-152.8 ± 16.2	-445.1 ± 50.3	69.3 ± 16.2	202.8 ± 50.3	8.5	-6.6 ± 0.1	
$B1_a$ (i)	-138.2	-408.6	84.0	238.3	11.8	-13.2	0.3
$B1_a$ (o)	-151.6	-444.5	70.6	202.4	9.2	-8.7	0.2
$B1_a$ $B3_a$ (i)	-135.2	-398.1	87.0	248.7	11.7	-12.7	0.2

Appendix D

Computing the probabilities from the law of mass action.

In this section, we explain how to compute the probability $p(S_i(k,l),T)$ by solving a set of coupled equations which reflect the assumption that the reaction

$$S_i(k,l) + N_\alpha(S_i) \rightleftharpoons N_\alpha(S_i)S_i(k,l)$$

is governed by the law of mass action. For the sake of clarity, we consider the particular case illustrated in Fig. D.1: a staple divided in two parts of equal length is inserted in the vicinity of another staple that holds together a portion of the scaffold. We consider three equilibriums:

The simultaneous binding of the two parts of S_i, with equilibrium constant:

$$K^{N_\alpha(S_i)}_{S_i(1,2)} = \exp\left(-\frac{\Delta G_{NN}(S_i(1,2),T) - T\Delta S_{bulge}(2L)}{kT}\right) \tag{D.1}$$

The binding of only one half $S_i(1,1)$ of the staple, with equilibrium constant:

$$K^{N_\alpha(S_i)}_{S_i(1,1)} = \exp\left(-\frac{\Delta F_{NN}(S_i(1,1),T)}{kT}\right) \tag{D.2}$$

The binding of the other half $S_i(2,2)$ of the staple:

$$K^{N_\alpha(S_i)}_{S_i(2,2)} = \exp\left(-\frac{\Delta F_{NN}(S_i(2,2),T)}{kT}\right) \tag{D.3}$$

More generally, for any set of equilibriums between S_i and its neighborhood $N_\alpha(S_i)$, the law of mass action reads:

$$\frac{[S_i N_\alpha(S_i)]}{[N_\alpha(S_i)][S_i(l,p)]} = K^{N_\alpha(S_i)}_{S_i(l,p)} = exp\left(-\frac{\Delta F(S_i(l,p),T+dT)}{k(T+dT)}\right) \tag{D.4}$$

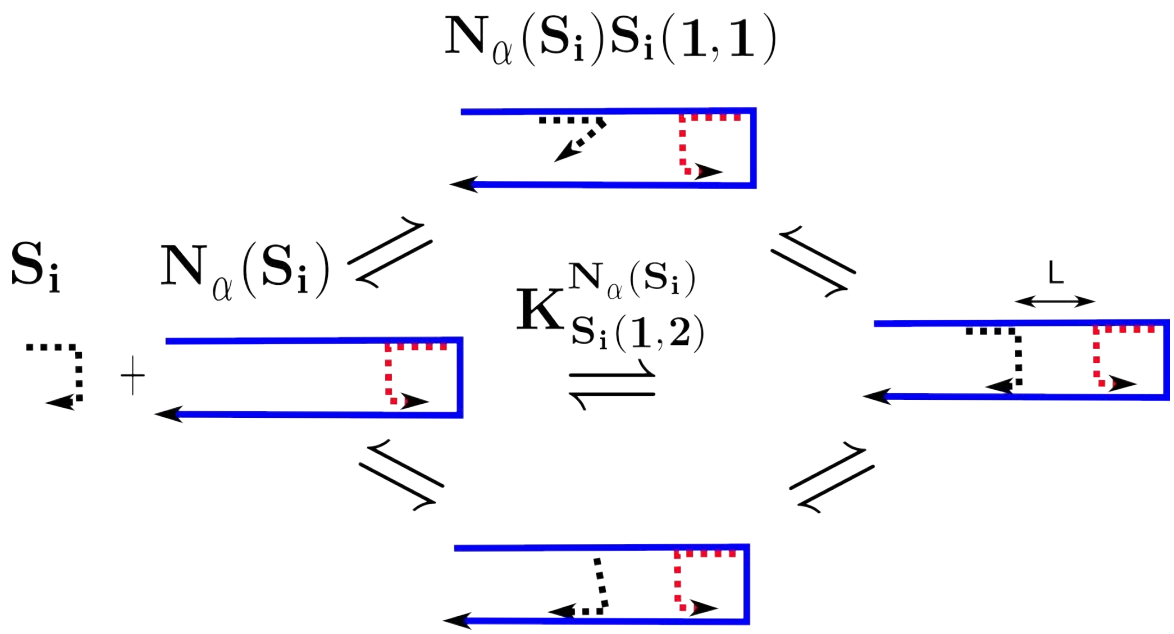


Figure D.1: Different possibilities for the staple S_i to bind in the neighbourhood $N_\alpha(S_i)$

The concentration of free staple in solution is given by

$$C_0\left(Exs - \sum_\beta \sum_{u,v,u\leq v} p(S_i(u,v)|N_\beta(S_i))p(N_\beta(Si))\right) \tag{D.5}$$

Exs is the excess of staple concentration compared to the initial concentration of scaffold C_0, $p(S_i(l,p)|N_\alpha(S_i))$ the conditional probability to observe $S_i(l,p)$, given the neighbour staples $N_\alpha(S_i)$ and $p(N_\alpha(Si))$ the probability of having the neighborhood N_α at the previous temperature step. As the excess is important, the concentration will variate weakly in the process of formation. It is then simplified to C_0Exs. This simplification is necessary to obtain a set of uncoupled equation. The concentration of N_α free in solution is given by:

$$C_0\left(p(N_\alpha(Si)) - \sum_{u,v,u\leq v} p(S_i(u,v)|N_\beta(S_i))p(N_\beta(Si))\right) \tag{D.6}$$

And the concentration of state $S_i(l,p)$ in $N_\alpha(S_i)$ is given by

$$C_0 p(S_i(l,p)|N_\beta(S_i))p(N_\beta(Si)) \tag{D.7}$$

The equilibrium constant gives:

$$K^{N_\alpha(S_i)}_{S_i(l,p)} = \frac{p(S_i(l,p)|N_\alpha(S_i))}{ExsC_0\left(p(N_\alpha(Si) - \sum_{u,v,u\leq v} p(S_i(u,v)|N_\alpha(S_i))p(N_\alpha(Si))\right)} \tag{D.8}$$

This set of equations can be solved by noticing that the denominator in D.8 is independent of $S_i(l,p)$. Thus, for any couple of configurations $(S_i(l,p), S_i(u,v))$,

$$\frac{p(S_i(l,p)|N_\alpha(S_i))}{p(S_i(u,v)|N_\alpha(S_i))} = \frac{K_{S_i(l,p)}^{N_\alpha(S_i)}}{K_{S_i(u,v)}^{N_\alpha(S_i)}} \tag{D.9}$$

Let $y_\alpha = \sum_{l,p,l\leq p} p(S_i(l,p)|N_\alpha(S_i))$ and $S_\alpha = \sum_{l,p,l\leq p} K_{S_i(l,p)}^{N_\alpha(S_i)}$. Then, by summing over every l,p eq. D.8 one obtains:

$$\frac{1}{C_0 Exs}\frac{y_\alpha}{(1-y_\alpha)} = S_\alpha \tag{D.10}$$

or:

$$y_\alpha = \frac{C_0 Exs S_\alpha}{1 + C_0 Exs S_\alpha} \tag{D.11}$$

y_α can also be expressed as:

$$y_\alpha = \frac{p(S_i(l,p)|N_\alpha(S_i))}{K_{S_i(l,p)}^{N_\alpha(S_i)}} S_\alpha \tag{D.12}$$

So the conditional probability $p(S_i(l,p)|N_\alpha(S_i)$ can be calculated. The total fraction of configuration $S_i(l,p)$ folded at $T + dT$ is then:

$$p(S_i(l,p), T+dT) = \sum_\alpha p(S_i(l,p)|N_\alpha(S_i))p(N_\alpha(S_i)) = \sum_\alpha p(N_\alpha(S_i))\frac{K_{S_i(l,p)}^{N_\alpha(S_i)} y_\alpha}{S_\alpha}$$

To summarise, the probability of a neighbourhood $p(N_\alpha(S_i)$ is know before changing the temperature. Once the temperature changed, given all the possible partial folding and all the possible neighbourhood, thanks to the simplification of the concentration of free S_i, the problem is reduced to finding the proportion of the different partial folding in a given neighbourhood $p(S_i(l,p)|N_\alpha(S_i)$. Then the total probability of having a partial folding is obtained by summing over all the neighbourhood $N_\alpha(S_i)$

D.1 Remarque on the small origami.

In the case of the small origami the experiments were not performed with an excess of staples. The approximation to simplify the concentration of staples is no longer correct.

We used for the concentration of staple in the configuration N_α the form:

$$C_0\left(Exs - \sum_{u,v,u\leq v} p(S_i(u,v)|N_\alpha(S_i))p(N_\alpha(Si))\right) \tag{D.13}$$

The difference is that the sum over the different neighbourhoods of the staple is not done. This form is correct for experiments with only one staple as there is only one neighbourhood. However when two staples are present there is two possible neighbourhoods. For the AT case as GC is already folded when AT fold this approximation is correct as there is only one neighbourhood at a given time. For the B1m B2m case, this approximation is no longer correct, and the correct result should be hopefully slightly different from our theoretical result.

Taking Eq. D.13 leads to slightly different equations. Instead of Eq. D.10 we have:

$$\frac{1}{C_0(Exs - y_{alpha})}\frac{y_\alpha}{(1-y_\alpha)} = S_\alpha \tag{D.14}$$

and:

$$y_{alpha} = \frac{1}{2}\left(E + 1 + \frac{2}{C_0 S_{alpha}} - \sqrt{(E + 1 + \frac{2}{C_0 S_{alpha}})^2 - 4E}\right) \tag{D.15}$$

Appendix E

Quenching intermediate states.

The model provides the hierarchical pathways that will fold (unfold) the origami during a cooling (heating) experiment. Therefore, at any temperature steps, one may have a good picture of the partially formed origami. Also, we predict the average folding of each staple at each T as well as a detailed description of its environment. In order to check the predictive ability of the model, it is of interest to compare the experimental structures with the calculated ones by quenching origami structures at a given T during a cooling (heating) event.

To do so, at a given temperature when cooling or heating, 3 μL of solution (containing 1nM of M13mp18 and an 10 fold excess of staple) were dropped on a mica surface, and after a short waiting time (for the oDNA to attach on the surface) a drop of 100 μL of buffer was added to quench the process. This drop by diluting the remaining staple in solution should quench the reaction. Then the deposit was studied by liquid AFM using peak force mode at 20 °C. In the following section we compare the experimental configuration to the expected theoretical one.

E.1 Study of the O1 origami

Thanks to the model it is possible to predict the mean configuration as a function of the temperature. In Fig. E.1 the configuration of O1 at different temperatures are represented on the annealing (a and b) and melting curve (c and d). A red color means that the part of the staple is folded at more than 50% and a green color means a probability of folding between 50% and 5%. The width of the colored line is also proportionnal to the percent of folding.

The mean space between strands has been set to 2.8 nm to respect the aspect ratio of the shape in the Mg Buffer. Indeed this value is a mean value between the minimal center to center distance between strands that is where the crossover are (1.8nm), and the maximun distance which is between two crossovers (3.9 nm) because of the electrostatic repulsion.

The path of the virus creates two separated blocks for this oDNA. An observation of the predicted structure is that the two blocs never attach one to the other even at low temperature.

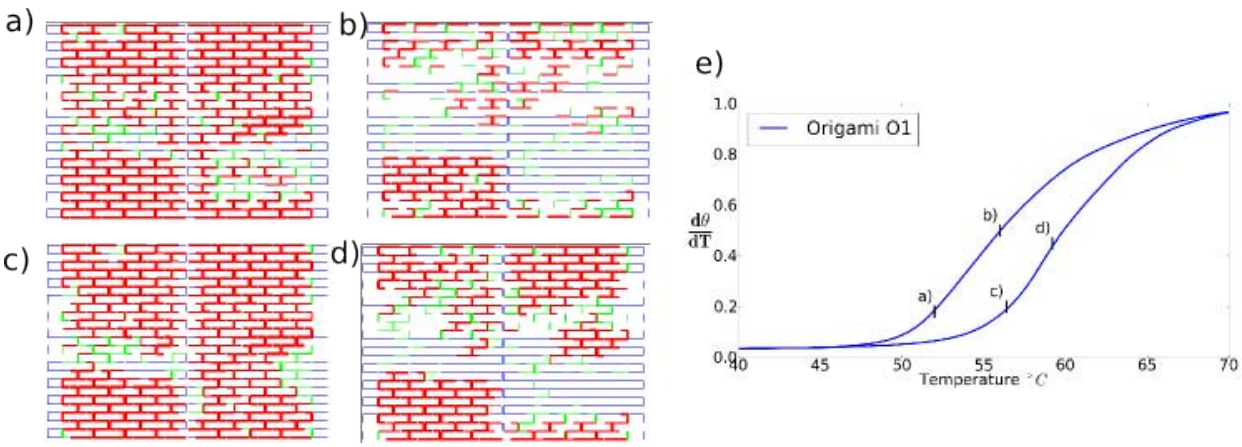


Figure E.1: Configuration at different temperature on the annealing curve: a) Cooling 52 °C; b) Cooling 56 °C; c) Heating 56°C; d) Heating 59°C e) theoretical melting curves.

This is linked to the notion of distance in the model. The notion of distance in the model is related to the notion of connectivity on the virus and not to the notion of physical distance. Fig. E.2 illustrate this notion. When looking at the two point A and B on the virus path, when no staple are folded the distance between A and B is $d_{AB} = \overline{AP_1} + \overline{P_2B}$. And an entropic penalty is added to the folding energy of the staple whose crossover goes by A and B. However when a staple fold at (Fig. E.2 right) the penalty is transformed to the penalty of a crossover as the distance $d_{AB} = \overline{AP_5} + \overline{P_6B} \approx 0$. For a staple connecting the two bases C and D the initial distance is $d_{CD} = \overline{CP_1} + \overline{P_2P_3} + \overline{P_4D} \approx 600$ b for O1 origami. Even when at the end all staples are folded , the distance throug the path is $d_{CD} = \overline{CP_5} + \overline{P_6P_7} + \overline{P_8D} \approx 128$ b. According to the model the associated penalty is still important. However, experimentaly the structures are going to be close when the left and right block will be formed, leading to a small distance between C and D and accordingly a small entropic penalty. The attachment between two blocks will probably be more close to a phenomenon of nucleation. So the connection between the two block is an unknown in our model.

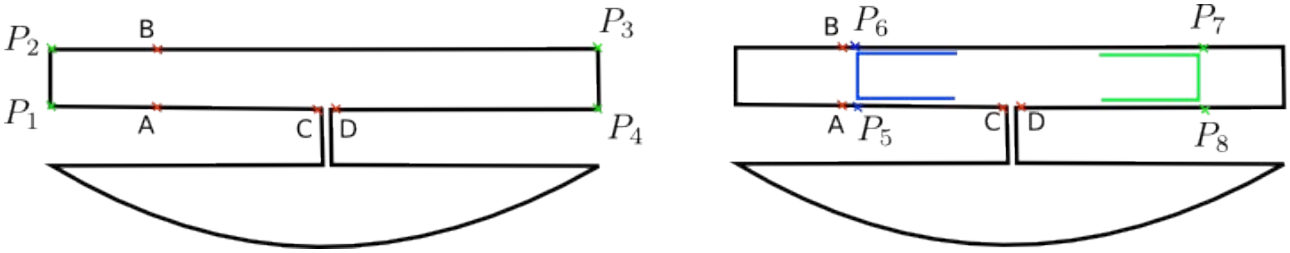


Figure E.2: Schematic representation of the notion of distance between two block compared to the notion of distance within one block. left) Initially when no staple are folded. right) at lower temperature when some staples are folded

Fig. E.3b) is an AFM image of quenched configurations in the annealing process at 50°C and Fig. E.3c) represents the expected configuration at this temperature. Several configurations were zoomed on E.3a). From the experimental configurations it can be observed that blocks smaller than the origami are first formed. Given the aspect ratio of the origami, the central weakness is present in some of the zoomed configurations , as well as a weakness in the middle of the oDNA but in the perpendicular direction.

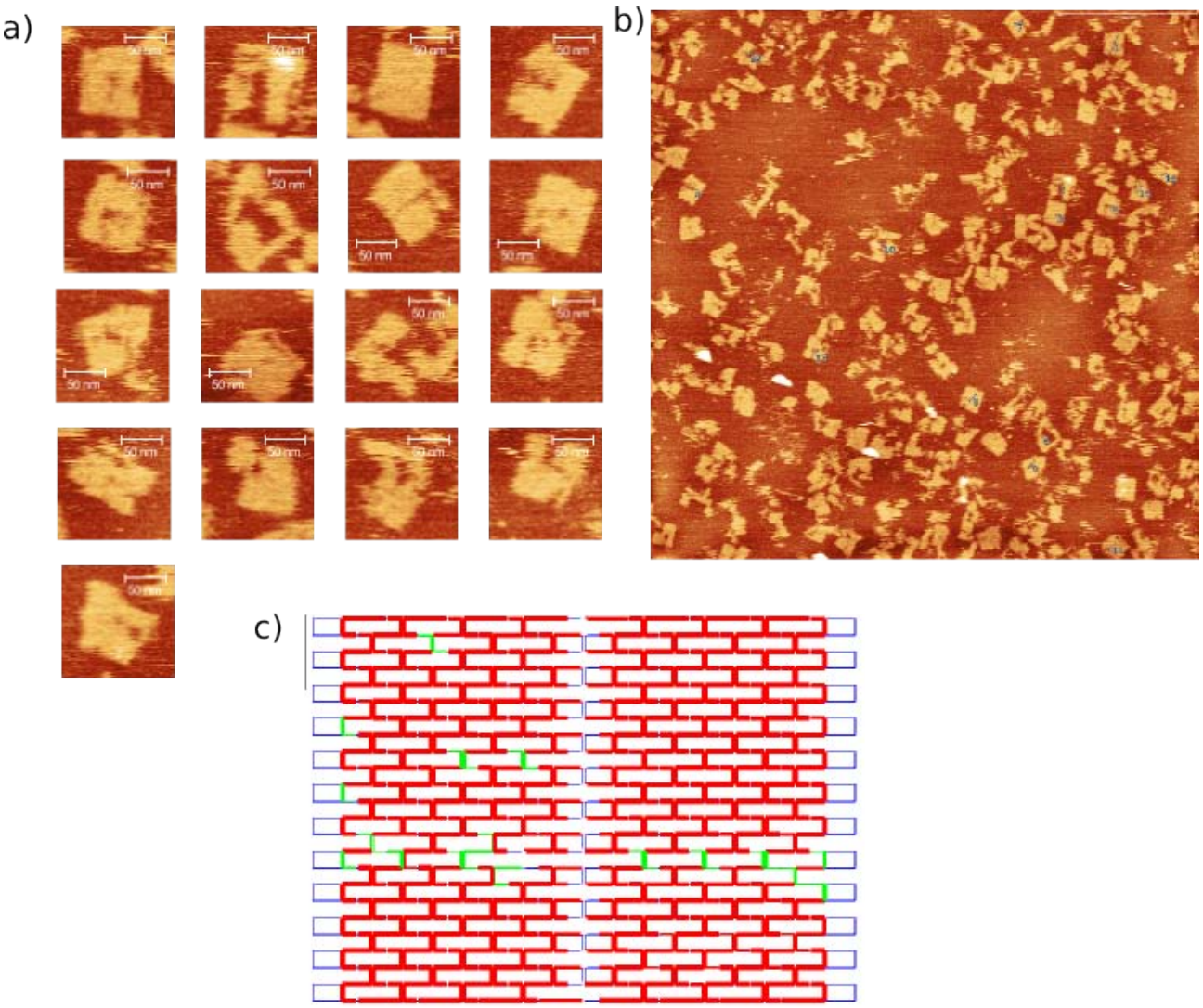


Figure E.3: Cooling: a) Several configuration extracted at 50°C; b) whole AFM image c) Expected configuration at 50°C

The same experiment was repeated at 52 °C (Fig. E.4). All the configurations seems to be blurry. This is due to parts of the oDNA that are not folded and fluctuate in solution. This

does not allow to reach a good resolution. All the zoomed configurations on the first line seems to be formed correctly in three corners with one corner hardly formed. At this temperature the expected configuration (Fig. E.4b)) is in agreement with the experimental data. Furthermore as observed at the previous temperature a weak line is present horizontally. This is due to the formation of no crossover between the two parts separated by the black line. Given the aspect ratio of some configurations this weakness seems to be present in some configuration as the one highlighted (Fig. E.4c).

