UNIVERSITÉ DE LYON

ENS DE LYON

THÈSE

en vue de l'obtention du grade de Docteur, délivré par

l'École Normale Supérieure de Lyon

Ecole Doctorale N°512

École Doctorale en Informatique et Mathématiques de Lyon

Discipline : Informatique

Soutenue publiquement le 17 janvier 2025, par :

Octave Hazard

Controlling DNA Origami Co-folding

Contrôle du co-repliement d'origamis ADN

Devant le jury composé de :

David Doty	Associate Professor, UC Davis	**Rapporteur**
Sergii Rudiuk	Chargé de Recherche, ENS Paris	**Rapporteur**
Barbara Saccà	Professor, University of Duisburg-Essen	**Examinatrice**
Raluca Tiron	Directrice de Recherche, CEA-Leti	**Examinatrice**
Olivier Bournez	Professeur, École polytechnique	**Examinateur**
Nicolas Schabanel	Directeur de Recherche, ENS de Lyon	**Directeur de thèse**

Abstract

DNA origami is a particularly robust technique for designing DNA structures, often on the order of hundreds of nanometers. Introduced in 2006 by Paul Rothemund, this technique relies on the controlled folding of a long *scaffold* DNA strand (a few thousand nucleotides long), using a set of carefully chosen shorter synthetic strands acting as *staples*. It is particularly well-suited for designing complex structures in two or three dimensions.

However, the size of these DNA origamis is limited by the length of the scaffold strand. Several studies propose assembling identical DNA origamis in multiple steps to achieve larger sizes (Rajendran et al., 2011, Tikhomirov et al., 2017, Wintersinger et al., 2023). We have sought to determine whether it is possible to design DNA origamis comprising several identical scaffold strands but each folding differently to form various components of the desired structure. This requires particular work on the design of the origami, but also an understanding of the folding process that leads to its correct assembly. This thesis therefore involves several aspects: 1) exploring algorithmic and geometrical techniques for designing DNA origamis with multiple scaffolds; 2) the experimental implementations associated with these techniques; and 3) modeling the folding and assembly process. In particular, I present and evaluate several algorithmic methods for selecting a small subset of staples, allowing the pre-formation of the origami scaffolds separately so that they differentiate properly during the final assembly. Furthermore, I explore the usage of this method for assembling frustrated shapes with DNA origami, as well as an alternative differentiation method based on an additional mold origami.

My thesis concludes with an ancillary work exploring theoretical and experimental approaches for assembling quasi-crystals with DNA strands algorithmically.

Résumé

L'origami ADN est une technique particulièrement robuste permettant de concevoir des structures en ADN, souvent de l'ordre de la centaine de nanomètres. Introduite en 2006 par Paul Rothemund, cette technique repose sur le repliement contrôlé d'un long brin d'ADN appelé *scaffold* (de quelques milliers de nucléotides), au moyen d'un ensemble de brins synthétiques plus courts bien choisis, appelés agrafes. Elle se prête particulièrement bien à la conception de structures complexes en deux ou trois dimensions.

Cependant, la taille de ces origamis ADN est limitée par la longueur du scaffold. Plusieurs travaux proposent d'assembler en plusieurs étapes un certain nombre d'origamis ADN identiques pour atteindre des tailles supérieures (Rajendran et al., 2011, Tikhomirov et al., 2017, Wintersinger et al., 2023). Nous avons, quant à nous, cherché à déterminer s'il était possible de concevoir des origamis ADN comprenant plusieurs brins scaffold identiques mais se repliant de façon différenciée pour former différentes composantes de la structure voulue. Cela nécessite un travail particulier de conception de l'origami, mais également de compréhension du processus de repliement qui mène à son assemblage correct. Cette thèse comporte donc plusieurs aspects : 1) l'exploration des techniques de design d'origamis ADN à plusieurs scaffolds; 2) les réalisations expérimentales associées à ces techniques; ainsi que 3) une proposition de modélisation du processus de repliement. Je propose notamment une méthode de sélection d'un petit sous-ensemble d'agrafes, permettant de pré-former séparément les scaffolds de l'origami afin qu'ils se différencient convenablement lors de l'assemblage final. J'explore également l'usage de cette méthode pour l'assemblage de formes frustrées en origami ADN, ainsi qu'une méthode alternative de différenciation, basée sur l'utilisation d'un origami supplémentaire comme moule.

Je présente enfin un travail annexe explorant les pistes théoriques et expérimentales pour la création de quasi-cristaux en brins d'ADN.

Résumé long en français

La programmation biomoléculaire vise à créer des nano-objets et des systèmes « intelligents » en utilisant des brins d'ADN. Elle s'appuie sur l'appariement de Watson-Crick : un brin d'ADN porte une séquence de nucléotide (A, T, G ou C) et peut s'appareiller à un brin portant une séquence (ou sous-séquence) complémentaire suffisamment longue. La technique qui nous intéresse est celle de l'origami ADN (figure 6.2). Elle consiste à replier un long brin d'ADN viral (nommé *scaffold*) selon une forme voulue en utilisant un ensemble de brins courts synthétiques jouant le rôle d'agrafes. Chaque agrafe va venir s'appareiller à plusieurs segments du brin long, afin de les rapprocher et permettre son repliement.

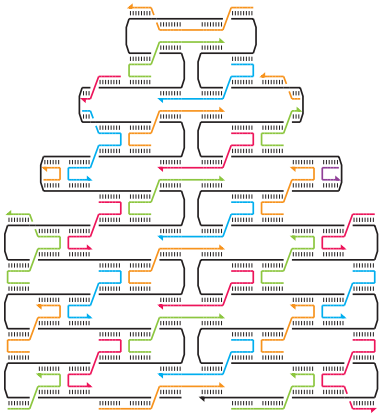


Figure 1: Représentation schématique d'un origami ADN (Rothemund, 2006). Le brin scaffold est représenté en noir et les agrafes en couleurs.

Seulement, une des principales limites de cette technique porte sur la taille des objets qu'il est possible de construire : la longueur du scaffold (de l'ordre de 7000 à 8000 nucléotides) permet de former des origamis d'au plus une petite centaine de nanomètres. Une approche intéressante serait de chercher à former des objets utilisant plusieurs copies du même scaffold dans la structure, chaque copie jouant un rôle différent. Cela nécessite de comprendre comment imposer à chaque agrafe de s'attacher correctement, alors que chacune a plus de configurations potentielles d'appariement que dans un design classique d'origami ADN.

Les deux premiers chapitres de ce manuscrit sont consacrés à une introduction au domaine des nanotechnologies ADN (chapitre 1) ainsi qu'à la technique de l'origami ADN (chapitre 2).

Je présente ensuite les motivations et les bases bibliographiques qui nous ont conduit à explorer la faisabilité d'un origami ADN utilisant plusieurs scaffolds (chapitre 3). J'évoque en particulier des études sur l'auto-assemblage simultané de plusieurs designs différents d'origami ADN, sur un origami ADN utilisant un scaffold répétitif ainsi que sur le contrôle de la nucléation dans la formation de structures complexes en brins d'ADN. Je poursuis avec la présentation de mes résulats expérimentaux préliminaires : la jonction de deux brins scaffolds par une ou deux agrafes de liaison, ainsi que la synthèse *one-pot* d'un double origami symétrique (chapitre 4). Je présente dans le chapitre 5 notre premier design asymmétrique d'origami ADN utilisant deux copie du scaffold M13mp18 dans sa structure : nous avons choisi une forme simple bidimensionnelle représentant un anneau et un carré placés côte à côte. Nos premières tentative de synthèse *one-pot* de cet origami présentent des structures chimériques mélangeant les deux moitiés du design, malgré de nombreux ajustements des

paramètres expérimentaux. Pour chercher à déterminer les agrafes jouant un rôle principal, nous avons donc décidé d'ajouter une étape de repliement partiel de l'origami avant l'étape finale d'auto-assemblage : chaque scaffold est d'abord recuit séparément avec un sous-ensemble (appelé *graine*) de ses agrafes, graine que l'on cherche à rendre la plus petite possible. Ce principe est présenté dans le chapitre 3, avec une première méthode de sélection des agrafes de la graine, ainsi que les résultats expérimentaux associés. Je propose chapitre 7 un algorithme plus fin pour la sélection de cette graine, basé sur une formulation des contraintes thermodynamiques associées sous la forme d'un problème d'optimisation linéaire en nombres entiers. Après la présentation de nos resultats expérimentaux issus de cet algorithme, je détaille dans le chapitre 8 notre méthode d'estimation du rendement pour comparer les différents choix de graine. Cette méthode s'appuie sur l'analyse d'image de microscopie à force atomique par un réseau de neurones pour détecter les origami ADN correctement repliés.

Nous avons également travaillé sur design d'origami ADN à double scaffold qui bénéficie pleinement de notre méthode d'assemblage en deux étapes (chapitre 10). Plus spécifiquement, ce design décrit une forme frustrée qui s'appuie sur le caractère incomplet de la première étape de pré-repliement, afin de s'auto-assembler correctement durant la seconde étape. Par soucis d'efficacité, ce design utilise un scaffold original, environ trois fois plus court que l'habituel M13mp18. Nous détaillons dans le chapitre 9 notre méthode de production de ce scaffold à partir d'un plasmide double brin.

Ayant exploré l'usage d'un sous-ensemble d'agrafes clées pour la différenciation des scaffold d'un origami multiple, nous avons ensuite décidé d'explorer une autre approche : utiliser un origami supplémentaire comme moule pour induire la différenciation du repliement simultané de deux brins scaffolds identiques (chapitre 11).

Afin de mener nos études sur le design et l'assemblage expérimental d'origamis ADN à plusieurs scaffolds, nous avions besoin de mieux comprendre son processus de repliement. À cette fin, j'ai travaillé sur un modèle stochastique qui pourrait nous aider à mieux prédire le repliement de nos conceptions. Après une recherche bibliographique approfondie visant à comprendre les divers aspects du repliement de l'origami ADN et sa modélisation (chapitre 12), je présente le modèle, sa mise en œuvre et ses résultats, ainsi que ses possibles améliorations et limitations (chapitre 13). Ce modèle utilise la méthode de Monte-Carlo cinétique pour simuler la séquence d'événements conduisant au repliement d'un origami ADN, en combinant une approche à l'échelle des nucléotides et à l'échelle des domaines (i.e. sous-séquence de nucléotides).

Enfin, les trois derniers chapitres de ce manuscrit sont consacrés à une étude annexe antérieure aux travaux précédemment présentés. J'y explore les approches théoriques et expérimentales pour assembler des quasi-cristaux avec des brins d'ADN de manière algorithmique. Je commence par introduire le procédé de *cut-and-project* pour la construction et l'étude de pavages apériodiques, ce qui nous permet de définir les règles locales nécessaires pour l'auto-assemblage de quasi-cristaux reproduisant ces pavages (chapitre 14). Ensuite, je présente une tentative expérimentale de construire un origami ADN représentant une portion du pavage de Penrose (chapitre 15) et quelques pistes pour concevoir des pavages apériodiques auto-assemblants (chapitre 16).

Remerciements/Acknowledgements

Tout d'abord, je tiens à remercier mon directeur de thèse, Nicolas, pour m'avoir fait découvrir le monde de la programmation bio-moléculaire. Merci pour l'intérêt que tu as éveillé chez moi en présentant ce domaine à ton cours au MPRI, et pour ton encadrement durant le stage et la thèse qui ont suivi. Ta vision et ton enthousiasme inébranlable et contagieux m'ont souvent aidé à retrouver de la motivation, notamment lors des nombreuses difficultés expérimentales que nous avons rencontrées.

I would also like to thank the members of my jury, David Doty, Sergii Rudiuk, Barbara Saccà, Raluca Tiron, and Olivier Bournez, for their interest in my work. I hope you will all enjoy this thesis defense (and the small buffet that follows). I am particularly grateful to Dave and Sergii, who agreed to read and review my manuscript and provided valuable feedback.

Un des principaux défis que j'ai dû relever durant cette thèse fut l'appropriation des méthodes et concepts de sciences expérimentales. Cette thèse s'appuie donc sur l'aide de nombreux·ses biologistes, physiciens·nes et chimistes. Sans prétendre pouvoir tous les citer, j'aimerais en remercier quelques uns: Olivier Gandrillon, pour nous avoir permis de réaliser nos manips au sein de ton équipe, Cendrine Moskalenko pour ton temps et ton expertise sur l'imagerie AFM à sec, Florence Lormière pour ton aide décisive dans l'élaboration d'un protocole expérimental (auquel un chapitre de cette thèse est consacré), et enfin Gaetan Bellot et son équipe, notamment Julie et Nicolas, pour votre expertise et votre accompagnement dans une bonne partie de mes manips. Je remercie plus généralement le Laboratoire de Physique de l'ENS de Lyon et le LBMC.

J'aimerai remercier celles et ceux que j'ai côtoyés au LIP, et qui contribuent à en faire un cadre agréable et accueillant. Tout d'abord, merci à mes co-doctorants : Pierre, pour ton accompagnement sur mes premières manips, les séances d'escalade et les apéros, et Nicolas L., pour m'avoir fait découvrir et vanté les bienfaits de Rust, ce langage qui occupe maintenant mes journées. Merci à Daria, pour ces moments passés à discuter, grimper ou skier ensemble. Merci à tous ces doctorants qui s'investissent pour faire vivre le labo, notamment Calvin, Joël, Malory, Meriem et Tom. Merci à la team des "futurs ex-grimpeurs" pour ces évasions verticales (presque) tous les vendredis midis.

J'aimerais également remercier les gestionnaires (notamment Laure, Marie B. et Marie N.), la direction (Nicolas T. et Isabelle) et le comité des thèses (Russ et Bora), pour leur investissement et leur aide considérable lors de fréquentes démarches et difficultés administratives. Le LIP a beaucoup de chance de vous avoir !

Ces (presque) quatre années à Lyon ne se sont pas limitées aux murs de l'ENS. Merci à Émile, Alice M., Angèle, Hugo, Adèle et le reste de la bande de Lokifer pour votre accueil à Lyon. Les soirées séries, balades à vélo et barbecues à Miribel ont su égayer mes débuts dans cette ville, où je ne connaissais alors pas grand monde et en période de restrictions sanitaires. Merci à Adèle et la bande des "bio" (Alice H, Justine, Nicoco, Nicolas L, Val, ...) pour les

soirées, randos et autres escapades Merci à l'ASVEL Ski Montagne et ses encadrants, pour ces weekends hors du temps à découvrir l'alpinisme et le ski de randonnée.

J'aimerais bien sûr remercier ma famille, qui me soutient depuis longtemps dans ce parcours. Merci Papa, de m'avoir montré ce chemin de la thèse de doctorat et ce monde de la recherche académique, même si je vais maintenant découvrir le monde de Maman, après avoir ressorti le diplôme d'ingénieur des cartons.

Et pour finir, j'aimerais te remercier Adèle, ma fidèle partenaire de vélo, de grimpe, de ski, de rando, (bientôt de vol 🪂 ?), d'achat immobilier... et partenaire tout court. Merci pour tout ton soutien pendant cette période, qui n'a pas toujours été facile.

Contents

Abstract iii

Résumé v

Résumé long en français vii

Remerciements/Acknowledgements ix

I Introduction 1

1 DNA nanotechnologies 3
1.1 The DNA molecule . . . 3
1.2 Artificial DNA self-assembly . . . 5

2 DNA origami 11
2.1 Principle and design of a DNA origami . . . 11
2.2 Classical fabrication and characterization process . . . 13
2.3 Advanced origami structures . . . 14

II A multi-scaffold DNA origami 17

3 Motivations and feasibility study 19
3.1 Motivations and challenges . . . 19
3.2 Supporting works . . . 20

4 Preliminary experiments 23
4.1 Step 1: Bindings two scaffolds . . . 23
4.2 Step 2: joining two identical DNA origamis . . . 23

5 First asymmetric test design and assembly attempts 27
5.1 Designing an asymmetric bi-origami . . . 27
5.2 First co-folding attempts of our bi-origami . . . 28

III Two-step annealing principle based on seed staple selection 33

6 Two-step annealing and greedy algorithm 35
6.1 Adding a step of partial annealing in separate pots . . . 35
6.2 Greedy algorithm . . . 35

6.3 Experimental evaluation 37

7 Linear program for selecting a seed **39**
7.1 A linear approximation for the Gibbs binding energy 39
7.2 An integer linear program to design seeding subsets 40
7.3 Experimental evaluation 40

8 Yield estimation based on AFM images using Deep Learning **43**
8.1 Neural network training and origami detection 43
8.2 Yield computation 44
8.3 Results and discussion 44

IV Assembling frustrated shapes **47**

9 Producing a smaller DNA scaffold strand **49**
9.1 Motivations and supporting works 49
9.2 Protocol elaboration 49
9.3 Final validation of the method: two DNA origamis 51

10 A geometrically frustrated DNA origami **53**
10.1 Principle and design ideas 53
10.2 Corrugated design 53
10.3 Experimental results 54

V Mold-assisted differentiation **59**

11 A shape-sorter-inspired DNA origami **61**
11.1 Principle and design options 61
11.2 Results and discussion 62

VI Simulation of DNA origami self-assembly **65**

12 Understanding DNA origami self-assembly **67**
12.1 Thermodynamics foundations 67
12.2 Nucleotide level and domain level models 68
12.3 Topological and geometrical considerations 70
12.4 Experimental data from the literature 72

13 Proposed model for our simulation **73**
13.1 A multi-level approach 73
13.2 Computing kinetic rates 74
13.3 Kinetic Monte-Carlo simulation and implementation 76
13.4 Results, perspectives and discussion 78

VII Self-assembly of aperiodic tilings 81

14 Quasicrystals and Cut-and-project aperiodic tilings 83
14.1 Cut-and-project principle 83
14.2 Atlas and window 85
14.3 Local rules and self-assembly 86

15 A Penrose tiling DNA origami 89
15.1 The Penrose origami design 89
15.2 Results and perspectives 91

16 Leads for designing self-assembling aperiodic tilings 93
16.1 Design approaches for implementing local rules 93
16.2 First hard-coded version and sequence design 94

Conclusion 97

A Materials and methods 99

Bibliography 101

Index 115

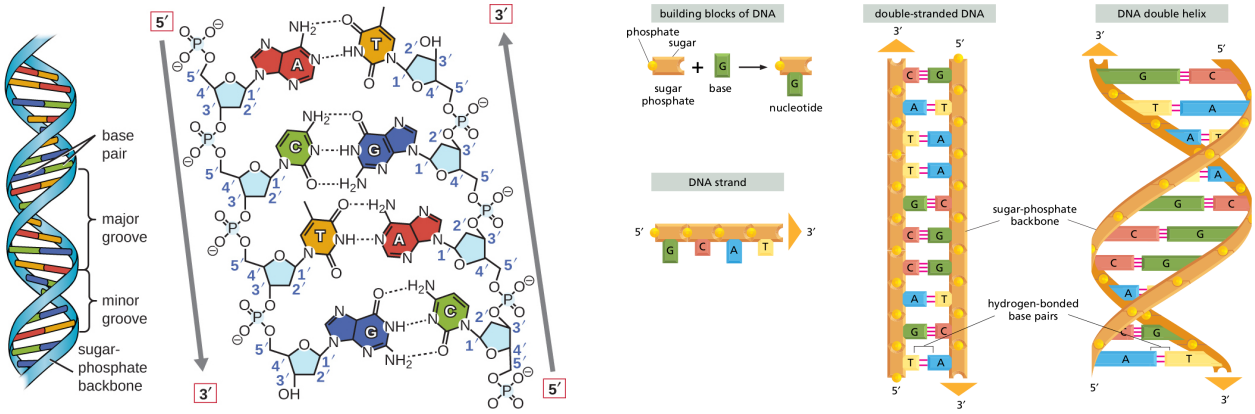
Part I

Introduction

Chapter 1

DNA nanotechnologies

1.1 The DNA molecule



(a) Atomic structure of the DNA molecule. From OpenStax [Par+16].

(b) Schematic representation of the DNA molecule building blocks. From [Alb08].

Figure 1.1: Structure of the DNA molecule.

Deoxyribonucleic acid (DNA) is a biological molecule, that has been extensively studied especially since the middle of the XXth century, after the discovery of its role in genetic information storage and transmission [AMM44]. It is a long macromolecule that generally consists in two chains of nucleotides coiled around each other into a double helix shape (see fig. 1.1). This helical structure was first discovered in 1952 by Rosalind Franklin and described by James Watson and Francis Crick in 1953 [WC53]. Each chain of nucleotides, or single-stranded DNA, is composed of nucleotide monomers linked together with covalent bonds. These nucleotides consist of a five-carbon sugar deoxyribose joined with a phosphate group and a nucleobase which can be of one of four types: adenine (A), thymine (T), guanine (G) or cytosine (C). The deoxyribose and phosphate groups are the building blocks of the sugar-phosphate backbone of the DNA strand: the deoxyribose group of one nucleotide covalently binds to the phosphate group of the next nucleotide, resulting in a chain equipped with directionality. By convention, each nucleotide is oriented from the phosphate group, referred to as *5'-end*, to the five-carbon sugar deoxyribose, referred to as *3'-end*[1].

[1]The terms *5'-end* and *3'-end* refer to the numbering of carbon atoms in the nucleotide's sugar group: the phosphate group is attached to the fifth carbon and the next nucleotide is attached to the third carbon.

Watson–Crick–Franklin base complementarity. Two nucleotide chains bind to each other using Watson–Crick–Franklin base complementarity: adenine and thymine can bind to each other by forming two hydrogen bonds, and guanine and cytosine can bind by forming three hydrogen bonds. As a consequence, two DNA strands can bind to each other if they present complementary sequences of nucleobases. They bind in opposite direction, the complementary sequence corresponds to the mirroring of the original sequence, where each nucleotide is replaced by its Watson–Crick-Franklin complement: adenine is replaced by thymine (and conversely), and guanine is replaced by cytosine (and conversely). As an example, the complementary sequence of 5'-GTAAC-3' is 5'-GTTAC-3'.

Stacking interactions between adjacent bases along a DNA strand also play an important (and even predominant) role in the stability of the DNA double helix [Sae84]. Nucleobases have a planar atomic structure and these planes, which are almost perpendicular to the direction of the double helix, stack at contact distance[2] (about 3.4 Ångström), thus largely contributing to the rigidity of the molecule.

Branched DNA. While the most common configuration of DNA is a long, sometimes cyclic, double helix involving two perfectly complementary strands, other more complex arrangements can be found in nature, such as branched DNA.

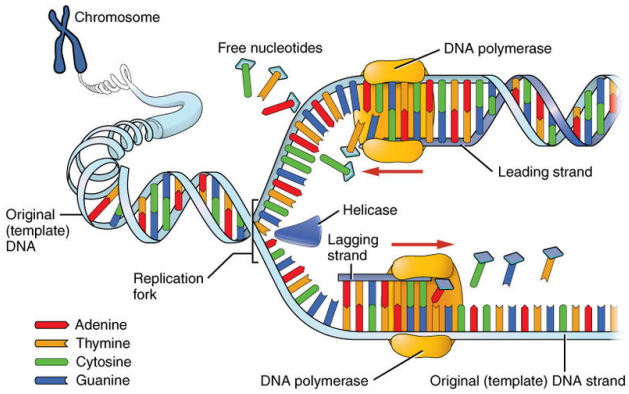


Figure 1.2: DNA replication fork. The strands from the initial double helix are separated and used as template for the synthesis of two new DNA strands. From OpenStax [Bet+22].

Branched DNA is a structure involving three or more DNA strands that only bind segments of their chain to each other. One classical example is the *replication fork* that forms in vivo during DNA replication (see fig. 1.2). During this process, the strands from an initial double helix are gradually separated while being used as template for the synthesis of new DNA strands. Thus, it forms an asymmetric three-way branching that propagates along the initial double strand.

Holliday juctions are another type of branching. It is a four-way junction primarily involved in the process of genetic recombination [Hol64]. This junction involves four DNA strands; each one is bound to two others using sequence complementarity (see fig. 1.3). In naturally-occurring Holliday junctions, the involved nucleotide sequences are largely identical (or complementary), thus resulting in a mobile junction: the junction can slide along the strands, in a process called branch migration (see fig. 1.3c). But in the early 1980s Ned

[2]The van der Waals contact distance corresponds to the distance between two atoms under which the van der Walls force becomes repulsive instead of attractive

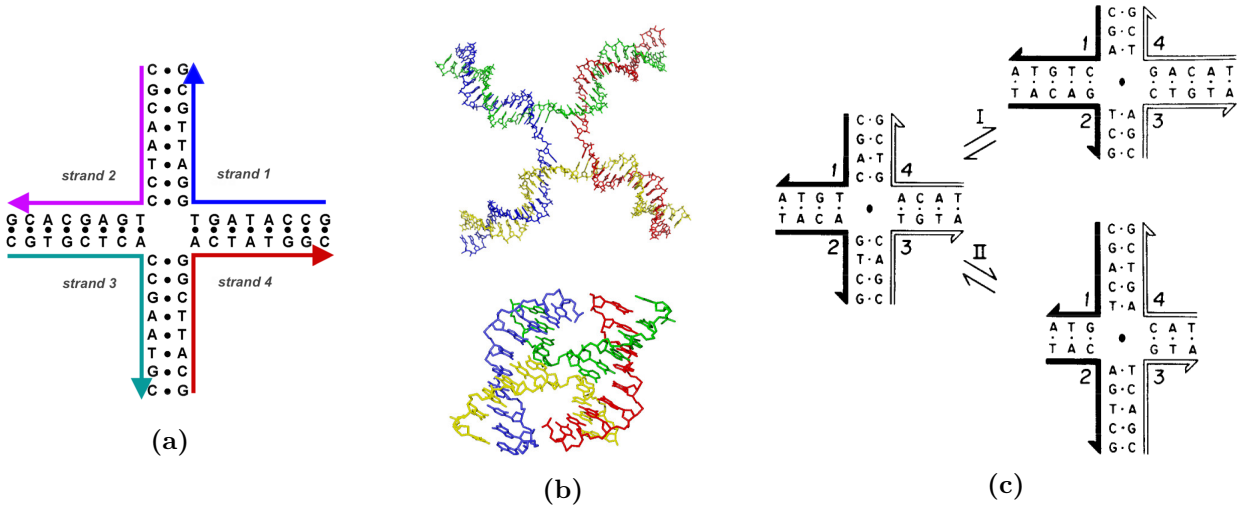


Figure 1.3: Holliday Junction. (a) Schematic representation of an example of (asymmetric) Holliday junction (image by Донор, distributed under a CC BY-SA 4.0 license, via Wikimedia Commons). (b) Molecular structure of an Holliday junction. Two main types of conformation are possible: open (first image) and stacked (second image). In the stacked configuration, two strands roughly maintain their straight helical form (the blue and red strands in the second image). (first image by Richard Wheeler, distributed under a CC BY-SA 3.0 license, via Wikimedia Commons, second image adapted from a public domain image by Tim Vickers, Wikimedia Commons) (c) Example of mobile Holliday junction from [See82]. Here, strands 1 and 3 present the same sequence of nucleobases. This symmetry allows the junction to migrate according to reaction I or II by breaking two facing base paring and immediately linking those bases again in a different configuration.

Seeman thought of using synthetic DNA in order to remove this symmetry, thereby allowing the formation on an immobile Holliday junction. This was one of the first seeds of DNA nanotechnology.

1.2 Artificial DNA self-assembly

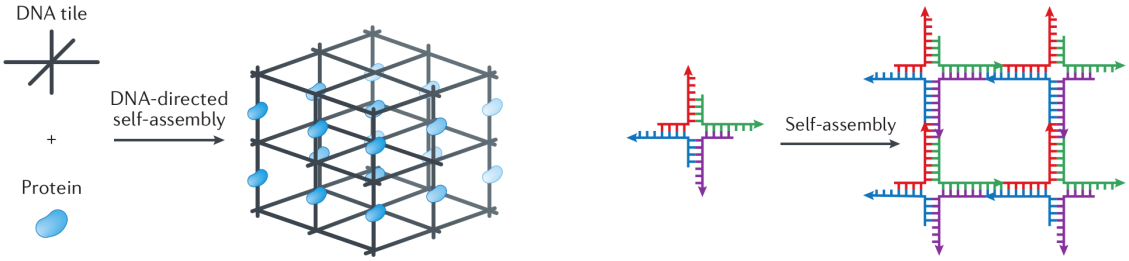
While DNA is mostly known and studied for its role as a genetic information carrier in biological organisms, the field of DNA Nanotechnology takes the DNA molecule out of its biological frame. It relies on natural and synthetic DNA molecules in order to build nanometer-scale structures, using a bottom-up method.



Figure 1.4: Woodcut *Depth* by M. C. Escher, 1955.

Bottom-up fabrication method. While the widely used "top-down" manufacturing approach starts from a large structure and reduces and shapes it by removing material, DNA Nanotechnology is based on "bottom-up" fabrication method. It takes advantage of the specificity of DNA base binding for guiding the self-assembly of small DNA molecules into larger structures.

Foundations of DNA Nanotechnologies. In 1980, Ned Seeman, inspired by the woodcut *Depth* (fig. 1.4) had the idea of using a 6-arm immobile DNA junction made of synthetic DNA strands in order to build small 6-branch tiles that would assemble into a three-



(a) DNA lattice as a template for protein crystallization

(b) Self-assembly of a DNA four-way junction

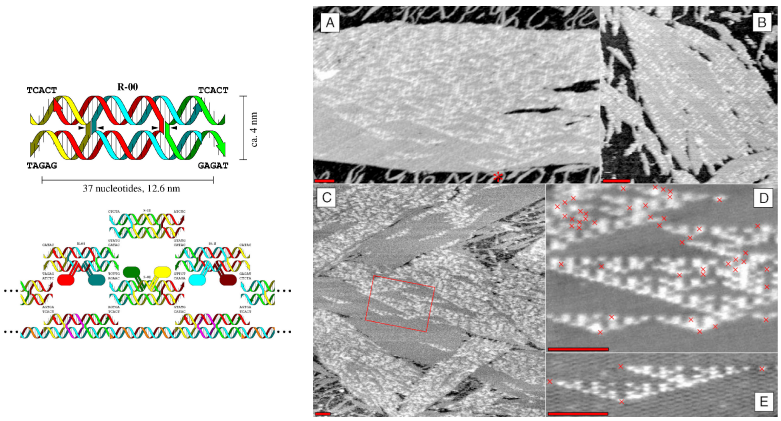
Figure 1.5: DNA self-assembly into a regular lattice. From [SS17]

dimensional lattice [See82]. His motivation at the time was to use this lattice as a template for facilitating protein crystallization, in the context of X-ray structure characterization (see fig. 1.5a).

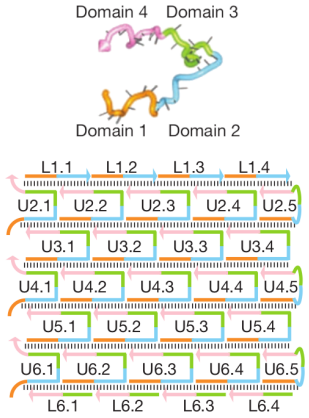
The first experimental realization of an immobile DNA junction was published in 1983 [KMS83]. During the next decade, this technique was used to build a 10-nanometer DNA cube involving 6 strands and 8 three-way junctions [CS91] and rigid DNA bricks using two stacked Holliday junctions: the double-crossover molecules [FS93; Li+96] (see fig. 1.6a). It attested the feasibility of building DNA nanostructures of various sizes with controlled geometry, connectivity and topology.

DNA tile assembly and algorithmic assembly. Double-crossover DNA complexes (or DX tiles) were used to grow the first predesigned DNA 2D crystals [Win+98]. In order to bind to each other, these tiles use so-called "sticky end": a single strand overhang at the end of a double helix that can attach to another overhang carrying its complementary sequence of nucleotides (see fig. 1.6a). This binding specificity allows these sticky ends to act as programmable glues, thus implementing chosen assembly rules, similar to Wang tiles. An early example of such algorithmic assembly is the fabrication of lattices presenting the fractal pattern of Sierpinski triangles [RPW04] (see fig. 1.6a). It demonstrated that engineered DNA self-assembly could be regarded as a Turing-universal biomolecular system, capable of implementing any desired algorithm for computation or construction tasks. Several innovations followed this path, including the idea of using proofreading tile sets to reduce assembly errors by introducing redundancy, and the development of smaller, non-rigid tiles known as single-stranded tiles (SST). Initially invented to construct DNA nanotubes with programmable circumference [Yin+08], SSTs were also used to robustly implement a wide variety of 6-bit algorithms [Woo+19] (see fig. 1.6c) and to assemble complex two-dimensional [WDY12] (see fig. 1.6b) and three-dimensional shapes [Ke+12].

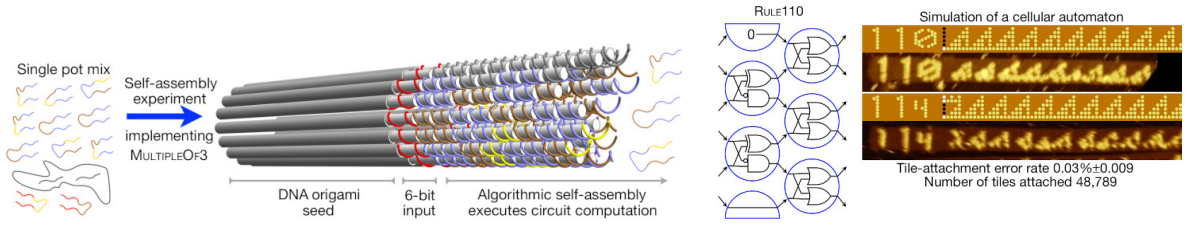
DNA origami is now the most common method for buiding DNA nanostructures of arbitrary, non-regular shapes. Invented in 2006 by Paul Rothemund [Rot06], this technique involves a long, natural virus strand referred to as a *scaffold*, which is folded into the desired shape using short, computationally designed *staple* strands. In terms of complexity, aperiodicity, and overall ease of self-assembly, the DNA origami approach was unprecedented at the time. Initially demonstrated with two-dimensional shapes like a smiley face or a coarse map of the Western Hemisphere, DNA origami has evolved to include solid three-dimensional structures. These can be created by arranging parallel DNA helices in a honeycomb pattern [Dou+09a], by folding two-dimensional faces into a hollow three-dimensional shape, similar to a cardboard box [And+09], or by describing a wireframe structure [Zha+15].



(a) DNA DX tile with four sticky ends and its assembly into Sierpinski triangles, from [RPW04]



(b) Schematic assembly of DNA single-stranded tiles into a rectangle, from [WDY12].



(c) Algorithmic assembly of DNA single-stranded tiles (SST), from [Woo+19]. The assembly rules given by the chosen sequences allow to simulate the execution of a cellular automaton.

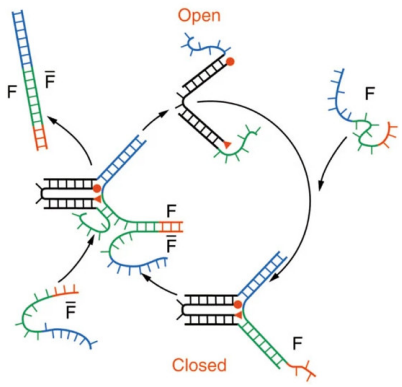
Figure 1.6: DNA tile assembly

The DNA origami method is central in this thesis and will be the subject of the next chapter.

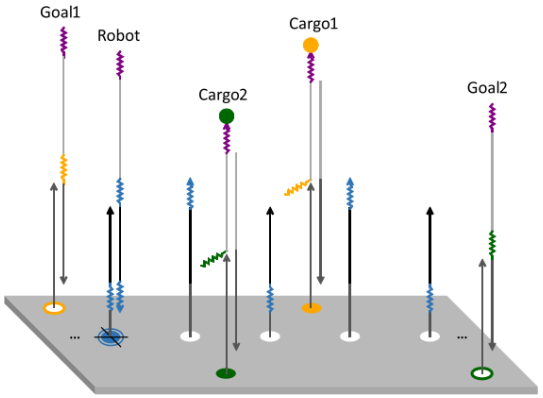
Strand displacement and nanomechanical devices. Dynamic DNA structures were made possible by the principle of *isothermal DNA strand displacement*, introduced in 2000 in a paper presenting a small tweezers-like DNA structure that could be opened and closed by the addition of auxiliary 'fuel' DNA strands [Yur+00] (see fig. 1.7a). It uses reversible strand binding on a short overhanging nucleotide sequence (called a *toehold*) followed by DNA branch migration in order to induce a conformational change. This principle allowed the creation of all kind of DNA structures with dynamic behavior, such as a 2D DNA lattice that can switch between a globally compact and stretched conformation [Fen+03], a hybridization chain reaction triggered by a short DNA single strand [DP04] and various realizations DNA 'walkers' that can autonomously follow a track [SS04] or transport and sort molecular cargo [Thu+17] (see fig. 1.7b).

Applications of DNA Nanotechnologies. While DNA Nanotechnology was born as a potential solution to protein crystallization, it rapidly encountered a broad range of possible applications.

First, DNA nanotechnology can play an important part in guiding the assembly of other inorganic nanostructures, such as metals and semi-conductor nanoparticles, nanowires and nanorods, a promising class of materials for optical, electronic, catalytic and sensing applications. For instance, DNA origamis have been used to precisely arrange gold nanoparticles into arbitrary patterns [Kuz+12] or to serve as a mold for growing gold nanostructures of



(a) DNA tweezers that can transition between an open form and a closed form using strand displacement (from [Yur+00]).



(b) A single-stranded DNA robot that can perform a cargo-sorting task on a DNA origami surface (from [Thu+17]).

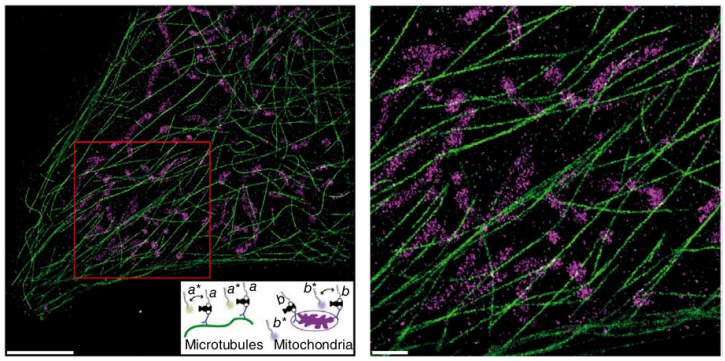
Figure 1.7: Strand displacement and nanomechanical devices

chosen shape [Sun+14].

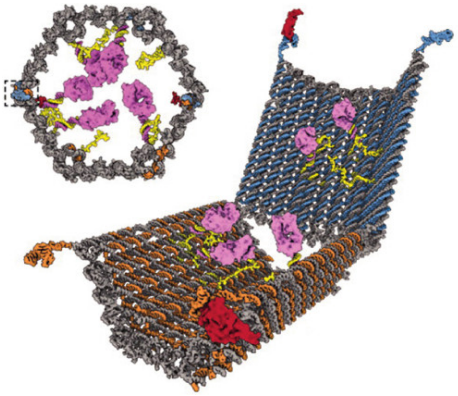
DNA nanostructures have also been used for understanding and controlling protein interactions, allowing advances in both fundamental biology and synthetic biology. Noteworthy studies include the creation of enzyme cascade by organizing metabolic proteins on DNA tiles [Wil+09b] or origami scaffolds [Fu+16], the study of motor proteins through their association with a programmable DNA origami cargo [Der+12] or the reproduction of protein structures and functions with designed DNA nanostructures, such as lipid membrane channels [Lan+12] or synthetic cells cytoskeletons [Zha+22].

Lastly, biological imaging techniques and biomedical applications have also benefited from DNA nanotechnology advances. An important example in microscopy is DNA point accumulation for imaging in nanoscale topography (DNA-PAINT) [Jun+10], a super-resolution technique that allows multi-color 3D imaging of proteins inside fixed cells [Jun+14] (see fig. 1.8a). Biomedical applications include diagnosis systems [Sun+19] and various approaches of targeted and adaptive drug delivery [DBC12; Lee+12; Edw+13] that take advantage of the biocompatibility of DNA nanostructures and the ability to program their drug release to respond to specific biological cues (see fig. 1.8b), as well as DNA origami traps that can act as virus-enveloping shells in order to prevent viruses from entering cells [Mon+22; Mon+23] (see fig. 1.8c).

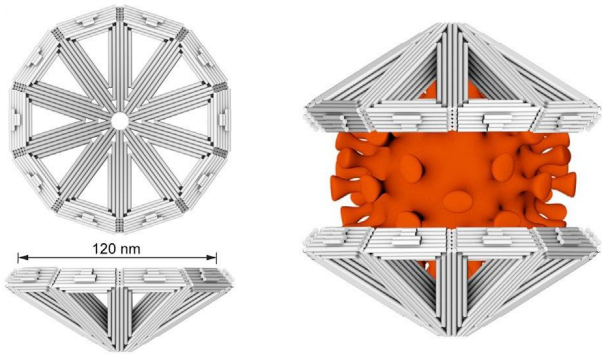
Conclusion. Throughout this chapter, we have explored the remarkable diversity of DNA nanotechnology approaches, ranging from branched DNA structures to complex self-assembled arrays and dynamic systems like strand displacement devices (a chronological overview can be found in fig. 1.9). These techniques illustrate the vast potential of DNA as a construction material for nanoscale structures and devices, with applications that span from molecular computing to biomedical engineering. Among these methods, the DNA origami technique, introduced by Paul Rothemund, stands out for its robustness and versatility in creating highly precise, programmable structures at the nanometer scale. By folding long DNA scaffolds into predefined shapes with the help of carefully designed staples, DNA origami offers unique advantages in terms of both structural complexity and ease of experimental implementation. This technique, with its ability to control the formation of large, intricate nanoscale assemblies, forms the primary focus of my PhD work. In the following chapters, we will delve deeper into the principles, challenges, and innovations in DNA origami, beginning



(a) Dual-color DNA-PAINT super-resolution image of microtubules and mitochondria inside a fixed HeLa cell, from [Jun+14]. Scale bars: 5 µm and 1 µm.



(b) A DNA nanorobot for targeted cellular delivery of antibodies [DBC12].



(c) DNA origami-based virus-enveloping shell, capable of preventing viruses from entering cells [Mon+23].

Figure 1.8: Biomedical and imaging applications of DNA nanotechnologies.

with an in-depth look at its design and fabrication processes.

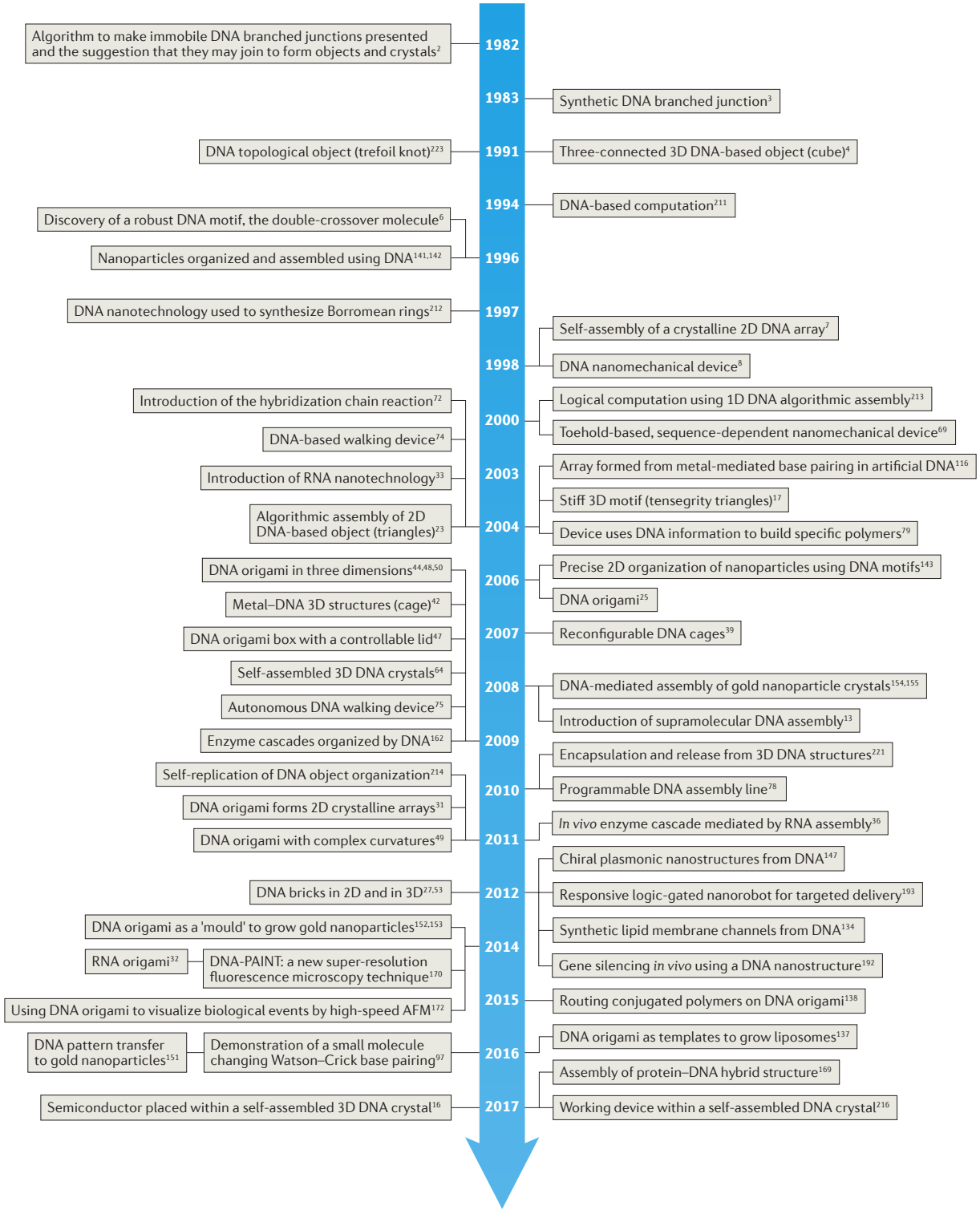


Figure 1.9: A timeline of the field of DNA nanotechnology. From [SS17].

Chapter 2

DNA origami

This PhD Thesis will mainly focus on the DNA origami technique, developed by Paul Rothemund in 2006 [Rot06]. In this chapter, I present this efficient and inexpensive technique to build reliably biocompatible nanostructures up to ∼100 nm wide with a ∼2 nm resolution.

2.1 Principle and design of a DNA origami

DNA origami is a method for directing the folding of a long, single strand of DNA (called the *scaffold strand*) into an arbitrary two- or three-dimensional shape, using numerous short single strands, called *staple strands*. The scaffold strand is typically a cyclic strand of viral origin, containing several thousand nucleotides, such as the M13mp18 phage (7249 nucleotides). The staple strands are carefully designed synthetic strands of 15 to 80 nucleotides, whose sequences are chosen so that each of them can bind to several domains of the scaffold strand, bringing them close to each other. This leads the scaffold strand to fold into a specific pre-defined shape (see fig. 2.1). Using a typical scaffold strand such as the M13mp18, the resulting structure is approximately 100 nm in diameter for a two-dimensional single-layer structure, and 20 to 50 nm in width for a three-dimensional design.

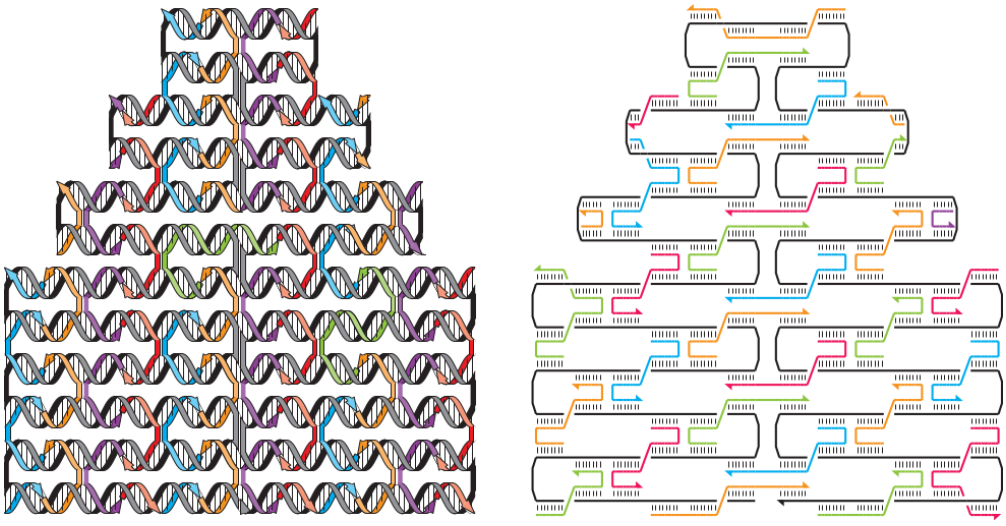


Figure 2.1: DNA origami principle Diagram of DNA origami (left), and a simplified representation often used for its readability (right). The scaffold strand is shown in black and the staple strands are in color. These two diagrams are from P. Rothemund's article [Rot06].

Design steps. A classical way to design a DNA origami is to first approximate the targeted shape with a set of cylinders of 2 nm diameter with a sub-nanometer gap between them (typically 0.6-1 nm for a 2D single-layer design). Those cylinders model the DNA double helices that will constitute the origami structure. Hence, they must be arranged so that one can find a closed path along those cylinders: the scaffold strand routing. This routing has to follow a right-handed helix on the surface of the cylinders according to the DNA structure (about 3.5 nm per turn) and transition between cylinders with the shortest possible connections (ideally bellow 1 nm)[1]. Once this scaffold strand routing is completed, the staple strands can be placed similarly, completing the DNA double helices and adding links between neighboring cylinders, in order to ensure the stability and rigidity of the resulting structure. In 2D designs, two parallel adjacent double helices are usually linked periodically by the staple strands every 3 helices turn, resulting in about one link every 31 nucleotides or 10.5 nm. In this case, staples are typically around 31 or 32 nucleotides in length, with either two equally long domains arranged in a 'C' shape, or one 15-16 nucleotide central domains linked to two shorter ones on neighboring double helices, describing an 'S' shape (as in fig. 2.1). As soon as the staple strands are positioned on the design and fully link to the scaffold strand, their sequences are fully determined by the scaffold sequences, using Watson-Crick-Franklin base complementarity. As most DNA origami use the M13mp18 cyclic strand as scaffold (or a variant of it), the only remaining design choice here is the position of the first nucleotide of its (fixed) sequence along its already chosen routing.

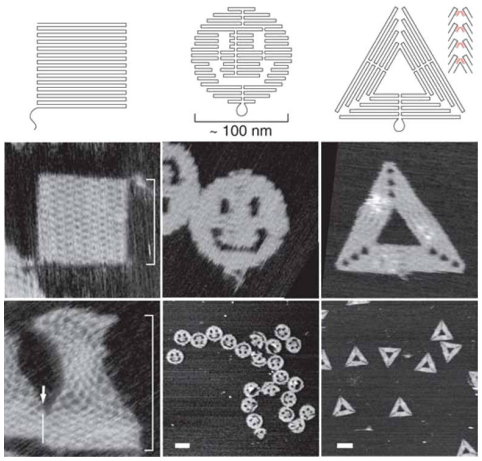


Figure 2.2: Some of the first examples of DNA origami from P. Rothemund's article [Rot06], along with their AFM images.

Those designing steps raise several challenges, that can largely be solved with the use of proper designing software programs, dedicated to DNA structures. Those challenges include: remaining consistent with the DNA double helix structure and mechanical properties in the design, finding a correct and effective scaffold routing, and choosing the right position for staple crossovers. The later can be particularly challenging, as an insufficient number of crossovers can lead to a weak and loose structure, while an excessive number means shorter binding domains and thus weaker bonds between the scaffold and the staples. Moreover, the asymmetric structure of the DNA double helix sometimes makes it difficult to just find a crossover position.

Designing software. I mainly used the ENSnano software, currently developed my colleagues at ENS de Lyon [LS21; Lev23] (see fig. 2.3). This software allows to precisely position, edit and link DNA double helices constituting an origami design. To this end, it includes both a 3D and a 2D interface, together with well-thought-out editing and visualization tools. However, several other DNA origami design tools are available, with various levels of automation and editing precision. Some mainly rely on a 2D abstract representation of the design for a rapid, nucleotide-level prototyping, such as Cadnano [Dou+09b] and SARSE-

[1] In reality, adjacent DNA double helices are not perfectly straight and parallel, as a crossover connects them by a distance of one phosphate group while electrostatic repulsion makes them bow away from each other between those crossovers. Thus, the distance between cylinders in the design corresponds to the targeted average distance between adjacent DNA double helices

DNA [And+08], while other offer a geometrically accurate 3D editable view like Tiamat [Wil+09a], vHelix [Ben+15] and Adenita [Lla+20]. In order to avoid repetitive and technical fine-tuning tasks, some programs also offer scriptable tools, like Scadnano [DLS20], or highly automated design of some specific classes of DNA origami, such as MagicDNA [Hua+21] for 3D wireframe multicomponent structures, and DNAxis [Fu+22] for 3D curved origami designs. Lastly, some simulation tools are available in order to evaluate the viability of the design, including physical simulation softwares such as CanDo [Kim+12], OxDNA [Doy+13] and MrDNA [MA20], as well as thermodynamic binding estimation software like NUPACK [Zad+11] and Vienna RNA [Gru+08].

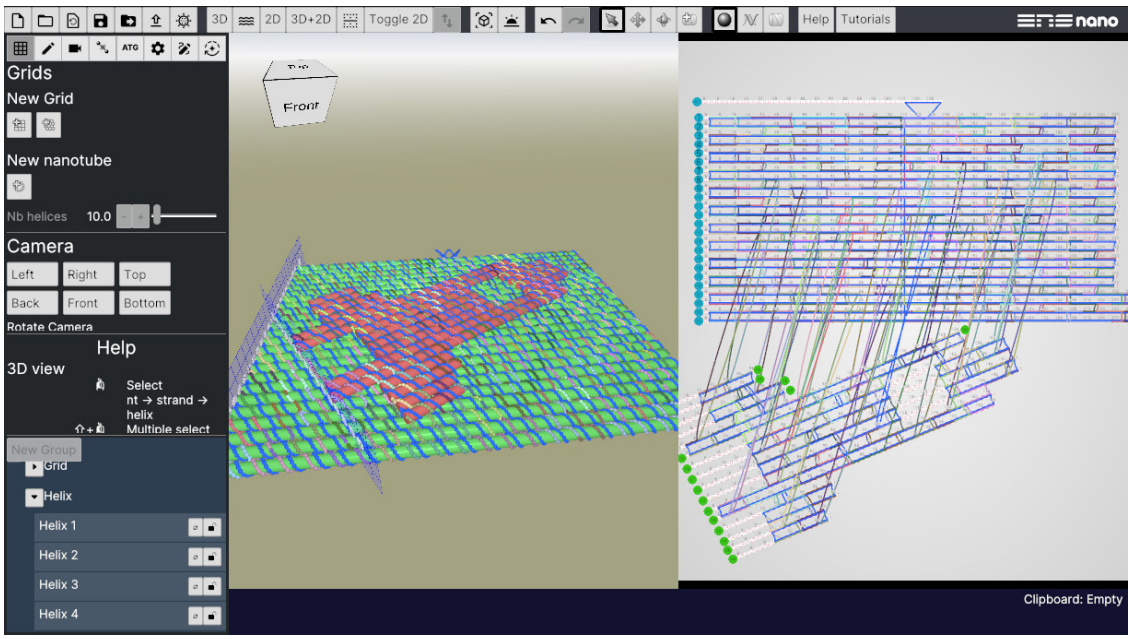


Figure 2.3: ENSnano software, presenting both a 2D and 3D visualization and editing interface. Here, the DNA origami design is a 2D rocket shape attached to a square layer.

2.2 Classical fabrication and characterization process

An important benefit of the DNA origami method is its ease of experimental assembly, making it accessible to researchers from various fields and without a lot of wet-lab expertise. A commonly used scaffold strand and the custom sequence DNA staple strands can be purchased from specialized companies such as Integrated DNA Technologies (IDT). Then, most assembly protocol involve a pH 8 Tris-acetate-EDTA (TAE) buffer with a concentration of Mg^{2+} ($MgCl_2$) of 5 to 20 mM, where the scaffold strand is diluted with a 2- to 100-fold excess of staple strands. This mix is then annealed in a thermocycler, following a decreasing temperature ramp, typically going from 80-95°C to 10-25°C in a few hours or days. Intuitively, this will ensure that the DNA strands are in a fully single-stranded state before the folding starts, then gradually attach to each other in a reversible way until they bind in the correct configuration, which is by design the most thermodynamically stable one.