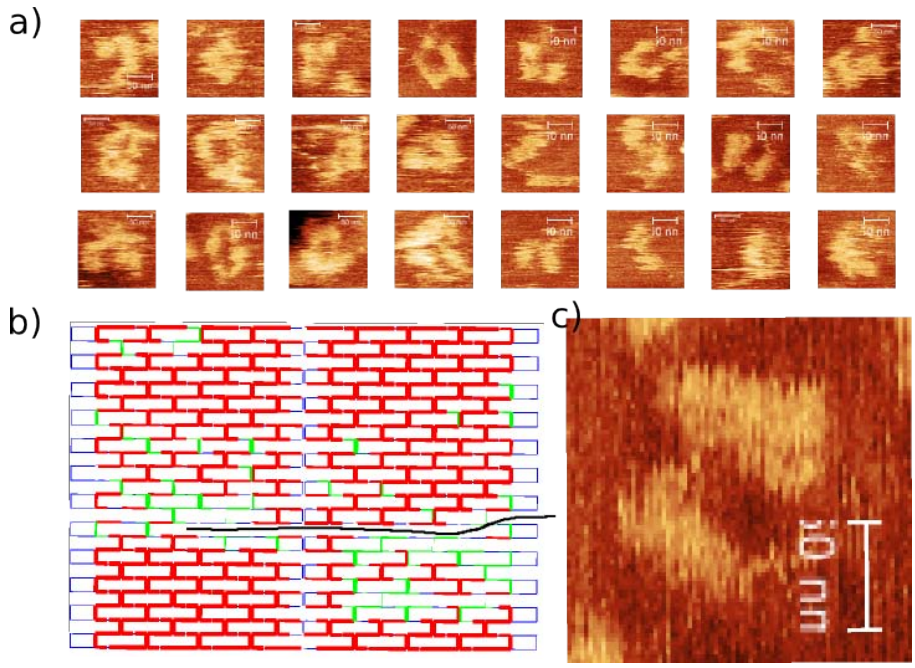


Figure E.4: Cooling: a) Several configuration extracted at 52°C; b) Expected configuration at 52°C c) Best agreement

Study of the melting process seems to prove that the configuration is less stable than expected (Fig. E.5). The expected configuration is (Fig. E.5c)) Only few of them are almost complete configuration (Fig. E.5b The expected configuration at 63°C (Fig. E.5d) is represented for information. Indeed it seems than in all the configurations we can see at least a small square which is the most stable part of this origami which is present experimentally and theoretically.

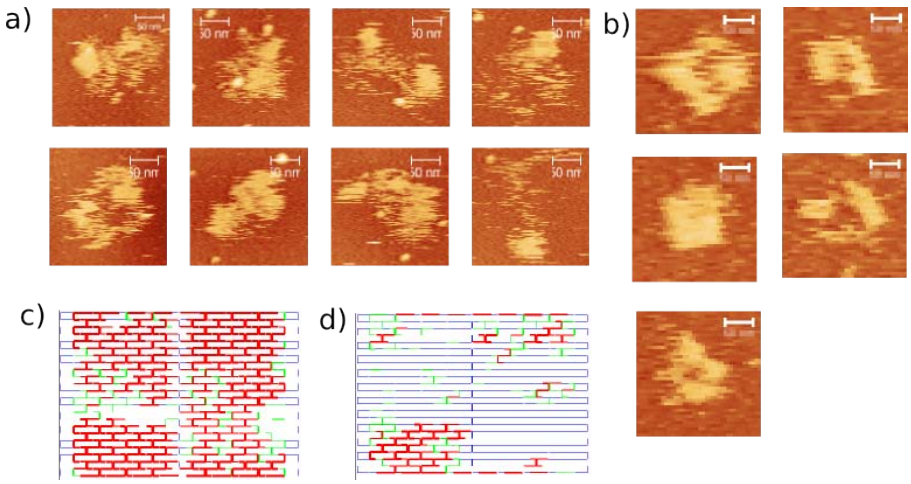


Figure E.5: Heating 57°C: a) Several configuration extracted at 57°C; b) almost complete configuration c) Expected configuration at 57°C; d) Expected configuration at 63°C

Finally the study of the configuration at 20°C shows that most of the oDNA are well formed (Fig. E.6). Some configurations have a hole in the middle of one block. This hole could corresponds to the last part folding in the theoretical model.

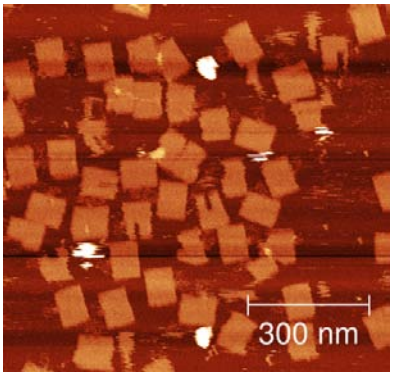


Figure E.6: Image at room temperature

E.2 Study of the O4 Origami

The process of cooling was also studied at 50°C and 53 °C(Fig. E.8a)) for the O4 origami.

At 50°C most of the configuration are close to the final configuration . We can notice a clear separation close to the middle of the several origamis (Fig. E.7). The expected configuration (Fig. E.7 c) is almost completely folded, at the exception of the bottom part. The configurations similar to the expected one were highlighted (Fig. E.7b). We also noticed several oDNA with an horizontal separation. As this temperature this is not predicted by the model. However when looking at a slightly higher temperature (53°C) the configuration exhibit several horizontal weakness which could explain the experimental structures.

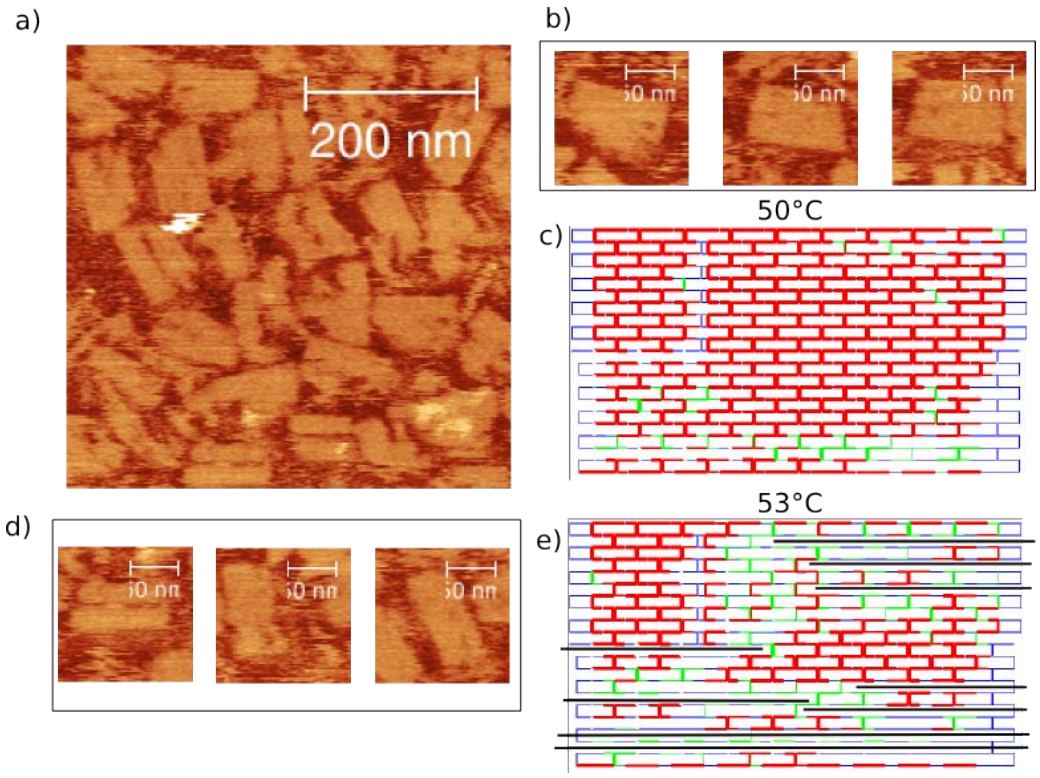


Figure E.7: Cooling 50°C a) AFM image °C b) Zoom on structure close to the final origami c) theoretical structure at 50 °C d) Zoom on structure partially formed. e) Theoretical structure at 53°C

The same experiment was repeated at 53°C. Several partial origamis were observable and several long and thin part are visible in agreement with the expected configuration at 53 °C (Fig. E.7e). The only almost complete origami was zoomed on and compared to the expected configuration at 51°C (Fig. E.8b). One can notice on this well formed configuration it is possible to reach a molecular resolution, whereas surrounding configurations that are a mix between ssDNA and dsDNA parts are blurry.

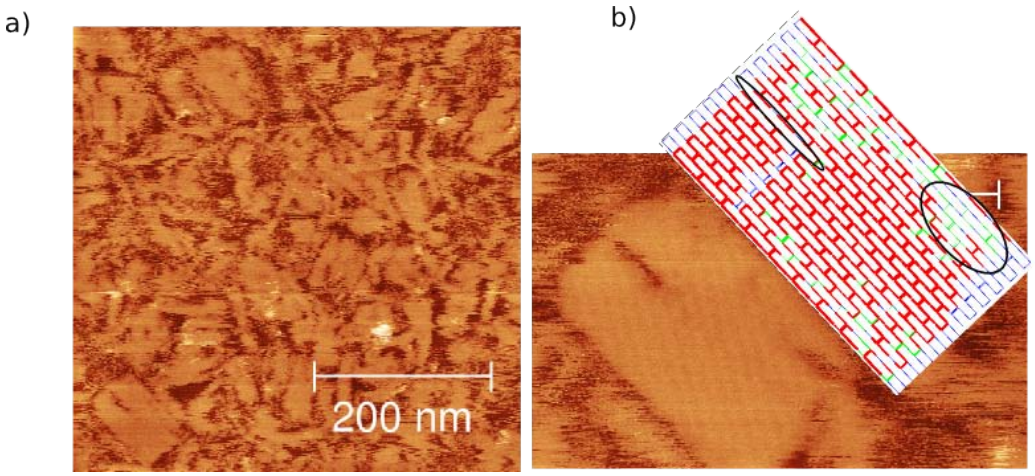


Figure E.8: Cooling 53°C a) AFM image; b) only configuration almost completely formed at 53 °C (compared to the configuration at 51°C)

The heating process was also studied. At 49 °C in agreement with the theoretical model(Fig. E.9b) most of the configuration are well formed (Fig. E.9c). Several configuration seemed to have a triangular shape (Fig. E.9d). The prediction of the model at higher temperature (Fig. E.9e) could partially explained the observed shape.

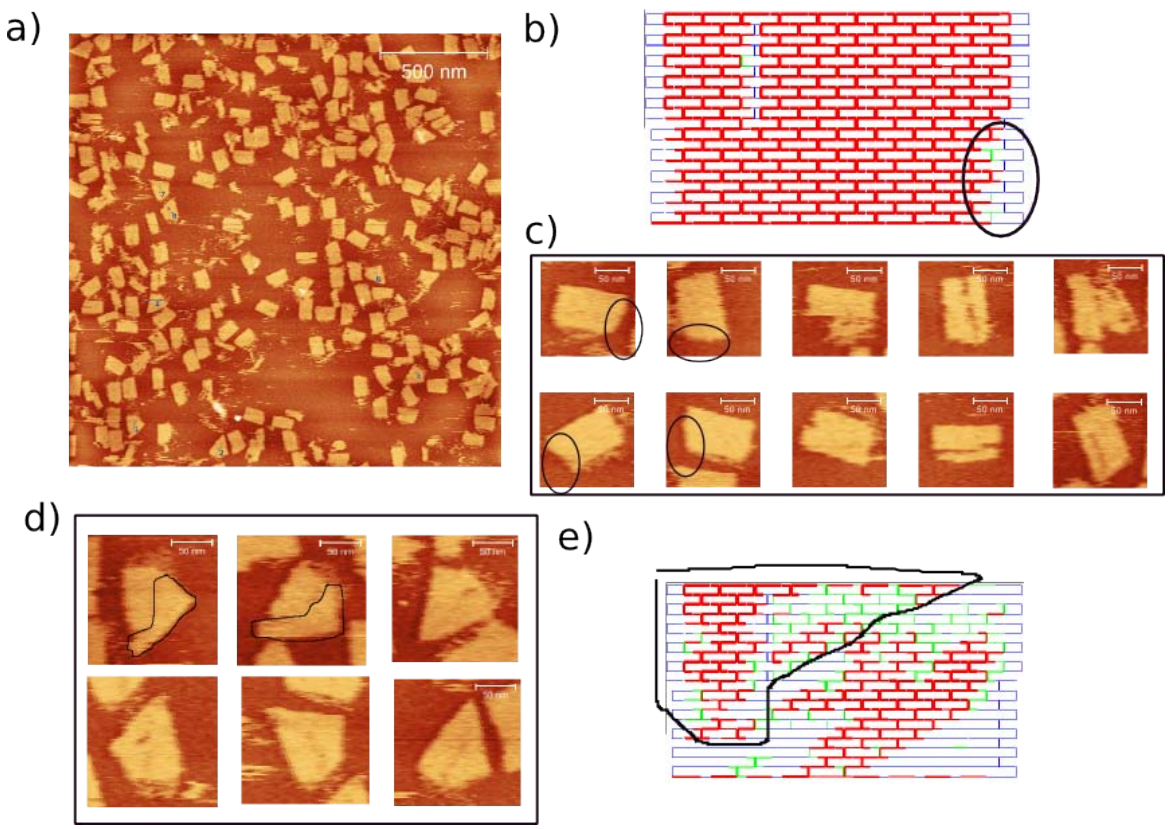


Figure E.9: Heating 49°C; a) AFM image; b) Theoretical configuration at this temperature. c) Zoom on several configuration; d) Triangular shaped configuration; e) Configuration at 57°C and 59°C that could explain the triangular shapes

At 57°C most of the configuration observed were about half the size of the regular O4 origami but still well formed Fig. E.10 a). All the configurations with a size close to the size of the final origami were extracted E.10c). The configuration expected at this temperature is Fig. E.10 b). So the predicted structure is most stable than expected.

At this temperature the model does not propose explanation for configurations close to half of the origami.

We then imaged the solution at 20 degrees (Fig. E.11 a, b c)). Most of the origami have a shape close to the expected one. However several small pieces can be observed. Some of them seem to be attached to the origami (Fig E.11b)), and some seem to be cutted in two pieces (Fig E.11c))

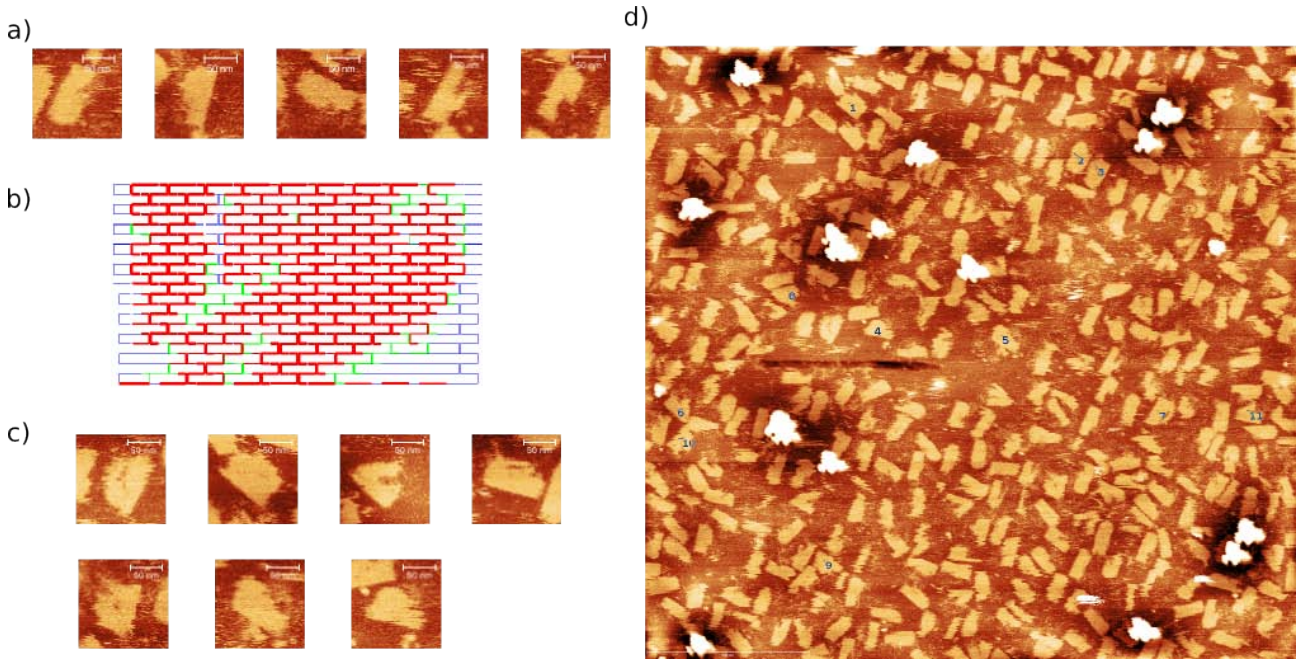


Figure E.10: Heating 57°C: a) Most frequent configuration b) AFM image c) Cooling configuration at 54°C; d) Heating configuration at 57°C e) Configuration possibly matching the predicted one

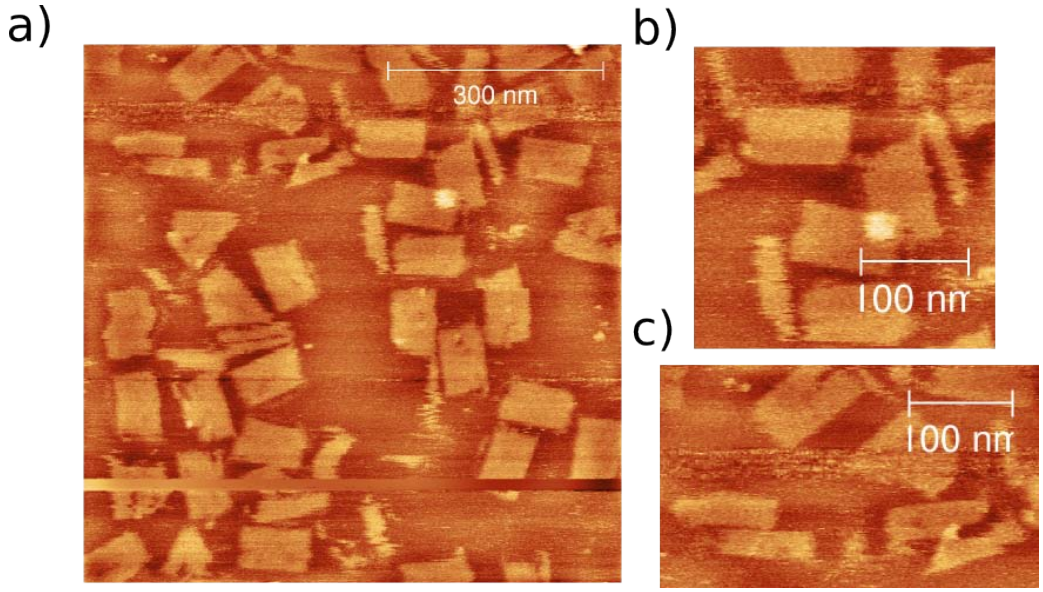


Figure E.11: Images of the O4 Origami in solution at 20°C

When looking at the process of formation it is possible to understand why small pieces were formed (Fig. E.12). In the process of formation some independent blocs are folded, once again in the sense that there are no crossover between them (Fig E.12 b and c)). In that case, as the path of the virus forms a bridge over the oDNA, one of the bloc could be flipped (Fig. E.12d). Then the probability to fold the missing peace and to flip back the flipped part is probably low. This shows the limits of the model that does not take into account a spatial representation of the origami.

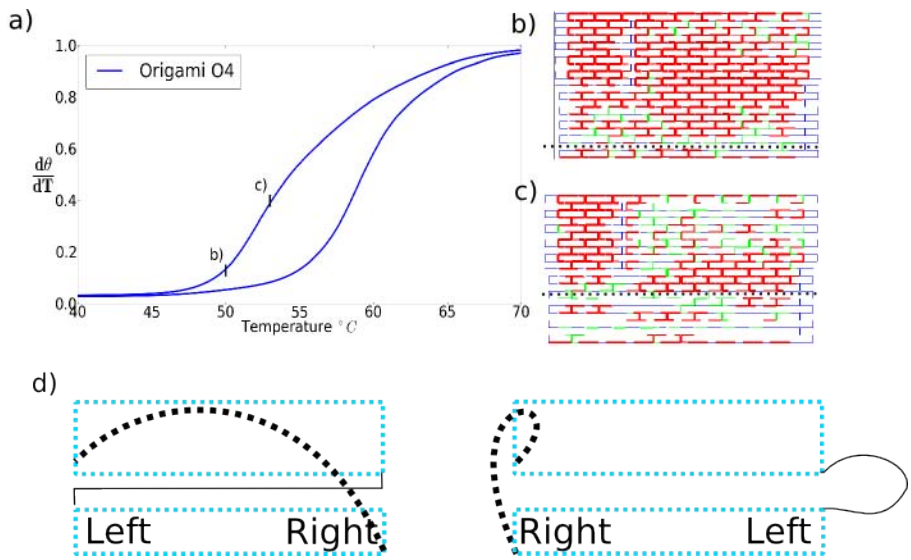


Figure E.12: a) annealing curve. b) Expected configuration at 50 °C. c) Expected configuration at 53°C. d) Schematic representation of a process leading to smaller independent parts

Appendix F

Calculation of the absorbance

As the experimental melting curves are followed by UV absorbance, it is necessary to calculate the absorbance of the partially folded origami at a given temperature.

The probability to find part$S_{i,j}$ of staple S_i folded is given by the sum of the probabilities of each configuration $S_i(k,l)$ that contains part$S_{i,j}$:

$$p(partS_{i,j},T) = \sum_{l,p,l\leq j\leq p} p(S_i(l,p),T) \tag{F.1}$$

And so the fraction of bases folded for the staple S_i is:

$$x_{S_i}(T) = \frac{1}{\sum_{k=0}^{N} |S_i(k,k)|} \sum_{k=0}^{N} p(partS_{i,k},T)|S_i(k,k)| \tag{F.2}$$

where $N+1$ is the number of parts of staple S_i.

The absorbance associated to any staple is:

$$Abs_{S_i}(T) = \left(Abs_0(S_i) + Abs_0(\bar{S}_i)\right)(1 - h_{S_i} x_{S_i}(T)) \tag{F.3}$$

where $Abs_0(S_i)$ is the absorbance of S_i and $Abs_0(\bar{S}_i)$ the absorbance of its complementary staple, calculated according to [101]. We used the following relation for the hypocromicity $h(S_i)$:

$$h_{S_i} = 0.287(1 - f_{gc}(S_i)) + 0.15 f_{gc}(S_i) \tag{F.4}$$

where $f_{gc}(S_i)$ is the fraction of GC content in the staple S_i. This formula has been slightly modified from [101] to improve the fit against the experimental data for the small origamis.