Once the annealing is completed, the resulting DNA origami structure can be purified and characterized with various methods:

- **The agarose gel electrophoresis** allows to both purify and evaluate the quality of a DNA origami folding. It relies on the negative electric charge of the DNA molecule

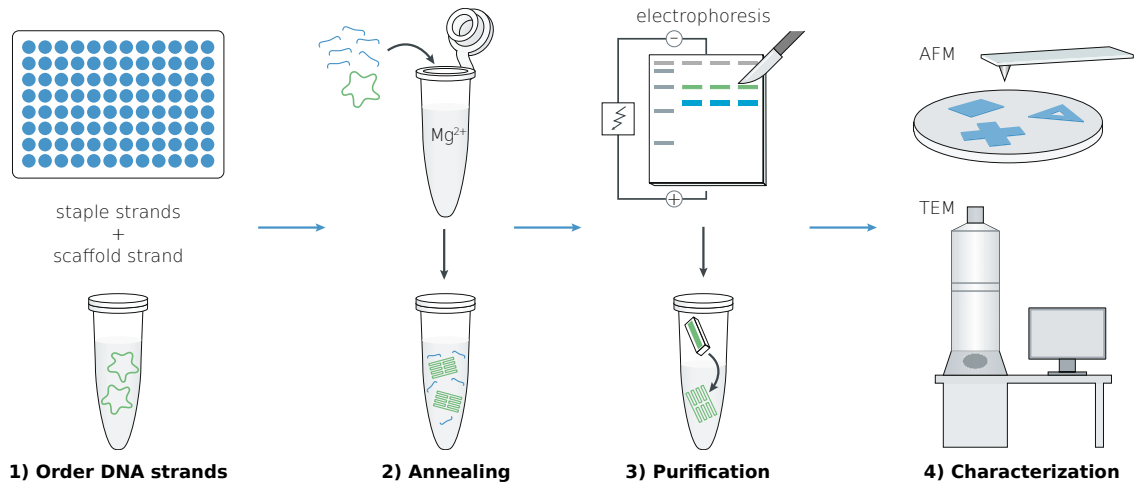


Figure 2.4: DNA origami assembly steps. From [Dey+21]

to separate DNA complexes that are placed on an agarose gel. An electric voltage is applied across the gel, causing the negatively charged DNA strands or structures to move towards the positive pole. The migration speed depends on the gel composition and the applied voltage, but also on the size and shape of the strand or strand assembly (more compact structures tend to migrate faster). This allows to separate different types of strands and DNA complexes in the solution based on their characteristics. Using a reference solution, one can measure, identify, and purify the DNA origami structure of interest.

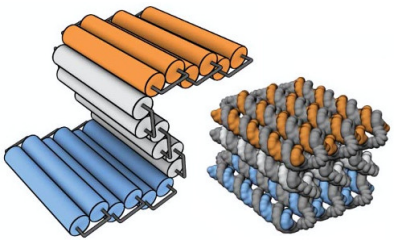
- **The Atomic Force Microscope (AFM)** [BQG86] is used to obtain high precision images of a relatively flat sample. It works by scanning the surface of the sample with an extremely sharp tip (about 1 nm), in order to reconstitute its three-dimensional shape in the form of a heightmap with a sub-nanometer resolution. It is particularly adequate for imaging flat DNA origami structures, that are first attached to a mica sheet surface.

- **The Transmission Electron Microscope (TEM)** form high resolution images by transmitting a beam of electrons through a sample. It can achieve a resolution of ~ 2 nm (or down to 0.1 nm in the case of cryo-EM), and is typically used for three-dimensional DNA origamis.

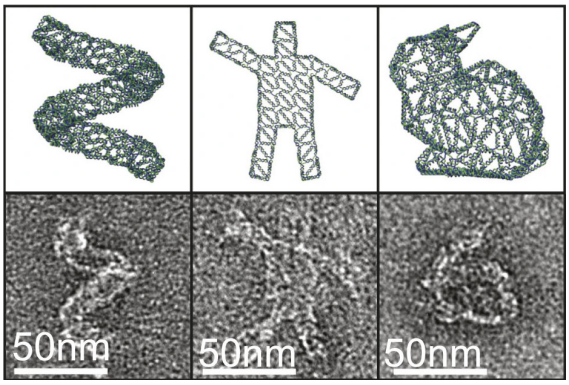
2.3 Advanced origami structures

While the first DNA origamis ever created were 2D single layer structures of various but simple shapes (see fig. 2.2), the space of achievable DNA structures with this technique has expanded quite a lot since its invention in 2006.

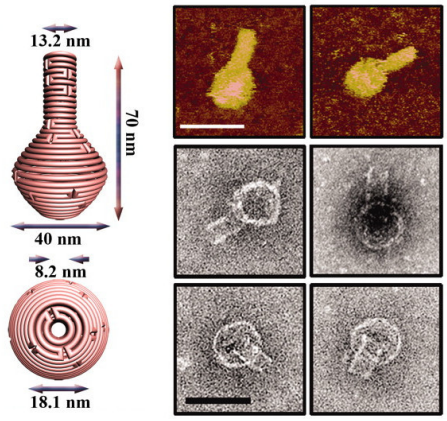
3D DNA origami structures first appeared in 2009 [Dou+09a], with designs where parallel DNA double helices are constrained to a honeycomb lattice (see fig. 2.5a). While this design method is still predominant, other classes of 3D designs were invented in the following year, including wireframe structures (fig. 2.5b) [Ben+15; Zha+15], shapes with complex curvatures (figs. 2.5c and 2.5d) [Han+11; Lev23] and some designs combining those methods [DDS09; Hua+21].



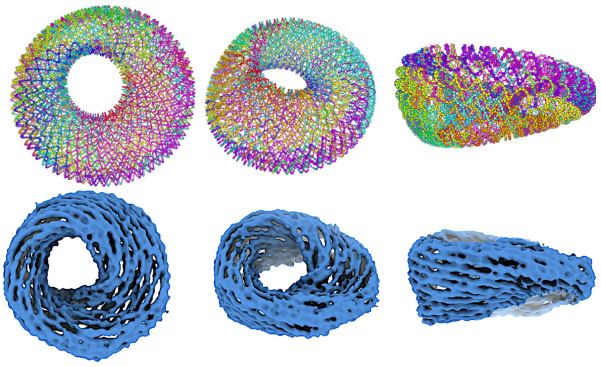
(a) Design principle of a 3D DNA origami using parallel double helices arranged into a honeycomb lattice. From [Dou+09a].



(b) Wireframe DNA origami [Ben+15], created using an automated method for converting 3D polyhedral meshed into a viable DNA origami design.



(c) DNA origami with 3D curvature, designed by stacking concentric DNA rings [Han+11].

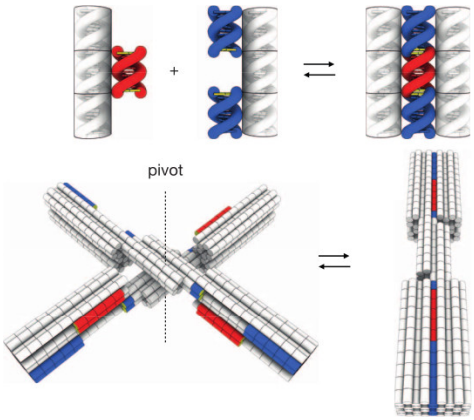


(d) Example of a 3D DNA origami with complex curvature, accurately designed using ENSnano software [Lev23]. Above is a 3D representation of the origami as designed in ENSnano and bellow is a 3D reconstruction from high resolution Cryo-EM data.

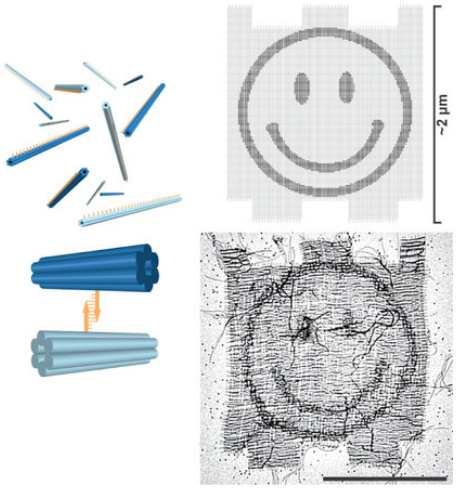
Figure 2.5: 3D DNA origami structures

Multicomponent DNA origami complexes were already explored in P. Rothemund's first publication on DNA origami [Rot06], but they progressively grew in size and complexity since then. Examples include 3D origami component assembled with additional connector staple strands [DDS09] or using shape complementarity and base stacking interactions between blunt-ended DNA double helices [WR11; Ger+15] (see fig. 2.6a). Large assemblies with uniquely addressable nanoscale features [TPQ17] or exceeding the gigadalton in mass [WSD17] have also been reported, and recently reached the multi-micron scale with assemblies based on controlled crisscross polymerization of DNA-origami slats [Win+23] (fig. 2.6b). Additionally, several examples of 2D [Liu+11] and 3D DNA origami crystals [Zha+18; Pos+24] have been developed, with promising applications as tunable photonic crystals (see fig. 2.6c).

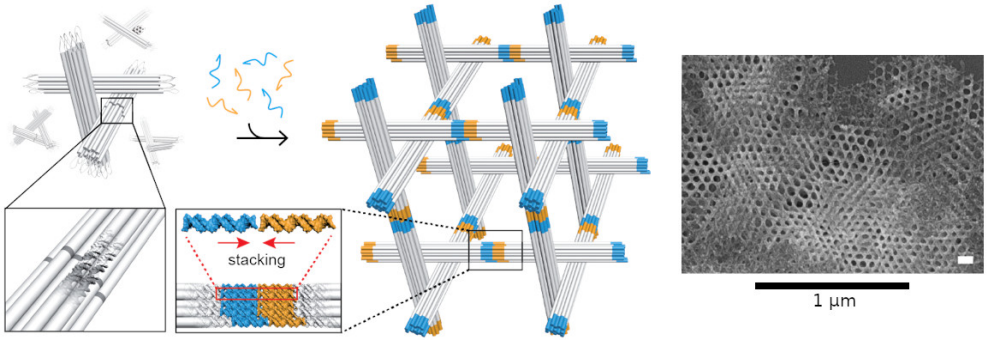
Other non-classical assemblies such as RNA origami [GRA14] can also be noted as ways of expanding and diversifying the DNA origami technique. A noteworthy advance in this direction is a method for designing and synthesizing a multi-kilobase DNA or RNA single strands that can fold into a user-prescribed complex, arbitrary shape, resulting in large single-stranded DNA or RNA origami [Han+17]. A different approach is the use of enzymes to enhance the DNA origami technique. Indeed, DNA ligase can be used to covalently bind staple strand ends in a fully folded DNA origami, in order to enhance its stability [Ram+19], and the use of DNA polymerase has also been reported as a way to complete the structure



(a) A dynamic nanodevice switchable between two conformations (open and closed) upon the competition between base stacking (arising from the complement of blue and red domains) and electrostatic repulsion, which is responsive to the change in temperature and/or Mg^{2+} concentration [Ger+15].



(b) Multi-micron DNA structures made by crisscross polymerization of DNA origami slats [Win+23]. Scale: 1 µm.



(c) 3D rhombohedral crystalline lattice made of DNA origami building blocks [Zha+18].

Figure 2.6: Multicomponent DNA origami.

of a DNA origami initially folded with a very small number of staple strand (only 18 in this example), thus drastically reducing its producing cost when the shape is simple enough [He+20].

Conclusion. Overall, it may be said that DNA origami is a versatile technique for constructing precise nanostructures, leveraging the programmability of DNA molecules. This method allows for the creation of intricate two- and three-dimensional designs, with broad applications from nanomedicine to molecular machines. But despite its remarkable versatility, limitations in scaffold size and assembly complexity persist. My interest lies in extending the capabilities of DNA origami beyond traditional single-scaffold designs. This opens new avenues for complex, multi-scaffold origamis, providing not only larger and more functional nanosystems but also offering insights into the self-assembly mechanisms at play. Throughout the remainder of this thesis, I investigate the experimental and computational methods used to address these challenges.

Part II

A multi-scaffold DNA origami

Chapter 3

Motivations and feasibility study for a double scaffold DNA origami

The idea of building a multi-scaffold DNA origami arose from the size limitation of the classical DNA origami technique. Here, we explore how to build larger DNA origami structures using only a single scaffold type (for instance the most commonly used one, the M13mp18) and a single annealing. Hence, the initial goal was to be able to fold several copies of the same scaffold strand into different shapes that will contribute for distinct parts of the same large structure. We call *k-origami* such a structure made of k instances of the same scaffold strand (or *bi-origami* for $k = 2$). Our goal is to find experimental methods and design guidelines that allow this multi-scaffold origami to fold entirely during one single annealing reaction.

3.1 Motivations and challenges

Simple one-pot single-annealing assembly is limited in size by the number of different available scaffolds which furthermore have largely overlapping sequences, leading to uncertain chimera structures. Custom orthogonal scaffolds can be designed and built using biomolecular processes, but it requires lots of skills, and it is a time-consuming process [Eng+19]. DNA origami surprising reliability remains a puzzle and the underlying assembly process is suspected to get its reliability from a combination of several factors [SMD19]: *thermodynamics* – staple strands tend to attach where they match perfectly, and longer domains tend to attach first and at higher temperatures; *topology* – which domains of the scaffold will be brought together by the staple strands – probably key at the beginning of the assembly; and *geometry* – the intermediate geometry of partially assembled origami helps discriminate between possible locations of binding for other staple strands – probably key at the later stages of the assembly. Understanding how these three driving processes play their parts could allow to take control over DNA origami folding to enforce a specific pathway as demonstrated in [Mar+16; JKC21].

Several challenges are to be dealt with in order to create multi-scaffold origami. We want to prevent a single scaffold to remain bound to competing staples (i.e. designed for distinct scaffold strands), which would result in chimeric structures. After joining together the desired number of scaffolds for our structure at an early stage of assembly, we need to break the symmetry between those scaffolds so that each one starts folding differently. The asymmetry must be preserved and propagated during the folding process, and we want to prevent undesired folding, for instance by controlling the folding pathway of each scaffold.

Another approach is to try to make each scaffold strand fold rapidly before starting to

attach together. It would need design strategies that force a scaffold strand to differentiate into one of the origami parts (if possible with equiprobable outcomes), without allowing the formation of chimeric structures.

3.2 Supporting works

Several papers inspired this project.

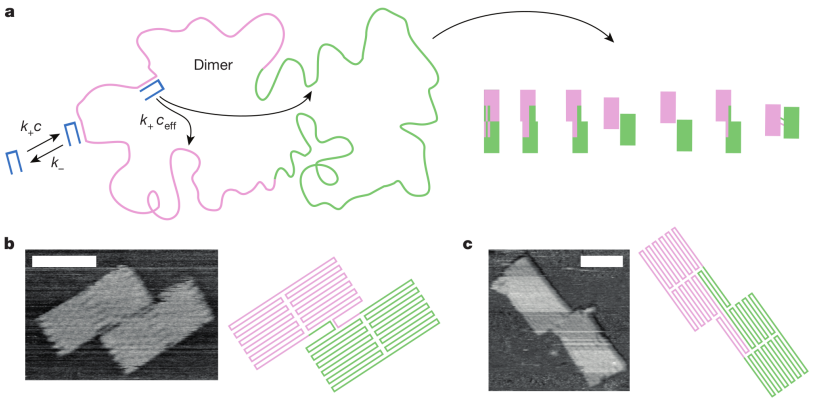


Figure 3.1: Repetitive scaffold origami design from "Guiding the folding pathway of DNA origami" (Dunn et al., 2015 [Dun+15]). Here the green and pink scaffold regions represent the same 2,646-nucleotides sequence. Image **a** shows the possible binding patterns of one staple and the seven possible well-folded shapes. Images **b** and **c** show AFM images of two possible shapes (with a scale bar of 50 nm) alongside a schematic of the fold.

A symmetric origami with repetitive scaffold. First, "Guiding the folding pathway of DNA origami" (Dunn et al., 2015 [Dun+15]) shows some experimental results on an origami with a repetitive scaffold which consists of two copies of the same 2,646-nucleotides sequence in head-to-tail repeat. Their design only uses one- and two-domains staple. Each domain of the scaffold corresponds to a two-domains staple. It can thus theoretically bind to two possible domains with that staple. The theoretical number of configurations (with no remaining single-stranded domain) is $2^{71} \simeq 2 \cdot 10^{21}$. But they demonstrated that this number is drastically reduced mainly because of the geometrical constraints that most of those configurations present. In fact, only seven different shapes were observed with reasonably high frequency (each shape corresponding to a few distinct staple configurations). Those outcomes are represented in figure 3.1. This article also gives some insights on how to model this kind of exotic DNA origami folding, taking into account the thermodynamics of staple bonds as well as their impact on the geometry of the resulting complex.

This design differs from our goal as our design should use two (or more) separated scaffold strands instead of a repetitive one, and we would like to fold each scaffold into a different shape in order to form each part of the desired structure. The following paper can help us to do so.

Co-folding multiple DNA origamis. In the article "Co-self-assembly of multiple DNA origami nanostructures in a single pot" (Johnson et al., 2021 [JKC21]), the authors managed to assemble distinct origami in a single annealing reaction and using only one scaffold type. But they only demonstrated it with two ad hoc designs: an 18 helix bundle and a 6 helix

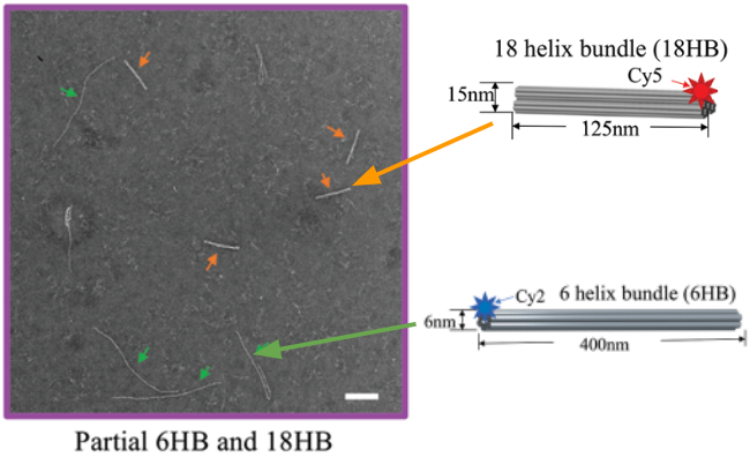


Figure 3.2: "Co-self-assembly of multiple DNA origami nanostructures in a single pot" (Johnson et al., 2021 [JKC21])

bundle. Moreover, they had to work a lot on adjusting the annealing parameters and staple concentration in order to obtain a low but reasonable yield (around 10-20% of each design simultaneously)

In our case, we would like to find a mechanism that helps the differentiation of a scaffold into a specific folded shape, and thus increase this yield. As our goal is to build large structures with multiple scaffold strands, an idea would be to use an asymmetric first staple bond between two scaffold strands, hoping that the induced symmetry breaking propagates inside both parts, in order to initiate the differentiation of those scaffolds into distinct shapes.

Controling the nucleation. Lastly, Evans et al. [Eva+24] propose a way to control the outcome of competitive nucleation phenomenons by adjusting individual strand concentrations in a mixture of 917 molecular species that can assemble into three distinct SST structures from one set of molecules.

Nucleation begins with the spontaneous formation of a critical seed that grows into a structure (see fig. 3.3). The rate at which a seed forms depends on the concentration of its components, and multiple seeds and pathways may lead to the same outcome. Thus, the overall rate of each structure formation is a complex function of the concentration pattern. Since the same components may contribute to various structures, competition for resources creates a winner-takes-all dynamic that emphasizes the discrimination between concentration patterns.

Our own objective is to extend this principle to multi-scaffold DNA origami, in order to guide the correct folding of each scaffold by controlling nucleation phenomenons.

We will see in the following chapters that nucleation-based methods don't seem to work so well for multi-scaffold DNA origamis. We took inspiration from the idea of a seed of carefully chosen staple strands that will help fold our designs. However, our seed acts as a barrier against unwanted conformations rather than as a nucleation seed.

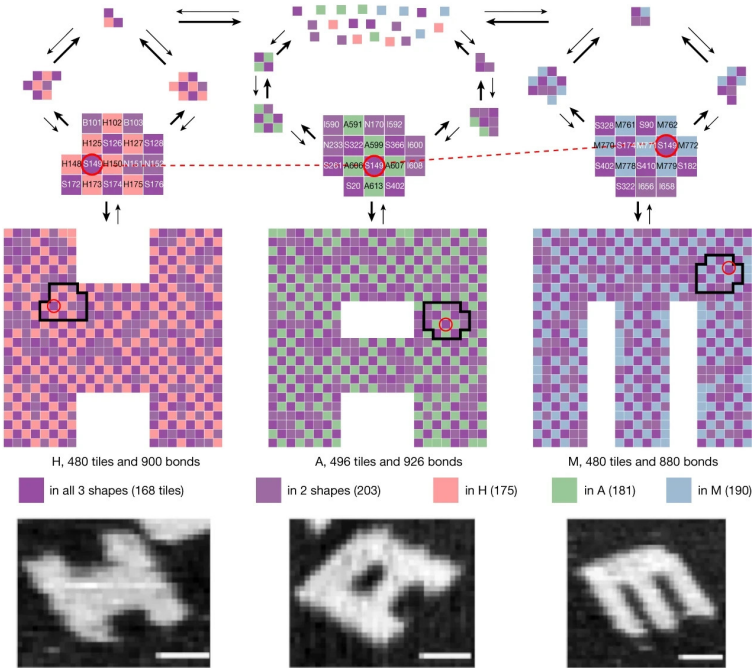


Figure 3.3: Controlling the nucleation of SST structures [Eva+24]. Three distinct structures (H, A or M) can be assembled from one pool of 917 single-stranded tiles. Purple tiles, such as S149, appear in multiple shapes. AFM images of each shape are shown in the lower row (scale bars: 50 nm).

Chapter 4

Preliminary experiments

Before designing and co-folding of an asymmetric bi-origami, we proceeded gradually by first validating the join of two identical scaffold strands, then the one-pot single-annealing of a bi-origami made of two identical components.

4.1 Step 1: Bindings two scaffolds

We want to attach two scaffold strands at high temperature so that it occurs before the other staples of the design bind to it. That should allow us to use this first binding to introduce asymmetry in the design and initiate the differentiation of each scaffold into a different shape. So these binding staples have to use longer domains than the other staples. Moreover, if a binding staple is composed of two long domains that bind to two distinct scaffold strands, then those domains have to be identical so that the staple can't bind its two domains to the same scaffold strand. Hence, we used 60-nucleotides repetitive strands (i.e. consisting in the concatenation of two copies of the same DNA sequence) as binding staples between two scaffold strands (fig. 4.1). We experimented with two staple variants that we put in stoichiometric amount relatively to the scaffold strands (i.e. the concentration of linkers was half that of the scaffold). Here is the reason behind this choice: in order to maximize the yield, we want that the complex with one scaffold strand and one link staple never represents more than 50% of the total number of scaffold. Indeed, the idea is that such complexes are too stable to dissociate themselves, so they need to encounter an unattached scaffold strand in order to form the desired structure. This reasoning was also validated numerically with NUPACK [Zad+11].

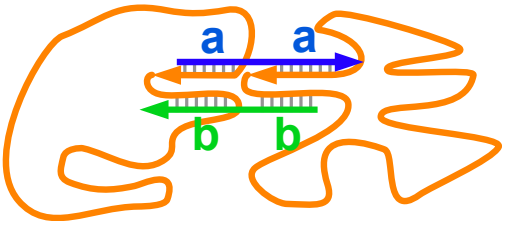


Figure 4.1: Principle of the two twin-domain binding staples aa and bb joining two scaffold strands together.

The gel electrophoresis in fig. 4.3 demonstrates the success of the join.

4.2 Step 2: joining two identical DNA origamis

The next step was to test a first symmetrical double-scaffold origami design, assembled in single pot. We used the double layer square origami design from [Thu+17] and two possible twin-domain binding staples. The resulting yield was quite good as 60% of the square structure were paired to another one (fig. 4.2). Surprisingly, we could not observe intertwined

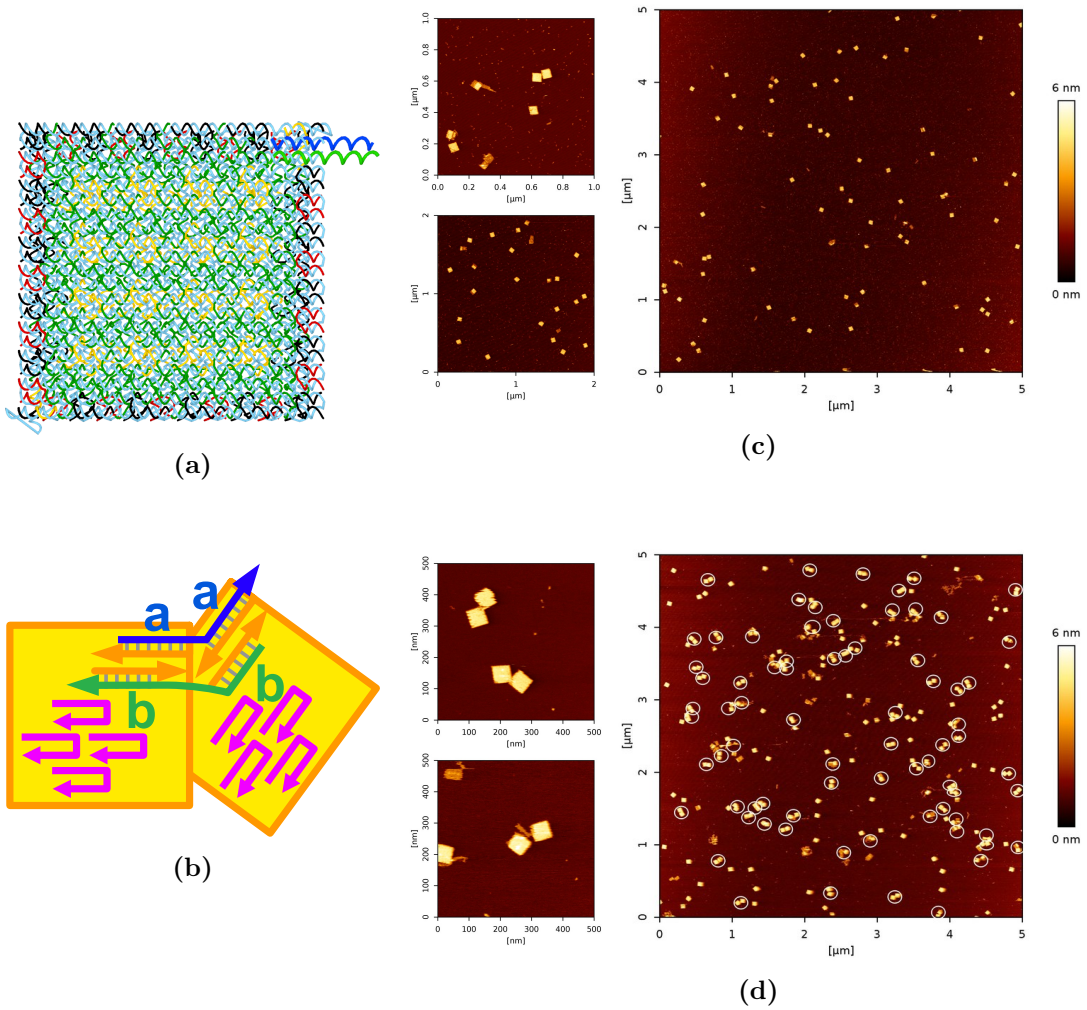


Figure 4.2: Binding two identical double-layer square origamis from [Thu+17]. (a) the two twin-domain link staples (aa in blue and bb in green); (b) schematic view of a linked pair of origamis (the scaffold strands are in orange); (c) Folding without link staples; (d) Folding with the two link staples in stoichiometric amount (1 : 2 scaffold strand). About 59% of the square origamis are paired in (d).

double-origami with non-binding staples joining the two components. We considered this as a very encouraging result towards the assembly of bi-origami.

We also compared by electrophoresis the yield for three different join configurations: with only one of the two twin-domain binding staples or both (fig. 4.3). As for the previous experiment, it appeared that the yield was maximum using both binding staple but was also satisfying with the second binding staple alone. But as some unwanted 3- and 4-origamis started to appear when both binding strands were added, we concluded that using the second binding staple alone would be the best option, and we adopted this alternative for most of our subsequent experiments.

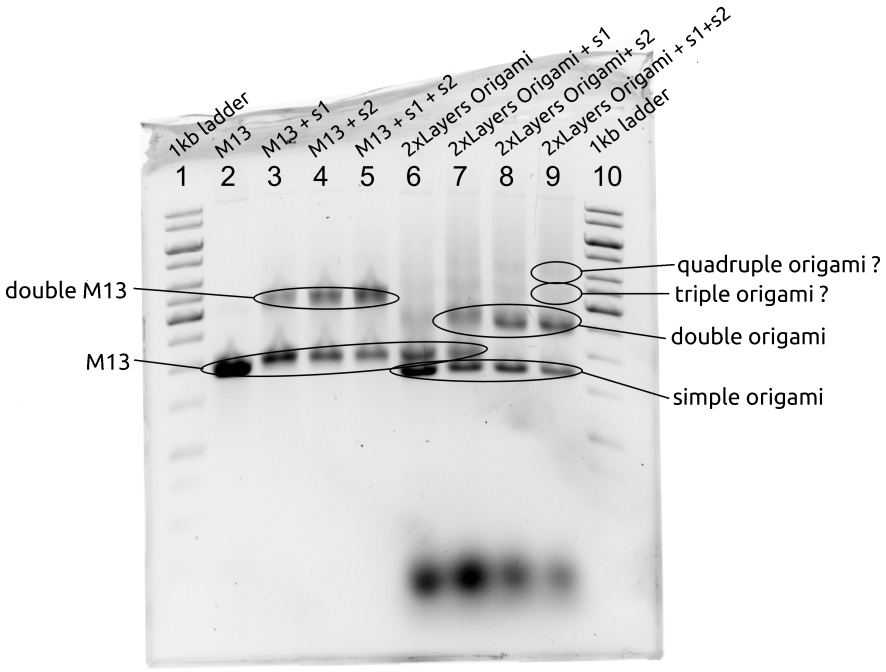
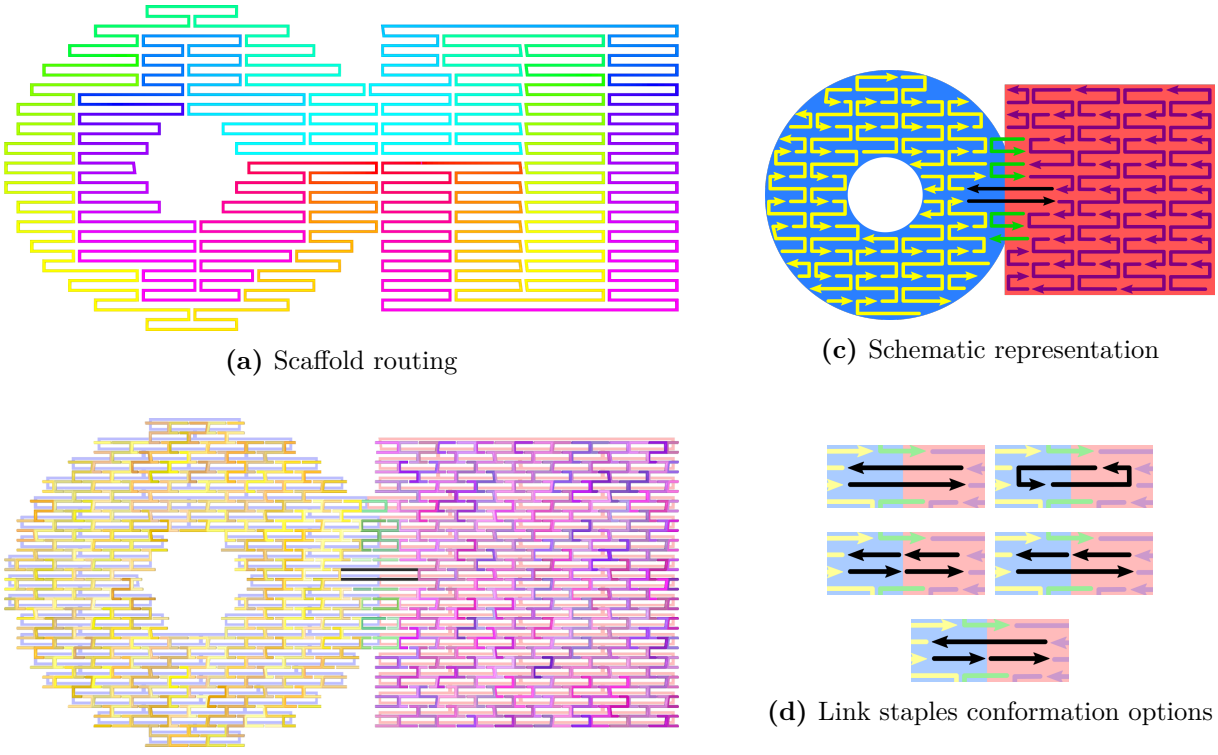


Figure 4.3: Agarose gel electrophoresis of the paired scaffolds and symetrical bi-origami. 1% agarose gel prepared from 0.5× TAE and 12.5 mM Mg^{2+} buffer with SYBER Safe DNA Stain. The gel was run at 50 V for 4 hours and 20 minutes (at ambient temperature). Wells 2-4: 40nM M13mp18 and 20nM of 60-nucleotides twin-domain binding strands (s1 and/or s2). Wells 6-9: 40nM M13mp18 with a 5-fold excess (i.e. 200nM) of origami staples and 20nM of 0, 1 or 2 twin-domain binding strands.

Chapter 5

First asymmetric test design and assembly attempts

5.1 Designing an asymmetric bi-origami



(a) Scaffold routing

(c) Schematic representation

(b) Staples outline

(d) Link staples conformation options

Figure 5.1: Ring-and-Square bi-origami design. (a-c) The base design is composed of two M13mp18 scaffold strands (in blue for the ring, and red for the square part), 237 ring staples (yellow), 230 square staples (purple), 2 long 64-nucleotides link staples (black) and 13 short link staples (green). (d) Four additional variations of the long links staples are considered.

We created our first bi-origami design with little knowledge of the best design strategies for this type of construction. We mainly based our choices on practicality and simplicity, in the hope that the experimental results would help us better understand those problems. With that in mind, we made a flat design composed of two simple and easily distinguishable shapes: a ring and a square (see fig. 5.1). Each shape consists of the folding of one copy of

the scaffold strand, and they are linked together with one or two long twin-domain binding staple strands (in black) and eleven shorter ones (in green).

We followed two main guidelines to route the two scaffold M13mp18 strands within our ◎□-origami. First, the half of the M13mp18 folding into the interior region of the ring component folds into the exterior part of the square component (and conversely). Our goal in proceeding that way was to break as early as possible the symmetry of the two components. A side motivation is that it simplifies the design of the two long twin-domain binding strands. Second, we chose the scaffold rotation that maximizes the binding strength of the twin-domain.

Staple details and notations for the bi-origami design:

- ◎□-origami: Ring-and-square bi-origami,
- ◎: Ring-shaped part of the bi-origami design,
- □: Square-shaped part of the bi-origami design,
- $S_{◎}$: set of staple strands that are exclusive to ◎ (in yellow in fig. 5.1c),
- $S_{□}$: set of staple strands that are exclusive to □ (in purple in fig. 5.1c),
- The main link staples (in black in fig. 5.1c) can join two domains of the ◎ scaffold (in blue) to the two sames domains of the □ scaffold (in red). We studied several configurations: either two parallel 60-nts long strands, or one long strand and the other one cut in two, both cut in two or two asymmetric staples with crossovers (see fig. 5.1d).
- The remaining staple strands that have domains in both components are represented in green in fig. 5.1c.

5.2 First co-folding attempts of our bi-origami

First, we checked that each half of the design folded properly as single scaffold origami. The results were entirely satisfactory (see fig. 5.2). However, annealing all staples strands in a one pot reaction only resulted in the formation of unwanted chimeric structures (fig. 5.3). Those structures range from single-scaffold ones to very large aggregates composed of more than ten scaffold strands. We also observed that square-like portions of those structures are overwhelming. It can probably be explained by the compactness of the square design that results in a more stable structure.

In order to counteract this predominance of square-like structures, we annealed with lower staple concentrations, ranging down to 0.75× the scaffold concentration, making sure that the square staple would not invade all the available scaffold strand (fig. 5.3). We then observed more structures identified as portions of the ring shape, and even some that resembled a portion of the expected bi-origami, however without managing to fully assemble any (table 5.1 offers a (non-exhaustive) of our one-pot folding attempts). We thus decided to proceed differently as exposed next.

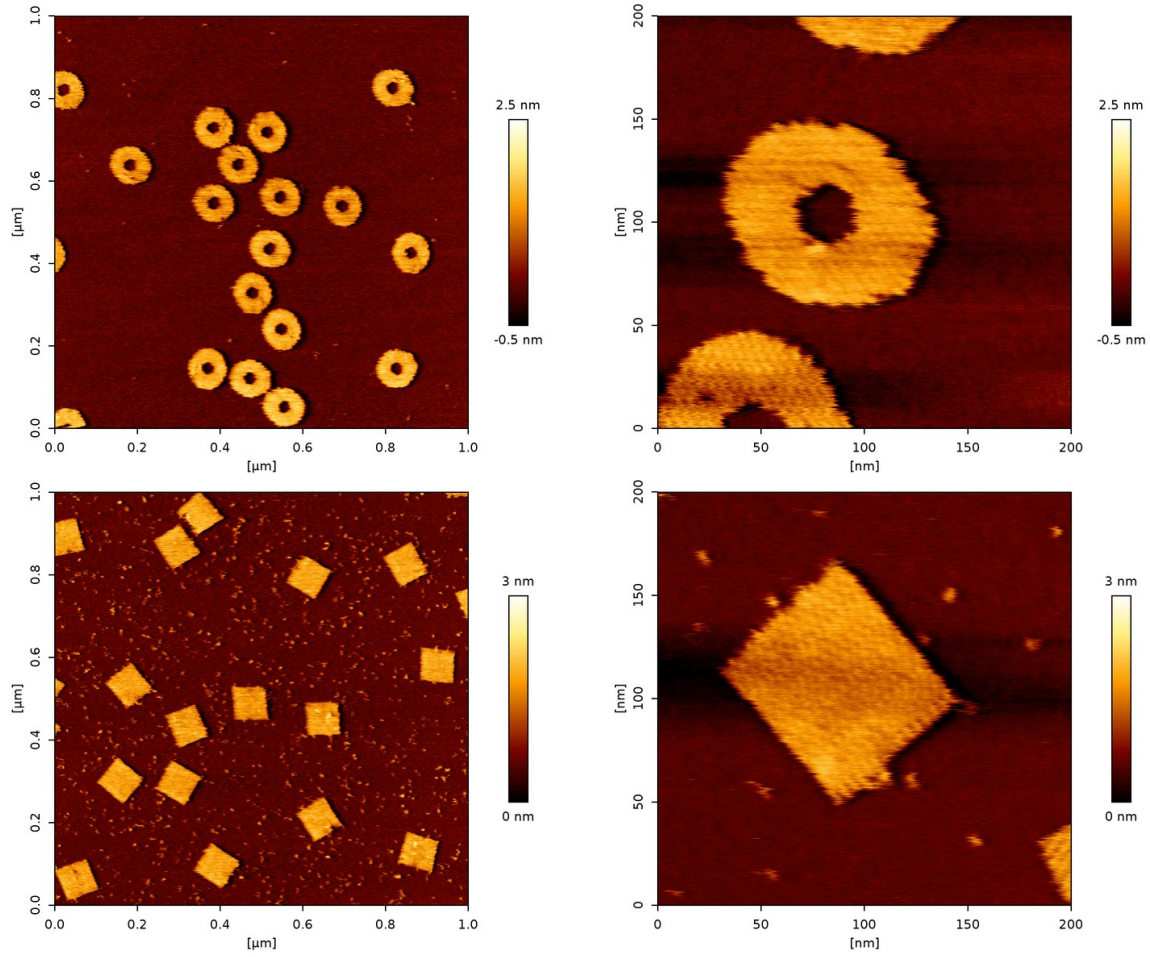


Figure 5.2: Separately folded ◎ and □ DNA origamis. Folding conditions: 40 nM of M13mp18 together with a 5-fold excess of yellow or purple staples and adapted black and green staples. Annealing: 90°C for 5 min, from 90°C to 70°C at the rate of −1°C every minute, then from 70°C to 45°C at the rate of −1°C every 2 minutes, then from 45°C to 20°C at the rate of −1°C every minute.

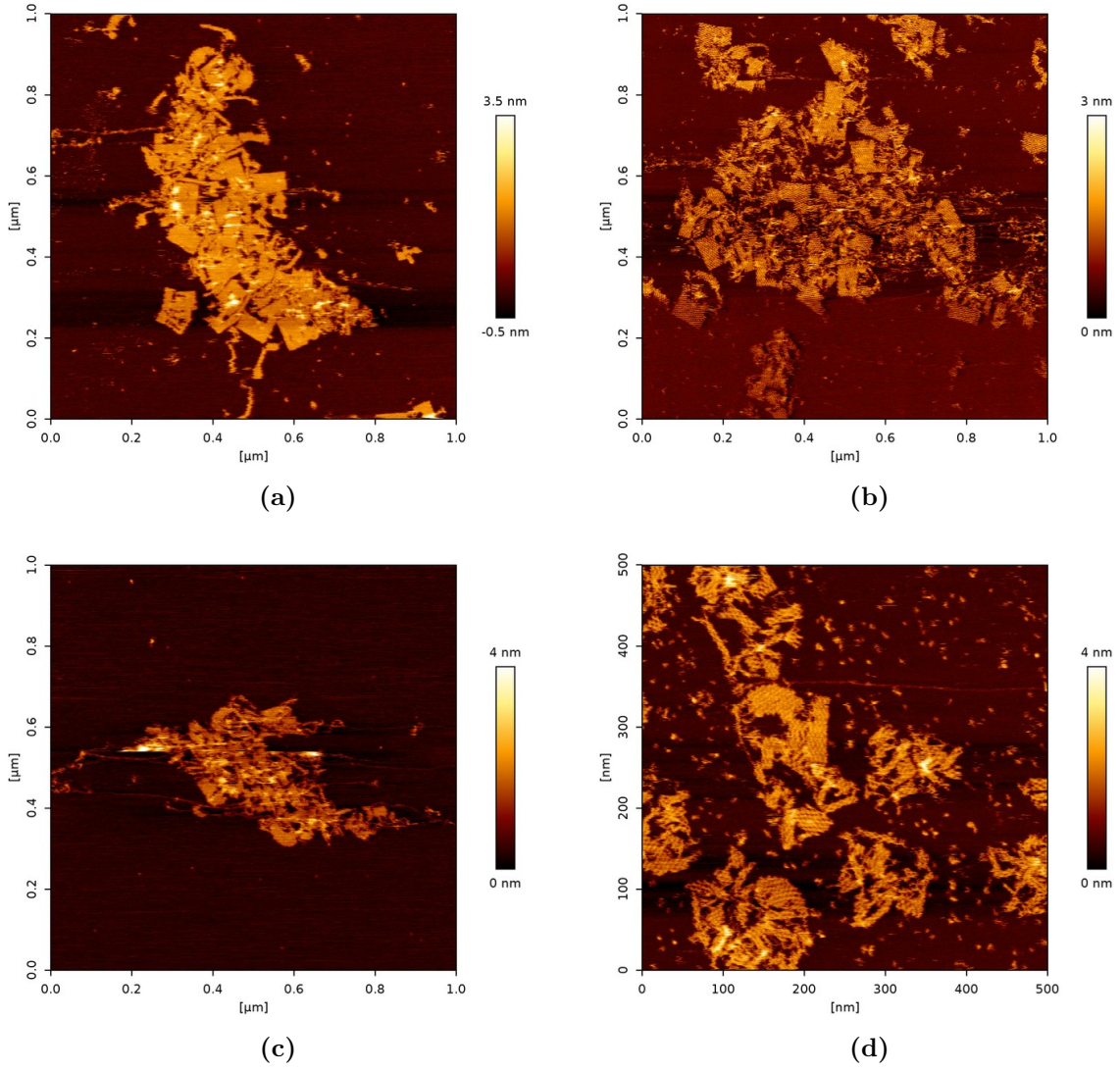
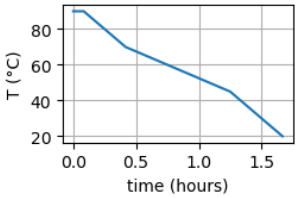
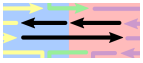
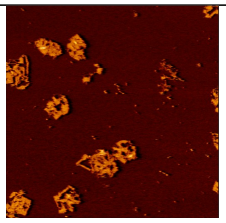
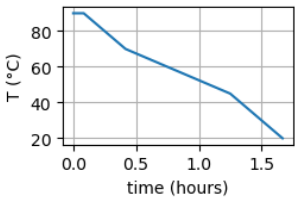
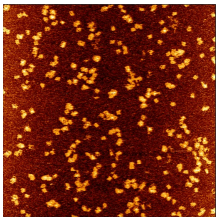
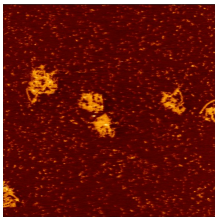
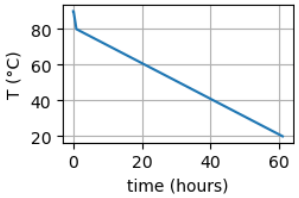
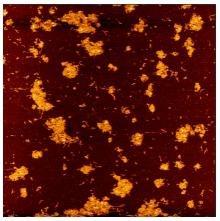
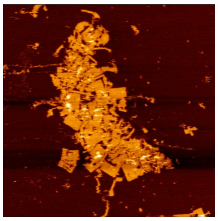
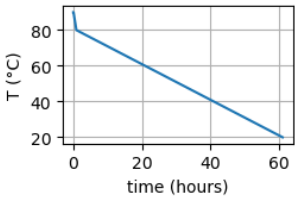
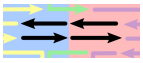
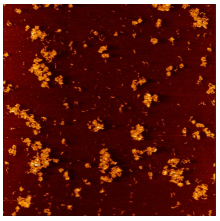
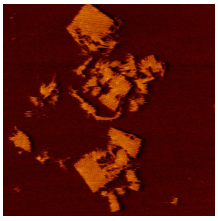
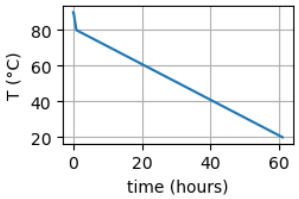
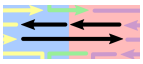
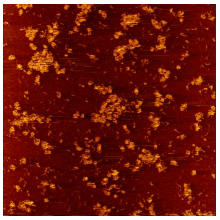
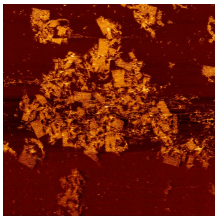
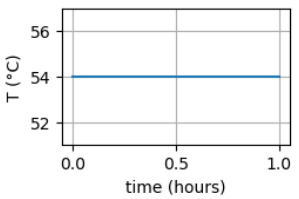
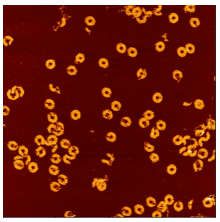
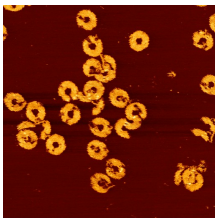
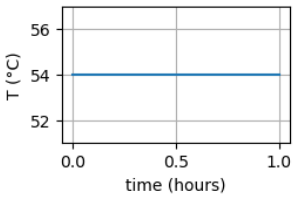
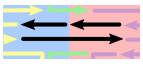
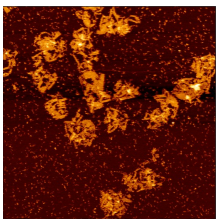
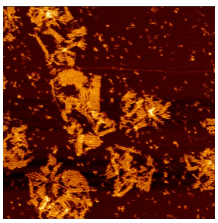
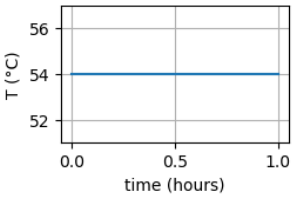
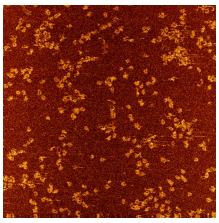
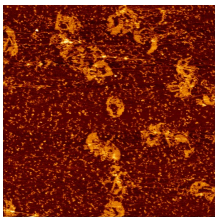
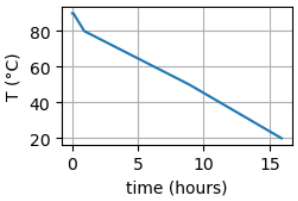
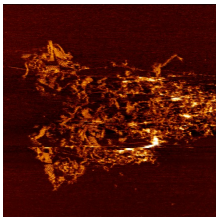
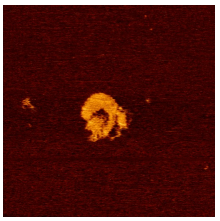
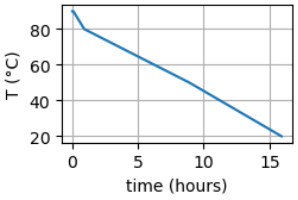
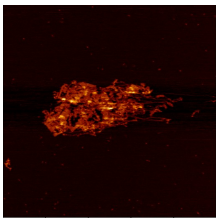
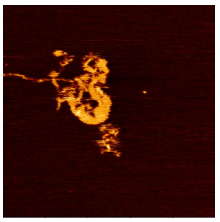
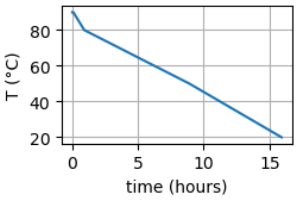
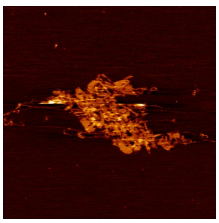
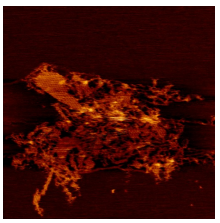
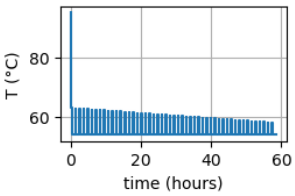
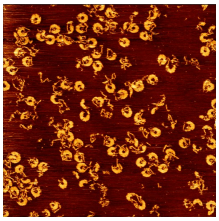
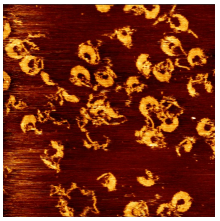
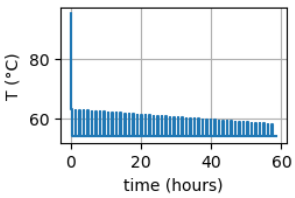
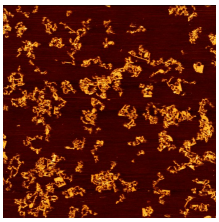
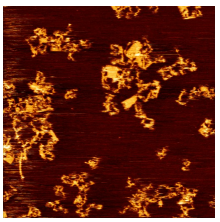


Figure 5.3: Chimeras obtained when co-folding the ◎□-origami. Several images with different relative staple concentrations. Ratio of yellow staples-to-M13 and purple staples-to-M13 are respectively, (a) 2.5 and 2.5, (b) 2.5 and 2.5, (c) 0.75 and 0.75, (d) 1 and 0.8.

Table 5.1: Some of the one-pot folding attempts for the ring-and-square bi-origami design. Annealing temperature in (l) and (m): 90°C for 10 minutes, then repeat for `T` in `[63, 62.9, 62.8, ..., 58.1]`: 10 minutes at 54°C then 10 minutes at `T`°C.

	Annealing	Staples to M13 ratio	Link staples	AFM results	
(a)		◎: 10 □: 10		$1 \times 1\,\mu m$ 	
(b)		◎: 10 □: 10		$5 \times 5\,\mu m$ 	$1 \times 1\,\mu m$ 
(c)		◎: 10 □: 10		$5 \times 5\,\mu m$ 	$1 \times 1\,\mu m$ 
(d)		◎: 10 □: 10		$5 \times 5\,\mu m$ 	$0.5 \times 0.5\,\mu m$ 
(e)		◎: 10 □: 10		$5 \times 5\,\mu m$ 	$1 \times 1\,\mu m$ 
(f)		◎: 1.6 □: 0	no link staples	$2 \times 2\,\mu m$ 	$1 \times 1\,\mu m$ 
(g)		◎: 2 □: 1.6		$1 \times 1\,\mu m$ 	$0.5 \times 0.5\,\mu m$ 

(h)		◎: 4 □: 2			
(i)		◎: 1.5 □: 1.5			
(j)		◎: 1.5 □: 1.5			
(k)		◎: 1.5 □: 1.5			
(l)		◎: 1.6 □: 0	no link staples		
(m)		◎: 0.8 □: 0.8	no link staples		

Part III

Two-step annealing principle based on seed staple selection

Chapter 6

Two-step annealing and greedy algorithm

6.1 Adding a step of partial annealing in separate pots

As observed in the previous chapter, when annealed together in a single pot reaction, staples from both components tend to attach to either component more-or-less at random, leading to massive chimeric mega-origamis. We propose to prevent this unwanted binding by adding one "prefolding" step where each component is partially assembled in a different pot using only a subset of its strands, called the *seeding strands*. The two pots containing the partially assembled components (and the seeding strands in excess) are then mixed together and annealed with the remaining staple strands of both parts in a second and final step, as illustrated in fig. 6.1.

Seeding strands are used to break symmetry to ensure that no strand will attach to the wrong side in the second annealing step where all the strands are present. Our goal is to select a minimum set of seeding strands in both sides to ensure a proper assembly of the bi-origami. Such minimum seeding sets would contain the staples that are the key to the differentiation between the two components and our hope is that understanding their structure will help understand how origamis assemble themselves. Their distribution in the design should help us determine which criteria prevails when between thermodynamics, topology and geometry.

Checkerboard seeding subsets. Staples in flat origami have a natural square grid structures. A natural geometric approach consists in selecting seed strands according to a checkerboard pattern in both sides, as [Eva+22] did in a similar context of differentiation. We evaluated two checkerboard patterns consisting respectively of 50 % and 25 % of the strands for our ◎□-origami. The results of their two steps assemblies is presented in the rows (c) and (d) in table 7.1. In spite of the apparently large coverage of the seeding strands and relatively small dispersion in the gel electrophoresis (col. cb50 in fig. 7.1), we could only observe ill formed bi-origamis. Staples from the other side seem to detach legitimate staples from their correct locations.

As this purely geometric strategy appears to fail for the ◎□-origami, and as strands seem to get kicked out by other's side strands, we concluded that thermodynamics had to be taken into account and designed algorithms accordingly to look for better seeding strands.

6.2 Greedy algorithm

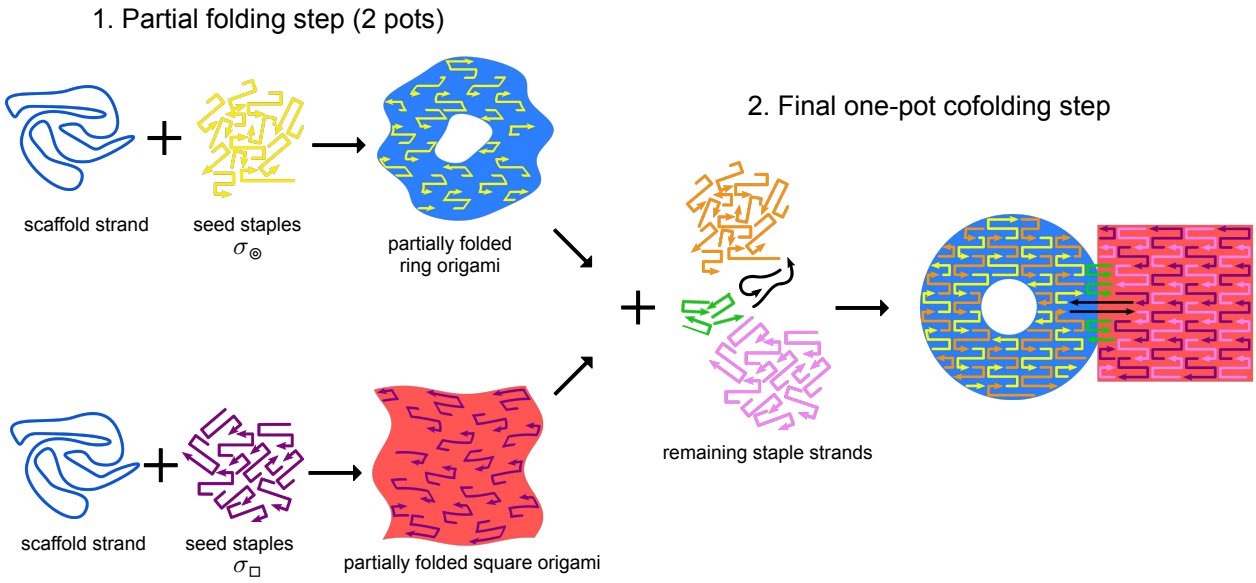


Figure 6.1: 2-steps cofolding annealing.

We investigate a first simple algorithm to select seeding strands on each side based on an ultra-simplified thermodynamic model.

Principle. We want staples to attach preferably to their proper side, ring or square. We propose then to select a subset of seeding strands on each side that will "screen" the scaffold so that every strand from the other side would receive a uniform penalty when trying to attach to the wrong side with respect to their expected location on the correct side. The following greedy algorithm implements this idea.