Appendix G

Temperature and salt corrections, calibration and results on small origamis

G.1 Temperature corrections

The computation of the nearest-neighbour contribution $\Delta G_{NN} = \Delta H_{NN} - T\Delta S_{NN}$ includes corrections to take into account temperature and salt variations.

Besides the nearest-neighbour contributions $(\Delta H_N^0, \Delta S_N^0)$ [27], we have also included temperature dependent corrections:

$$\begin{aligned}\Delta H_{NN}(S_i(l,p)) &= \Delta H_N^0(\sum_{k=l}^{p} seqA_{i,k}) + C_p(l,p)(T - T_{ref}) \\ \Delta S_{NN}(S_i(l,p)) &= \Delta S_N^0(\sum_{k=l}^{p} seqA_{i,k}) + C_p(l,p)\ln(\frac{T}{T_{ref}})\end{aligned}$$

where $C_p(l,p) = C_p \sum_{k=l}^{p} |S_i(k,k)|$ with $C_p = -42cal/mol/K/bases$ and $T_{ref} = 53°C$ according to [99].

G.2 Salt correction

The parameters of the NN model [27] are given for standard salt concentrations ([Na]=1M, [Mg]=0). Different salt conditions can be taken into account using the correcting terms in [100]. These corrections apply when Mg is dominant. We assume that for each attachment of the staple S_i on the configuration $S_i(l,p)$ (of length $|S_i(l,p)|$), the relation

$$\frac{1}{T_m(Mg, Na, C, S_i(l,p))} = \frac{1}{T_m(0, Na, C, S_i(l,p))} + f(Mg, Na, f_{GC}, |S_i(l,p)|)$$

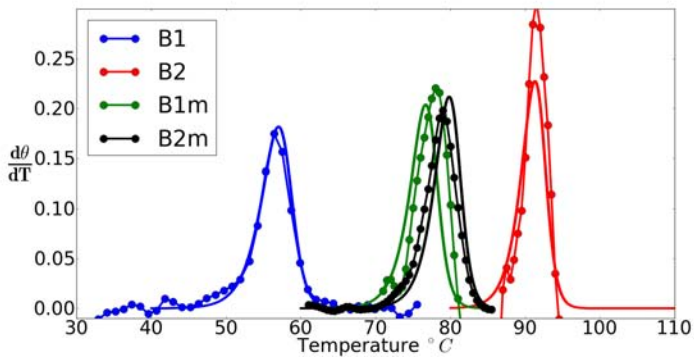


Figure G.1: (dotted line = experiment, solid line = theoretical) Comparison between the model and the melting experiments for four different dsDNA

between the melting temperatures T_m at different concentrations of Mg and Na holds (f_{GC} is the fraction of GC content of $S_i(k,l)$).

The melting temperature corresponding to the point where $\Delta G(T_m) = 0$:

$$T_m(0, Na, C, S_i(l,p)) = \frac{\Delta H_{NN}(S_i(l,p))}{\Delta S_{NN}(S_i(l,p)) + Rln(C_T/4)} \tag{G.1}$$

it can be deduced that the salt corrections are taken into account by an entropic correction given by:

$$SC(Mg, Na) = f(Mg, Na, f_{GC}, |S_i(l,p)|)\Delta H \tag{G.2}$$

where ΔH is the sum of all the contributions previously cited (nearest-neighbour, temperature corrections, topological contributions). In this paper, we modified the function $f(Mg, Na, f_{GC}, |S_i(l,p)|)$ as given in [100] by a small additive term to take into account low AT content strands.

For this, we compared the predictions of [100] with the experimental results obtained with the small origamis. In [100], the authors calibrated their model against 17 different dsDNA, involving a wide range of salt concentrations, and obtained a mean deviation $-1.7 \pm 0.7\ °C$. We deal here with a more restricted range of salt concentrations, it is therefore expected that the model [100] can be improved. We used a correction to the f function of [100] for the low fractions f_{GC} of GC content, to obtain a mean deviation of -0.4 ± 0.3. The corrected f function reads:

$$f_{new} = f + (f_{gc} - 0.5)0.00008 \quad \text{if} \quad f_{gc} < 0.5.$$

It gives the following results for the strands involved in the differents small origamis (Fig G.1)

Appendix H

Manip Supplementary

H.1 Effect of the salt concentration

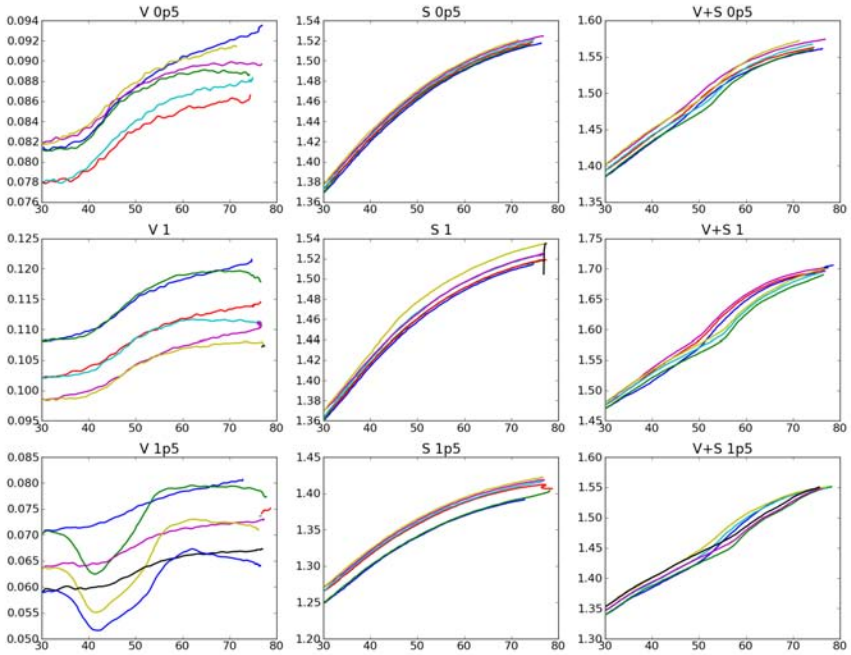


Figure H.1: Melting curves of the origami O1 for different concentrations in magnesium. V means M13mp18 strand alone, S means mix of the 200 staple strands alone, and V+S both mixed. 0p5 = 6.2 mM Mg, 1 = 12.5 mM Mg and 1p5 = 18.7 mM Mg. The M13mp18 concentration is 1.4 nM. The staples are in a 15 fold excess (by staple).

H.2 Effect of the excess on O1

Comparison of the annealing-melting curve for a 5 and 10 fold excess of staple with respect to the M13mp18 phage whose concentration is 1nM. The buffer is a TAE buffer with Mg (18.25 mM).

For an excess of 10 the maximum is at 54.2 and for an excess of 5 it is at 52.7. The width of this peaks is approximatly the same For the heating process the shift is smaller: around 0.4-0.5 degrees (Fig: H.3)

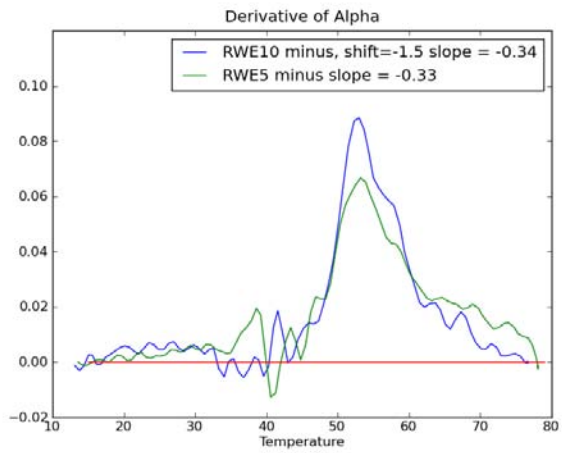


Figure H.2: Annealing: To compare the curve, the curve of the 10 fold excess has been shifted by -1.5 degrees. RWE10 means O1 with an excess of 10 and RWE5 means O1 with an excess of 5

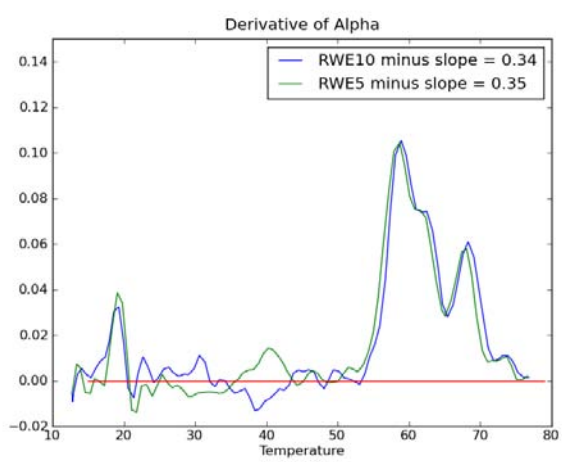


Figure H.3: Comparaison of the cooling curve for a 5 and a 10 times excess

H.3 Manip with origami O4 at slower rate.

Comparison of two temperature gradient (0.4 and 0.1 degree by minutes) (Fig. H.4) The concentration are: M13mp18 = 1 nM and a 10 times excess of staples. The buffer is a TAE buffer with [Mg] = 18 mM. The difference is more important for the cooling part (Fig. H.4) than for the heating part (Fig. H.5) that seems to be identical at both rates.

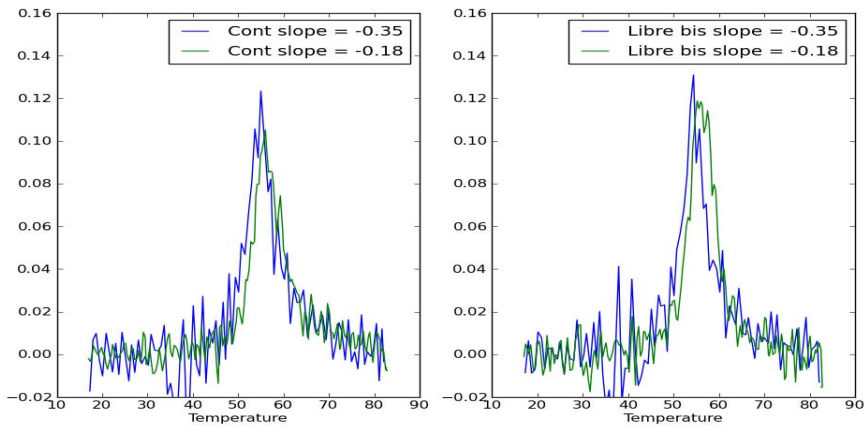


Figure H.4: Cooling the origamis with differents gradients. (left) O4 origami. (right) O4 origami with the four staples of the reservoir modified to constrain the bridge to a size of 200 bases

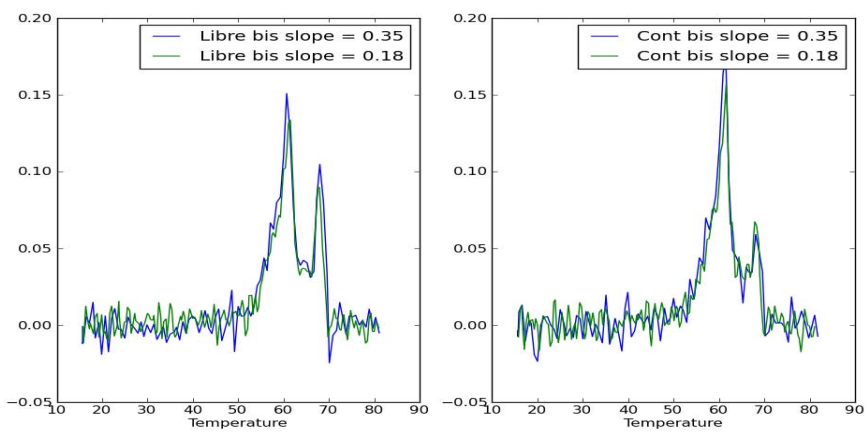


Figure H.5: Heating the origamis with different gradients (left) O4 origami. (right) O4 origami with the four staples of the reservoir modified to constrain the bridge to a size of 200 bases

H.4 Manip on one half of the strand.

A melting experiment was performed on a origami where only half of the staple were mixed. The concentration are: M13mp18 = 1 nM and a 10 times excess of staples. The buffer is a TAE buffer with [Mg] = 18 mM. The selection of the staple was done such as no adjacent staple was selected. (One staple on two was removed) "One half" corresponds to the first half and "half comp" corresponds to the complementary origami (with the other half).

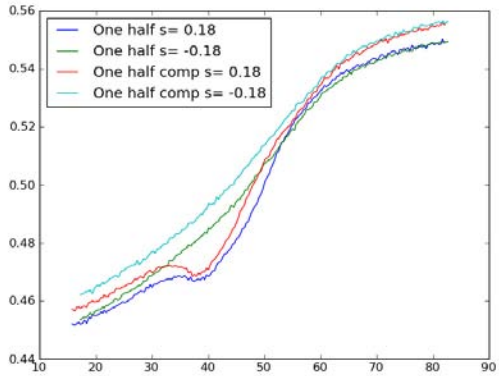


Figure H.6: Absorbance curves

(Fig H.6) shows that the absorbance decrease around 40 degrees when heating. This phenomenon is similar to the one observed with the virus alone in appendix H.1 for a salt concentration. It could mean that the origami did not formed at all. The derivative were extracted (Fig. H.7).

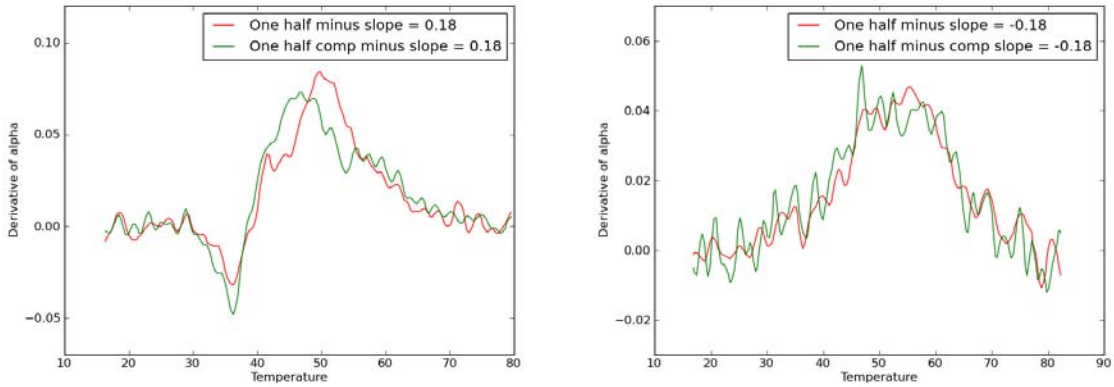


Figure H.7: (left)Heating;(right) Cooling

H.5 Manip with a ratio 1 for 1

O4 origami was formed, then purified of the staple excess. It was then heated in the UV spectro. After being heated the annealing and melting curves were recorded (Fig H.8)

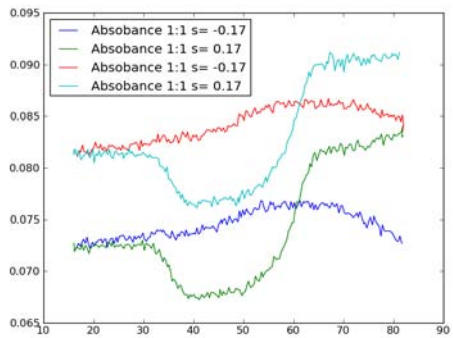


Figure H.8: Absorbance origami excess 1:1

Once again a weird thing happen around 40 degrees. It is almost impossible to extrac datas from the cooling curve because of the poor baseline (low concentration in object and evaporation. . .)

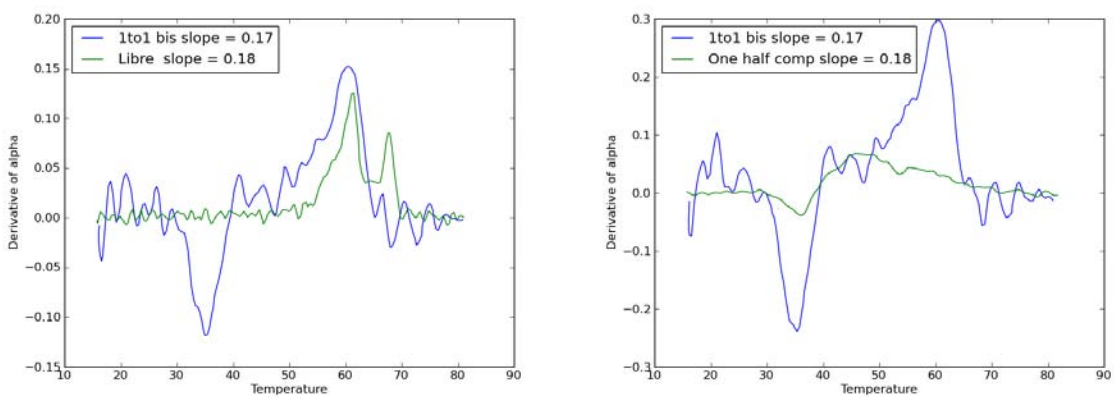


Figure H.9: (left) Heating origami excess 1:1 comparaison with normal origami (right) comparaison with one half of strands

When compared to the regular melting curve only the first peak is observed (Fig H.9)(left). When compared to the experiment where only one half of the staple was mixed (Fig. H.9(right) one can see than around 40 degrees there is a similarity of behaviour

Appendix I

Treatments of absorbance curves

Herein we detail the different steps of the treatment of the UV experiments in this thesis, from the raw absorbance to the normalised θ.

I.1 Generality

First one has to substract the absorbance at 310 nm to remove any particularity due to the cuvet and to the buffer ($Abs_{DNA260} - Abs_{DNA310}$) For regular dna construct 260 nm is the length of maximal absorption and at 310nm it do not absorb.

The second step is to substract the buffer signal ($Abs_{Buffer260} - Abs_{Buffer310}$), because the buffer evolve with temperature. This step is not always done because the effect is small.

Most of the times the absorbance curves have some noise. This one is suppressed by convoluating the curves according to the formula:

$$Abs_{convol}(T) = \frac{Abs(T - T_{step}) + 2Abs(T) + Abs(T + T_{step})}{4} \tag{I.1}$$

For most of the curves this formula is applied once or twice. For curves with low absorbance of small T_{step} it is applied 3 or 4 times.

I.2 Detail on the treatment of the curves for oDNA experiments.

Here all the steps of the treatment of an oDNA experiment are explained. Usually one experiment consists in doing several cycles of temperature on different samples. One cuvet contains the buffer, the second contains the phage (M13mp18) alone, the third cuvet contains the staples alone, and the fourth the mix phage plus staples.

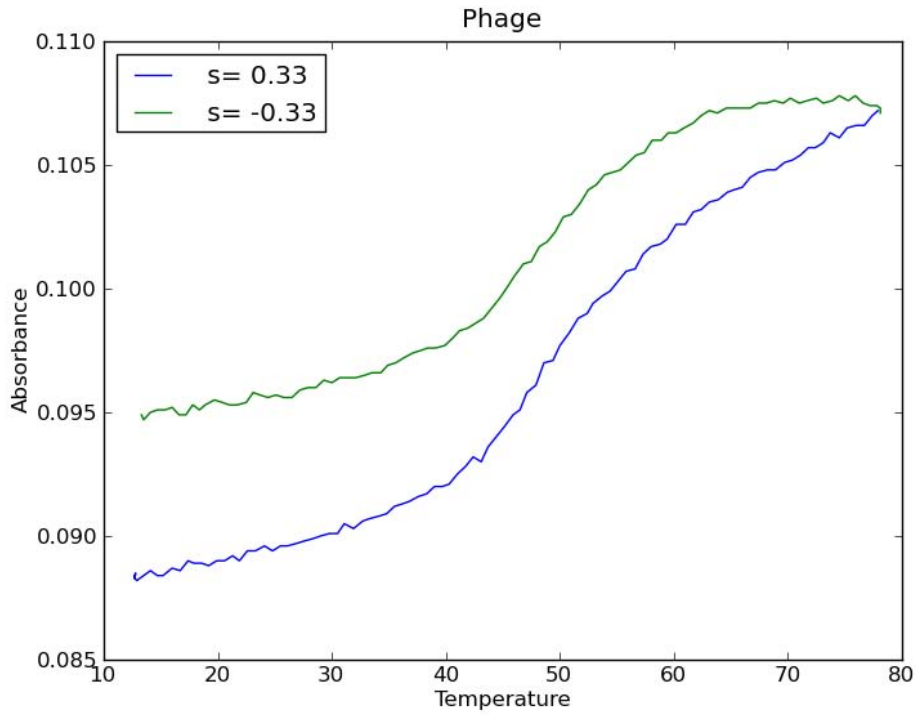


Figure I.1: Phage alone at 1 nM for one temperature cycle: heating at s = 0.33°/min , cooling s = -0.33°/min

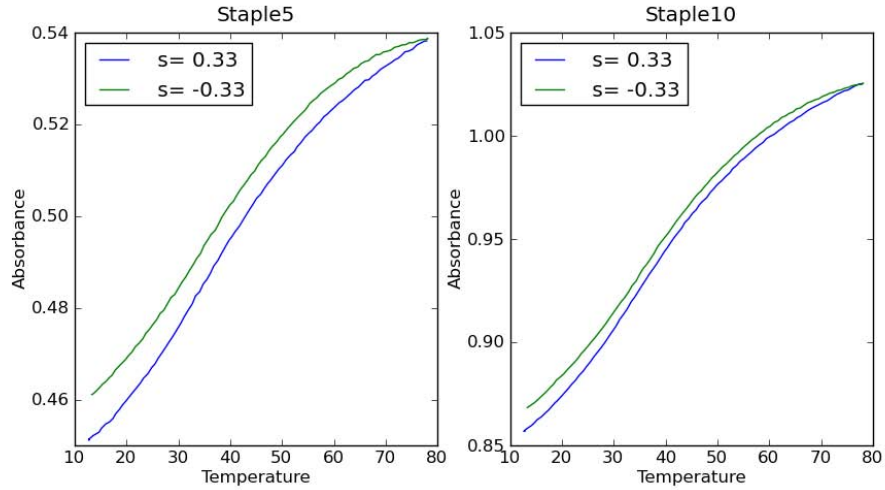


Figure I.2: Staple alone at two different excess. left a 5 fold excess, right a 10 fold excess

The absorbance of the phage alone show a small transition around 45 degrees (Fig: I.1) For the staples alone (Fig I.2) there is no clear transition

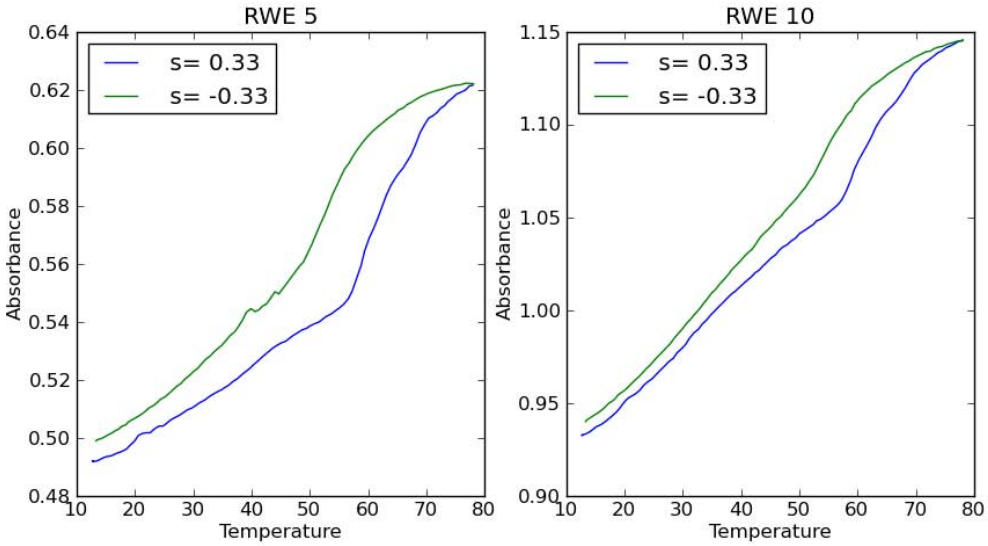


Figure I.3: Phage plus Staple at two excess of staple 5 and 10

The curves with positive s or slope are heating curves, and respectively the other are folding curves (Fig: I.1). For the mix phage plus staples one can see than there is an hysteresis between the cooling and the heating. It is difficult to compare the curve for the origami with an excess of five and an excess of ten (Fig I.3). Indeed the signal due to the excess of staples have an important contribution to the curve and we will have to remove it.

The usual way to study melting curves is to evaluate

$$\theta(T) = (L0(T) - Abs(T))/(L0(T) - L1(T)) \tag{I.2}$$

or

$$\alpha(T) = (Abs(T) - L0(T))/(L1(T) - L0(T)) \tag{I.3}$$

with $L0(T)$ the baseline at low temperature and $L1(T)$ the baseline at high temperature and $Abs(T)$ the experimental absorbance. For a simple experiment on a dsDNA α corresponds to the fraction of base pair unfolded. So at low temperature α equal 0 and at high temperature it is equal to 1. To estimate this quantity one has to choose a baseline for the curve.

The choice of the baseline can greatly change the aspect of the $\alpha(T)$ curves. For example one can see to different curves $\alpha(T)$ for two different choices of baseline on the same experiment. (Fig I.4 and Fig I.5) However the choice of the baseline is less visible on the derivative of α. (Fig I.6 and Fig I.7) The derivative of α is important because the maximun corresponds to the melting temperature for a simple experiment on a dsDNA.

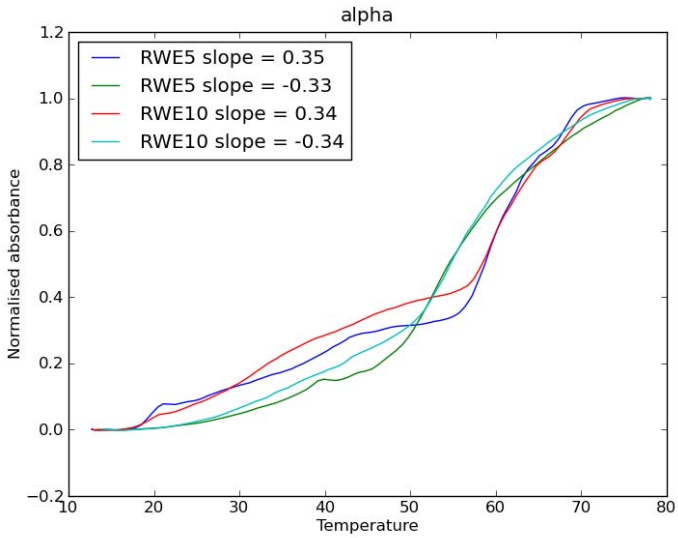


Figure I.4: Phage plus Staple at two excess of staple 5 (blue = melting, green = annealing) and 10 (red=melting and cyan annealing)

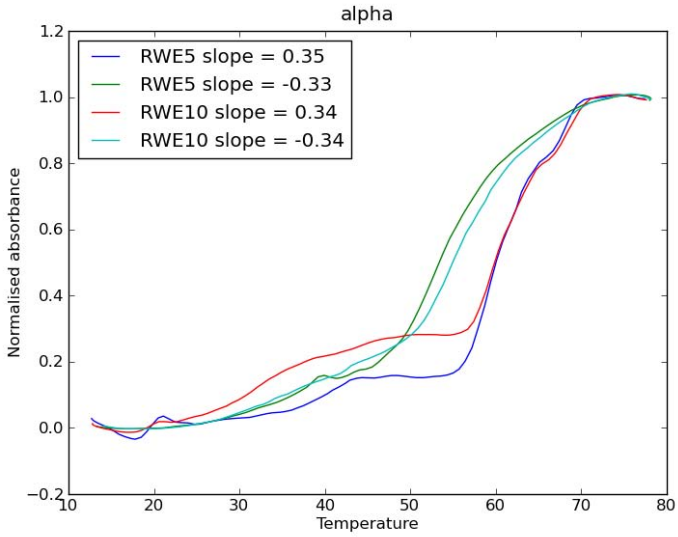


Figure I.5: Same experiment than Fig I.4 but with a different choice of baselines

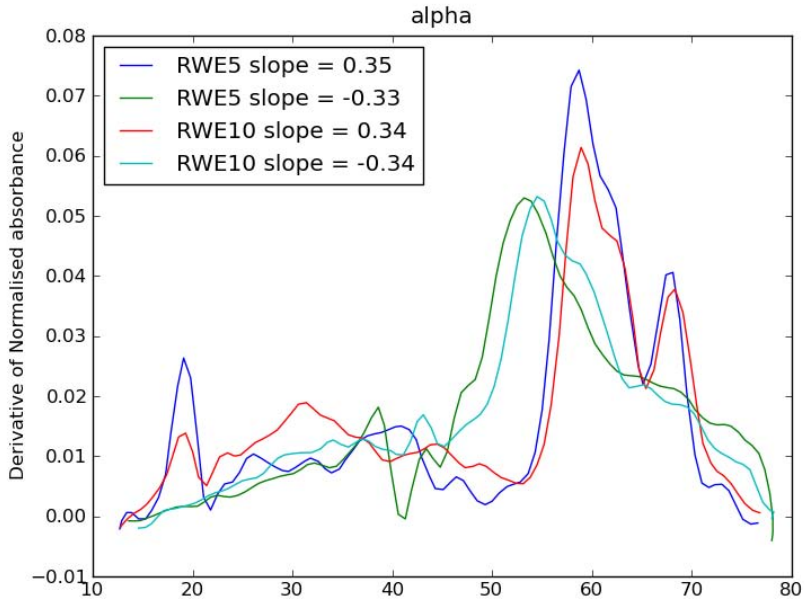


Figure I.6: Derivative of Fig. I.4

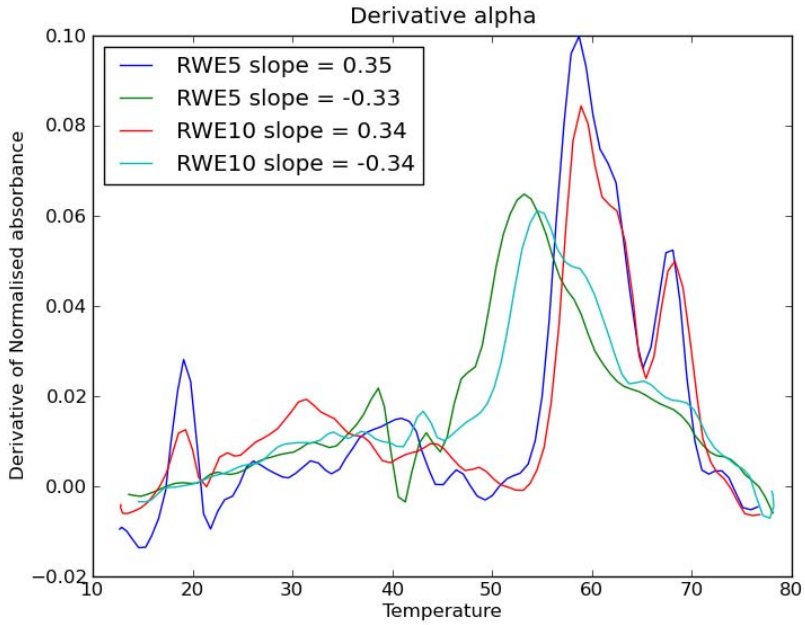


Figure I.7: Derivative of Fig. I.4. We can see that these curves are very similar to the ones of Fig. I.6

Those experiments are realised with an excess of staples. It is interesting to substract this excess of staples in the hypothesis that the excess of staples can be considered as non interacting.

For the origami with a 10 time excess of staples, at high temperature the sum of the absorbance of the cuvets staples alone + phage alone is close to the one of the oDNA mix. (Abs(oDNA mix) =1.1 and Abs(oDNA mix - staple alone - phage alone) = 0.01) The absorbance that corresponds to the hybridisation of the staples with the phage is given by: Abs(oDNA) = Abs(oDNA mix) - 0.9 Abs(staples alone). In Fig I.8 the original curve (Abs(oDNA mix)) is shown in blue and the subtracted one (Abs(oDNA)) in green. (the maximum of the curve has been set to zero for comparaison). We can notice that for the melting part of the subtracted curve (Fig I.8 (left green)) the low temperature absorbance is constant, showing that the oDNA is stable. Once this step done, one has to choose the high and low baseline.

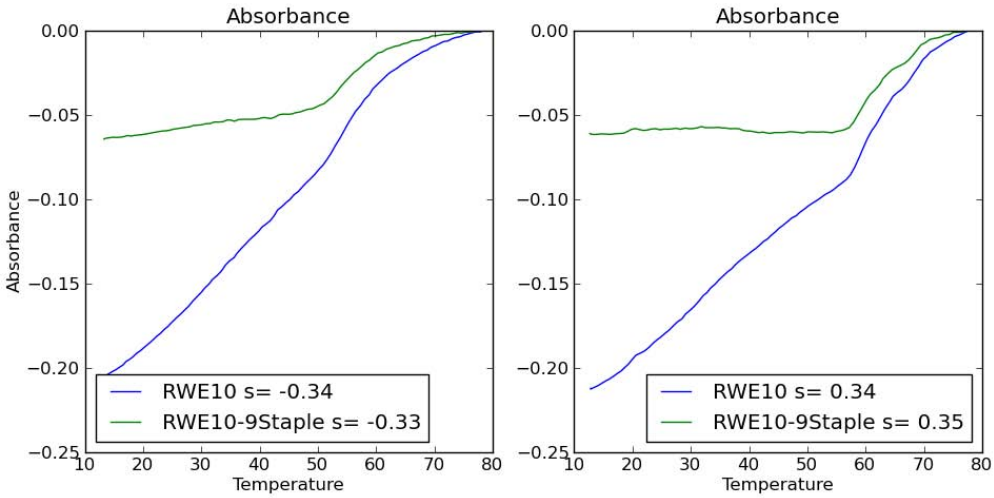


Figure I.8: Comparaison Absorbance with and without substraction (superposated at high temperature). Left: Cooling, Right: Heating

The baseline at high temperature (L1(T)) can be chosen with the same slope for the annealing and melting curves (Fig: I.9). At low temperature the baselines seem to be differents. However, as we shown previously, the choice of the baseline has not much influence on the derivative of α.

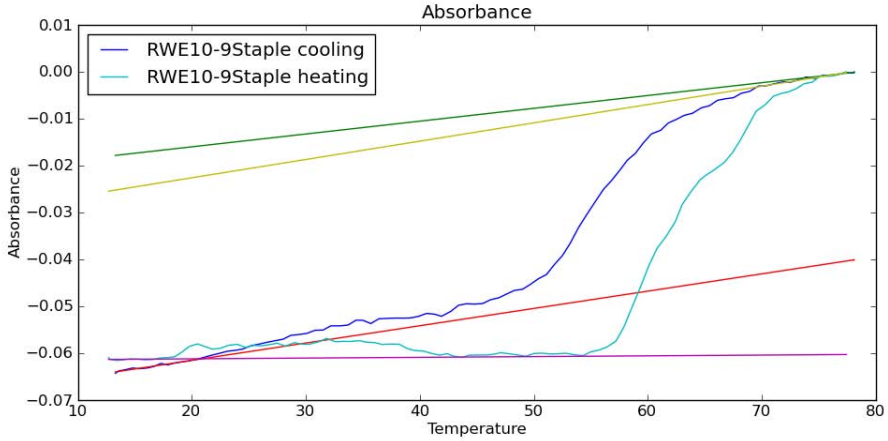


Figure I.9: Choice of the baseline for the absorbance Abs(oDNA)

In Fig I.10 left) (resp right) we compare the normalised absorbance for the annealing (resp melting) curve with (green curve) and without (blue curve) subtracting the excess of staple.

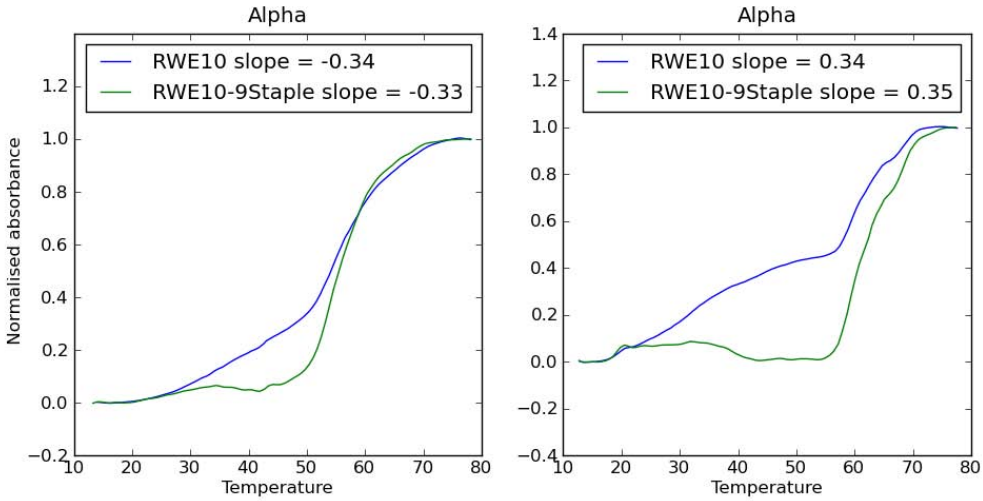


Figure I.10: Comparaison Normalised absorbance with (Abs(oDNA) = green) and without substraction (Abs(oDNA mix) = blue). Left: Cooling, Right: Heating

We then show the influence of subtracting the excess on the derivative of those curves. On can see than subtracting the excess of staples amplify the derivative of alpha (Fig: I.11).

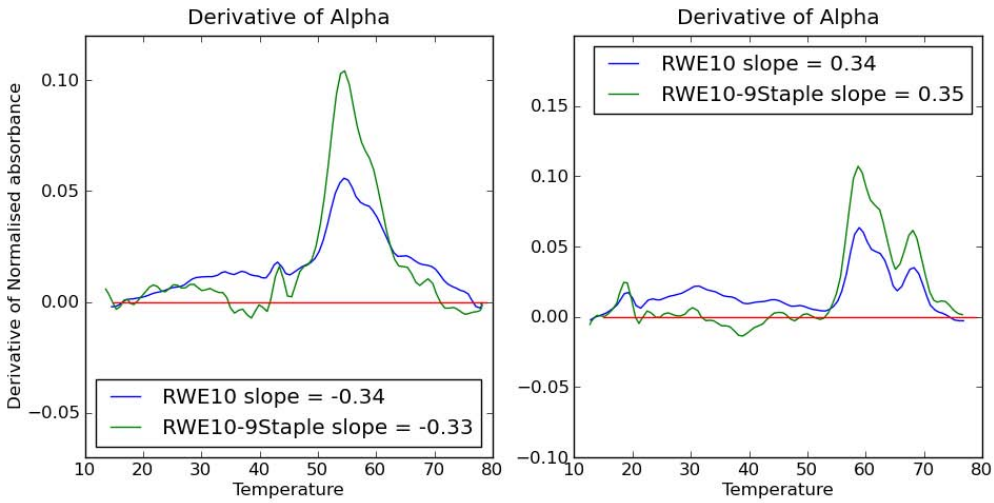


Figure I.11: Comparaison Derivative of the normalised absorbance with (Abs(oDNA) = green) and without substraction (Abs(oDNA mix) = blue). Left: Cooling, Right: Heating

We then checked the reproductibility of the experiments by doing two cycles of annealing-melting experiment on the same samples. Given (Fig: I.12) it seems that the reproductibility of the experiment is good. However when cooling (Fig: I.12 left) , at high temperature the behavior of the two curves are slightly differents.

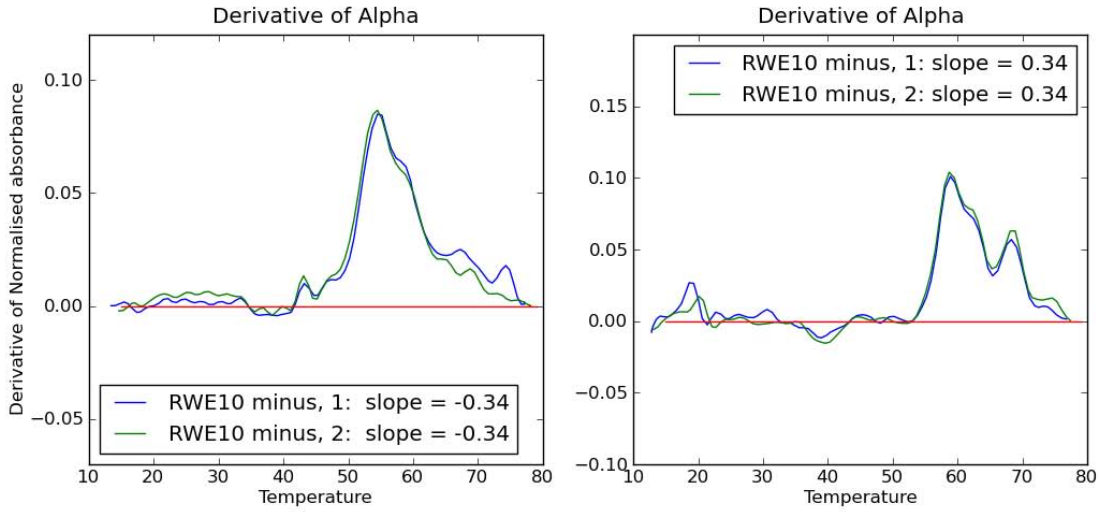


Figure I.12: Reproductibility of the experiment two temperature cycles for the same sample for Abs(oDNA) Left: Cooling, Right: Heating; blue = first cycle, green = second cycle

Appendix J

Strands used in the differents structures

J.1 Small Origami

For the first small origami with B1 outer and B2 inner: The staples are given from 5to 3:

B0	TAATAATAATATAATTGGCGGCCGGGGCCGGCCGGCGCGGCGGCCGGCATTAATATAATATAAT
B1	ATTATATTATATTAATAATTATATTATTATTA
B2	GCCGGCCCCGGCCGCCGCCGGCCGCCGCGCCG
$\overline{B1}$	TAATAATAATATAATTATTAATATAATATAAT
$\overline{B2}$	CGGCGCGGCGGCCGGCGGCGGCCGGGGCCGGC

When B1 is in the 'inner' position the sequence of the scaffold is:

$C1231$	CGGCGCGGCGGCCGGCATTAATATAATATAATTAATAATAATATAATTGGCGGCCGGGGCCGGC

For the B1m B2m structure:

B0	TTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTT
B1m	AAGGGAGCCCCCGATTCGAGAAAGGAAGGGAA
B2m	AAGCCGGCGAACGTGGTAGAGCTTGACGGGGA
$\overline{B1m}$	TTCCCTTCCTTTCTCGAATCGGGGGCTCCCTT
$\overline{B2m}$	TCCCCGTCAAGCTCTACCACGTTCGCCGGCTT

J.2 Origami O1 and O2

O1 is the same Origami used by Rothemund [1]. The staples list can be found in SI [2]