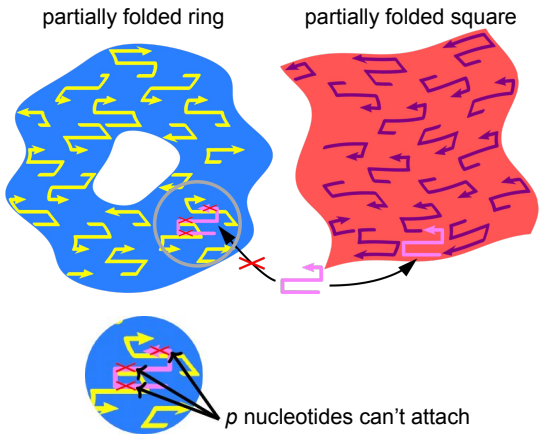


Figure 6.2: Illustration of the penalty p. A staple from $S_{\square}$ (in pink) is prevented form attaching to the partially folded ring because at least p nucleotides of its corresponding domains on the ring scaffold (in blue) are already connected to the seed staples $\sigma_{\circledcirc}$ (in yellow).

The greedy algorithm. The two seeding subsets $\sigma_{\circledcirc}$ and $\sigma_{\square}$ for each side $\circledcirc$ and $\square$ are computed independently according to the same greedy algorithm.

We fix a positive integral *penalty* p which is defined as a minimum number of nucleotides p covered by the seeding strands that a strand of the other side has to detach in order to attach on the wrong side. This implies that each strand from the other side has a strong incentive to attach on its legitimate location rather than to attach on the opposite side.

Formally for $(a, b) \in \{(\circledcirc, \square), (\square, \circledcirc)\}$, we start with $\sigma_a := \varnothing$. The nucleotides of the M13mp18 scaffold strand are numbered from 1 to 7249. Given a staple strand $s \in S_a$ (resp. S_b) we denote by $\mathsf{dom}(s) \subset \{1, \ldots, 7249\}$ the binding domains of s on the scaffold strand in its corresponding origami a (resp. b). Given a strand $s \in S_b$, its *remaining number of nucleotides to cover* is defined as:

$$\mathsf{rnntc}(s, \sigma_a) = \max\Big(0, p - \sum_{s' \in \sigma_a} |\,\mathsf{dom}(s) \cap \mathsf{dom}(s')|\Big),$$

i.e. the number of nucleotides in its binding domains that still need to be covered by the seed σ_a so that at least p of them are covered. The total rnntc of seed σ_a is defined as:

$$\mathsf{rnntc}(\sigma_a) = \sum_{s \in S_b} \mathsf{rnntc}(s, \sigma_a).$$

rnntc can be seen as a cost function, and if $\mathsf{rnntc}(\sigma_a) = 0$ then the seed σ_a meets the requirement given by the penalty p. The greedy algorithm proceeds as follows:

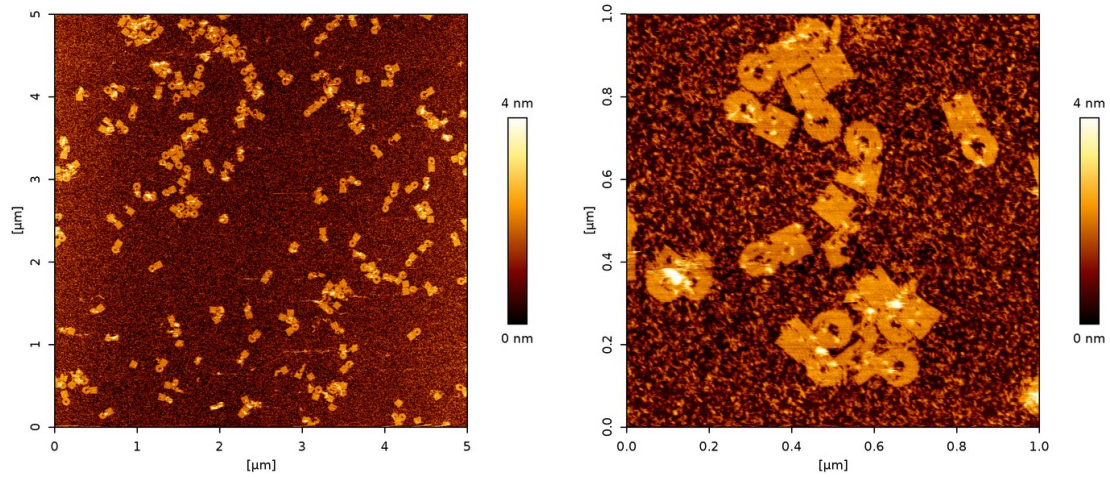
for $(a, b) \in \{(\circledcirc, \square), (\square, \circledcirc)\}$ **do**
  Start with $\sigma_a := \varnothing$
  while $\mathsf{rnntc}(\sigma_a) > 0$ **do**
    Select the staple strand $s \in S_a \setminus \sigma_a$ that minimizes $\mathsf{rnntc}(\sigma_a \cup \{s\})$ (ties are broken arbitrarily), and add s to σ_a: $\sigma_a := \sigma_a \cup \{s\}$
  end
end
return $(\sigma_{\circledcirc}, \sigma_{\square})$

6.3 Experimental evaluation

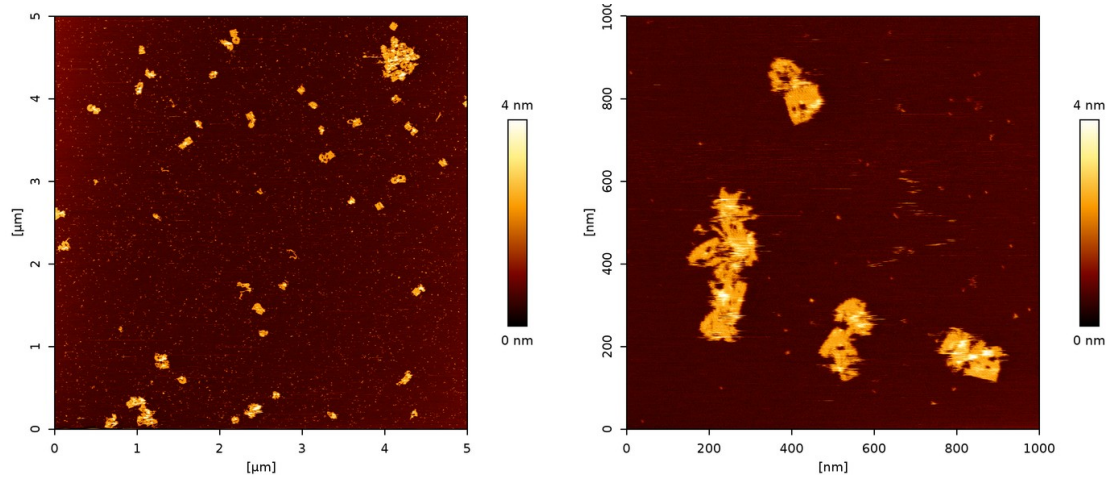
We evaluated the greedy algorithm for two values of the penalty on the ⊚□-origami: $p = 8$ and $p = 4$ corresponding respectively to seeding subsets of 47 % and 34 % of the total number of staple strands. The resulting seeding subsets are displayed in rows (e) and (f) resp. in table 7.1, together with AFM images of the assemblies obtained by the seeded co-folding process (also in fig. 6.3).

One can observe that the greedy seeding staples are surprisingly organized according to some zebra random-looking pattern, with large uncovered areas on both sides.

The seeding subset for $p = 8$ performs however quite well as correctly co-folded structures could be observed. Note that the corresponding band in the gel eletrophoresis (col. 8cov in fig. 7.1) is also better discriminated than for the checkerboard pattern with a similar number of seeding strands. The seeding subset with $p = 4$ performs however quite poorly, which led us to develop a finer seed selection algorithm, presented in the next chapter.



(a) Penalty value: $p = 8$ (47 % of the total number of staple strands)



(b) Penalty value: $p = 4$ (34 % of the total number of staple strands)

Figure 6.3: AFM result of the greedy algorithm. ⊙□-origami was annealed in two steps. First annealing (20 nM of M13mp18 and a 5-fold excess of yellow/purple seed staples 12h): 90°C for 15 min, from 90°C to 65°C at the rate of −1°C every 2 minutes, then from 65°C to 45°C at the rate of −1°C every 30 minutes, then from 45°C to 25°C at the rate of −1°C every 2 minutes. Second annealing (mixing the results of the first annealing and adding a 5-fold excess of all staples): 51°C for 30 min, then from 51°C to 20°C at the rate of −1°C every 3 minutes.

Chapter 7

Linear program for selecting a seed

In this chapter, we propose an integer linear programming formulation for determining the seeding subsets. We then use IBM Cplex solver (version 22.1) to output seeding strands for each side of the ⊙□-origami which are experimentally evaluated at the end of this chapter.

As opposed to the previous chapters where we looked for seeding strands that would apply a uniform binding penalty to every strand from the other side, here we want to ensure that no strand from the other side will bind with a Gibbs binding energy smaller than a fixed energy threshold. This should ensure that all staple should be "equally unlikely" to bind to the wrong side of the bi-origami, independently of their length and sequence. This should furthermore ease finding the optimal experimental co-folding conditions as there should exist a temperature threshold above which no strand would bind to the wrong side in the second step of the folding.

7.1 A linear approximation for the Gibbs binding energy

We choose to approximate the Gibbs binding energy ΔG° of a staple strand as a linear function of its binding domains, meaning that it is equal to the sum of the Gibbs binding energies evaluated for each of it domains (independently of the subdivision). More precisely, we approximate the Gibbs binding energy to a given domain of the scaffold by substituting the ΔG°(init.) term in the nearest-neighbor model in [San98] with the half of the sum of the binding energy of the two stacks at both $5'$- and $3'$-ends of the domain:

$$\frac{1}{2}\Big[\Delta G^\circ(\text{5'-end stack}) + \Delta G^\circ(\text{3'-end stack})\Big].$$

Note that one must take into account the two nucleotides surrounding the domain, which might be incorrect if the domain ends at a scaffold crossover. However, the resulting error in the calculation is only marginal and can be safely neglected for our purposes. For example, the binding energy of the domain spanning from nucleotide 2 to 6 in M13mp18 scaffold, whose sequence is underlined in A<u>ATGCT</u>A, is evaluated as:

$$\begin{aligned}\widetilde{\Delta G}^\circ_{2\to 6} \;=\; & \frac{1}{2}\Delta G^\circ(\text{AA/TT}) + \Delta G^\circ(\text{AT/TA}) + \Delta G^\circ(\text{TG/AC}) + \Delta G^\circ(\text{GC/CG}) \\ & + \Delta G^\circ(\text{CT/GA}) + \frac{1}{2}\Delta G^\circ(\text{TA/AT}).\end{aligned}$$

As the stacking interaction between two domains is counted twice in half, this expression allows to define the binding energy of a given domain as the sum of the binding energy of

any of its partitions into subdomains. Similarly, we evaluate the binding energy to a set of disjoint domains as the sum of their $\widetilde{\Delta G}^\circ$.

For $i \in S_\circledcirc$ and $j \in S_\square$, we denote by $\widetilde{\Delta G}^\circ_{ij}$ the binding energy of the intersection of the scaffold domains that bind to staple i and to staple j. The binding free energy of staple s or s' to the scaffold strand with no other staple is then:

$$\widetilde{\Delta G}^\circ_i = \sum_{j \in S_\square} \widetilde{\Delta G}^\circ_{ij} \quad \text{and} \quad \widetilde{\Delta G}^\circ_j = \sum_{i \in S_\circledcirc} \widetilde{\Delta G}^\circ_{ij}.$$

Note that if the binding domains of ring staple i and square staple j do not intersect, then $\widetilde{\Delta G}^\circ_{ij} = 0$. Note furthermore, that the values $\widetilde{\Delta G}^\circ_{ij}$ are fixed once the both staple sets $S_\circledcirc$ and $S_\square$ are fixed.

7.2 An integer linear program to design seeding subsets

When a ring-seed $\sigma_\circledcirc$ is entirely bounded to a scaffold strand, the Gibbs binding energy of a given square staple $j \in S_\square$ to this partially assembled origami becomes:

$$\widetilde{\Delta G}^\circ_{j,\sigma_\circledcirc} = \sum_{i \in S_\circledcirc \setminus \sigma_\circledcirc} \widetilde{\Delta G}^\circ_{ij}; \quad \text{and similarly } \widetilde{\Delta G}^\circ_{i,\sigma_\square} = \sum_{j \in S_\square \setminus \sigma_\square} \widetilde{\Delta G}^\circ_{ij}$$

for a ring strand $i \in S_\circledcirc$ binding to a partially assembled origami where all the staples in $\sigma_\square$ are attached. Given a threshold $E_{\min}$, we want to find the smallest possible seeds $\sigma_\circledcirc$ and $\sigma_\square$ such that for any staple strand of a given side binds to the partially assembled opposite side with energy larger than $E_{\min}$, that is such that:

$$(\forall i \in S_\circledcirc) \sum_{i \in S_\circledcirc \setminus \sigma_\circledcirc} \widetilde{\Delta G}^\circ_{ij} \geqslant E_{\min} \quad \text{and} \quad (\forall j \in S_\square) \sum_{j \in S_\square \setminus \sigma_\square} \widetilde{\Delta G}^\circ_{ij} \geqslant E_{\min}.$$

Introducing indicator variables $(X_i) \in \{0,1\}^{S_\circledcirc}$ and $(Y_j) \in \{0,1\}^{S_\square}$, determining such minimum-size seeds can be formulated as two independent integer linear programs:

$$\begin{cases} \text{Minimize} & \sum_{i \in S_\circledcirc} X_i \\ \text{such that} & (\forall j \in S_\square) \sum_{i \in S_\circledcirc} (1 - X_i) \widetilde{\Delta G}^\circ_{ij} \geqslant E_{\min} \end{cases}$$

$$\text{and} \quad \begin{cases} \text{Minimize} & \sum_{j \in S_\square} Y_j \\ \text{such that} & (\forall i \in S_\circledcirc) \sum_{j \in S_\square} (1 - Y_j) \widetilde{\Delta G}^\circ_{ij} \geqslant E_{\min} \end{cases}$$

Recall that as the staple set $S_\circledcirc$ and $S_\square$ are given by the bi-origami, the coefficients $\widetilde{\Delta G}^\circ_{ij}$ are fixed once the temperature for the second stage annealing is decided. The optimal seed sets $\sigma^*_\circledcirc$ and $\sigma^*_\square$ are then obtained by selecting all the ring staple strands i with $X_i = 1$ and all the square staple strands j with $Y_j = 1$:

$$\sigma^*_\circledcirc = \{i \in S_\circledcirc : X_i = 1\} \quad \text{and} \quad \sigma^*_\square = \{j \in S_\square : Y_j = 1\}.$$

7.3 Experimental evaluation

We set the temperature for the computation of the coefficients $\widetilde{\Delta G}^\circ_{ij}$ at $T = 51°\text{C}$ which corresponds to the plateau at the beginning of the second-stage annealing. The value of

$E_{\min}$ was chosen to be the q-th percentile of the strand binding energies (meaning that the $q\,\%$ most strongly bound staples should have their binding energy increased to be higher than $E_{\min}$ when binding to the wrong partially folded origami). The two integer programs were solved using IBM Cplex 22.1 in a few seconds on a regular laptop. We evaluated two values for q: $q = 85$ and $q = 70$ corresponding respectively to selecting $46\,\%$ and $32\,\%$ of the strands.

The obtained seeding subsets and AFM images of the resulting assembly are displayed in rows (g) and (h) of table 7.1. As for the greedy seeds, we observe that the seeding subsets display puzzling zebra random-like patterns with large uncovered areas on both sides. However, as opposed to the greedy seeds, we have obtained a consistent proportion of well-formed ◎□-origami even in the case of seeds as little as $32\,\%$. This was confirmed by the gel electrophoresis (col. lp85 and lp70 in fig. 7.1). This demonstrates the soundness of our integer linear program-based thermodynamic approach.

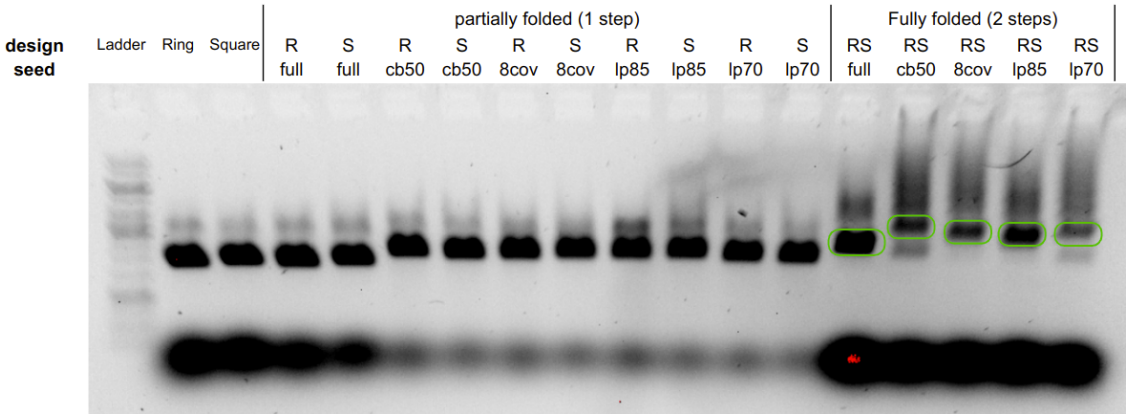


Figure 7.1: Gel electrophoresis of the different stages of assembly. First annealing (20 nM of M13mp18 and a 5-fold excess of adequate staples, no green link staples, 2 h total): 90°C for 5 min, from 90°C to 65°C at the rate of -1°C every minute, then from 65°C to 45°C at the rate of -1°C every 3 minutes, then from 45°C to 20°C at the rate of -1°C every 90 seconds. Second annealing (only for the last five wells, mixing partially folded Ring and Square and adding a 5-fold excess of all staples): 51°C for 30 min, then from 51°C to 20°C at the rate of -1°C every 3 minutes. The (possibly ill-formed) ◎□-origami bands are circled in green.

Conclusion. We demonstrated that systematic algorithmic techniques can be successfully used to guide DNA origami folding pathways and facilitate differentiation of the various parts of a multi-components DNA origami sharing the same scaffold strand. Similar algorithmic techniques may be derived from them to save the experimenter from tedious experimental condition manual adjustments [Mar+16; JKC21]. We expect our 2-steps co-folding method to allow to build reliably larger DNA origami, in particular with more than 2 components.

Our study allowed us to investigate how to influence the folding pathway of a DNA origami. As shown by our first attempt to fold the ◎□-origami in a one-pot single annealing, nucleation and staple cooperation alone does not seem to prevent staples to attach to the wrong side. Our checkerboard seed selection seems as well to demonstrate that geometry alone cannot enforce the desired pathway. Our most successful attempt involves selecting key seed staples using a fairly sophisticated thermodynamic-based integer linear program, suggesting an important role of individual staple binding thermodynamics in scaffold differentiation.

	Seed algorithm	Ring seed	Square seed	2-steps co-folding AFM images		
(a)	Full seed	100 % (237)	100 % (230)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(b)	No seed	0 % (0)	0 % (0)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(c)	Checker-board seed (50 %)	50 % (118)	50 % (115)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(d)	Checker-board seed (25 %)	25 % (60)	24 % (56)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(e)	8nt-Greedy seed (47 %)	45 % (106)	50 % (112)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(f)	4nt-Greedy seed (34 %)	33 % (78)	35 % (80)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(g)	LP 85 % (46 %)	46 % (109)	47 % (107)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm
(h)	LP 70 % (32 %)	32 % (76)	33 % (75)	$5 \times 5\,\mu$m	$1 \times 1\,\mu$m	500 × 500 nm

Table 7.1: Summary of the experimental results. Experimental conditions are described in fig. 6.3.

Chapter 8

Yield estimation based on AFM images using Deep Learning

I wanted to find a reliable and most importantly unbiased way of computing the yield of bi-origami production for each set of seeding strands. The method I chose is based on the automated detection of well-folded ◎□-origamis on AFM images. It uses the detection method from [Chi+22], who provides a straightforward procedure for identifying correctly formed DNA nanostructures with high fidelity on AFM images, using You Only Look Once (YOLO) object detection CNN framework [Red+16]. After identifying all well-folded ◎□-origamis in the AFM images of a sample, the yield of the sample is computed as the ratio of well-folded origami surface to the total surface of folded origami (partial or chimeric, as well as well-folded ones).

8.1 Neural network training and origami detection

Available data and annotation. We use 5×5 µm AFM images of the 8 samples presented in table 7.1. There are between 4 and 8 AFM images per sample, totaling to 43 images (more images were taken for samples with less origami density on the mica), and up to 300 ◎□-origamis per image. The 7 images from the full seed sample were used to train the model, as they present a high number of correctly folded ◎□-origamis (1895 in total).

Raw AFM images were manually annotated using the open-source data annotation tool labelImg.

Data augmentation and Network training. In order to increase the size of our training dataset, we used the data augmentation tools provided by [Chi+22]. After converting all images to greyscale, it allowed us to generate three variants of each image with modified contrast and brightness. The data augmentation tools also include a rotational transformation tool, but after some testing I realized that in our case, the rotational augmentations didn't improve the resulting detection of ◎□-origami.

From the 7 annotated images, 5 were used for the training data set and 2 for the validation set of YOLOv5s model, and each data set was augmented by a factor of 3 using the above method. Then, following [Chi+22] method, a YOLOv5s model was trained for 1000 epochs with an image size of 1280 and a batch size of 10, using the default training hyper-parameters.

◎□-origami detection. The trained weights were used to perform inference on all 43 AFM images with various values of confidence threshold and intersection over union (IoU) threshold. Both parameters are taken between 0 and 1, and the IoU parameter is a measure

of the overlap between detection boxes. The numerical value of the IoU threshold represents the maximum possible overlap between two bounding boxes, where the overlap is computed as the ratio between the intersection area and the union area of those bounding boxes. A low IoU threshold helps eliminate multiple detections of the same object.

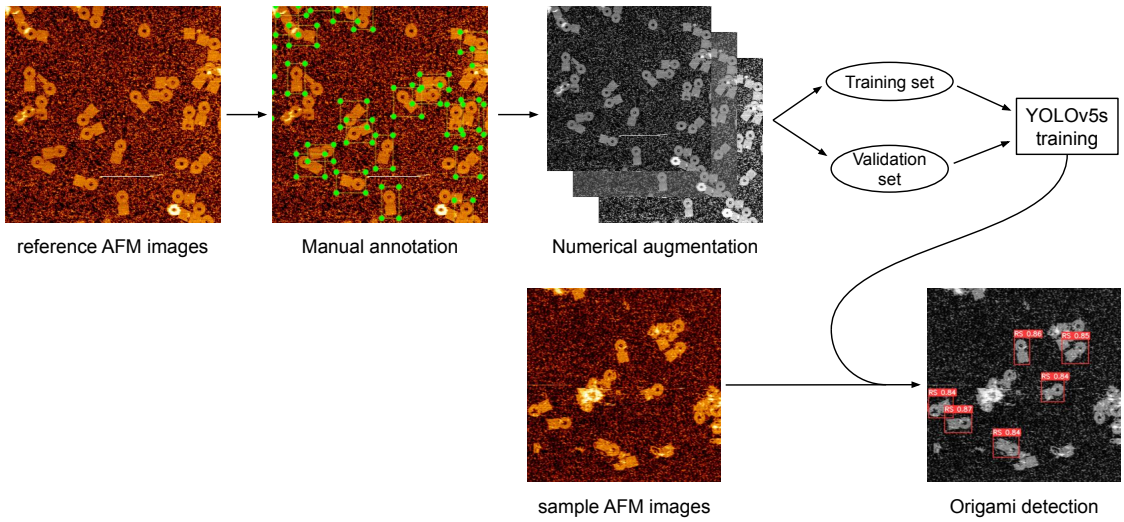


Figure 8.1: Neural network training and origami detection.

8.2 Yield computation

Once the number of correctly folded ◎□-origami is estimated in each AFM image using YOLOv5s model, the yield of a sample is computed as the ratio of well-folded origami surface to the total surface of folded origami (partial/chimeric as well as well-folded ones). Those surfaces are computed using image manipulation libraries.

Starting from a raw AFM image, we first convert it to black and white, using the Otsu algorithm for finding the appropriate threshold. The resulting image often presents some remaining background noise due to excess staples strands on the mica. In order to remove this noise, we use morphological erosion then dilatation of white areas. The erosion step erases white pixels that are at distance 2 or less from a black pixel, and conversely for the dilatation step (see fig. 8.2a). The surface of folded DNA is then simply given by the number of white pixels.

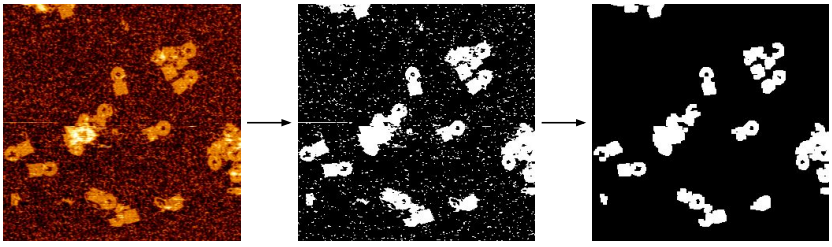
Lastly, the surface of a well-folded ◎□-origami is a constant that we compute using the same method on a reference image presenting 17 perfectly folded ones (see fig. 8.2b). Hence, the estimated yield of a sample is given by:

$$\text{yield} = \frac{n_{◎□} \times S_{\text{ref}}}{S_{\text{sample}} \times 17}$$

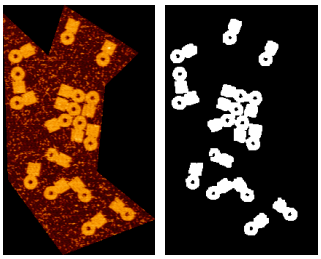
Where $n_{◎□}$ is the number of detected ◎□-origami across all the AFM images of the sample, S_{sample} is the total folded DNA surface from those images and S_{ref} is the folded DNA surface of the reference image containing 17 ◎□-origamis (fig. 8.2b).

8.3 Results and discussion

The yields were computed using the YOLOv5s model with an IoU of 0.3 and a confidence threshold of 0.7 (see table 8.1). In particular, those results show the efficiency of the linear-



(a) Folded DNA surface computation on AFM images



(b) Reference image for computing the surface of a well-folded ⊙□-origami.

Figure 8.2: Surface computation for folded DNA and reference ⊙□-origamis.

optimization-based seed selection method (chapter 7), as it maintains a $> 10\,\%$ yield even with a seed as little as $32\,\%$, while our first greedy algorithm (section 6.2) drops to a $2\,\%$ yield for a $34\,\%$-staples seed.

However, those results and especially the separated yield values should be treated cautiously.

Limitations of the method. Although unbiased and scalable, this yield estimation method presents some limitations. First, in our case, the YOLOv5s model doesn't perform nearly as well as an experienced user and would probably require a bigger and more diverse training set, or a finer tweak of the model parameters. A notable consequence is the high estimated yield of the sample with no seed staples ($11.15\,\%$), even though it doesn't present any well-folded ⊙□-origami. It is due to it particularly small sample size and several false positive predictions that occurred on a large chimeric aggregate. Moreover, we use a method that was first demonstrated for the detection and classification of classical 2D DNA origamis, with a training data set containing almost exclusively perfectly-folded ⊙□-origamis. Hence, it might need some adjustment for being used on a set of partially defectives nanostructures, for example by classifying the detected nanostructures according to their level of imperfection.

A possible manual but unbiased alternative would be to split each AFM images into pieces containing just a few ⊙□-origamis (well-folded or chimeric ones), for example into $\sim 0.5 \times 0.5$ µm squares, then displaying those small images in a random order to a user who will count well-folded ⊙□-origamis in each image, without knowing the source sample. The user could use fractional numbers to take into account defective ⊙□-origamis.

Seed algorithm	Seed size (% of staples)	Images count	Detected ⊙□	Yield
full seed	100 %	7	1636	63.12 %
no seed	0 %	3	12	11.15 %
cb50 seed	50 %	4	56	7.42 %
cb25 seed	25 %	8	1	0.20 %
cov8 seed	47 %	4	245	29.72 %
cov4 seed	34 %	5	4	2.26 %
lp85 seed	46 %	5	285	29.87 %
lp70 seed	32 %	7	32	10.10 %

Table 8.1: Computed yield results for each seed algorithm. The ⊙□-origami detection by YOLOv5s model was performed with an IoU of 0.3 and a confidence threshold of 0.7.