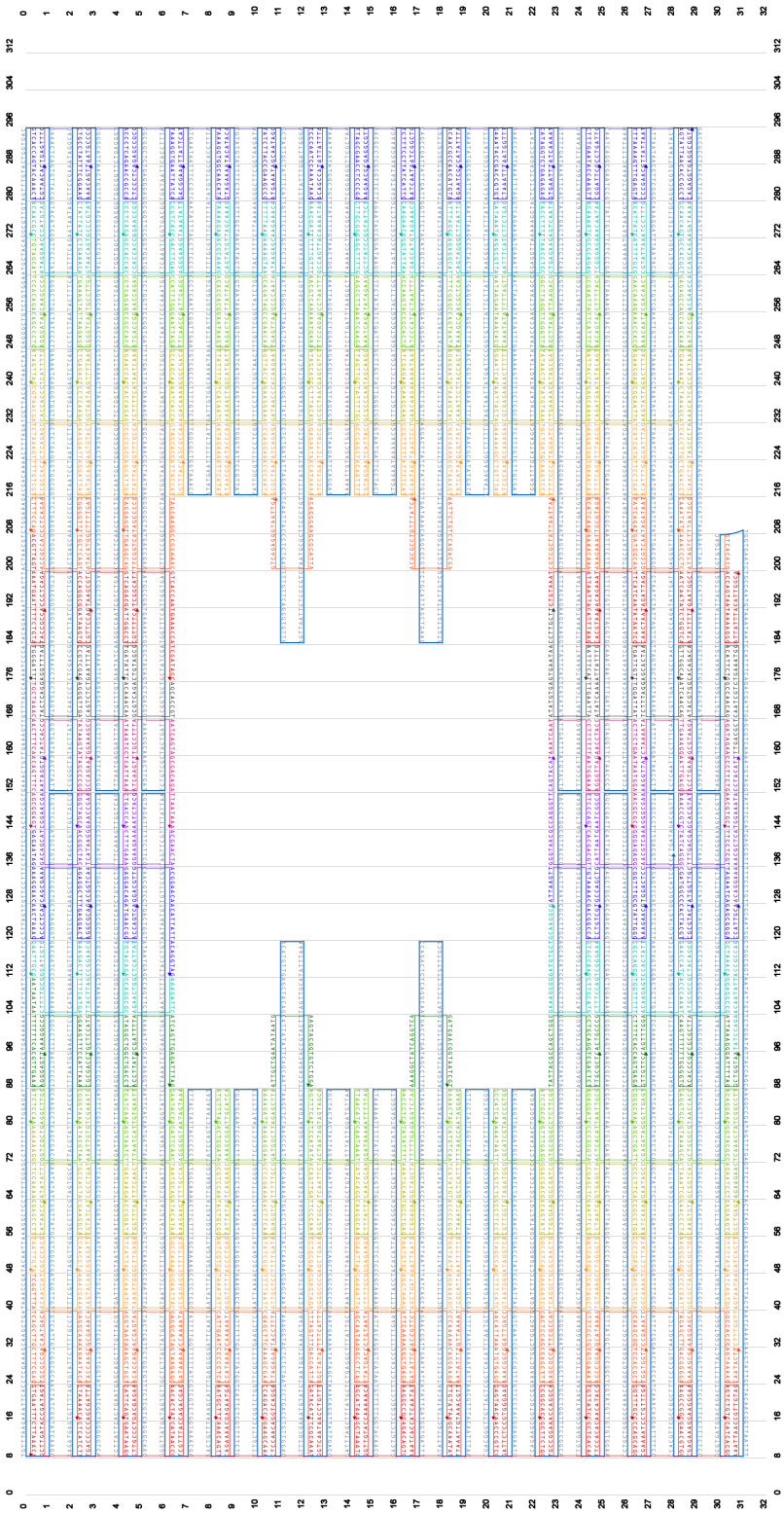


Figure J.1: Origami 02, it was furnished by Thao Tran from Jean-louis Mergny laboratory

J.3 Origami O3

Same staple than in O1. The staple connected from O1 are listed here

177p175	TTTCATGAAAATTGTGTCGAAATCTGTACAGACCAGGCGCTTAATCATTGTGAATTACAGGTAG
174p176	TTTCAACTATAGGCTGGCTGACCTTGTATCATCGCCTGATGGAAGTTTCCATTAAACATAACCG
165p163	AGAAAAGCAACATTAAATGTGAGCATCTGCCAGTTTGAGGGAAAGGGGGATGTGCTAGAGGATC
162p164	CAGCTGGCGGACGACGACAGTATCGTAGCCAGCTTTCATCCCCAAAAACAGGAAGACCGGAGAG
34p32	AGCGCCAACCATTTGGGAATTAGATTATTAGCGTTTGCCACCTCAGAGCCGCCACCGATACAGG
55p57	CACCAGAGTTCGGTCATAGCCCCCGCCAGCAAAATCACCAAATAGAAAATTCATATATAACGGA
46p44	ACAGAAATCTTTGAATACCAAGTTCCTTGCTTCTGTAAATCATAGGTCTGAGAGACGATAAATA
67p69	TAACCTCCATATGTGAGTGAATAAACAAAATCGCGCAGAGATATCAAAATTATTTGACATTATC

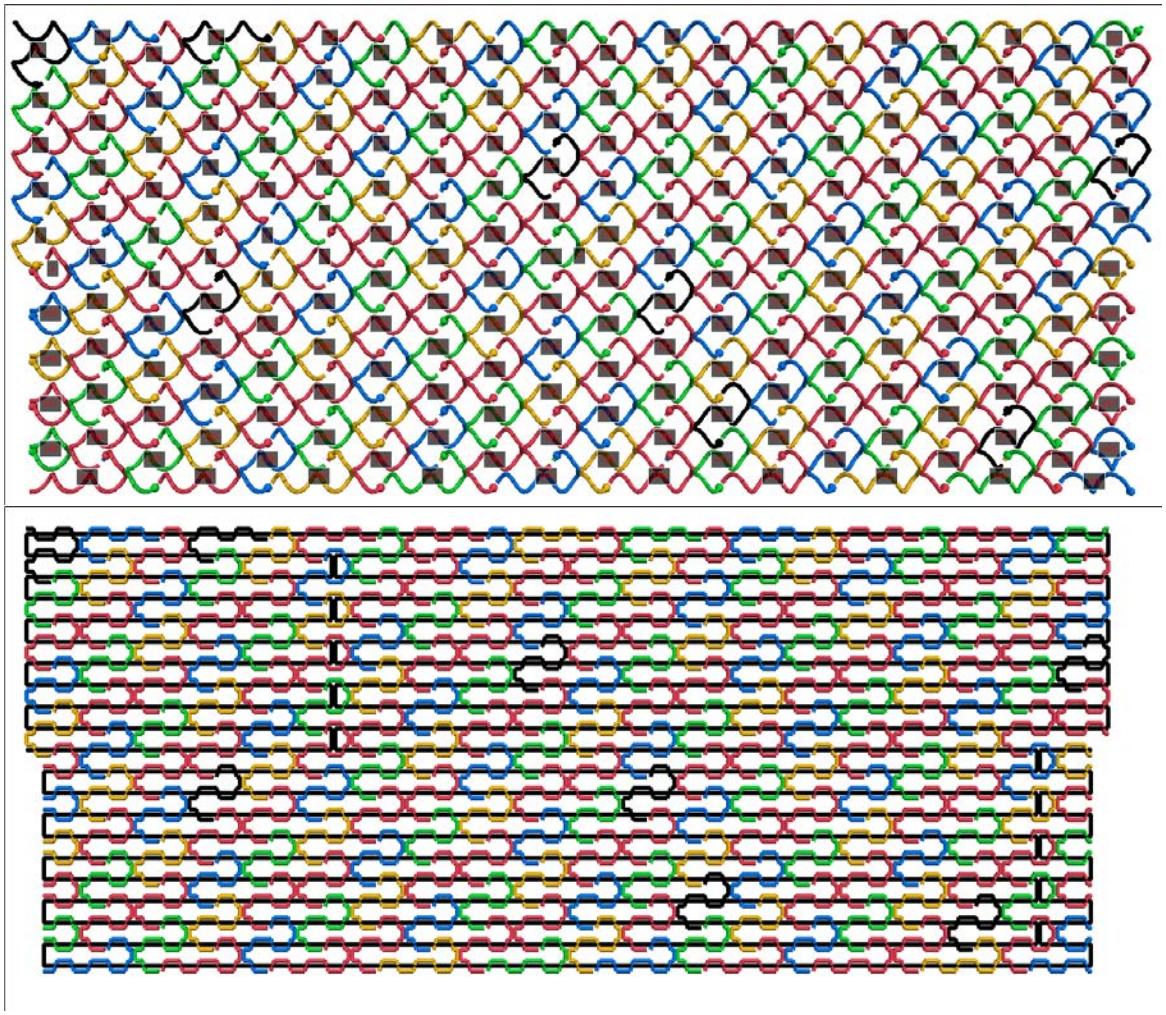


Figure J.2: Schem of Origami 4 without (a) and with the path of the scaffold

J.4 Origami O4

The staples are given from 5to 3. 3 is represented by a small ball in (Fig. J.2)

1	GAAGGGAAGACTCCAACGTCAAAGACTATTAA
2	TGGTAATAAGTCGAGAAAG
3	GACGGGGATCAGGGCGATGGCCCATTGAGTGT
4	CGTCATACCCTTGAGTAACAGTGCTAGAGCTT
5	GAACCCTACCAAATCAAGTTTTTTCTTATAAA
6	AGCGCAGTTGCCCCCTGCCTATTTCTAAATCG
7	GGTAGTGAAGCGGGCCGTAAAGCACGGAACCT
8	AATAAAGTGAAACCGGCAAAATCCGGGGTCGA
9	ACCATCACAAGGGAGCCCCCGATTCCGTATAA
10	CAGGCGAAATAGCCCGAGATAGGGCTACGTGA
11	ACCGTCTAAAGCCGGCGAACGTGGTTTAACGG
12	TGCAGCAATTTGGAACAAGAGTCCGGCGAAAA
13	AGAACGTGTTCACCGCCTGGCCCTCAACAGCT
14	TGTTCCAGGCGGTCCACGCTGGTTAGGGTGGT
15	TCAAAAGAAATCCTGTTTGATGGTCGCGGGGA
16	TGAGATTTGTTGAATCGGCCAACGGGTTCCGA
17	AGTCGGGATTGCGTATTGGGCGCCTGCCCCAG
18	ATTAATTGTCACCAGTGAGACGGGGAGAGAGT
19	GATTGCCCTGCCTAATGAGTGAGCAGTGTAAA
20	TTTTCTTTCGTTGCGCTCACTGCCAATTCCAC
21	GAGGCGGTAACCTGTCGTGCCAGCCTGTTTCC
22	AATGGTCATAAGACATGGTCATAGTGCATTAA
23	AGAGGATCATTGTTATCCGCTCACCGCTTTCC
24	GTGCCAAGCGAGCCGGAAGCATAATAACTCAC
25	GCCTGGGGCACGACGTTGTAAAACCCAGGGTT
26	ACAACATACTTGCATGCCTGCAGGGATGTGCT
27	TGTGTGAACCCGGGTACCGAGCTCCTCTTCGC
28	CGAAAGATAGGTATCGGTGCGGGCGAATTCGT
29	CGCCATTCCAGCTGGCGAAAGGGGTCGACTCT
30	TTCTGGTGATTAAGTTGGGTAACGGACGGCCA
31	TTCCCAGTCTCCAGCCAGCTTTCCCCTCAGGAAGATCGCA
32	GCAAGGCGCCGGAAACCAGGCAAACTGCCAGT
33	TATTACGCAGGCTGCGCAACTGTTTTGGTGTA
34	TTGACCGTAATGGGATAGGTCACGGGGAAGGG
35	GATGGGCGCATCGTAACCGTGCATGCGCCATT
36	TTGAGGGGACGACGACAGTATCGGGGCACCGC
37	TAATGTGTTCAAAAGGGTGAGAAAACGGCGGA
38	CCGTCGGATTCTCCGTGGGAACAAGGCCGGAG
39	AACCCTCAATCACCATCAATATGAGTAACAAC
40	TTCATCAACATTAAATGTGAGCGATATTCAAC
41	CTTTATTTCTGATAAATTAATGCCAGCCAGCT
42	ATAATTCGCGTCTGGCCTTCCTGTGGAGAGGG
43	TTATGACCTTTGAGAGATCTACAACATCAAAA
44	TCATTTTTTAACCAATAGGAACGCAGGCTATC
45	AGCATAAAGCCTGAGAGTCTGGAGAAATCAGC
46	TAAAATTCGCATTAAATTTTTGTTCAAACAAG
47	AGAATTAGTGAACGGTAATCGTAATATTTTGT
48	AAATATTTAAATTGTAAACGTTAAAACTAGCA
49	TTAACATCATATGTACCCCGGTTGGTATAAGC
50	CAAAAACAGGAAGATTATAATCAG
51	AAAAGCCCTCAATTCTACTAATAGAGCTGAAA
52	TGTCAATCCAATAAATCATACAGGACCTGTTT
53	AGAATCGACAAAATTAAGCAATAAGACCATTA
54	AGGTCATTGCTAAATCGGTTGTACTTCCCAAT
55	TAGCTATTCTGTAATACTTTTGCGTGTCTGGA
56	CGTTCTAGCAACGCAAGGATAAAACAACATGT
57	ACAGTCAATATATTTTAAATGCAAAGAGCTTA
58	CCTTTTGATTTTTGCGGATGGCTTTGCCTGAG
59	CAGGTCAGATATAATGCTGTAGCTATTTTTAG
60	TCAAAGCGTGCAACTAAAGTACGGGGAGAAGC
61	AAGACTTCTCCATATAACAGTTGACAAAAACA
62	GATTGCATCGAGTAGATTTAGTTTAGCCTCAG
63	TTACCCTGTCGCAAATGGTCAATACAAGGCAA
64	ACGAGAATTTTCATTTGGGGCGCGTAGTAGCA
65	AGGTGGCACAAATGCTTTAAACAGTTCATTGA
66	AGCTATATGACCATAAATCAAAAATAGCGTCC
67	GATACATTACTATTATAGTCAGAAAGAGGGGG
68	TCTGCGAACAAAAAGATTAAGAGGTAGCGAGA
69	AGTTTCATAAATATCGCGTTTTAACCTCGTTT
70	TTTAAATAAACCAGACCGGAAGCAACGAGGCA
71	ATTGCTGAGATTAGAGAGTACCTTACTAATGC
72	ATTCATCAAGGAATACCACATTCATAATTGCT
73	ACTAACGGAACGCCAAAAGGAATTAACTCCAA
74	GTTGGGAAGCAACACTATCATAACTTCGAGCT
75	TTTAAGAAACGATAAAAACCAAAAAAGCCCGA
76	AACTTTAACAAAAGAAGTTTTGCCGCAAAGCG
77	GTAGTAAAAAATGTTTAGACTGGATCAGGTCT
78	AAGGCTTGGGAATCGTCATAAATATTCAGAAA
79	ATCCCCCTAACAAAGCTGCTCATTTTACCCAA
80	AATACTGCCCCTGACGAGAAACACCAAGAGTA
81	TAATAGTATTGGGCTTGAGATGGTCAGACCAG
82	GGCTTTTGTCATTGTGAATTACCTCCAACTTT
83	ACCAGACGCTGGCTCATTATACCAGCAGACGG
84	TAGTAAGAGAAAAATCTACGTTAAACCTGCTC

85 AGATACATAACAACATTATTACAGTCATCGCC
86 CCAAGCGCACAACGGAGATTTGTAGTAGAAAG
87 AAAACACTTGTGTCGAAATCCGCGTAAAACGA
88 AACCTAAATTAGCCGGAACGAGGCGTCAGGAC
89 GGTAAAATAAGGGAACCGAACTGATATGCGAT
90 AGACTTTTACAGATGAACGGTGTATTAATTTC
91 GTAGCAACGCTGGCTGACCTTCATCAGAACGA
92 CCTCAGCAAAGAACCGGATATTCACAGTGAAT
93 ATCAACGTAGGCCGCTTTTGCGGGGCTTGCAGGGAGTTAA
94 ATCTTGACGCGAAAGACAGCATCGCCCACGCA
95 GCGCATAGGGCTACAGAGGCTTTGGATAGTTG
96 GAAAGAGGTCATGAGGAAGTTTCCTTCGAGGT
97 TCAATCATACGTAATGCCACTACGAGCCTTTA
98 CATGTTACACGAAAGAGGCAAAAGTTGAAAAT
99 TGATAAATCATCTTTGACCCCCAGCTAAAGGA
100 ACAGTTTCGAATAGAAAGGAACAACGATTATA
101 TTTCTGTATAATAATTTTTTCACGAATACACT
102 TTTGTCGTAAAAGGCTCCAAAAGGAAGGCACC
103 CAGCCCTCGGTTTATCAGCTTGCTATTAAACG
104 TACAAACTTAAACAGCTTGATACCAGGACTAA
105 CCATGTACATGACAACAACCATCGGAACGAGG
106 TCATTTTCATATTCGGTCGCTGAGATCGTCAC
107 TAACCGATAGGGATAGCAAGCCCATCAGAACC
108 CGCCGACACGTAACACTGAGTTTCGTTTAGTA
109 GAATTTCTACAACGCCTGTAGCATGCCCGGAA
110 ATTGTATCATAGTTAGCGTAACGATAAGTGCC
111 CTCCAAAACTTTCCAGACGTTAGTGATTAGCG
112 ATTGCGAATGGGATTTTGCTAAACAAGAGGCT
113 ACAAATAAGAAACATGAAAGTATTAACTTTCA
114 GCAGGTCATCAAGAGAAGGATTAGAAATGAAT
115 AGAGCCGCCTCAGTACCAGGCGGATCTAAAGT
116 CCACCCTCGGTTGATATAAGTATATCCACAGA
117 CGCCACCCTCACCGTACTCAGGAGGTCACCAG
118 AGAGCCACCTCAGAACCGCCACCCATAGGAAC
119 GCCATCTTCAGAGCCACCAGATAACAGAACCC
120 GCCACCCTTTCATAATCAAAATCATTTTCGGT
121 CCGCCACCCACCGGAACCGCCTCCTAGCGTCA
122 TAGGTGTATCAGAACCGCCACCCTGTAATCAG
123 GTCGAGAGAGAGCCGCCACCAGAAACGTCACC
124 GGGTTTTGCGCCAGCATTGACAGGCAGTAGCA
125 GAGACTCCGACGATTGGCCTTGATAGCCATTT
126 ATTATTCTATCCTCATTAAAGCCAATTAAAGG
127 ACAGTTAACTCTGAATTTACCGTTGGAGGGAA
128 GGTCAGTGATGGCTTTTGATGATAAAAGACAA
129 TTACCAGCGCCCAGGAGTGTAC
130 TTATTTTGCATTCAACCGATTGAGCCAGTAAG
131 TAAAAGAATTGACGGAAATTATTCGAATGGAA
132 GTAGAAAACACCGTCACCGACTTGATTCACAA
133 AAGACTCCGAGCCAGCAAAATCACAGGTTGAG
134 ATAACGGAATTAGCAAGGCCGGAACCACCACC
135 AGTTACCACATCGATAGCAGCACCCAGAGCCA
136 TTTTTAAGGAATCAAGTTTGCCTTCTCAGAGC
137 ATGAAATACGCGTTTTCATCGGCACCGGAACC
138 TGAGTTAACCTTATTAGCGACCGCTTAACTTT
139 CATAGCCCGCCCAATAATAAGAGCGCTAATAT
140 GACTGTAGGCAATAGCTATCTTACACACCCTG
141 TAGCGACAAAAAGTAAGCAGATAGAGGGAAGC
142 AATGAAACGAAGGAAACCGAGGAAAGCAGCCT
143 CCATTACCATACCCAAAAGAACTGACGATTTT
144 GGGAATTATTATTACGCAGTATGTAGCCATAT
145 TGAATTATTACATACATAAAGGTGCCAGAGCC
146 GGTAAATAACGCAAAGACACCACGCCTGAATC
147 AAGGGCGATCACAATCAATAGAAAAGTTGCTA
148 AAATCAAGATTATTCATATGGT
149 CGAACCTCCCAGCTACAATTTTATGAATAAGT
150 ATCCGGTAGCTAACGAGCGTCTTTGCAACATA
151 CCAATAGCCAGTTACAAAATAAACTAGCAAAC
152 TTTTATTTCCAATCCAAATAAGAAGCATGATT
153 AGTACCGCCGTCAAAAATGAAAATACGCAATA
154 TCCTTATCAGAATAACATAAAAACCCGAACAA
155 CATGTAGACGGGAGAATTAACTGACGAAGCCC
156 AAGAAAAACAGAGGGTAATTGAGCAAGAAACA
157 GCCTGTTTTAACCCACAAGCAGCAGCCAGAAT
158 CAGAGAGAATCAACAATAGATAAGTAAACAAC
159 AACAAAGTTAATATCCCATCCTAAAAGGTAAA
160 GCATTAGAAACCAATCAATAATCGGCCAGTAA
161 TTACAGAGATTCCAAGAACGGGTACCAACATG
162 TTGTTTAAACTCATCGAGAACAAGGGCTTAAT
163 TATTTATCTCATCGTAGGAATCATCTTACCAG
164 TAATTTGCAAGCAAATCAGATATAGCCTGTTT
165 TTACCAACTTCTAAGAACGCGAGGAATAAACA
166 TTTTGCACCCGACTTGCGGGAGGTCCGACCGT
167 GGTTTGAAATATTTGAAGCCTT
168 TTTTAGTTTAAGGCGTTAAATAAGCGTTTTAG