Part IV

Assembling frustrated shapes

Chapter 9

Producing a smaller DNA scaffold strand

9.1 Motivations and supporting works

In order to expand our design possibilities while reducing development time and cost, we wanted to design DNA origamis using a shorter scaffold than the most widely used M13mp18. The production of such scaffold should also be time-and-cost-effective, while using a protocol that doesn't require a lot of wet-lab expertise.

One first thought was to use a portion of the M13mp18 strand. The protocol principle would be to amplify the desired portion of M13mp18 with a polymerase chain reaction (PCR), then to remove one strand from the double-stranded PCR product using an exonuclease. This would require to add a biotin molecule at the $5'$ end of one of the primers, in order to 'protect' the corresponding strand by preventing the exonuclease (e.g. T7 or Lambda exonuclease), that degrades in a $5' \to 3'$ direction, from initiating the digestion of this strand. However, we searched for an even simpler method, that would directly produce a circular scaffold strand instead of a linear one (linear scaffold strands are much less common in the literature, and it is unclear to us whether they are as effective as circular ones).

We ultimately decided to use a (double-stranded) pUC19 plasmid that we made single-stranded via selective single-strand digestion, taking inspiration from [Erk+14] and [Dun+15]. Indeed, [Erk+14] provides a detailed 5-step method for preparing a single-stranded circular DNA scaffold of chosen size from a double-stranded M13mp18 plasmid (see fig. 9.1). In our case, we only used the two final steps, starting from the already short enough pUC19 plasmid (see fig. 9.2). pUC19 is a 2686-base-pair plasmid, commonly used as a cloning vector, and utilized by [Dun+15] in a slightly modified variant as a DNA origami scaffold.

We elaborated this protocol with the help of Florence Lormières, from the Biology Department of ENS de Lyon.

9.2 Protocol elaboration

As shown in fig. 9.2, in order to obtain a single-stranded pUC19 DNA scaffold (sspUC19) we first need to cleave one strand of the plasmid then use an exonuclease to digest this strand.

For the first step, we need a nicking endonuclease that cleaves only one strand of DNA on a double-stranded DNA substrate. Its nicking site must appear at least once on the pUC19 plasmid and in one orientation only, so that exactly one strand is cleaved (possibly in several places). Therefore, we chose the Nt.BspQI enzyme (an engineered BspQI variant), as its

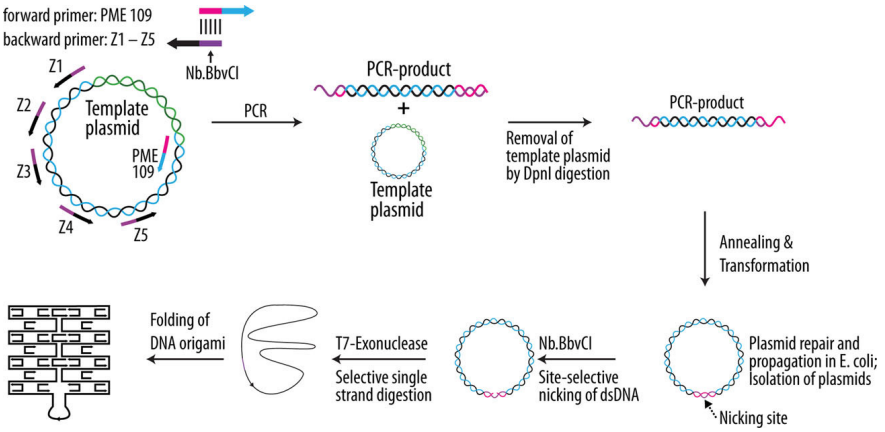


Figure 9.1: Origami scaffold preparation from [Erk+14]

nicking site appears exactly once along the pUC19 plasmid.

For the second step, we can use the same exonuclease as in [Erk+14]: the T7 exonuclease, a double-stranded specific exonuclease that initiates at the 5′ terminus of a DNA strand and digest this strand in the 5′ to 3′ direction. However, as the T7 doesn't use the same buffer as Nt.BspQI, we decided to experiment with both buffer and compare the results at each step of the process, using gel electrophoresis (see fig. 9.3). This allowed us to validate the protocol with the use of NEB4 buffer, and we were able to further corroborate this completion by folding the produced sspUC19 into two simple rectangle DNA origamis, differing in their scaffold routing and staple design (see fig. 9.4).

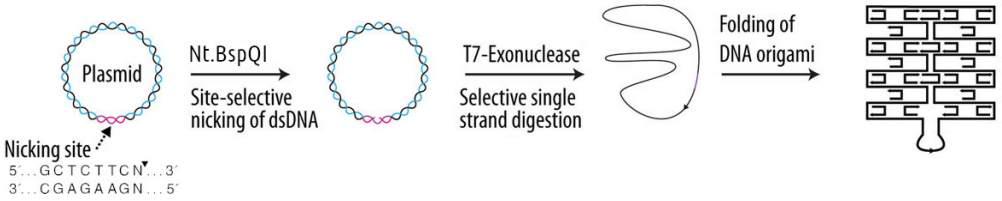


Figure 9.2: Origami scaffold preparation inspired from [Erk+14]

After some optimizations for the incubation durations and the final purification step, we ended up with this protocol:

1. Prepare a 100 µL mix of 0.25 mg/mL pUC19 and 500 units/mL Nt.BspQI in NEBuffer4,
2. Incubate at 50°C for 2 hours,
3. Add 5 µL of T7 exonuclease at 10,000 units/mL,
4. Incubate at 25°C for 1 hour,
5. Add 6 µL proteinase K 8 units/mL (100-fold dilution of the stock, 800 units/mL proteinase K),
6. Incubate at 37°C for 30 minutes,
7. Incubate at 80°C for 20 minutes,
8. Use a Monarch RNA Cleanup Kit to purify the produced sspUC19, using 25µL of nuclease-free water in the final elution step.

We used a NanoDrop™ spectrophotometer to quantify the resulting concentration of sspUC19, which was of 110.5 ng/µL.

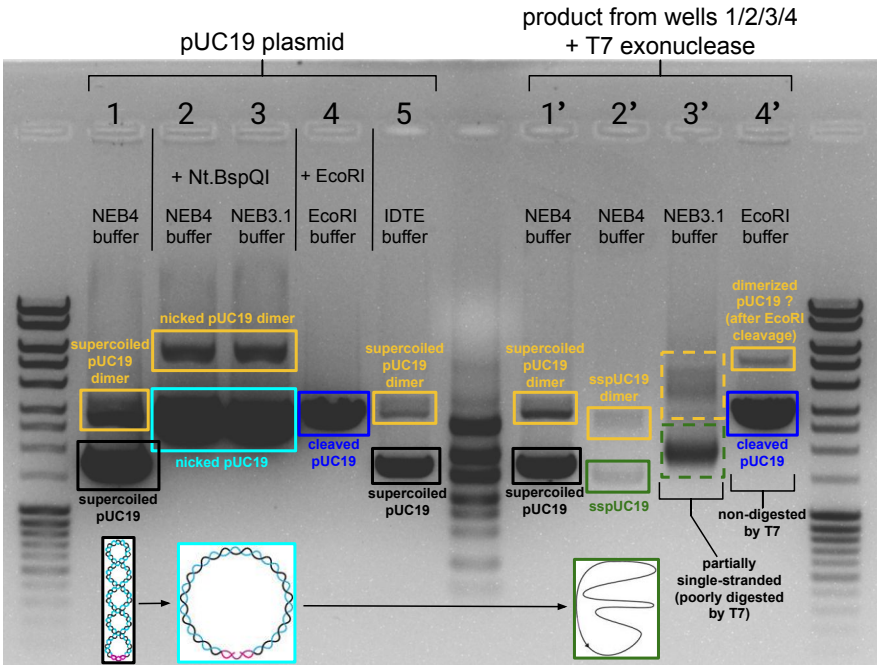


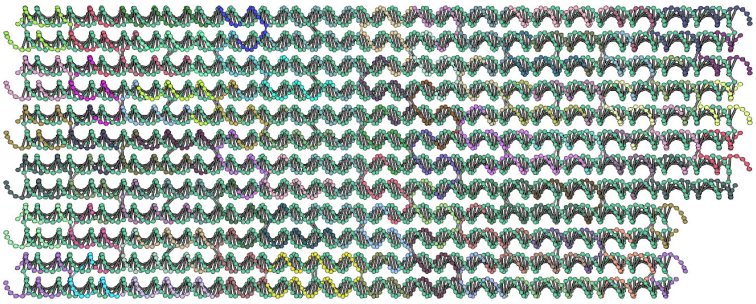
Figure 9.3: Gel electrophoresis at each step of the production of single-stranded pUC19 plasmid. Wells 1 to 5 contain pUC19 plasmids that have been incubated at 50°C for 18 hours, in different buffers and without enzyme (wells 1 and 5), with Nt.BspQI nicking enzyme (wells 2 and 3) or with EcoRI cleaving enzyme (well 4). Wells 1′ to 4′ contains the product of this first incubation, after an additional incubation at 25°C for 22 hours with T7 exonuclease. Our interpretation of the electrophoresis result is the following : The pUC19 plasmid is initially in a supercoiled form, with a minority in a dimeric form (wells 1 and 5), and Nt.BspQI successfully nicked both forms in NEB4 and NEB3.1 buffers (wells 2 and 3). Then, the T7 exonuclease successfully digested the cleaved strand in NEB4 buffer (well 2′), but only partially in NEB3.1 buffer (well 3′). Therefore, the NEB4 buffer seems adequate for our usage.

9.3 Final validation of the method: two DNA origamis

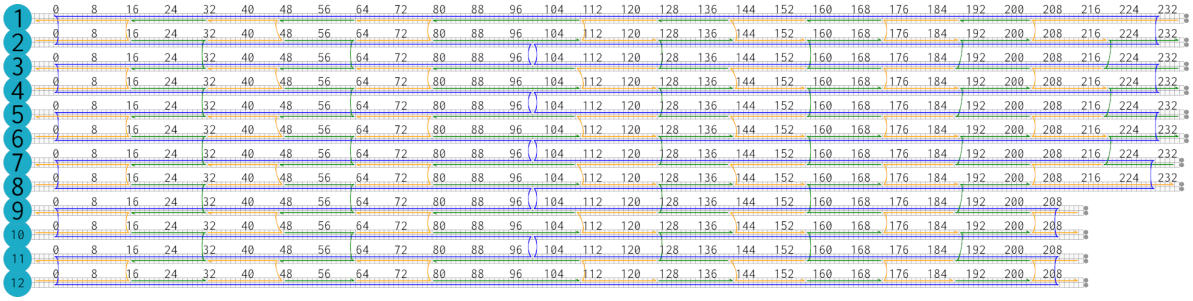
In order to fully validate our scaffold strand producing method, I designed a simple flat rectangle DNA origami in ENSnano [LS21] using the sspUC19 scaffold. It is a 77×31 nm rectangle with a 10×7 nm corner notch. Two versions of the origami were designed and assembled:

- A first variant with 92 staples (mostly 'C' shaped) and a central seam on the scaffold routing (fig. 9.4b),
- A second variant with 87 staples (mostly 'S' shaped) and three seams on the scaffold routing (fig. 9.4c).

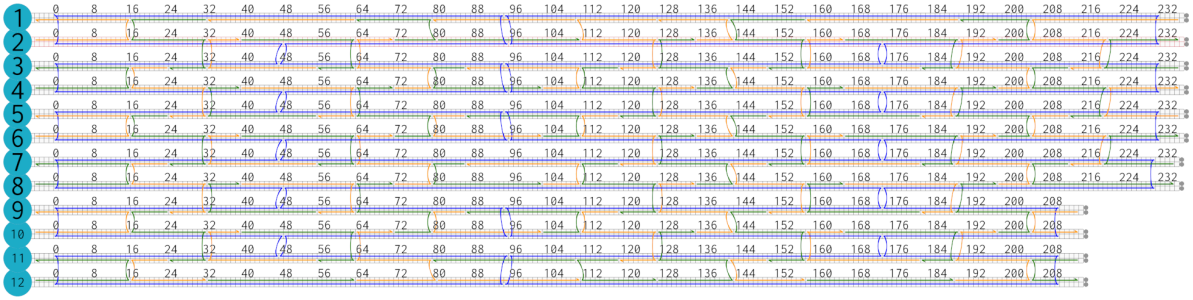
Both versions were prepared with 58 nM of scaffold strand and 200nM of staple strands and annealed as follows: 90°C for 5 min, from 90°C to 70°C at the rate of -1°C every minute, then from 70°C to 40°C at the rate of -1°C every 5 minutes, then from 40°C to 20°C at the rate of -1°C every minute. The resulting AFM images show the successfully folding of both design, proving the correctness of our scaffold strand producing method (see figs. 9.4d and 9.4e).



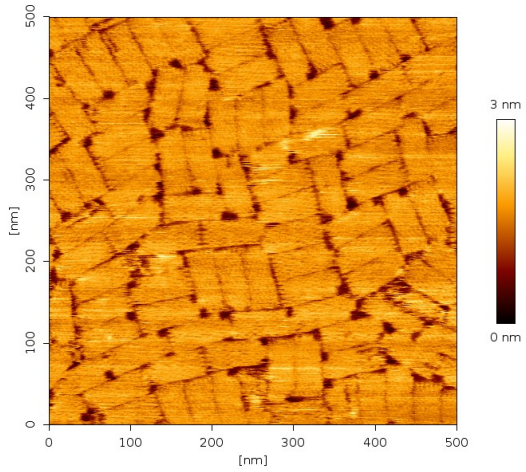
(a) ENSnano 3D view (second variant).



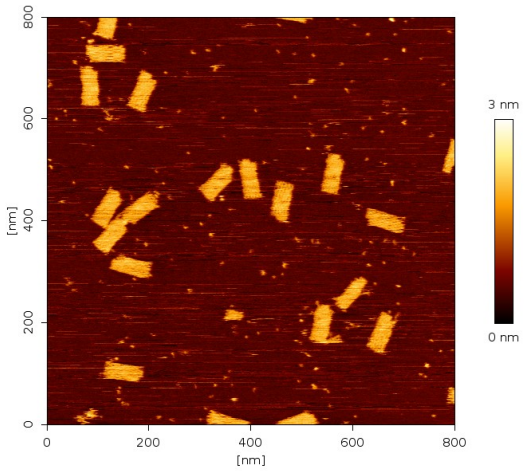
(b) First variant with ‘C’ shaped staples and one central seam.



(c) Second variant with ‘S’ shaped staples and three seams.



(d) AFM image of the first origami variant.



(e) AFM image of the second origami variant.

Figure 9.4: A DNA origami using sspUC19. The second DNA origami sample (e) was further diluted on the mica after we observed a saturation of the mica surface with the first origami sample (d).

Chapter 10

A geometrically frustrated DNA origami

10.1 Principle and design ideas

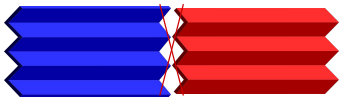
We want to design a bi-origami that fully takes advantage of our two-step folding method. It relies on the flexibility induced by the incompleteness of the pre-folding step in order to correctly self-assemble in the second annealing step. More specifically, it is a frustrated shape that shouldn't be able to assemble correctly if the two scaffold strands were already fully folded.

The goal of this design is not to provide an additional example for our seed strand selection method, but rather to show that the two-step assembly process we presented in chapter 6, that includes a partial folding step, can be particularly useful for some bi-origami designs. In this case, our design should be able to fold correctly only if each scaffold strand is first partially folded. Similarly to our experiments in chapters 6 and 7, it is important to fold each scaffold separately (at least partially), with a large enough set of staple strands, in order to avoid the formation of defective or chimeric structures during the final assembly step. But the main difference here is that if too many staple strands are used during this first pre-folding step and if the scaffold strands are nearly entirely folded, then the halves of the bi-origami design won't be able to correctly attach to each other because of mechanical/geometrical constraints that prevent them from reaching the right conformation. Therefore, one has to find the right set of seed staple strands in order to sufficiently pre-fold each half of the design while avoiding introducing too much stiffness to the structure.

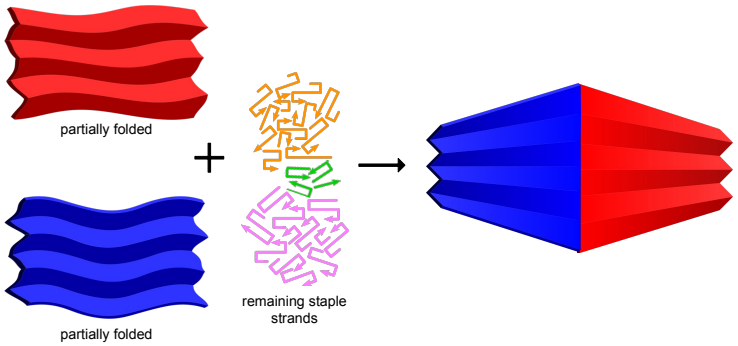
The first rough outline that we had for this design is the shape of two paper fans that can bind to each other when open, but whose mechanically stable conformation (when separated) is the close one (see fig. 10.1).

10.2 Corrugated design

We use the sspUC19 scaffold for this design (see chapter 9). Double helices are arranged into a corrugated shape, and we use a honeycomb lattice as framework in order to facilitate a precise placement of crossovers between double helices (as strands on a DNA double helix are roughly 120° apart around the helix axis). The resulting design is presented in fig. 10.2. We call this design *corrugami*, with 1-corrugami being any of the single-scaffold halves of the design (thus a classical DNA origami). The two 1-corrugamis can bind using 4 link staples at the bounding sites indicated in green in fig. 10.2, forming a 2-corrugami that presents



(a) The two halves can't bind if fully formed because of a misalignment.



(b) The two halves can bind in two steps.

Figure 10.1: Paper fans bi-origami design sketch.

an open conformation (fig. 10.2b). Other double helix ends are equipped with 4-Thymine tails in order to avoid any stacking interactions between blunt-ended helices [Rot06]. Both 1-corrugamis of the design have exactly the same shape and staple positioning, and they only differ by their scaffold rotation, and thus by the staple sequences.

We considered two different approaches for selection the seed staples strands. The first thermodynamic approach uses our method presented in chapter 7 for finding key differentiating strands of each 1-corrugami. The second approach is a simple, geometrical one, as it simply selects the staple strands that are far enough from the junction between both scaffold strands. This should allow enough structural flexibility in the binding region of each 1-corrugami in order to ensure the correct placement of binding staples, thus the apparition of the frustrated open conformation (fig. 10.2b).

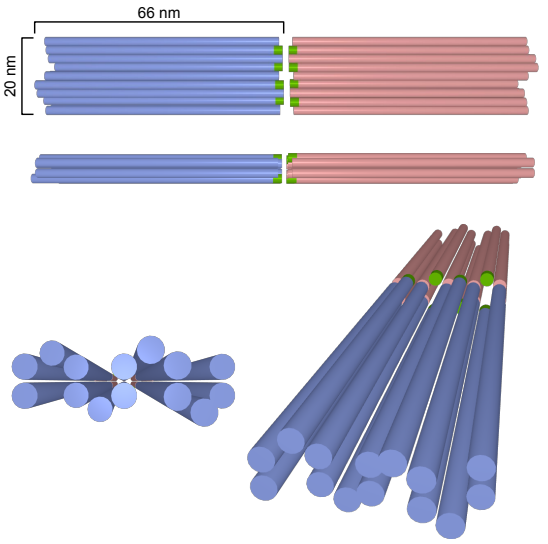
10.3 Experimental results

Experimental challenges. This design can't be properly imaged using atomic force microscopy, as it systematically becomes flat on the mica surface. Moreover, using gel electrophoresis for recognizing well-folded 2-corrugamis is challenging, as two non-open 1-corrugamis can still potentially be bound together. Indeed, fully-formed closed 1-corrugamis could attach to two binding sites (instead of four) without altering their shape. In this case, it might not be possible to distinguish them from a well-folded open 2-corrugami by just looking at the gel, due to their similar shape and size. In order to avoid this issue, one could try to precisely tweak the design parameters (e.g. the length of the link staple strands) and the experimental settings (buffer, concentration of staples and sspUC19), so that two binding sites are insufficient for irreversibly attaching two 1-corrugamis, while the target open configuration of the 2-corrugami (that uses the four binding sites) remains stable enough.

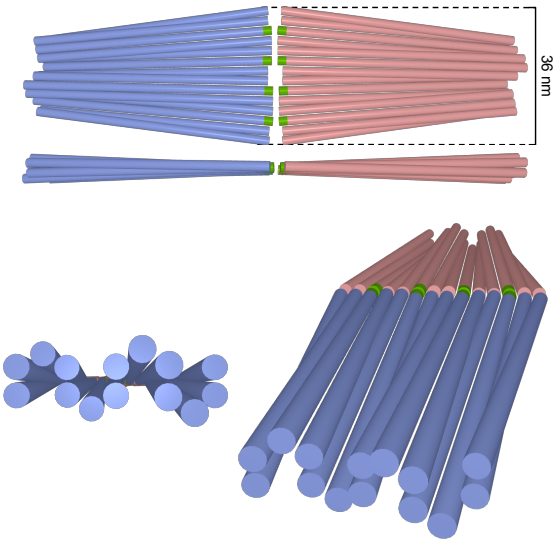
Results and conclusion. We started by folding and imaging a 1-corrugami using Transmission Electron Microscopy in order to confirm its corrugated shape (see fig. 10.4). But after a few attempts at adjusting the experimental conditions (Mg concentrations, and annealing temperature) and testing several seeds, we weren't able to determine the right set of parameters by only using gel electrophoresis and AFM. However, the cost, duration and poor resolution (that I managed to obtain) of TEM made it unsuitable to use for testing all those parameters. In the end, we still managed to show the correct corrugated shape and aspect ratio of the 1-corrugami, and it was the opportunity for me to take theoretical and practical training in transmission electron microscopy. Moreover, the flattened assemblies we

observed by AFM still support our two-step folding and seed selection method (fig. 10.3c).

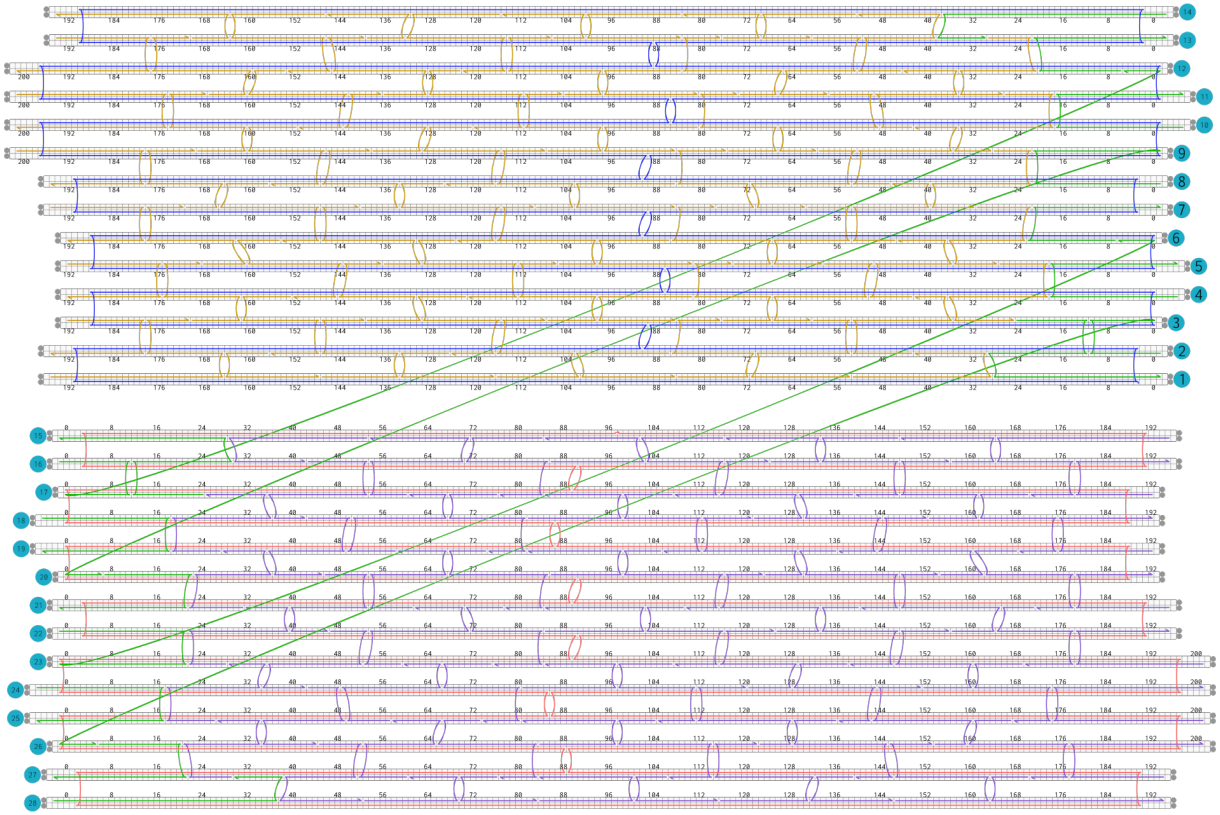
We ultimately decided to leave this project aside as controlling geometrical and mechanical constraints revealed itself to be more difficult than expected. For a future iteration, one could choose a design that uses topological constraints (for example, taking inspiration from [Mar+16]) instead of mechanical and geometrical ones, as they seem easier to control and less dependent on experimental conditions.



(a) Closed conformation.

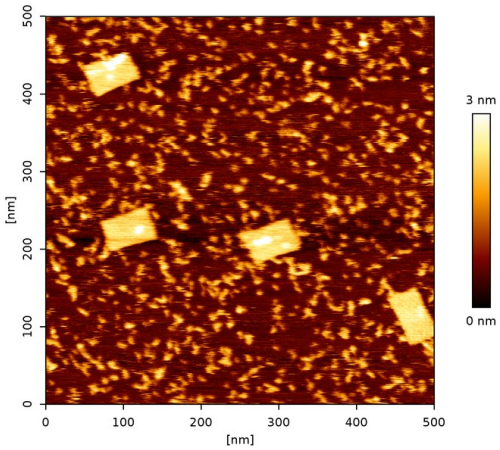


(b) Open conformation

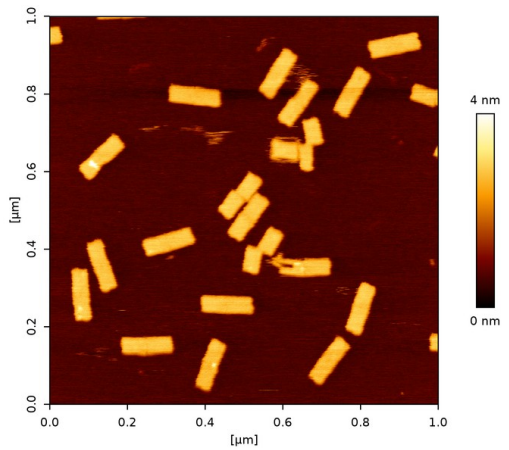


(c) Detailed representation of scaffold routing and staple strands.

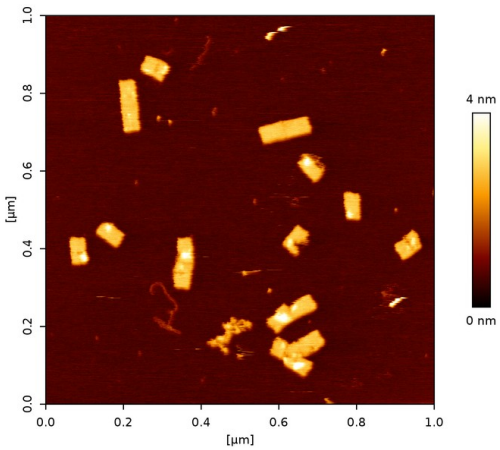
Figure 10.2: Corrugami design. (a) and (b): Top, side, front and diagonal views of the design, on the closed (a) and open (b) conformations. Green ends indicate the bounding sites between the two halves of the design.



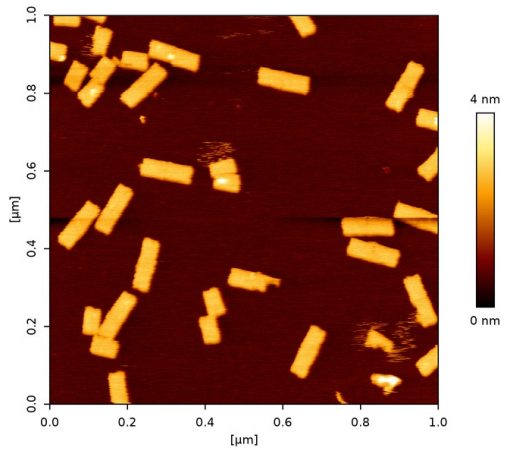
(a) 1-corrugami.



(b) 2-corrugami, using all staples as seed.



(c) 2-corrugami using our LP strand selection method for the seed (48 % of staples).



(d) 2-corrugami, using staple strands that are far enough from the junction as seed (84 % of staples).

Figure 10.3: AFM images of the corrugami design. (a) Folding conditions: 14 nM of sspUC19 scaffold together with a 5-fold excess of staples. Annealing (∼2.5 hours) : 90°C for 5 min, from 90°C to 65°C at the rate of −1°C every 30 seconds, from 65°C to 45°C at the rate of −1°C every 5 minutes, from 45°C to 25°C at the rate of −1°C every minute.
(b-d) 2-corrugamis were assembled in two steps. First annealing (20 nM of sspUC19 and a 5-fold excess of seed staples, ∼3.5 hours): 90°C for 5 min, from 90°C to 65°C at the rate of −1°C every 2 minutes, then from 65°C to 45°C at the rate of −1°C every 5 minutes, then from 45°C to 25°C at the rate of −1°C every 2 minutes. Second annealing (mixing the results of the first annealing and adding a 5-fold excess of all staples): 51°C for 30 min, then from 51°C to 20°C at the rate of −1°C every 3 minutes.
One can notice that the corrugami flattens on the mica surface as these images show no corrugation on the DNA origami structures and the 1-corrugamis in (a) present an aspect ratio below 2:1 while the corrugated conformation should yield an aspect ratio above 3:1 (66 to 20 nm).

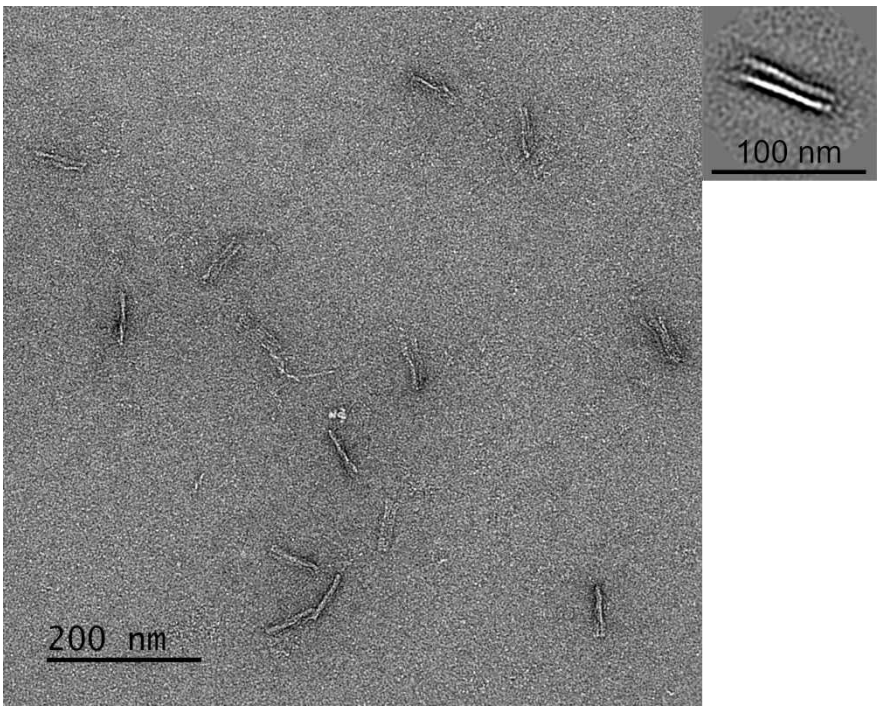


Figure 10.4: TEM image of the 1-corrugami and 2D class average. Folding conditions: 40 nM of sspUC19 scaffold together with a 10-fold excess of staples in a 5 mM Tris-HCl, pH=8.0, 1 mM EDTA, 18 mM $MgCl_2$ buffer. Annealing ($\sim$2.5 hours) : 90°C for 5 min, from 90°C to 70°C at the rate of -1°C every minute, from 70°C to 40°C at the rate of -1°C every 5 minutes, from 40°C to 10°C at the rate of -1°C every minute.
Gel-purified origamis were visualized by adsorption onto glow-discharged carbon-coated grid (Quantifoil Micro tools GmbH, Germany), stained for 60 s with a 2 % (w/vol) aqueous uranyl acetate (Merck, France) solution, and then dried with ashless filter paper (VWR, France). Observations of EM grids were carried out on a JEM-1400Flash Tungsten microscope working at 120 kV, equipped with a GATAN One View camera. Two-dimensional class averages were computed using EMAN2 software. Images were taken by Julie Finkel, at the *Centre de Biologie Structurale* in Montpellier.

Part V

Mold-assisted differentiation

Chapter 11

A shape-sorter-inspired DNA origami

In this chapter, we slightly diverge from multi-scaffold DNA origami design in order to focus on how to co-fold two distinct DNA origamis using identical scaffold strand. To this end, we thought of using an additional helper origami that should induce the differentiation of each scaffold into the desired shape. We also decided to return to a flat, AFM-compatible design, as this microscopy technique is easier and more available to us that TEM.

11.1 Principle and design options

After relying on a seed of key differentiating strands to fold a double scaffold origami (part III), we decided to explore another approach: using an additional origami as a mold for inducing the differentiation of two initially identical scaffold strands.



Figure 11.1: A shape-sorter toy.

I created a simple flat design, with two little DNA origamis, the *mini-ring* and the *mini-square*, together with a M13pm18-based rectangular mold origami. The mini-ring and mini-square use the sspUC19 scaffold strand (see chapter 9) shaped as reduced versions of the ◎□-origami components (chapter 5), while the mold is a M13mp18-based origami evoking a shape sorting toy (see fig. 11.2). The goal of this design is to explore several ways of co-folding the mini-ring and mini-square DNA origamis using this mold origami, and potentially being able to detach later those small origamis from the mold using toehold mediated strand displacement. Indeed, I designed those DNA origamis with two adjustable tools for the scaffold differentiation:

- Using a chosen number of single-stranded overhang of the mold (in light blue in fig. 11.2) that can attach to an sspUC19 segment and place it at the correct location along the mold. Those sspUC19-complementary overhangs are in fact prolonged with a 10-nucleotides toehold (highly orthogonal to all other sequences in use), in addition to their 8-nucleotides sspUC19-binding domain, in order to be able to detach them from a fully formed mini-ring or mini-square using strand displacement.
- Using stacking interactions and shape complementarity to encourage the folding into either of the two correct shapes. However, this option may not suffice on its own, and it is incompatible with the release of the mini-ring and mini-square origamis by strand displacement.

To this end, we ordered staple strand so that it is possible to adjust the number and positions of sspUC19-binding overhang of the mold from 0 to 9 per hole. And each double helix end

of the mold holes can be chosen to be blunt-ended or equipped with a 4-Thymine tail.

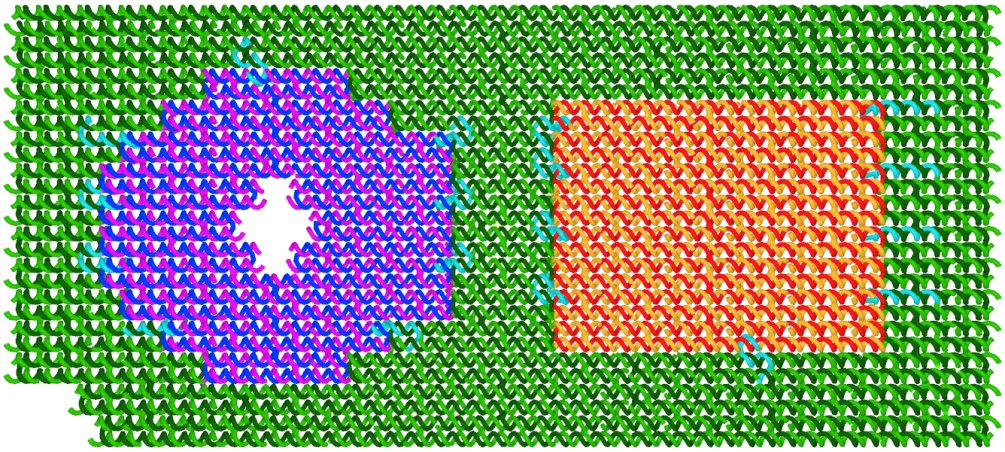
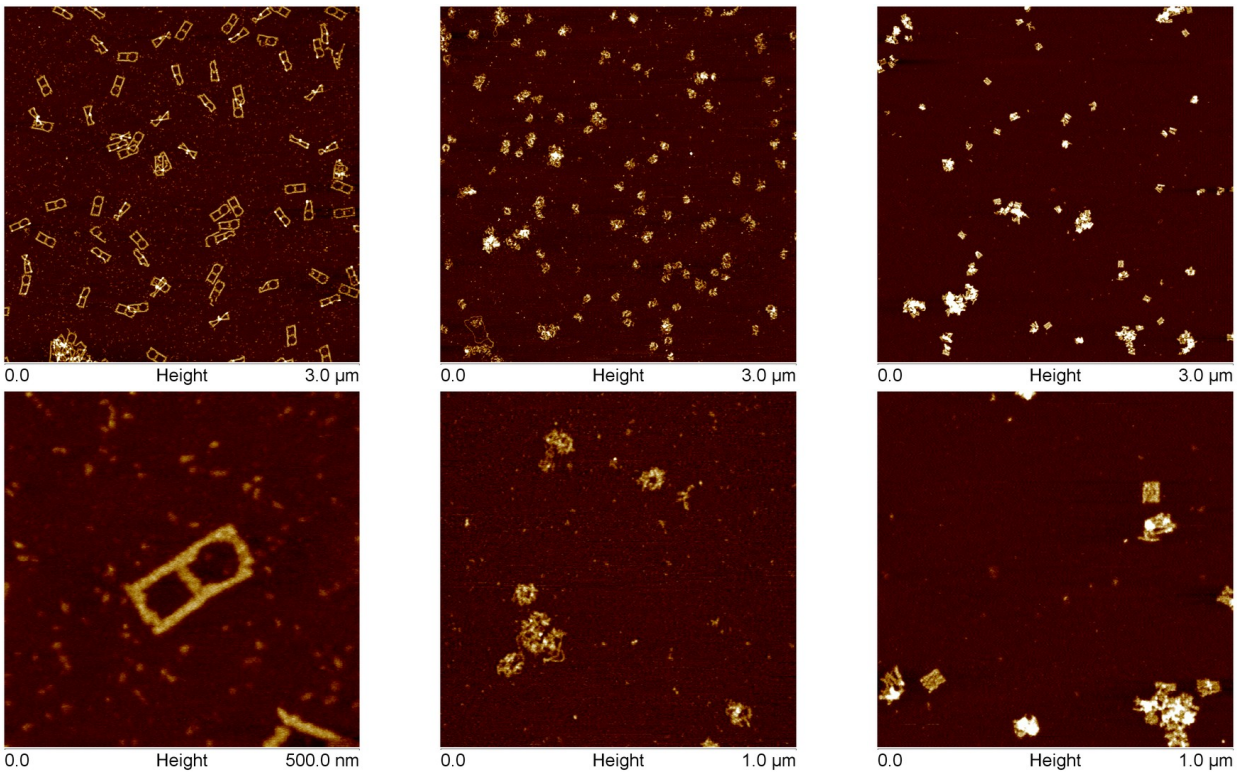


Figure 11.2: Mold, mini-ring and mini-square origami design. The mold origami is composed of an M13pm18 (7249 nucleotides) scaffold strand (in dark green) folded using 254 staple strands (light green) plus 18 additional link staple strands (light blue) that act as binding sites for the mini-ring and mini-square origamis. The mini-ring is composed of an sspUC19 (2686 nucleotides) scaffold strand (dark blue) folded by 86 staple strands, and the mini-square uses an identical sspUC19 scaffold strand (red) together with 87 staple strands (orange).

11.2 Results and discussion

As a preliminary validation, we first attempted to fold each of the three DNA origamis separately. However, although the mold origami self-assembled properly, neither the mini-ring nor the mini-square folded correctly, despite all our effort (see fig. 11.3). This seemed quite puzzling to us, as the design method (i.e. scaffold routing and staple strand placement) for a flat origami is a simple and proven approach that worked on all our previous designs. In order to rule out any protocol or wet-lab error, I tested several annealing curves (slow, fast and constant temperature) that usually all yield to correctly folded origamis for this kind of design; I tried using a new staple mix extracted from the 96 wells plates we ordered; and I used a pool of sspUC19 that had already been used with success on previous designs, but also attempted with a newly synthesized one. As I didn't manage to determine the source of the issue and in addition to some timing and AFM issues, those complications led us to put this project on hold.



(a) Mold origami **(b)** Mini-ring origami **(c)** Mini-square origami

Figure 11.3: AFM images of each component of the design. Folding conditions: 10 nM of scaffold strand (either M13mp18 or sspUC19) together with a 5-fold excess of staples. Annealing: (a) 90°C for 15 min, from 90°C to 65°C at the rate of −1°C every 2 minutes, then from 65°C to 45°C at the rate of −1°C every 30 minutes, then from 45°C to 20°C at the rate of −1°C every 2 minutes. (b-c) 51°C for 30 min then from 50°C to 20°C at the rate of −1°C every 3 minutes.
As our usual AFM was unavailable at that time, samples were rinsed, dried then imaged on a Bruker Multimode 8 AFM, using ScanAsyst-Air-HR probes. AFM imaging was performed by Cendrine Moskalenko, from the Physics Lab of ENS de Lyon.
Unfortunately, we weren't able to correctly fold the mini-ring and mini-square origamis (even by varying experimental conditions, notably by using the same annealing ramp as the mold origami), nor to diagnose the origin of this problem.

Part VI

Simulation of DNA origami self-assembly

Chapter 12

Understanding DNA origami self-assembly

The details of the folding process of a DNA origami are currently poorly understood, as well as the reasons why this technique is so efficient and reliable. During the design process, the staple strand sequences are chosen so that each staple domain is complementary to its corresponding scaffold domain, but nothing prevents this complementary sequence to appear at several positions along the scaffold sequence. This possibility is amplified by the viral (or bacterial) origin of most origami scaffold strands, as it presents a lot more regularity than a random sequence. For example, the M13mp18 scaffold contains repeated sequences up to 42 nucleotides and 20 % of its 8-nucleotides subdomains are non-unique (instead of ∼10 % for a random sequence of the same length), some of them being repeated up to 8 times. Also, both scaffold strands and staple strands can present secondary structures: non-intersecting complementary domains within a strand can bind to each other at a low entropy cost.

One can reckon that the geometry of a forming DNA origami can lead neighboring staple strands to cooperate for folding the scaffold strand into the designed shape, depicting a kind of nucleation phenomenon, and the designed origami structure is generally conjectured to be the most thermodynamically stable (i.e. the global minimum of the system's Gibbs free energy landscape). However, as we observed previously during our folding attempts for our bi-origami design (chapter 5), this nucleation phenomenon and conjectured stability are not always enough to correctly fold the desired structure when it deviates from classical origami designs. Therefore, we wanted to better understand the process of DNA origami formation, and the detailed succession of events that leads to the correct or defective folding of the nanostructure. To this end, we explored ways of developing a stochastic model and simulation of DNA origami folding. This chapter 12 aims at understanding the various aspects of DNA origami folding and its modeling, through a literature review. Then, I present the model I worked on in chapter 13, together with its results, limitations and possible refinements.

12.1 Thermodynamics foundations

Three simple principles are often at the base of DNA thermodynamics intuition when designing DNA nanostructures:

- Double-stranded DNA separates into single-stranded at high enough temperature,
- Cytosine / guanine base-pairing is generally stronger than adenine / thymine base-pairing due to an additional hydrogen bond between each other,

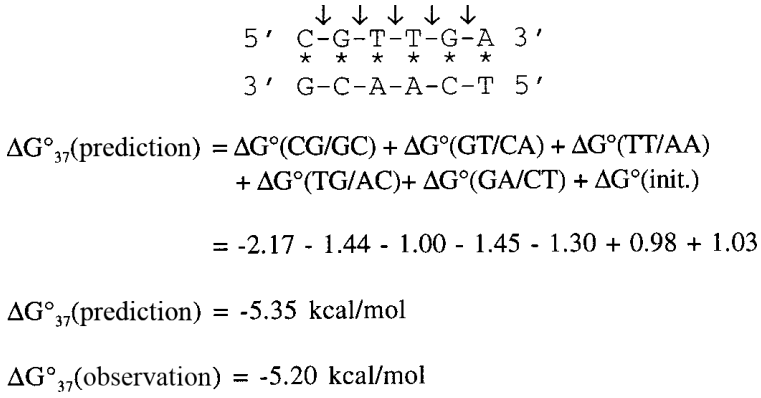


Figure 12.1: Application of the unified nearest neighbor model, from [San98]. Nearest neighbor dimer duplexes are represented with a slash separating strands in antiparallel orientation (e.g. GT/CA means 5′-GT-3′ Watson–Crick base-paired with 3′-CA-5′). ΔG°(init.) represents an initiation parameter, which depend on terminal base pairs of the duplex (here C·G on the left and A·T on the right). Here $T = 37°$C, hence the "37" indices everywhere.

- Salt concentration enhance the stability of DNA nanostructures, as the negatively charged DNA strands interact with positive ions (typically Na^+ or Mg^{2+}).

Those principles are quantified in the nearest-neighbor thermodynamic model presented in [San98]. It allows to compute an estimation of the free energy ΔG° of forming a DNA double-helix from two complementary single strands. This model bases the computation of ΔG° on the interaction between neighboring base pairs along the double helix. Using a set of experimentally determined values for ΔH°_{37} and ΔS°_{37} (respectively the standard enthalpy change and standard entropy change at 37°C) for each couple of neighboring base pairs, one can estimate $\Delta G^\circ_{\mathrm{T}}$ at a chosen temperature T using the equation:

$$\Delta G^\circ_{\mathrm{T}} = \Delta H^\circ_{\mathrm{T}} - T\Delta S^\circ_{\mathrm{T}} \approx \Delta H^\circ_{37} - T\Delta S^\circ_{37} \tag{12.1}$$

This approximation is reasonable for the range of temperatures that we consider as they don't deviate too much from 37°C. Indeed, $\Delta H^\circ_{\mathrm{T}}$ and $\Delta S^\circ_{\mathrm{T}}$ are nearly temperature independent and eq. (12.1) remain quite accurate due to compensating errors [San98; SH04]. Once $\Delta G^\circ_{\mathrm{T}}$ is known for any base pair dimer, it can be computed for the associations of two complementary DNA strands, as illustrated in fig. 12.1.

An empirical salt correction [San98; FPP20] can be applied to the value of ΔS° in eq. (12.1), in order to account for conditions that deviate from the default (1 M of $[NA^+]$). This correction is given in cal.K^{-1}.mol^{-1} (with n_{pb} being the total number of base pairs in the duplex):

$$\Delta S^\circ = \Delta S^\circ\left([\mathrm{Na}^+] = 1\mathrm{M}\right) + 0.368 n_{\mathrm{pb}} \ln\left([\mathrm{Na}^+] + [\mathrm{K}^+] + [\mathrm{NH}_4^+] + 3.3[\mathrm{Mg}^{2+}]^{1/2}\right)$$

One can notice that ΔS° is an increasing function of ions concentrations, hence with eq. (12.1) it illustrates that salt concentration increases the stability of DNA Watson-Crick-Franklin binding while temperature decreases it.

12.2 Nucleotide level and domain level models

Studies on DNA thermodynamics such as SantaLucia's work primarily provide a static model that can be used to describe the equilibrium state of a system composed of interacting DNA strand species. In order to understand the dynamics of such system and the sequence of

events during the folding of a DNA origami, one needs to develop a kinetic model that can capture the multiple binding and unbinding reactions that occur during this folding.

While atomic resolution models can be used to predict the conformations of a fully folded DNA origami [MA20; MYA16], they typically start from an already formed DNA nanostructure that presents the exact designed 3D conformation (e.g. given by the CAD software used for its design) and simulates its conformational changes for a few nanoseconds to a few hundreds of nanoseconds and without breaking nor creating any base pair. Modelling the whole folding process of a DNA origami has to be done at the nucleotide scale or even the domains level, in the hope of being able to simulate a reaction that can take minutes, or even hours.

Nucleotide level models probably are the most common ones for studying DNA nanostructures, as they allow to capture nucleation-zipper and strand displacement mechanisms. One example of such model is presented in [MKT21]. This model uses a kinetic Monte Carlo simulation to describe the association and dissociation mechanisms of short DNA strands. It simulates the evolution of a system through a set of possible states together with possible transitions. Here, the system is a set of short DNA strands and a state is defined by the set of base pairs. Transitions between states correspond to adding or removing a base pair.

Each possible transition is associated to the kinetic rate of the corresponding chemical reaction and, in the Monte Carlo simulation, this rate is proportional to the probability of applying the associated transition to the system. Forward transition rates (i.e. for forming a new base pair) are computed using the theory of diffusion-controlled reactions for bimolecular association [CD83] and an experimentally fitted simplified model (based on the theory of relaxation kinetics [CCD71]) for the propagation of base pairs association (i.e. forming a base pair consecutively to an already existing one). Backward transition rate can then be computed for each reaction, using an estimate of ΔG and two expressions of the equilibrium constant K, linking it to both the kinetic rates and ΔG:

$$\frac{k_{\text{forward}}}{k_{\text{backward}}} = K = e^{\frac{-\Delta G}{RT}} \quad \text{hence} \quad k_{\text{backward}} = k_{\text{forward}} \times e^{\frac{\Delta G}{RT}} \tag{12.2}$$

An overview of this model, is given in fig. 12.2, applied to two 5′-CATATG-3′ strands.

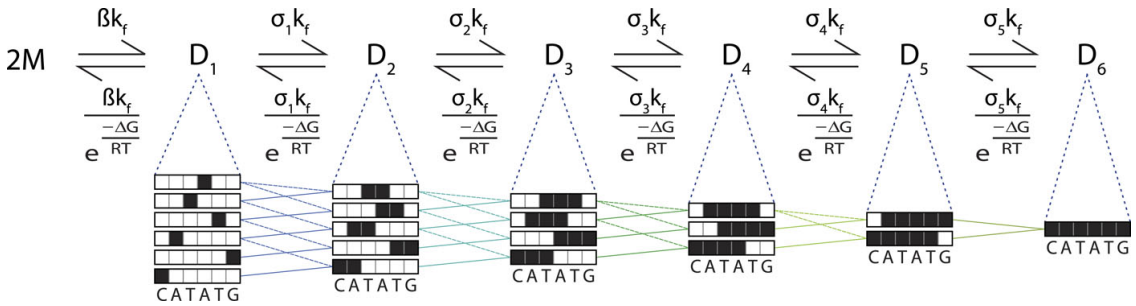


Figure 12.2: Kinetic model reaction scheme from [MKT21], for two 5′-CATATG-3′ strands. Each allowed configuration is represented by a row of six base pairs that are either intact (black) or broken (white), and the lines between states denote the allowed moves for transitioning between those configurations. Here, the model only allows states were the two DNA strand domains are properly aligned, and the intact base pairs form an uninterrupted sequence.

This kind of model is used in [MKT21] to study each phase of the nucleation-zipper mechanism in DNA association: two complementary DNA strands first bind with a single base pair, then enter a pre-equilibrium phase consisting of a rapid succession of metastable

states that can easily revert to the dissociated state, and if the system forms a "critical nucleus" of intact base pairs, it becomes thermodynamically favored to rapidly zip up the remaining bases. More generally, it can be used to explore the secondary structure energy landscape for sets of multiple interacting DNA strands, as in the simulation tool *Multistrand* [STW15].

This type of approach can also capture strand displacement and branch migration mechanisms, which are conjectured to be an essential part of DNA origami folding, especially for removing unwanted domain bindings or secondary structures [Rot06]. However, using such a nucleotide level model for studying an entire DNA origami seems unrealistic. Indeed, for each domain binding reaction, the rapid succession of states in the pre-equilibrium phase, along with its oscillation with the dissociated state, contains a long sequence of extremely short-lived states, which is too costly to simulate in the context of an origami folding reaction that lasts on the order of minutes (at least).

Domain level models can solve this limitation by proposing state transitions that directly attach or detach entire complementary domains. This is what suggests Dannenberg et al. In [Dan+15], they present a simulation model of DNA origami folding whose only state transitions correspond to attaching or detaching a half-staple domain (see fig. 12.3). Their model is specifically designed to study cooperative interactions between staples and the effect or experimental conditions changes on the folding of a simple two-dimensional DNA origami that only contains one- or two-domains staples (i.e. staple strands that only bind to one or two sites of the scaffold).

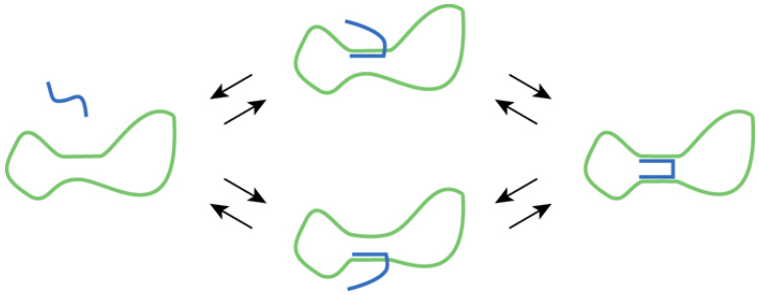


Figure 12.3: Transitions in the domain level model from [Dan+15]. A two-domains staple strand has four possible bonding configurations, with two possible bi-molecular reactions for the first binding.

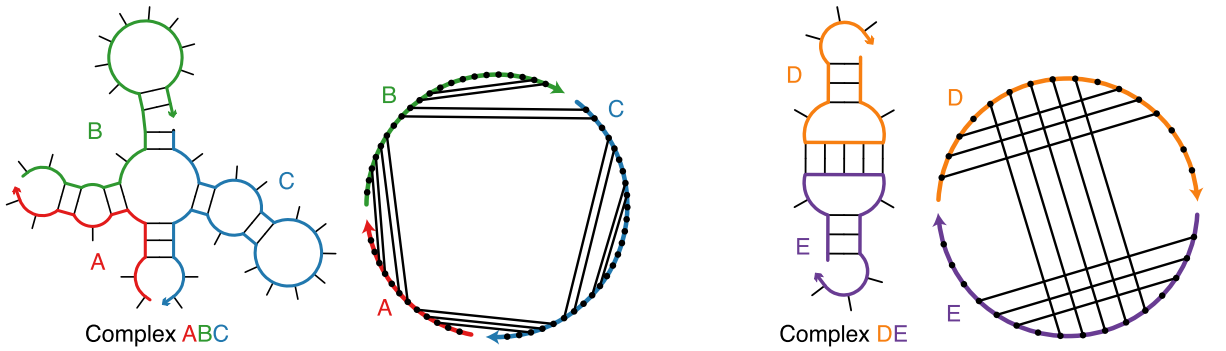
However, only allowing bindings between pre-defined domains given by the origami design can lead to miss a lot of metastable conformations, which are critical to capture chimeric assembly. This isn't thus the right option for us. Moreover, even though the article takes into account inter-strand stacking interactions and the current shape of the forming DNA origami in the estimation of the Gibbs free energy and kinetic rates, the validity of those estimations is limited to very simple 2D origami designs.

12.3 Topological and geometrical considerations

As presented in section 12.1, simple thermodynamic models for bimolecular reactions of complementary DNA single-strands are well-established in the literature [San98]. Their kinetics are well understood through the theory of diffusion-controlled reactions [CD83]. But up to this point, I didn't really give any details on the ways to estimate the Gibbs free

energy (and kinetic rates) for more complex conformations. Several tools can be used to perform those calculations, with various levels of precision and complexity.

NUPACK [Zad+11; FPP20] is a widely used software suite for the analysis and design of nucleic acid structures, in particular for computing the Gibbs free energy of intricate conformations of DNA strand complexes. Even thought it is an advanced software for this kind of computation, its model only applies to unpseudoknotted secondary structures (see fig. 12.4), which is never the case in practice for DNA origami structure (even with a linear scaffold strand). Also, this kind of model is not necessarily suitable for the study of DNA origami, as its strength relies on accurately predicting complex secondary structures (containing nucleotide level motives such as mismatches, bulges and internal loops) without any information on its geometry in a 2D or 3D space, while DNA origami present rather simple secondary structures in the sense that most staple strands are designed to fully bind the scaffold using a few distinct domains, but the resulting rigidity and size of such structures makes its global 2D or 3D shape central to the folding process. While NUPACK can often be used for local analysis of a DNA origami design, such as individual strand secondary structures or local interactions between a few strands, it is thus not intended for a more global one.



(a) Unpseudoknotted secondary structure. **(b)** Pseudoknotted secondary structure.

Figure 12.4: DNA strand complexes, with and without pseudoknot from [FPP20]. Each secondary structure can be represented using a *polymer graph*: strands are ordered around a circle and base pairs are connected with straight lines. In (a), the secondary structure is unpseudoknotted as there exists a strand ordering for which the polymer graph has no crossing lines, which is not the case for (b).