169 ACAAAGAAATAATTACTAGAAAAAGAAGGCTT
170 TATGTAAAATGCGTTATACAAATTTACCGCGC
171 TAACCTCCCAACGCTCAACAGTAGCAAGCCGT
172 ATCAAAATGCCATATTTAACAACGTTAAACCA
173 TTAAGACGGCAGAGGCATTTTCGAGCTGTCTT
174 TAGAATCCTATAAAGTACCGACAATTTACGAG
175 CTGTAAATGTCCAGACGACGACAATCCTGAAC
176 AAATCAATCTAATGCAGAATGAACCTTGCCGC
177 ATGTTCAGATATGTGAGTGAATAAGAATTACC
178 GTAATTCTCGTCGCTATTAATTAAAACAAAAT
179 TAAGAGAATTGAAAACATAGCGATGCAAAAGA
180 TAATTTAGCTGAGAAGAGTCAATAGCAGAGGC
181 TGAGAATCCATAGGTCTGAGAGACTGATTGCT
182 TATAAAGCGGCTTAGGTTGGGTTATTTACATC
183 AGTATCATTGCTGATGCAAATCCATAACGTCA
184 CCGGAATCCGCGAGAAAACTTTTTAGAAATAA
185 GTGATAAAAATTTCATCTTCTGACCCTACCAT
186 AAGGGTTAGAACTAAATTTAAT
187 ATCCTGATTATTTGCACGTAAAACCAAATATA
188 TGATTATCCGTAGATTTTCAGGTTATCGCAAG
189 CACCAGAATACAGTAACAGTACCTTATAACTA
190 AGTAACATCAATAACGGATTCGCCTACCTTTT
191 TTTGCCCGCAAGTTACAAAATCGCGTGAATTT
192 TTACAAACCATTTCAATTACCTGAAGCTTAGA
193 TAATACATAAACAAACATCAAGAATTTTCCCT
194 CTAACAACTTTAACAATTTCATTTCCTTGCTT
195 GAGGAAGGTGGAAACAGTACTGGTAATATCAT
196 TTTTTTAATTATCTAAAATATCTTTAGGAGCA
197 TAATTACATAATAGATTAGAGCCGTCAATAGA
198 AGATGATGTTGAGGATTTAGAAGTATTAGACT
199 GAATTATTAATTCGACAACTCGTATTAAATCC
200 TTGAATACAACGTTATTAATTTTAAAAGTTTG
201 GGGAGAAATATCATTTTGCGGAACAAAGAAAC
202 GATGAATAGGAGCGGAATTATCATCATATTCC
203 AGAAATTGAGATGATGGCAATTCATCAATATA
204 ATCAAAATTGTTTGGATTATACTTCTGAATAATGG
205 TCAACAGTTGAAAGGAATT
206 AACCCTCAATCCAGTTGGCAAA
207 TAAAGCATCACCTCAAATATCA
208 CCACGCTGAGAAATGAAAAATC
209 GGCGGTCAGTACTGCAACAGTG
210 CGAACCACCAGAACAGAGGTGA

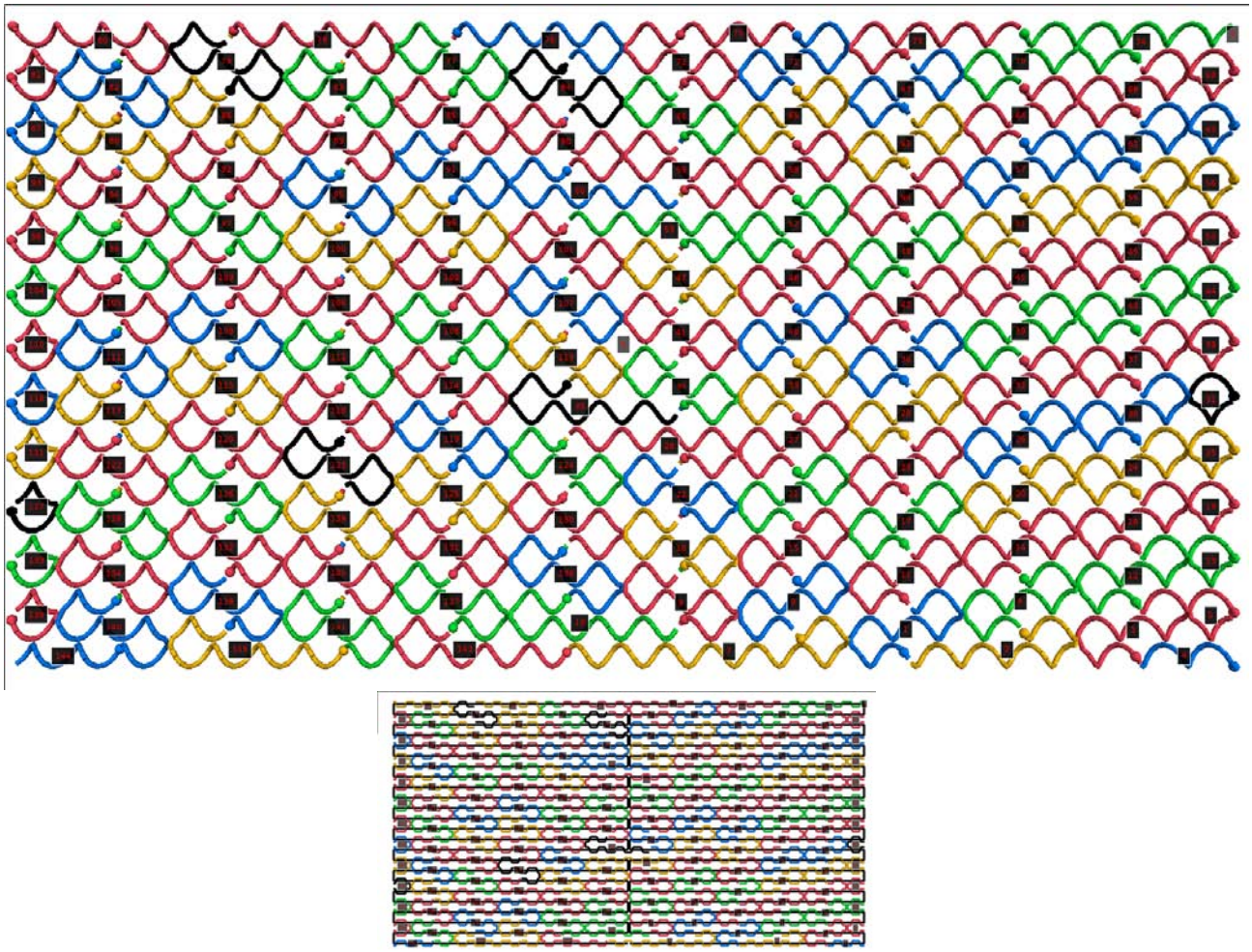


Figure J.3: Schem of Origami 5 without (a) and with the path of the scaffold

J.5 Origami O5

The staples are given from 5to 3. 3 is represented by a small ball in (Fig. J.3)

1 ACAGTGCCACGCTATACGTGGCACAGACAATATTTTTGAACAACAGGAAAAACG
2 CTCATGGAAATACCTACATTTTGACGCTCAATCGTCTGAAGAACCCTTCTGACC
3 TGAACCTCAAATAAGGGACATTCTGGCCAACAGAGATAATGGATTATTTACA
4 TTGGCAGATTCACCAGTCACACG
5 ACCAGTAATAAATCAAACCCTC
6 TGAAAGCGTAAGAGAGAGCCAGCAGCAAATGAAAAATCTAAGGAATTGAGGAAG
7 GAAAGGAAGGGAACAAACTATCGGCCTTGCTGGTAATATCCAGAACAATATTACCGCCAGCCATTGTGGCTATTAGTCT
8 TTAATGCGCGAACTAAAACAGAGGTGAGGCGGTCAGTATTTAGAGCCGTCAA
9 TATTAGACTTTACAATACCGAACGAACCACCAGCAGAAGATGATAGCCCTAAAA
10 CATCGCCATTAAATGGTGGTTCCGAAATCGGCAAAATCCCGCCCCAGCAGGCGA
11 TTATCATTTTGCGTCTTTAGGAGCACTAACAACTAATAGATAACACCGCCTGCA
12 TCCTGATTATCAGGTCAGTTGGCAAATCAACAGTTGAAAAGCATCACCTTGC
13 AATCAATATCTGATGATGGCAA
14 GTTATCTAAAATAGAACAAAGAAACCACCAGAAGGAGCGGTGAATAATGGAAGG
15 TAGATAATACATTCTTTGCCCGAACGTTATTAATTTTAAGAAATAAAGAAAT
16 ATGAATATACAGTAAACAATTCGACAACTCGTATTAAATCTGAGGATTTAGAAG
17 AAATCGCGCAGAGATATCAAAATTATTTGCACGTAAAACAAAGTTTGAGTAACA
18 GAAACAAACATCAATCCTGATTGTTTGGATTATACTTCAATTATCATCATAT
19 TTCATCAATATAAGAAAACAAA
20 GTTAGAACCTACCGCGAATTATTCATTTCAATTACCTGAGTTTAATGGAAACAG
21 TGCGTAGATTTTCAATAACGGATTCGCCTGATTGCTTTGTAAATCGTCGCTA
22 AAAACATAGCGATAACAGTACCTTTTACATCGGGAGAAACAGGTTTAACGTCAG
23 CTTTTTAACCTCCATATGTGAGTGAATAACCTTGCTTCTGAATACCAAGTTACA
24 AAATCCAATCGCTTAACAATTTCATTTGAATTACCTTTCAAAAGAAGATGAT
25 ATTAATTACATAAGACAAAGAA
26 TACATAAATCAATGGCTTAGGTTGGGTTATATAACTATATCATCTTCTGACCTA
27 TTAATTAATTTTCCAATAGTGAATTTATCAAAATCATAGCGTTAAATAAGAA
28 CATTAAATTTTTGAGCTTAGATTAAGACGCTGAGAAGAGTCCTTAGAATCCTTG
29 ATCGCCATATTTAGAAATACCGACCGTGTGATAAATAAGGGTCTGAGAGACTAC
30 AGTAATAAGAGATTTTTCAAATATATTTTAGTTAATTTGTAAATGCTGATGC
31 CGCGAGAAAACATATAAAGTAC
32 AATTTAATGGTTTACAACGCCAACATGTAATTTAGGCAGAATAAACAACATGTT
33 TAAACACCGGAATCAGTATAAAGCCAACGCTCAACAGTAGTCCTGAACAAGA
34 GCATGTAGAAACCTATCATATGCGTTATACAAATTCTTACCATAATTACTAGAA
35 AAAGCCTGTTTAGAAGGCCGGAGACAGTCAAATCACCATCTGTAGGTAAAGATT
36 GAGAACAAGCAAGACGCGCCTGTTTATCAACAATAGATAAGGGCTTAATTGAGA
37 AGCAAGCAAATCAAAGTAATTCTGTCCAGACGACGACAGGCATTTTCGAGCC
38 CGACAAAAGGTAGATATAGAAG
39 CAGCTAATGCAGACCGTTTTTATTTTCATCGTAGGAATCAGAACCTCCCGACTT
40 AAAATAATATCCCCATTCCAAGAACGGGTATTAAACCAATTTGCACCCAGCT
41 CTAACGAGCGTCTAATCAATAATCGGCTGTCTTTCCTTATATCCTAATTTACGA
42 CGATTTTTTGTTTAAGCCTTAAATCAAGATTAGTTGCTATGTACCGCACTCATC

43 CATAAAAACAGGATTCTAAGAACGCGAGGCGTTTTAGCTTACCGCGCCCAAT
44 GCTTATCCGGTGAAGCGCATTA
45 GCGGGAGGTTTTGAACGTCAAAAATGAAAATAGCAGCCTTAGGGTAATTGAGCG
46 ACAATTTTATCCTAATAAACAGCCATATTATTTATCCCACAATAATAAGAGC
47 CTTACCGAAGCCCTTCCAGAGCCTAATTTGCCAGTTACAAGAATCTTACCAACG
48 ACGGAATACCCAAGATAACCCACAAGAATTGAGTTAAGCCATCCAAATAAGAAA
49 AGCAAACGTAGATAACTGAACACCCTGAACAAAGTCAGTACAGAGAGAATAA
50 GACGGGAGAATAAATACATACA
51 CTAATATCAGAGAAAGAACTGGCATGATTAAGACTCCTTAACGGAATAAGTTTA
52 AAGAAACAATGAAAACAAAGTTACCAGAAGGAAACCGAGGCCAAAGACAAAA
53 GCATAGTAAGAGCTTTTTAAGAAAAGTAAGCAGATAGCCGATAGCAATAGCTAT
54 AGCAAAATCACCAAATAGAAAATTCATATGGTTTACCAGCGAAACGCAATAATA
55 GAAACCATCGATACATATAAAAGAAACGCAAAGACACCTTACGCAGTATGTT
56 TAAAGGTGGCAAGCAGCACCGT
57 TTTTGTCACAATCGTAGCACCATTACCATTAGCAAGGCCGAGACTGTAGCGCGT
58 GGGCGACATTCAATATCACCGTCACCGACTTGAGCCATTTGCCATCTTTTCA
59 CCACCGGAACCGCGACGGAAATTATTCATTAAAGGTGAATCCGATTGAGGGAGG
60 GAAGGTAAATATTCATTCAGTGAATAAGGCTTGCCCTGACTACCCAAATCAACG
61 ACCACCACCAGAGTTCGGTCATAGCCCCCTTATTAGCGTTTGGGAATTAGAGCC
62 TTGGCCTTGATAACAGAATCAAGTTTGCCTTTAGCGTCGAAACGTCACCAAT
63 AATCAGTAGCGTTCACAAACAA
64 TTTCATCGGCATTCCGCCGCCAGCATTGACAGGAGGTTGATCTGAATTTACCGT
65 TAATCAAAATCACCGCCACCCTCAGAGCCACCACCCTCAGTACTGGTAATAA
66 ACAGTGCCCGTATCTCCCTCAGAGCCGCCACCCTCAGAACCGGAACCAGAGCCA
67 AGACTCCTCAAGAATACATGGCTTTTGATGATACAGGAGTGAGCCGCCACCAGA
68 ATAAGTGCCGTCTTAAAGCCAGAATGGAAAGCGCAGTCGGCAGGTCAGACGA
69 ATAAATCCTCAGAGAGGGTTGA
70 TCCAGTAAGCGTCGAAGGATTAGGATTAGCGGGGTTTTGCAGGAGGTTTAGTAC
71 GTTTTAACGGGGTGGAACCTATTATTCTGAAACATGAAAAGAGCCACCACCC
72 GGAACCCATGTACAAACAGTTAATGCCCCCTGCCTATTTCCAGTGCCTTGAGTA
73 CGCCACCCTCAGAACCGCCACCCTCAGAACCGCCACCCTCGTATTAAGAGGCTG
74 TATAAGTATAGCCCGGAATAGGTGTATCACCGTACTCTCAGTACCAGGCGG
75 TCATTTTCAGGGATAGCAAGCCCAATA
76 GCTTGCTTTCGAGCGTAACACTGAGTTTCGTCACCAGTACAAACTACAACGCCT
77 GTAGCATTCCACAAAAAGGCTCCAAAAGGAGCCTTTAATTATAGTTGCGCCGAC
78 AAATCTCCAAAAGACAGCCCTCATAGTTAGCGTAACGATCTAAAGTTTTGTC
79 GTCTTTCCAGACGAACAACTAAAGGAATTGCGAATAATAAGGTCGCTGAGGCTT
80 AGAATAGAAAGGTTAGTAAATGAATTTTCTGTATGGGATTTTGCTAAAC
81 AACTTTCAACACAGCAGCGAAA
82 CGGCTACAGAGGGGCCGCTTTTGCGGGATCGTCACCCTGTTTCAGCGGAGTG
83 AACGGGTAAAATATCGCCCACGCATAACCGATATATTCTTTTTTCACGTTGA
84 GCAAAAGAATACAGTGAATTTCTTAAACAGCTTGATACCGGTATCGGTTTATCA

85 AATGACAACAACCACGTAATGCCACTACGAAGGCACCAACGATTATACCAAGCG
86 GCAGGGAGTTAAACTTTGAGGACTAAAGACTTTTTCATGAAAATTGTGTCGAAA
87 GACAGCATCGGCAGACGGTCAA
88 TTTGAAAGAGGACCATGTTACTTAGCCGGAACGAGGCGAACGAGGGTAGCAA
89 CTTCATCAAGAGAACGGAGATTTGTATCATCGCCTGATGGAAGTTTCCATTA
90 TAACAAAGCTGCTCTAAAACACTCATCTTTGACCCCCAGCCTAAAACGAAAGAG
91 CGAAACAAAGTACTAATCTTGACAAGAACCGGATATTCATGAGAAACACCAGAA
92 TCCGCGACCTGCTCAGATGAACGGTGTACAGACCAGGCGCTTAATCATTGTGAA
93 TCATAAGGGAATCAGGACGTTG
94 CGAACTAACGGATTTTAAGAACTGGCTCATTATACCAGCCGAACTGACCAAC
95 AGGAATACCACAGGGCTTGAGATGGTTTAATTTCAACTATAGGCTGGCTGAC
96 CGAGTAGTAAATTTTCAACTAATGCAGATACATAACGCCAACGACGATAAAAAC
97 TTACCTTATGCGAACAACATTATTACAGGTAGAAAGATTCAGGGGGTAATAGTA
98 GGAAGAAAAATTCATAAATATT
99 AACAGTTCAGAAGGATAGCGTCCAATACTGCGGAATCGCTACGTTAATAAAA
100 ACTATTATAGTCGGCTTTTGCAAAAGAAGTTTTGCCAGATCAGTTGAGATTT
101 CGAAAGACTTCAAAACACTATCATAACCCTCGTTTACCAGAAAGGAATTACGAG
102 CAAAATAGCGAGAAGAAGCAAAGCGGATTGCATCAAAAAGGCGAACCAGACCGG
103 AAATGTTTAGACTAACGAGAATGACCATAAATCAAAAATCATTGCTCCTTTTGA
104 CATTGAATCCCTGAATATAATG
105 GCAACTAAAGTATTGCGGATGGCTTAGAGCTTAATTGCCCTCAAATGCTTTA
106 ATTCTGCGAACGCAGGTCAGGATTAGAGAGTACCTTTAAGGTCTTTACCCTG
107 CAATAACCTGTTTATATCGCGTTTTAATTCGAGCTTCAAAATTAAGAGGAAGCC
108 AAGCAAACTCCAAAGTAGATTTAGTTTGACCATTAGATACTGAAAAGGTGGCAT
109 TAAGAGGTCATTTCGGTGTCTGGAAGTTTCATTCCATATACATACAGGCAAGGC
110 CTGTAGCTCAAAAAGCTAAATC
111 TGTAATACTTTTAATTAAGCAATAAAGCCTCAGAGCATCATGTTTTAAATAT
112 AGAACCCTCATAAGTAGTAGCATTAACATCCAATAAATACAGTTGATTCCCA
113 CAAAAGGGTGAGAAGCTATATTTTCATTTGGGGCGCGAGCATTTCGCAAATGGT
114 CAATTCTACTAATTATTTTAAATGCAATGCCTGAGTAATGAATATGATATTCAA
115 AAAGAATTAGCAAGCGGGAGAAGCCTTTATTTCAACGCAATTTTTGAGAGATCT
116 GGTTGTACCAAACAAGAGAATC
117 ATGTCAATCATAGGTCATTGCCTGAGAGTCTGGAGCAAAAACATTATGACCC
118 GATTGTATAAGCTAAATTAATGCCGGAGAGGGTAGCTAGGATAAAAATTTTT
119 CCGTTCTAGCTGAAAATATTTAAATTGTAAACGTTAATATGGAACGCCATCAAA
120 ACAAAGGCTATCATGTACCCCGGTTGATAATCAGAAAAGCAACATTAAATGTGA
121 GATGAACGGTAGCGGATTGACC
122 AGATGGGCGCATCGTCGGATTCTCCGTGGGAACAAACGATCGTAAAACTAGC
123 TCGGCCTCAGGATGGCCTTCCTGTAGCCAGCTTTCATCCCCAAAAACAGGAA
124 CGGAAACCAGGCATTAAATCAGCTCATTTTTTAACCAATATTTGTTAAAATTCG
125 AATAATTCGCGTCAGATCGCACTCCAGCCAGCTTTCCGGCAACTGTTGGGAAGG
126 GCGAGTAACAACCCGTAACCGTGCATCTGCCAGTTTGAGGGAAAGGGGGATGTG

127 GTAATGGGATAAGTCACGACGT
128 TTGCATGCCTGCAAGTTGGGTAACGCCAGGGTTTTCCCGGTCACGTTGGTGT
129 GTAATCATGGTCGCCTCTTCGCTATTACGCCAGCTGGCGGACGACGACAGTA
130 CACACAACATACGAAGCGCCATTCGCCATTCAGGCTGCGCACCGCTTCTGGTGC
131 GCGATCGGTGCGGATAGCTGTTTCCTGTGTGAAATTGTTAGGGTGCCTAATGAG
132 CTGCAAGGCGATTAGGTCGACTCTAGAGGATCCCCGGGTACTTTCCAGTCGGGA
133 TGTAAAACGACGCGGGGAGAGG
134 GGTTTTTCTTTTAGCTGCATTAATGAATCGGCCAACGCGGCCAGTGCCAAGC
135 GGCCCTGAGAGAATTAATTGCGTTGCGCTCACTGCCCGCCGAGCTCGAATTC
136 AAATCCTGTTTGAAGCCGGAAGCATAAAGTGTAAAGCCTGTCCGCTCACAATTC
137 TGAGCTAACTCACGTTGCAGCAAGCGGTCCACGCTGGTTTTTATAAATCAAAAG
138 AACCTGTCGTGCCCACCAGTGAGACGGGCAACAGCTGATTACAAGAGTCCACTA
139 CGGTTTGCGTAGTCTATCAGGG
140 CCCAAATCAAGTACTCCAACGTCAAAGGGCGAAAAACCTTGGGCGCCAGGGT
141 TAAAGGGAGCCCAGGGTTGAGTGTTGTTCCAGTTTGGAGCCCTTCACCGCCT
142 AATAGCCCGAGATCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGA
143 TTAAAGAACGTGGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCC
144 CGATGGCCCACTACGTGAACCATCA

Appendix K

Raman spectra of different molecules

All the abscissa are given in cm^{-1}. The ordinate is given in number of photon hit.

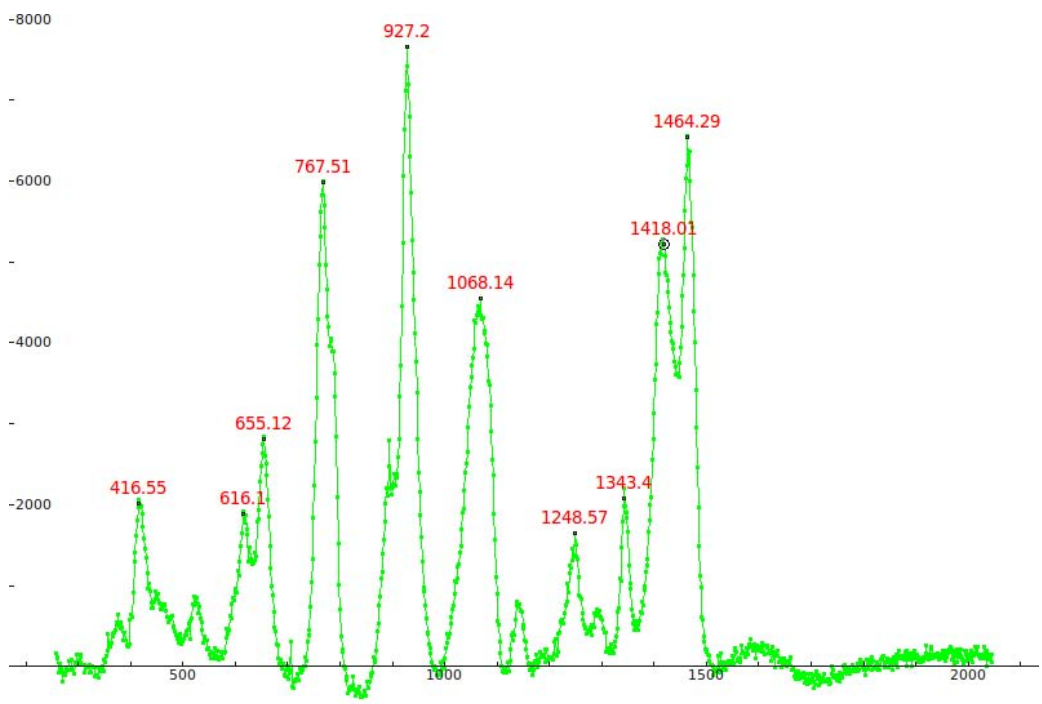


Figure K.1: tampon TAE (Tris acetate EDTA)

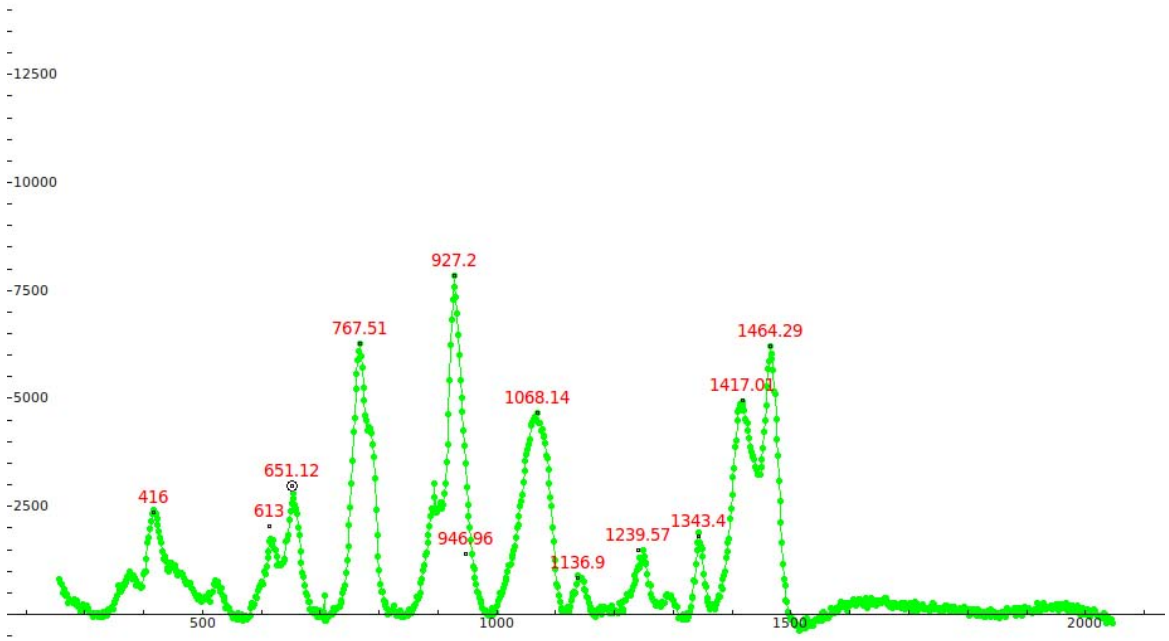


Figure K.2: tampon TBE (Tris borate EDTA)

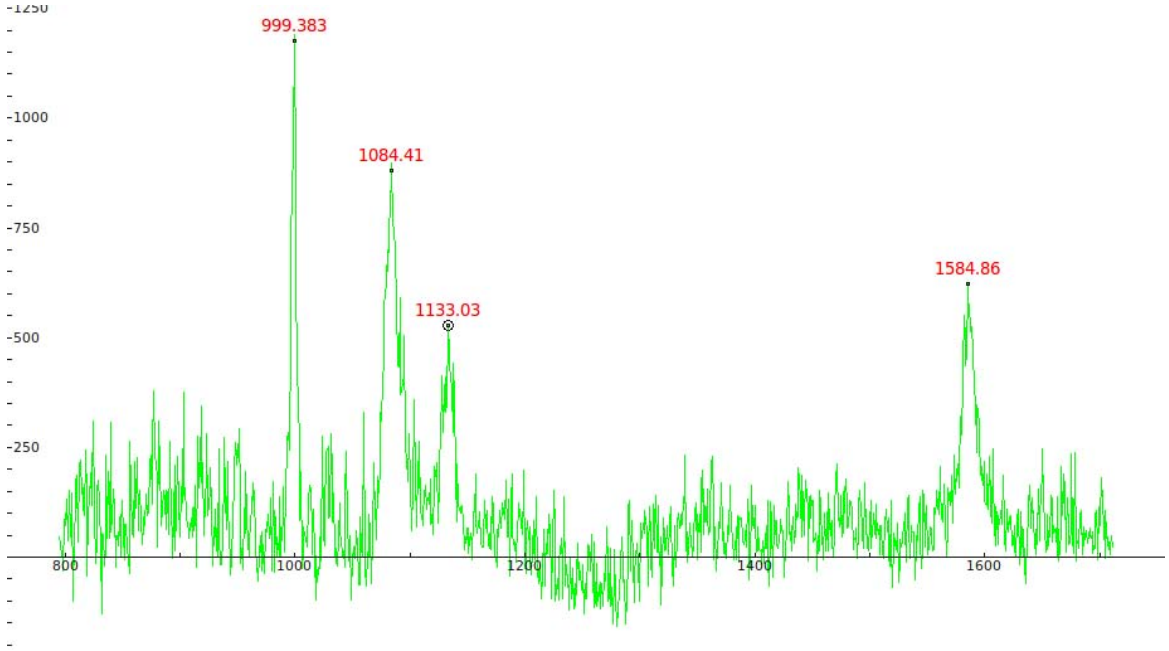


Figure K.3: tampon BSPP

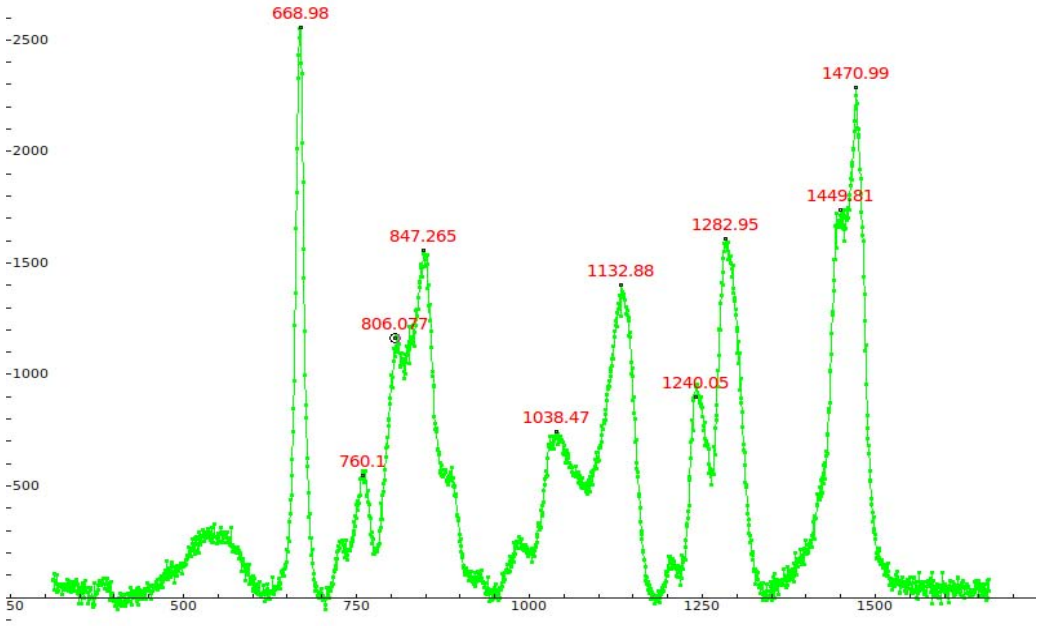


Figure K.4: PEG

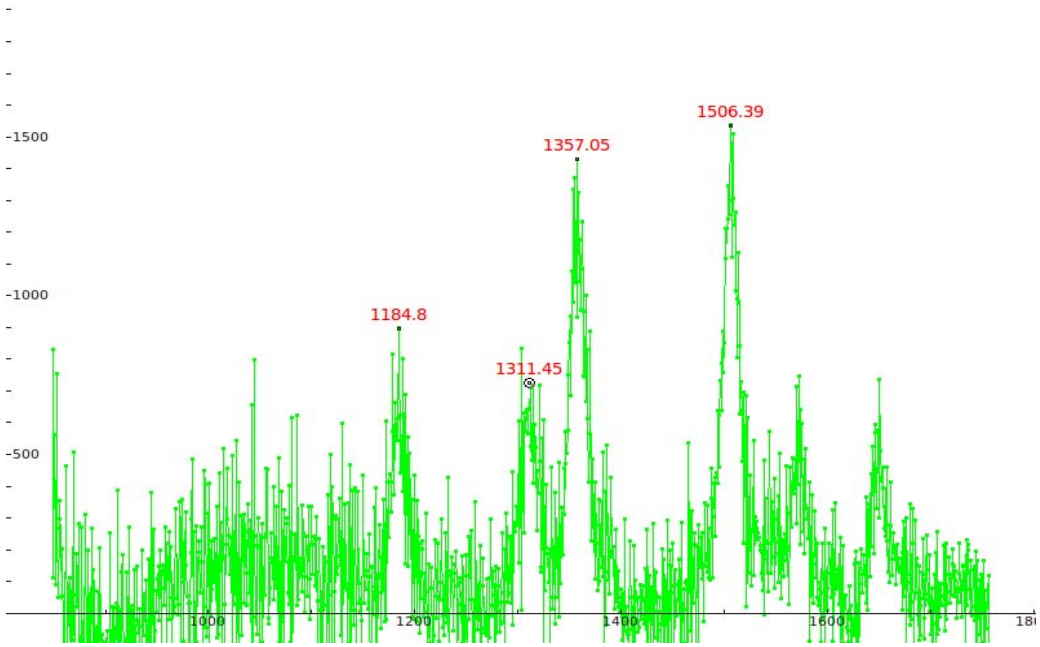


Figure K.5: Rhodamine 6G 100 μM (Mq)

Bibliography

[1] Rothemund, P. (2006)Folding DNA to create nanoscale shapes and patterns . *Nature* **440**, 297–302. i, 3, 5, 34, 60, 86, 163

[2] Ke, Y, Lindsay, S, Chang, Y, Liu, Y, & Yan, H. (2008)Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays. . *Science* **319**, 180–3. i, 6, 7, 123, 163

[3] Subramanian, H, Chakraborty, B, Sha, R, & Seeman, N. (2011)The Label-Free Unambiguous Detection and Symbolic Display of Single Nucleotide Polymorphisms on DNA Origami . *Nano letters*. i, 6

[4] Gu, H, Chao, J, Xiao, S, & Seeman, N. (2010)A proximity-based programmable DNA nanoscale assembly line . *Nature* **465**, 202–205. i, 7, 123

[5] Dietz, H, Douglas, S, & Shih, W. (2009)Folding DNA into twisted and curved nanoscale shapes . *Science* **325**, 725. i, 5, 35, 102

[6] Han, D, Pal, S, Nangreave, J, Deng, Z, Liu, Y, & Yan, H. (2011)DNA origami with complex curvatures in three-dimensional space . *Science* **332**, 342. i, 5

[7] Mergell, B, Ejtehadi, M, & Everaers, R. (2003)Modeling DNA structure, elasticity, and deformations at the base-pair level . *Physical Review E* **68**, 21911. ii, 10, 15, 17, 25, 34

[8] Kauert, D, Kurth, T, Liedl, T, & Seidel, R. (2011)Direct mechanical measurements reveal the material properties of 3D DNA-origami . *Nano Letters*. ii, 14, 45, 47

[9] Andersen, E. S, Dong, M, Nielsen, M. M, Jahn, K, Lind-Thomsen, A, Mamdouh, W, Gothelf, K. V, Besenbacher, F, & Kjems, J. (2008)DNA origami design of dolphin-shaped structures with flexible tails. . *ACS Nano* **2**, 1213–8. 5

[10] Andersen, E, Dong, M, Nielsen, M, Jahn, K, Subramani, R, Mamdouh, W, Golas, M, Sander, B, Stark, H, Oliveira, C, et al. (2009)Self-assembly of a nanoscale DNA box with a controllable lid . *Nature* **459**, 73–76. 5, 7

[11] Yurke, B, Turberfield, A, Mills Jr, A, Simmel, F, & Neumann, J. (2000)A DNA-fuelled molecular machine made of DNA . *Nature* **406**, 605–608. 5

[12] Chhabra, R, Sharma, J, Ke, Y, Liu, Y, Rinker, S, Lindsay, S, & Yan, H. (2007)Spatially addressable multiprotein nanoarrays templated by aptamer-tagged DNA nanoarchitectures . *J. Am. Chem. Soc* **129**, 10304–10305. 6, 11

[13] Rinker, S, Ke, Y, Liu, Y, Chhabra, R, & Yan, H. (2008)Self-assembled DNA nanostructures for distance-dependent multivalent ligand–protein binding . 7, 11

[14] Voigt, N, Tørring, T, Rotaru, A, Jacobsen, M, Ravnsbæk, J, Subramani, R, Mamdouh, W, Kjems, J, Mokhir, A, Besenbacher, F, et al. (2010)Single-molecule chemical reactions on DNA origami . *Nature nanotechnology* **5**, 200–203. 7

[15] Shen, W, Zhong, H, Neff, D, & Norton, M. (2009)NTA Directed Protein Nanopatterning on DNA Origami Nanoconstructs . *Journal of the American Chemical Society* pp. 427–431. 7

[16] Kuzuya, A, Kimura, M, Numajiri, K, Koshi, N, Ohnishi, T, Okada, F, & Komiyama, M. (2009)Precisely Programmed and Robust 2D Streptavidin Nanoarrays by Using Periodical Nanometer-Scale Wells Embedded in DNA Origami Assembly. . *Chembiochem: a European journal of chemical biology.* 7

[17] Maune, H, Han, S, Barish, R, Bockrath, M, Goddard III, W, Rothemund, P, & Winfree, E. (2009)Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates . *Nature Nanotechnology* **5**, 61–66. 7

[18] Gerdon, A, Oh, S, Hsieh, K, Ke, Y, Yan, H, & Soh, H. (2009)Controlled Delivery of DNA Origami on Patterned Surfaces* . *Small (Weinheim an der Bergstrasse, Germany).* 7

[19] Kershner, R, Bozano, L, Micheel, C, Hung, A, Fornof, A, Cha, J, Rettner, C, Bersani, M, Frommer, J, Rothemund, P, et al. (2009)Placement and orientation of individual DNA shapes on lithographically patterned surfaces . *Nature Nanotechnology* **4**, 557–561. 7

[20] Lund, K, Manzo, A, Dabby, N, Michelotti, N, Johnson-Buck, A, Nangreave, J, Taylor, S, Pei, R, Stojanovic, M, Walter, N, et al. (2010)Molecular robots guided by prescriptive landscapes . *Nature* **465**, 206–210. 7

[21] Gu, H, Chao, J, Xiao, S, & Seeman, N. (2009)Dynamic patterning programmed by DNA tiles captured on a DNA origami substrate . *Nature Nanotechnology.* 7

[22] Wickham, S, Endo, M, Katsuda, Y, Hidaka, K, Bath, J, Sugiyama, H, & Turberfield, A. (2011)Direct observation of stepwise movement of a synthetic molecular transporter . *Nature Nanotechnology* **6**, 166–169. 7

[23] Zhang, D & Seelig, G. (2011)Dynamic DNA nanotechnology using strand-displacement reactions . *Nature chemistry* **3**, 103–113. 7

[24] Han, D, Pal, S, Liu, Y, & Yan, H. (2010)Folding and cutting DNA into reconfigurable topological nanostructures . *Nature nanotechnology* **5**, 712–717. 7

[25] Soloveichik, D, Seelig, G, & Winfree, E. (2010)DNA as a universal substrate for chemical kinetics . *Proceedings of the National Academy of Sciences* **107**, 5393. 7

[26] Qian, L, Winfree, E, & Bruck, J. (2011)Neural network computation with DNA strand displacement cascades . *Nature* **475**, 368–372. 7

[27] SantaLucia Jr, J & Hicks, D. (2004)The thermodynamics of DNA structural motifs. . *Annual review of biophysics and biomolecular structure* **33**, 415. 9, 66, 94, 147

[28] Bloomfield, V. (1997)DNA condensation by multivalent cations . *Biopolymers* **44**, 269–282. 10, 26, 97

[29] Zhang, Y, Zheng, W, Liu, J, & Chen, Y. (1997)Theory of DNA melting based on the Peyrard-Bishop model . *Physical Review E* **56**, 7100. 10

[30] Poland, D & Scheraga, H. (1970) *Theory of helix-coil transitions in biopolymers.* (Acad. Pr.). 10, 91

[31] Poland, D & Scheraga, H. (1966)Phase transitions in one dimension and the helix-coil transition in polyamino acids . *The Journal of chemical physics* **45**, 1456. 10, 71

[32] Everaers, R, Kumar, S, & Simm, C. (2007)Unified description of poly-and oligonucleotide DNA melting: Nearest-neighbor, Poland-Sheraga, and lattice models . *Physical Review E* **75**, 041918. 10, 71, 72

[33] Essevaz-Roulet, B, Bockelmann, U, & Heslot, F. (1997)Mechanical separation of the complementary strands of DNA . *Proceedings of the National Academy of Sciences* **94**, 11935. 10

[34] Narang, P, Bhushan, K, Bose, S, & Jayaram, B. (2005)A computational pathway for bracketing native-like structures for small alpha helical globular proteins . *Phys. Chem. Chem. Phys.* **7**, 2364–2375. 11

[35] Jungmann, R, Steinhauer, C, Scheible, M, Kuzyk, A, Tinnefeld, P, & Simmel, F. (2010)Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami . *Nano letters.* 11

[36] Yin, Y & Zhao, X. (2011)Kinetics and Dynamics of DNA Hybridization . *Accounts of Chemical Research*. 11

[37] Liedl, T, Högberg, B, Tytell, J, Ingber, D, & Shih, W. (2010)Self-assembly of three-dimensional prestressed tensegrity structures from DNA . *Nature nanotechnology* **5**, 520–524. 14

[38] Rubinstein, M & Colby, R. (2003) *Polymer physics*. (Oxford University Press New York) Vol. 105. 15, 16

[39] Wiggins, P, van der Heijden, T, Moreno-Herrero, F, Spakowitz, A, Phillips, R, Widom, J, Dekker, C, & Nelson, P. (2006)High flexibility of DNA on short length scales probed by atomic force microscopy . *Nat. Nanotechnol* **1**, 137–141. 15, 22

[40] Kornyshev, A, Lee, D, Leikin, S, & Wynveen, A. (2007)Structure and interactions of biological helices . *Reviews of Modern Physics* **79**, 943–996. 15, 26

[41] Allahyarov, E, Gompper, G, & Lowen, H. (2005)DNA condensation and redissolution: interaction between overcharged DNA molecules . *Journal of Physics-Condensed Matter* **17**, 1827. 15, 26

[42] Kim, D, Kilchherr, F, Dietz, H, & Bathe, M. (2011)Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures . *Nucleic Acids Research*. 15, 47

[43] Doi, M & Edwards, S. (1988) *The theory of polymer dynamics*. (Oxford University Press, USA) Vol. 73. 16

[44] Cui, S, Yu, Y, & Lin, Z. (2009)Modeling single chain elasticity of single-stranded DNA: A comparison of three models . *Polymer* **50**, 930–935. 16, 18

[45] Rivetti, C, Guthold, M, & Bustamante, C. (1996)Scanning Force Microscopy of DNA Deposited onto Mica: EquilibrationversusKinetic Trapping Studied by Statistical Polymer Chain Analysis . *Journal of molecular biology* **264**, 919–932. 17, 24

[46] Cocco, S, Marko, J, & Monasson, R. (2002)Theoretical models for single-molecule DNA and RNA experiments: from elasticity to unzipping . *Comptes rendus-Physique* **3**, 569–584. 17, 21

[47] Smith, S, Cui, Y, & Bustamante, C. (1996)Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules . *Science* **271**, 795. 17, 18

[48] Bustamante, C, Smith, S, Liphardt, J, & Smith, D. (2000)Single-molecule studies of DNA mechanics . *Current Opinion in Structural Biology* **10**, 279–285. 18, 21

[49] Sain, A, Chen, J, & Ha, B. (2006)Persistency of single-stranded DNA: The interplay between base sequences and base stacking . *Physica A: Statistical Mechanics and its Applications* **369**, 679–687. 18

[50] Tinland, B, Pluen, A, Sturm, J, & Weill, G. (1997)Persistence Length of Single-Stranded DNA . *MACROMOLECULES* **30**, 5763–5765. 18

[51] Williams, L & Maher III, L. (2000)Electrostatic mechanisms of DNA deformation, Annu. Re V . *Biophys. Biomol. Struct* **29**, 497–521. 20

[52] Calladine, C, Drew, H, Luisi, B, & Travers, A. (2004) *Understanding DNA: the molecule & how it works*. (Academic press). 20

[53] Valle, F, Favre, M, De Los Rios, P, Rosa, A, & Dietler, G. (2005)Scaling Exponents and Probability Distributions of DNA End-to-End Distance . *Phys. Rev. Lett.* **95**, 158105. 20

[54] Konrad, M & Bolonick, J. (1996)Molecular dynamics simulation of DNA stretching is consistent with the tension observed for extension and strand separation and predicts a novel ladder structure . *J. Am. Chem. Soc* **118**, 10989–10994. 21

[55] Todd, B, Parsegian, V, Shirahata, A, Thomas, T, & Rau, D. (2008)Attractive Forces between Cation Condensed DNA Double Helices . *Biophysical Journal* **94**, 4775. 26

[56] Strey, H, Podgornik, R, Rau, D, & Parsegian, V. (1998)Dna-dna interactions . *Current opinion in structural biology* **8**, 309–313. 26, 97

[57] Maffeo, C, Schöpflin, R, Brutzer, H, Stehr, R, Aksimentiev, A, Wedemann, G, & Seidel, R. (2010)DNA–DNA Interactions in Tight Supercoils Are Described by a Small Effective Charge Density . *Physical review letters* **105**, 158101. 26, 45

[58] Israelachvili, J. (1991) *Intermolecular and surface forces*. (Academic press London). 26

[59] Catenaccio, A, Daruich, Y, & Magallanes, C. (2003)Temperature dependence of the permittivity of water . *Chemical physics letters* **367**, 669–671. 26

[60] Kean, S. (2011) *The Disappearing Spoon: And Other True Tales of Madness, Love, and the History of the World from the Periodic Table of the Elements*. (Doubleday). 28

[61] Newman, M & Barkema, G. (1999) *Monte Carlo methods in statistical physics*. (Oxford University Press, USA). 29, 30, 31, 32

[62] Gorin, A, Zhurkin, V, & Wilma, K. (1995)B-DNA twisting correlates with base-pair morphology . *Journal of molecular biology* **247**, 34–48. 34

[63] Ke, Y, Douglas, S, Liu, M, Sharma, J, Cheng, A, Leung, A, Liu, Y, Shih, W, & Yan, H. (2009)Multilayer DNA origami packed on a square lattice . *Journal of the American Chemical Society* **131**, 15903–15908. 34, 35

[64] Maiti, P, Pascal, T, Vaidehi, N, Heo, J, & Goddard, W. (2006)Atomic-Level Simulations of Seeman DNA Nanostructures: The Paranemic Crossover in Salt Solution . *Biophysical journal* **90**, 1463–1479. 35, 65

[65] Sa-Ardyen, P, Vologodskii, A, & Seeman, N. (2003)The flexibility of DNA double crossover molecules . *Biophysical journal* **84**, 3829–3837. 43

[66] Marky, L, Kallenbach, N, McDonough, K, Seeman, N, & Breslauer, K. (1987)The melting behavior of a DNA junction structure: a calorimetric and spectroscopic study . *Biopolymers* **26**, 1621–1634. 43, 75, 77

[67] Podtelezhnikov, A, Mao, C, Seeman, N, & Vologodskii, A. (2000)Multimerization-cyclization of DNA fragments as a method of conformational analysis . *Biophysical journal* **79**, 2692–2704. 43

[68] Landau, L & Lifshitz, E. (2003)Theory of Elasticity, 202 . 46

[69] Marko, J & Siggia, E. (1995)Stretching dna . *Macromolecules* **28**, 8759–8770. 46

[70] Douglas, S, Marblestone, A, Teerapittayanon, S, Vazquez, A, Church, G, & Shih, W. (2009)Rapid prototyping of 3D DNA-origami shapes with caDNAno . *Nucleic acids research* **37**, 5001. 47

[71] Chandler, D. (1987)Introduction to modern statistical mechanics . *Introduction to Modern Statistical Mechanics, by David Chandler, pp. 288. Foreword by David Chandler. Oxford University Press, Sep 1987. ISBN-10: 0195042778. ISBN-13: 9780195042771* **1**. 51

[72] Virnau, P & Müller, M. (2004)Calculation of free energy through successive umbrella sampling . *The Journal of chemical physics* **120**, 10925. 51

[73] Blossey, R & Carlon, E. (2003)Reparametrizing the loop entropy weights: Effect on DNA melting curves . *Phys. Rev. E* **68**, 061911. 60

[74] Tinoco Jr, I. (1960)Hypochromism in Polynucleotides1 . *Journal of the American Chemical Society* **82**, 4785–4790. 60, 63

[75] Jost, D & Everaers, R. (2009)A Unified Poland-Scheraga Model of Oligo-and Polynucleotide DNA Melting: Salt Effects and Predictive Power . *Biophysical journal* **96**, 1056–1067. 60, 91

[76] Arbona, J. M, Aime, J.-P, & Elezgaray, J. (2012)Folding of small origamis . *The Journal of Chemical Physics* **136**, 065102. 62

[77] Mergny, J & Lacroix, L. (2003)Analysis of thermal melting curves . *Oligonucleotides* **13**, 515–537. 63, 85, 86

[78] Spink, C. (2008)Differential scanning calorimetry . *Methods in cell biology* **84**, 115–141. 63

[79] Castro, C, Kilchherr, F, Kim, D, Shiao, E, Wauer, T, Wortmann, P, Bathe, M, & Dietz, H. (2011)A primer to scaffolded DNA origami . *Nature Methods* **8**, 221–229. 63

[80] Kypr, J, Kejnovska, I, Renvciuk, D, & Vorlivckova, M. (2009)Circular dichroism and conformational polymorphism of DNA . *Nucleic Acids Research* **37**, 1713–1725. 63

[81] Wildes, A, Theodorakopoulos, N, Valle-Orero, J, Cuesta-López, S, Garden, J, & Peyrard, M. (2011)Thermal Denaturation of DNA Studied with Neutron Scattering . *Physical Review Letters* **106**, 48101. 63

[82] Morrison, L & Stols, L. (1993)Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution . *Biochemistry* **32**, 3095–3104. 63

[83] Saccà, B, Meyer, R, Feldkamp, U, Schroeder, H, & Niemeyer, C. (2008)High-Throughput, Real-Time Monitoring of the Self-Assembly of DNA Nanostructures by FRET Spectroscopy . *Angewandte Chemie International Edition* **47**, 2135–2137. 63, 65

[84] Völker, J, Blake, R, Delcourt, S, & Breslauer, K. (1999)High-resolution calorimetric and optical melting profiles of DNA plasmids: Resolving contributions from intrinsic melting domains and specifically designed inserts . *Biopolymers* **50**, 303–318. 64

[85] Bernal-Méndez, E & Leumann, C. (2002)Stability and kinetics of nucleic acid triplexes with chimaeric DNA/RNA third strands . *Biochemistry* **41**, 12343–12349. 64

[86] Rougée, M, Faucon, B, Mergny, J, Barcelo, F, Giovannangeli, C, Garestier, T, & Hélène, C. (1992)Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatched . *Biochemistry* **31**, 9269–9278. 64

[87] Fu, T & Seeman, N. (1993)DNA double-crossover molecules . *Biochemistry* **32**, 3211–3220. 65

[88] Spink, C, Ding, L, Yang, Q, Sheardy, R, & Seeman, N. (2009)Thermodynamics of forming a parallel DNA crossover . *Biophysical journal* **97**, 528–538. 65

[89] Maiti, P, Pascal, T, Vaidehi, N, & Goddard III, W. (2004)The stability of Seeman JX DNA topoisomers of paranemic crossover (PX) molecules as a function of crossover number . *Nucleic acids research* **32**, 6047. 65

[90] Nangreave, J, Yan, H, & Liu, Y. (2011)DNA Nanostructures as Models for Evaluating the Role of Enthalpy and Entropy in Polyvalent Binding . *Journal of the American Chemical Society*. 65

[91] SantaLucia, J. (1998)A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics . *Proceedings of the National Academy of Sciences* **95**, 1460. 71

[92] Ouldridge, T, Johnston, I, Louis, A, & Doye, J. (2009)The self-assembly of DNA Holliday junctions studied with a minimal model . *The Journal of chemical physics* **130**, 065101. 71, 76

[93] Ouldridge, T, Louis, A, & Doye, J. (2010)DNA nanotweezers studied with a coarse-grained model of DNA . *Physical review letters* **104**, 178101. 71, 85

[94] Gillespie, J, Mayne, M, & Jiang, M. (2009)RNA folding on the 3D triangular lattice . *BMC bioinformatics* **10**, 369. 72, 73

[95] Garel, T & Orland, H. (2004)Generalized Poland–Scheraga model for DNA hybridization . *Biopolymers* **75**, 453–467. 72, 75

[96] Frenkel, D & Smit, B. (1996)Understanding molecular simulation: from algorithms to applications . 73

[97] Holliday, R. (1964)A mechanism for gene conversion in fungi . *Genet. Res* **5**, 282–304. 75

[98] Endo, M, Katsuda, Y, Hidaka, K, & Sugiyama, H. (2010)Regulation of DNA methylation using different tensions of double strands constructed in a defined DNA nanostructure . *Journal of the American Chemical Society* **132**, 1592–1597. 86

[99] Hughesman, C, Turner, R, & Haynes, C. (2011)Correcting for Heat Capacity and 5'-TA Type Terminal Nearest Neighbors Improves Prediction of DNA Melting Temperatures Using Nearest-Neighbor Thermodynamic Models . *Biochemistry*. 94, 147

[100] Owczarzy, R, Moreira, B, You, Y, Behlke, M, & Walder, J. (2008)Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations . *Biochemistry* **47**, 5336–5353. 97, 147, 148

[101] Tataurov, A, You, Y, & Owczarzy, R. (2008)Predicting ultraviolet spectrum of single stranded and double stranded deoxyribonucleic acids . *Biophysical Chemistry* **133**, 66–70. 97, 145

[102] Douglas, S, Dietz, H, Liedl, T, Hogberg, B, Graf, F, & Shih, W. (2009)Self-assembly of DNA into nanoscale three-dimensional shapes . *Nature* **459**, 414. 102

[103] Hansen, M, Zhang, A, Rangnekar, A, Bompiani, K, Carter, J, Gothelf, K, & LaBean, T. (2010)Weave Tile Architecture Construction Strategy for DNA Nanotechnology . *Journal of the American Chemical Society.* 102

[104] Rajendran, A, Endo, M, Katsuda, Y, Hidaka, K, & Sugiyama, H. (2011)Photo-cross-linking-assisted thermal stability of DNA origami structures and its application for higher-temperature self-assembly . *Journal of the American Chemical Society.* 102

[105] Song, J, Arbona, J.-M, Zhang, Z, Liu, L, Xie, E, Elezgaray, J, Aime, J.-P, Gothelf, K. V, Besenbacher, F, & Dong, M. (2012)Direct visualization of transient thermal response of a dna origami . *Journal of the American Chemical Society* **134**, 9844–9847. 102

[106] Jung, J, Chen, L, Lee, S, Kim, S, Seong, G, Choo, J, Lee, E, Oh, C, & Lee, S. (2007)Fast and sensitive DNA analysis using changes in the FRET signals of molecular beacons in a PDMS microfluidic channel . *Analytical and bioanalytical chemistry* **387**, 2609–2615. 105

[107] Lord, R & Thomas Jr, G. (1967)Raman studies of nucleic acids:: II. Aqueous purine and pyrimidine mixtures . *Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis* **142**, 1–11. 105

[108] Benevides, J, Overman, S, & Thomas Jr, G. (2005)Raman, polarized Raman and ultraviolet resonance Raman spectroscopy of nucleic acids and their complexes . *Journal of Raman Spectroscopy* **36**, 279–299. 105

[109] De Gelder, J, De Gussem, K, Vandenabeele, P, & Moens, L. (2007)Reference database of Raman spectra of biological molecules . *Journal of Raman Spectroscopy* **38**, 1133–1147. 105

[110] Moskovits, M. (2005)Surface-enhanced Raman spectroscopy: a brief retrospective . *Journal of Raman Spectroscopy* **36**, 485–496. 106, 110

[111] Aroca, R & Corporation, E. (2006) *Surface enhanced vibrational spectroscopy.* (Wiley Online Library). 106

[112] Braun, G, Lee, S, Dante, M, Nguyen, T, Moskovits, M, & Reich, N. (2007)Surface-enhanced Raman spectroscopy for DNA detection by nanoparticle assembly onto smooth metal films . *Journal of the American Chemical Society* **129**, 6378–6379. 106

[113] Faulds, K, Smith, W, & Graham, D. (2004)Evaluation of surface-enhanced resonance Raman scattering for quantitative DNA analysis . *Analytical chemistry* **76**, 412–417. 106

[114] Graham, D, Thompson, D, Smith, W, & Faulds, K. (2008)Control of enhanced Raman scattering using a DNA-based assembly process of dye-coded nanoparticles . *Nature nanotechnology* **3**, 548–551. 106

[115] Kneipp, K, Wang, Y, Kneipp, H, Perelman, L, Itzkan, I, Dasari, R, & Feld, M. (1997)Single molecule detection using surface-enhanced Raman scattering (SERS) . *Physical Review Letters* **78**, 1667–1670. 106

[116] Nie, S & Emory, S. (1997)Probing single molecules and single nanoparticles by surface-enhanced Raman scattering . *Science* **275**, 1102. 106

[117] Rodriguez-Lorenzo, L, Alvarez-Puebla, R, Pastoriza-Santos, I, Mazzucco, S, Stéphan, O, Kociak, M, Liz-Marzán, L, & Garcia de Abajo, F. (2009)Zeptomol detection through controlled ultrasensitive surface-enhanced Raman scattering . *Journal of the American Chemical Society* **131**, 4616–4618. 106

[118] Lim, D, Jeon, K, Kim, H, Nam, J, & Suh, Y. (2009)Nanogap-engineerable Raman-active nanodumbbells for single-molecule detection . *nature Materials* **9**, 60–67. 106

[119] Wustholz, K, Henry, A, McMahon, J, Freeman, R, Valley, N, Piotti, M, Natan, M, Schatz, G, & Duyne, R. (2010)Structure- Activity Relationships in Gold Nanoparticle Dimers and Trimers for Surface-Enhanced Raman Spectroscopy . *Journal of the American Chemical Society*. 106, 107

[120] Mirkin, C, Letsinger, R, Mucic, R, & Storhoff, J. (1996)A DNA-based method for rationally assembling nanoparticles into macroscopic materials . 106

[121] Alivisatos, A, Johnsson, K, Peng, X, Wilson, T, Loweth, C, Bruchez, M, & Schultz, P. (1996)Organization of'nanocrystal molecules' using DNA . 106

[122] Zanchet, D, Micheel, C, Parak, W, Gerion, D, & Alivisatos, A. (2001)Electrophoretic isolation of discrete Au nanocrystal/DNA conjugates . *Nano Letters* **1**, 32–35. 106

[123] Zanchet, D, Micheel, C, Parak, W, Gerion, D, Williams, S, & Alivisatos, A. (2002)Electrophoretic and structural studies of DNA-directed Au nanoparticle groupings . *The Journal of Physical Chemistry B* **106**, 11758–11763. 106

[124] Fu, A, Micheel, C, Cha, J, Chang, H, Yang, H, & Alivisatos, A. (2004)Discrete nanostructures of quantum dots/Au with DNA . *Journal of the American Chemical Society* **126**, 10832–10833. 106

[125] Deng, Z, Tian, Y, Lee, S, Ribbe, A, & Mao, C. (2005)DNA-Encoded Self-Assembly of Gold Nanoparticles into One-Dimensional Arrays . *Angewandte Chemie* **117**, 3648–3651. 106

[126] Zheng, J, Constantinou, P, Micheel, C, Alivisatos, A, Kiehl, R, & Seeman, N. (2006)Two-dimensional nanoparticle arrays show the organizational power of robust DNA motifs . *Nano Letters* **6**, 1502–1504. 106

[127] Bidault, S, de Abajo, F, & Polman, A. (2008)Plasmon-based nanolenses assembled on a well-defined DNA template . *Journal of the American Chemical Society* **130**, 2750–2751. 107

[128] Li, K, Stockman, M, & Bergman, D. (2003)Self-similar chain of metal nanospheres as an efficient nanolens . *Physical review letters* **91**, 227402. 107

[129] Busson, M, Rolly, B, Stout, B, Bonod, N, Larquet, E, Polman, A, & Bidault, S. (2011)Optical and Topological Characterization of Gold Nanoparticle Dimers Linked by a Single DNA Double Strand . *Nano letters*. 107, 117

[130] Dieringer, J, Wustholz, K, Masiello, D, Camden, J, Kleinman, S, Schatz, G, & Van Duyne, R. (2008)Surface-enhanced Raman excitation spectroscopy of a single rhodamine 6G molecule . *Journal of the American Chemical Society* **131**, 849–854. 107

[131] Elezgaray, J, Berguiga, L, & Argoul, F. (2010)Optimization of branched resonant nanostructures illuminated by a strongly focused beam . *Applied Physics Letters* **97**, 243103–243103. 107

[132] Schreiber, R, Kempter, S, Holler, S, Schüller, V, Schiffels, D, Simmel, S, Nickels, P, & Liedl, T. (2011)DNA Origami-Templated Growth of Arbitrarily Shaped Metal Nanoparticles . *Small*. 107

[133] Samano, E, Pilo-Pais, M, Goldberg, S, Vogen, B, Finkelstein, G, & LaBean, T. (2011)Self-assembling DNA templates for programmed artificial biomineralization . *Soft Matter* **7**, 3240–3245. 107

[134] Pilo-Pais, M, Goldberg, S, Samano, E, LaBean, T, & Finkelstein, G. (2011)Connecting the Nanodots: Programmable Nanofabrication of Fused Metal Shapes on DNA Templates . *Nano letters*. 108

[135] Pal, S, Varghese, R, Deng, Z, Zhao, Z, Kumar, A, Yan, H, & Liu, Y. (2011)Site-Specific Synthesis and In Situ Immobilization of Fluorescent Silver Nanoclusters on DNA Nanoscaffolds by Use of the Tollens Reaction . *Angewandte Chemie* **123**, 4262–4265. 108

[136] Liu, J, Geng, Y, Pound, E, Gyawali, S, Ashton, J, Hickey, J, Woolley, A, & Harb, J. (2011)Metallization of branched DNA origami for nanoelectronic circuit fabrication . *ACS nano*. 108

[137] Geng, Y, Liu, J, Pound, E, Gyawali, S, Harb, J, & Woolley, A. (2011)Rapid metallization of lambda DNA and DNA origami using a Pd seeding method . *J. Mater. Chem.* **21**, 12126–12131. 108

[138] Ding, B, Deng, Z, Yan, H, Cabrini, S, Zuckermann, R, & Bokor, J. (2010)Gold Nanoparticle Self-Similar Chain Structure Organized by DNA Origami . *J. Am. Chem. Soc* **132**, 3248–3249. 108, 112

[139] Kuzyk, A, Schreiber, R, Fan, Z, Pardatscher, G, Roller, E, HAkgele, A, Simmel, F, Govorov, A, & Liedl, T. (2011)DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response . *Arxiv preprint arXiv:1108.3752*. 108, 112, 122

[140] Shen, X, Song, C, Wang, J, Shi, D, Wang, Z, Liu, N, & Ding, B. (2012)Rolling up gold nanoparticle-dressed DNA origami into three-dimensional plasmonic chiral nanostructures . *Journal of the American Chemical Society*. 108

[141] Tan, S, Campolongo, M, Luo, D, & Cheng, W. (2011)Building plasmonic nanostructures with DNA . *Nature Nanotechnology* **6**, 268–276. 108, 111

[142] Xing, H, Wang, Z, Xu, Z, Wong, N, Xiang, Y, Liu, G, & Lu, Y. (2012)DNA-Directed Assembly of Asymmetric Nanoclusters Using Janus Nanoparticles . *ACS nano*. 108

[143] Schrader, B. (2008) *Infrared and Raman spectroscopy*. (Wiley-VCH). 110

[144] Haynes, C, McFarland, A, & Duyne, R. (2005)Surface-enhanced Raman spectroscopy . *Analytical Chemistry* **77**, 338–346. 110

[145] Ghosh, S & Pal, T. (2007)Interparticle coupling effect on the surface plasmon resonance of gold nanoparticles: from theory to applications . *Chemical reviews* **107**, 4797–4862. 110

[146] Xu, H, Aizpurua, J, Käll, M, & Apell, P. (2000)Electromagnetic contributions to single-molecule sensitivity in surface-enhanced Raman scattering . *Physical Review E* **62**, 4318. 111, 121

[147] Douglas, S, Bachelet, I, & Church, G. (2012)A logic-gated nanorobot for targeted transport of molecular payloads . *Science's STKE* **335**, 831. 124