Geometry aware models also exist for studying DNA origami self-assembly, often including necessary simplifications or approximations, such as a lattice-based or coarse-grained representations, in order to reduce the configuration space. For example, Cumberworth et al. present a DNA origami self-assembly model where each binding domain is represented as a particle on a cubic lattice [CRF18; CFR22]. This notably allows for the simple inclusion in the model of constraints such as the double-helical twist and the rigidity of double helices. Such lattice models are also used at the nucleotide level to study oligonucleotide hybridization, where each strand is represented as a self-avoiding walk of the polymer chain [ST19].

Coarse-grained modeling is also a common approach for simulating DNA structure dynamics and self-assembly in a continuous 3D space, using various granularity levels. A noteworthy simulation tool that uses this method at the nucleotide level is oxDNA [Doy+13], which was reported to allow for the direct simulation of small DNA origami [Sno+16] or

integrated in a multiscale model for simulating the assembly of big SST structures [Fon+18]. A coarser model has also been recently reported for the simulation of DNA origami folding, up to several kilobases in size [DeL+24]. It uses a mesoscopic bead-chain representation of DNA (with 7 or 8 nucleotides per bead) and a brownian dynamics simulation model.

12.4 Experimental data from the literature

The understanding and modelling of DNA origami self-assembly rely on a wide variety of experimental studies. For example, the folding pathways of DNA origami have been extensively studied through atomic force microscopy to identify intermediate structures, revealing not only the diversity of possible pathways [Wan+22] but also the potential for selecting and controlling them [Lee+16].

Another approach at the level of staples is used in [SMD19]: this article presents experimental data on the mean folding time of each staple strand on a multi-layer DNA origami, during constant-temperature assembly (see fig. 12.5). In particular, it shows that there is no correlation between the total ΔG° for binding a staple and its folding time (i.e. the duration between the beginning of the annealing reaction and the moment the staple strand is fully bound to the scaffold). On the other hand, they show that a permutation of the scaffold sequence in the origami design has little effect on the mean folding times of staples. Specifically, they observed a strong correlation between the mean folding times of staples from the original and the permuted version of the origami. Those results suggest that the assembly pathways of DNA origami are more dependent on the geometry and connectivity of the design than on the specific DNA sequences of binding domains. Some of those results are shown in fig. 12.5.

Beyond the various observations one can make from analyzing the produced data, these articles also provide a strong experimental basis for testing, validating and adjusting any simulation model of DNA origami self-assembly.

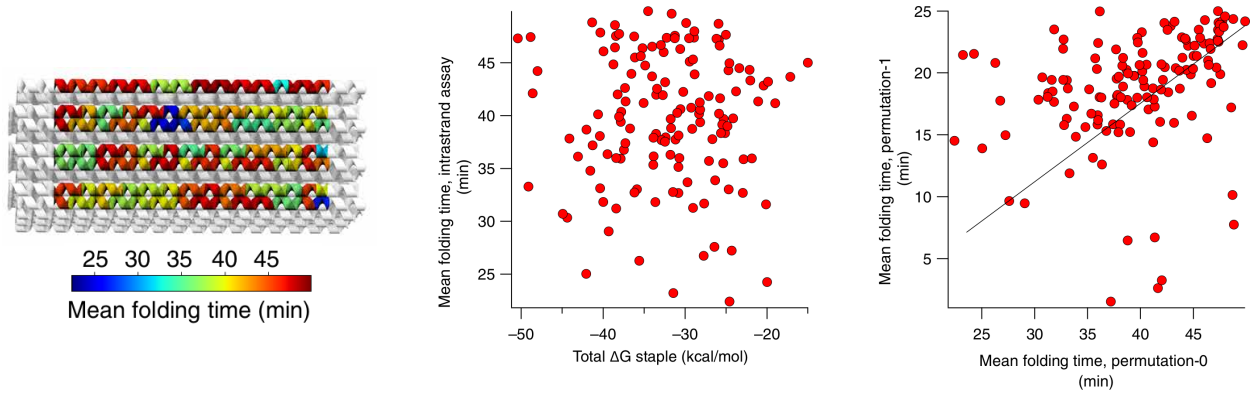


Figure 12.5: Experimental data on individual staple folding time from [SMD19].

Chapter 13

Proposed model for our simulation

13.1 A multi-level approach

We want to build a model that is specifically designed for simulating DNA origami self-assembly, but also intended for the study of multi-scaffold origamis. This model should allow us to study:

- the viability of a (multi-scaffold) origami design and the formation of chimeric or ill-formed origami,
- the multiplicity of folding pathways and there details (such as nucleation phenomenons and staple cooperation or competition),
- the influence of design choices such as scaffold routing, crossover placements and staple designs,
- the influence of experimental parameters such as individual strand concentration, ions concentrations and temperature.

A domain level model (similar to [Dan+15]) is too restrictive as pre-defined attachable domain pairs would probably not allow the system to reach some unpredicted conformations. Moreover, as we want the model to leave those conformations reachable, we also need to allow the mechanisms that participate to correcting wrongly attached staple strands and other faulty metastable states. That is why strand displacement and branch migration should be possible in this model. However, we saw that nucleotide level model allow to simulate a long sequence of extremely short-lived states that are inconsequential for the global evolution of the system. This can make it unsuitable for simulating a whole DNA origami self-assembly. I thus decided to use a multiscale approach: a kinetic Monte Carlo simulation that includes both nucleotide-level and domain-level state transitions, with some optimizations to allow for an efficient simulation while preserving realism.

Model state space and transitions. We chose to model a folding DNA origami as a finite set of interacting DNA strands. The strand concentration is a global parameter C (expressed in mol/L) that corresponds to the concentration of each individual strand of the system. Thus, the model doesn't directly allow to have some strands of the system at a higher concentration than others. A simple workaround is to include several copies of the same strand in the system (but at an obvious performance cost).

A state of the system is defined by the set of base pairs that are formed between DNA strands, with no information on the 2D or 3D positions of the nucleotides or strands. Four state transitions are possible in the model (see fig. 13.1):

- **Binding** two domains (in the most general sense of strand segment), that are complementary, non-intersecting and containing no paired nucleotides,

- **Unbinding** two domains that are fully paired and maximal (i.e. they are not strict sub-domains of two longer fully paired domains),
- **Elongating** an existing domain pairing by one base pair, when the corresponding nucleotides are unpaired and complementary,
- **Shortening** an existing domain pairing by one base pair.

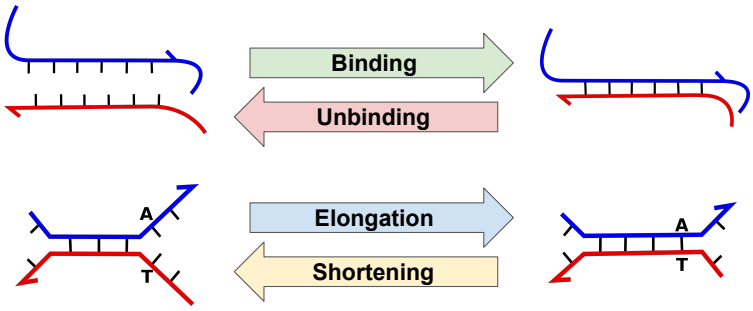


Figure 13.1: State transitions allowed in our model.

Domain level transitions (*binding* and *unbinding*) were inspired by the work of Dannenberg et al. [Dan+15]. As I wanted to allow incorrect domain hybridization, I was led to allow any complementary domain hybridization. But some metastable incorrect configurations are then accessible, and their resolution could need strand displacement or branch migration mechanism. Therefore, the *elongation* and *shortening* transitions were added to the model and allow the system to perform a kind of local search on the state space that might lead it to leave incorrect configurations by ending up in a more stable (and hopefully design-compatible) configuration.

The model doesn't include any direct information on the role of each strand (staple or scaffold), but I left the possibility of restricting allowed state transitions, in particular the binding reactions. For example, one can truncate the graph of accessible states by only allowing binding reactions between a staple domain and a scaffold domain. This would prevent the formation of any secondary structure of a staple-only complex, while still allowing chimeric or ill-formed origami structures. This approximation is based on the assumption that, as each staple domain is designed to be complementary to a specific scaffold domain, binding reactions between staple domains are rare enough or weak enough to have little to no influence on the global folding pathway.

More generally, as this model is specifically designed for DNA origami and notably intended to study multi-scaffold DNA origami, it is intentionally a bit simplistic and is based on the assumption that DNA origami, due to their specific design, are systems that evolve in far simpler and smaller state spaces than an identical amount of random-sequence DNA strands. Thus, we make the hypothesis that this state space and associated transitions are simple enough to be simulated with approximated transition rates, and without any information about the nucleotide positions but only their connectivity.

13.2 Computing kinetic rates

The most complex aspect of designing a model for DNA origami self-assembly certainly is the estimations of the kinetic rate associated with each possible state transition. A common subdivision of this problem is to first determine the forward kinetic rates (that corresponds here to the binding and elongation reactions), and then estimate the Gibbs free energy change ΔG of the forward reaction and use eq. (12.2) in order to deduce the kinetic rate of the backward reaction (unbinding or shortening in our case). Indeed, we have:

$$\frac{k_{\text{binding}}}{k_{\text{unbinding}}} = e^{\frac{-\Delta G_{\text{binding}}}{RT}} \quad \text{and} \quad \frac{k_{\text{elongation}}}{k_{\text{shortening}}} = e^{\frac{-\Delta G_{\text{elongation}}}{RT}}$$

As seen in chapter 12, the estimation of both the kinetic rates and ΔG are usually based on approximate, simplified or even empirical models.

Forward kinetic rates. Three different kinds of forward kinetic rates can be distinguished and associated to distinct computation methods:

- **The elongation rate $k_{\text{elong.}}$** (of unit s^{-1}): we chose it to be a constant of the model. In [MKT21] this rate depends on the ratio of intact base pairs to the total number of base pairs of complementary domains (which is not well-defined in our case), used in a relatively arbitrary formula whose parameter fitted values range across two orders of magnitude. In order to still match the requirements expressed in [MKT21], we could also have made this rate an increasing function of the (paired) domain length, with some constant k_f as limit at $+\infty$. For example, we could modify the formula from [MKT21] into $k_{\text{elong.}} = k_f \tanh\left(\frac{N_{\text{bp}}}{\alpha}\right)$ where N_{bp} is the number of base pairs of the considered domain and α is a parameter that can be seen as the number of base pairs above which $k_{\text{elong.}}$ is close to k_f.
- **The bimolecular binding rate k_{bi}** (of unit $\text{M}^{-1}{\cdot}\text{s}^{-1}$) for the binding reaction between two domains that were previously unconnected (i.e. the corresponding strands were part of two distinct connected components of the system). As in [Dan+15], we chose it to be a fixed constant $k_{\text{bi}} = 10^6\ \text{M}^{-1} \cdot \text{s}^{-1}$, as it seems that there is a consensus to use this value as reasonable approximation [CCD71; WB04; Dan+15]. It translates into a concentration dependent transition rate : $C \cdot k_{\text{bi}}$.
- **The unimolecular binding rate k_{uni}** (of unit s^{-1}) for the binding reaction between two domains that already are on the same connected component. This rate is very difficult to correctly estimate as it strongly depends on the global shape and structural properties of the component. While I left it as a model constant for now, Dannenberg et al. compute it for their domain level model [Dan+15] using a very convenient yet reasonable assumption: "the unbinding rate is equal to the temperature- and sequence-dependent unbinding rate for the corresponding isolated duplex". Thus the only difficulty is to compute the ΔG and the kinectic rates are then rather easily deduced:

  $$k_{\text{unbinding uni}} = k_{\text{unbinding duplex}} = k_{\text{bi}} \times e^{\frac{\Delta G_{\text{binding duplex}}}{RT}}$$

  Thus:

  $$k_{\text{uni}} = k_{\text{bi}} \times e^{\frac{\Delta G_{\text{binding duplex}} - \Delta G_{\text{binding uni}}}{RT}}.$$

Gibbs free energy. In the current implementation of the model, I only use SantaLucia's nearest-neighbor model [San98] for the computation of ΔG:

- **The value of $\Delta G_{\text{binding}}$** is computed as in section 12.1, thus it only depends on the domain sequence and experimental parameters (temperature and salt concentrations).
- **The value of $\Delta G_{\text{elongation}}$** is simply computed as the difference of ΔG for binding the new (longer) domain and ΔG for binding the old (shorter) domain. For example, for the elongation of a domain pairing of the form $\begin{smallmatrix} 5'\text{ATGCT}\text{A} \\ 3'\dot{\text{T}}\dot{\text{A}}\dot{\text{C}}\dot{\text{G}}\dot{\text{A}}\dot{\text{T}} \end{smallmatrix} \rightarrow \begin{smallmatrix} 5'\text{ATGCTA} \\ 3'\dot{\text{T}}\dot{\text{A}}\dot{\text{C}}\dot{\text{G}}\dot{\text{A}}\dot{\text{T}} \end{smallmatrix}$ we would have:

  $$\Delta G_{\text{elongation}} = \Delta G_{\text{binding}}(\text{ATGCTA}) - \Delta G_{\text{binding}}(\text{ATGCT}).$$

This approximation of $\Delta G_{\text{binding}}$ obviously isn't ideal as it doesn't take into account the current state of the system, apart from the domains in question. For example, two complementary single-stranded domains that are on the same connected component should be more likely to bind to each other if they are close than if they are far apart, which is not taken into account in the current model.

Moreover, this approximation is not entirely self-coherent: attaching a long (e.g. 10 nucleotides) domain has a ΔG that is not equal to the sum of ΔGs for attaching its first half, then the other half of the domain, but it is equal to the ΔG for attaching the first half then zipping up the remaining complementary bases using elongation reaction (which is far more likely than the first scenario).

Actual values used in my implementation. In this first implementation of the model, unimolecular transition rates (including elongation and binding rates) where artificially and arbitrarily reduced (by orders of magnitude) in order to reduce the several-orders-of-magnitude gap with the bimolecular transition rate at classical concentrations (1 to 100 nM). This choice was made both for performance reasons and for a better readability of the simulations traces, but also to check (and debug) the correct operation of the simulation and each transition type. For our test simulations, we used:

- $C = 100$ nM,
- $k_{\text{bi}} = 10^6\ \text{M}^{-1} \cdot \text{s}^{-1}$, hence a transition rate $C \cdot k_{\text{bi}} = 0.1\ \text{s}^{-1}$,
- $k_{\text{uni}} = 50Ck_{\text{bi}} = 5\ \text{s}^{-1}$,
- $k_{\text{elong.}} = 60k_{\text{uni}} = 300\ \text{s}^{-1}$.

13.3 Kinetic Monte-Carlo simulation and implementation

The kinetic Monte Carlo (KMC) method is used to simulate a stochastic process where a system evolves in a discrete state space with known transition rates between states. In particular, it is a very common approach for simulating and studying complex chemical reaction networks. At each step, the system is initially in some state S_i and the simulation picks among a finite set of possible transitions of the form $S_i \xrightarrow{k_{ij}} S_j$ (of rate k_{ij}) with the probability of choosing transition $S_i \xrightarrow{k_{ij}} S_j$ being proportionate to k_{ij}. This is equivalent to drawing an occurrence delay τ_{ij} according to an exponential distribution of parameter k_{ij} for each possible transition $S_i \xrightarrow{k_{ij}} S_j$, then choosing the first occurring transition (i.e. S_j such τ_{ij} is minimal).

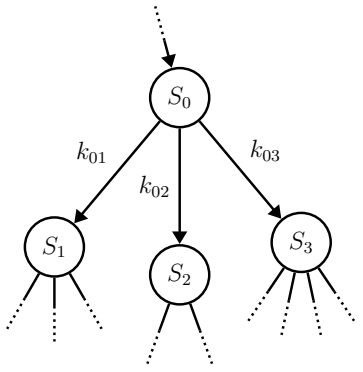


Figure 13.2: KMC scheme. At each step, the system can jump from its current state S_0 into several subsequent states (here S_1, S_2 or S_3), stochastically chosing the corresponponding transition according to transition rates (k_{ij}).

In our case, we describe each state transitions τ as a local change of the system, affecting a small subset of nucleotides, together with a transition rate k_τ (in s^{-1}) which is equal to 0 if the transition is not applicable to the current state. For example, a binding transition τ_b is defined by the strand segments it binds. Its transition rate k_{τ_b} notably depends on whether the reaction is bimolecular or not, and is set to 0 if the corresponding strand domains are not

fully unpaired and complementary. With this local definition of a transition, only the rate of the transition τ depends on the current state of the system, and this rate often remains unchanged when the system undergos transitions involving nucleotides or strands that are not involved in τ.

For this reason, we chose to implement our KMC simulation by drawing in advance an occurrence delay Δt_τ for every possible transition τ (i.e. with $k_\tau > 0$), and storing them in a priority queue H (implemented as a binary min-heap). This allows a fast, constant time access to the next transition event of the simulation, together with its occurrence time. At each step, one must update the priority queue H by removing some transitions that are now impossible (i.e. with $k_\tau = 0$), adding new possible transitions and updating transitions with changed rate (lines 11 to 22 in Algorithm 1).

The pseudocode description of our algorithm is given bellow, using the notation $\mathrm{Exp}(\lambda)$ as the exponential distribution of parameter λ, i.e. of density function:

$$x \mapsto \begin{cases} \lambda e^{-\lambda x} & x \geqslant 0, \\ 0 & x < 0. \end{cases}$$

Algorithm 1: KMC algorithm

Input: Initial state S
Output: Sequence of values taken by (t, S)

```
1  t ← 0
2  H ← empty min-heap
3  for every transition τ with k_τ > 0 do
4  |  Draw Δt_τ ~ Exp(k_τ)
5  |  Push (t + Δt_τ, τ) in H
6  end
7  while True do
8  |  Pop first transition (t_τ, τ) from H
9  |  t ← t_τ
10 |  Apply transition τ to S
11 |  for every transition τ with changed k_τ do
12 |  |  if k_τ = 0 then
13 |  |  |  Remove (t_τ, τ) from H
14 |  |  else
15 |  |  |  Draw Δt_τ ~ Exp(k_τ)
16 |  |  |  if (t_τ, τ) ∈ H then
17 |  |  |  |  Replace (t_τ, τ) by (t + Δt_τ, τ) in H
18 |  |  |  else
19 |  |  |  |  Push (t + Δt_τ, τ) in H
20 |  |  |  end
21 |  |  end
22 |  end
23 end
```

This algorithm was implemented in `Python3` using distinct classes for the min-heap data structure, state description, energy model, etc.

13.4 Results, perspectives and discussion

Results. The model and its implementation didn't reach a mature enough state to provide sensible insight on the self-assembly of realistic DNA origami, but we did use and validate the simulation on an extremely simple test design made of three DNA strands. This allowed us to observe some interesting (but predictable) mechanisms and offer a glimpse of what this kind of simulation could do, its possible refinements and limitations. A detailed example of simulation execution is given in fig. 13.3. One can observe several already discussed phenomenons, such as oscillations between unstable states and a metastable one, until the system reach another more stable state (partly detailed in steps 1-36, 38-52 and 119-122), nucleation zipper (detailed in steps 52-59) and the preponderance of unimolecular interactions over bimolecular ones. But one can also notice during the last phase (steps 153 to 168) two domain bindings that would be topologically impossible in reality, due to the helical shape of double-stranded DNA. Also, the temperature value used for this simulation was of -20°C, which should be far too cold for the self-assembly of a DNA origami, but I hypothesize that the folding still succeeded because I only allowed staple-scaffold domain bindings in this run.

While this simulation model reached a first functional stage of development, it still needs improvements, in both realism and performance.

Resolving coherence and performance issues of domain binding reactions. Our initial motivation for allowing any domain binding was to reduce the performance cost of nucleotide level event for the simulation, while leaving some unpredicted conformations accessible. But allowing any-length domain binding gives rise to a two-folds issue.

First, this means that, for example, two 10-nucleotides complementary domains that are initially unpaired (configuration A) have multiple ways of reaching a fully bound configuration B, typically starting by binding one of the 55 possible subdomains (10 of length 1, 9 of length 2, etc) then zipping up the remaining base pairs. However, all those 55 possible paths from A to B (and more) are already implicitly accounted for in the calculation of the kinetic rates for directly binding/unbinding the full domains (i.e. the reaction $A \to B$). Thus, the resulting effective rate of fully binding the domains is roughly the sum of the effective rate of every direct reaction paths (computed using the steady-state approximation), which causes a coherence and realism issue: in practice, domain binding reactions such as $A \to B$ are incorrectly favored in the model, to an extent that increases with domain length.

Secondly, allowing any-length domain binding leads to a high number of single-nucleotide domain binding reactions, which defeats the performance purpose of a domain level model.

For those reasons, I thought of restricting the allowed length for domain binding (and partially implemented it), in order to prevent pathway redundancy in the model, but also quickly reversed reactions such as single-nucleotide domain binding. This rule can be expressed in the following way: a domain binding/unbinding reaction is allowed if and only if the corresponding $-\Delta G$ (or the ratio $\frac{k_{\text{binding}}}{k_{\text{unbinding}}}$, or domain length) is above some predefined threshold and no strict sub-domain binding is allowed. One simple example of such rule would be to only allow length-four domain binding/unbinding.

Realism of the model rates. Our currently simplistic rate computation method doesn't take into account the variety of interactions between the components of a partially folded origami structure. Several enhancements are worth considering:

- Taking intramolecular distance into account into the computation of ΔG for unimolecular binding reactions, or using a loop-based energy model such as presented in [Dan+15],

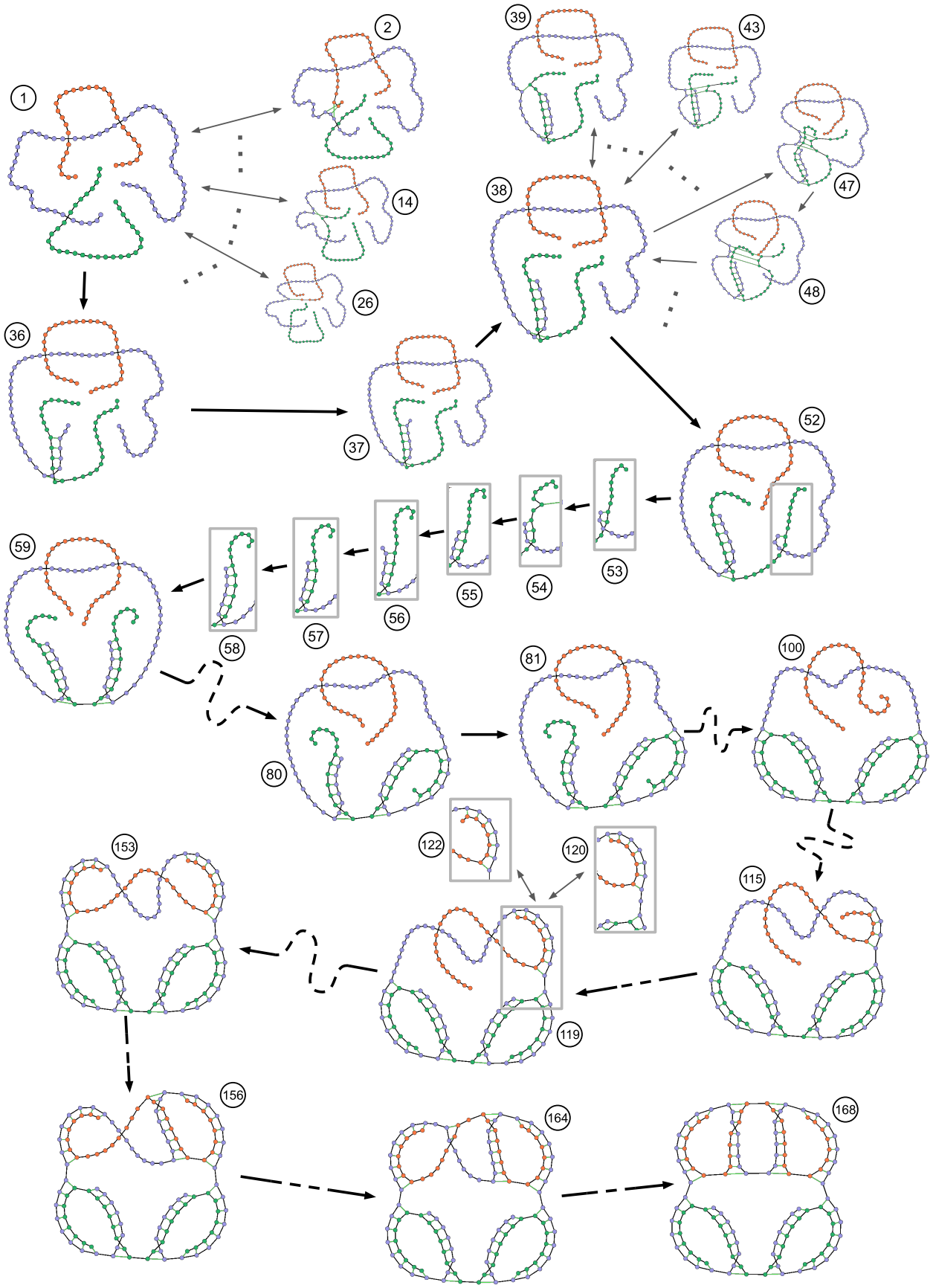


Figure 13.3: Example of simulation execution. Images where generated with Gephi, but the simulation model doesn't include any information on nucleotides/strands positions. Domain bindings were only allowed between a staple domain and a scaffold domain for this simulation. Simulation steps are numbered (with the rank of the state's first occurrence) and dots or doted arrows indicate the omission of a series of simulation steps.

- Excluding the binding of intermediary domains when a staple is already bound at both ends (such as in steps 153-168 in fig. 13.3), which is topologically impossible due to the helical shape of DNA,
- Taking the important rigidity difference between single- and double-stranded DNA into account for ΔG or distance estimation.

However, those refinements would need proper formalization before being implemented, and some could result in a considerable performance drop of the model, as rates would need to be updated much more often.

Stacking interactions between double-stranded domains. I didn't include any additional stacking term in the estimation of ΔG in the current state of the implemented model, mainly because it would require information on the geometry of the current structure. However, [Dan+15] suggests adding the term $\Delta G^{\text{stack}}(T)$, that only depends on the temperature (and a model parameter), for every domain binding reaction involving a scaffold domain that is adjacent to an already bound domain. This method doesn't take into account the fact that two adjacent scaffold domains are not necessarily aligned in the origami design, due to bends in the scaffold routing, which can prevent their stacking interactions. But as most DNA origami designs tend to minimize the number of scaffold bends in its raster-filling routing, most of its adjacent binding domains are in fact aligned (for example, the simplest flat, square 30-helices design using 7249-nucleotides M13mp18 and ∼240 (mostly) 3-domains 'S' shaped staple strands, will have 58 scaffold turns for ∼720 binding domains). This support the soundness of adding this $\Delta G^{\text{stack}}(T)$ term, and, in the case of our model, it should also be added with the same conditions for the elongation reaction.

Performance of the implementation. The use of a heap for our implementation of the kinetic Monte Carlo simulation is not very conventional, but allow some optimizations, such as the pre-computation of all transition rates and occurrence delay. But in a lot of situations, a more classical implementation seems more efficient, as it usually only needs to evaluate two random variables per step. Indeed, our implementation requires at each step to draw as many occurrence delays as the number of transition rates that changed to a non-zero value, which doesn't seem optimal when those rates strongly depend on the current state of the system. A classical implementation, such as Gillepsie algorithm [Gil77], would instead draw independently, at each step, a transition τ (with probability proportionate k_τ), and an occurrence delay Δt_τ according to an exponential distribution of parameter $\sum_\tau k_\tau$.

Conclusion. While the advancement of this simulation model didn't really reach a level where it was directly useful for my designing and experimental work, it still showed promising first results by capturing several interesting mechanisms, such as nucleation zipper and rapid oscillations around a metastable state until reaching a more stable one. It allowed me to better understand the mechanisms behind DNA origami folding, but also, more generally, the thermodynamics of the DNA molecule and the methods to quantify it. By balancing nucleotide- and domain-level interactions, the model offers a feasible route to capture a wide variety of behaviors in DNA folding pathways, which is a challenging aspect in simulating such large-scale structures. Addressing some of its limitations could further enhance the model's applicability in experimental validation and refinement for DNA nanotechnology applications.

Part VII

Self-assembly of aperiodic tilings

Chapter 14

Quasicrystals and Cut-and-project aperiodic tilings

This chapter, along with chapters 15 and 16 are mainly based on my M2 internship work and the very beginning of my PhD thesis (2021). This was my first introduction to DNA nanotechnologies (and secondarily to aperiodic tilings).

The goal of these chapters is to explore the theoretical and experimental leads for the creation of DNA quasiperiodic crystal.

14.1 Cut-and-project principle

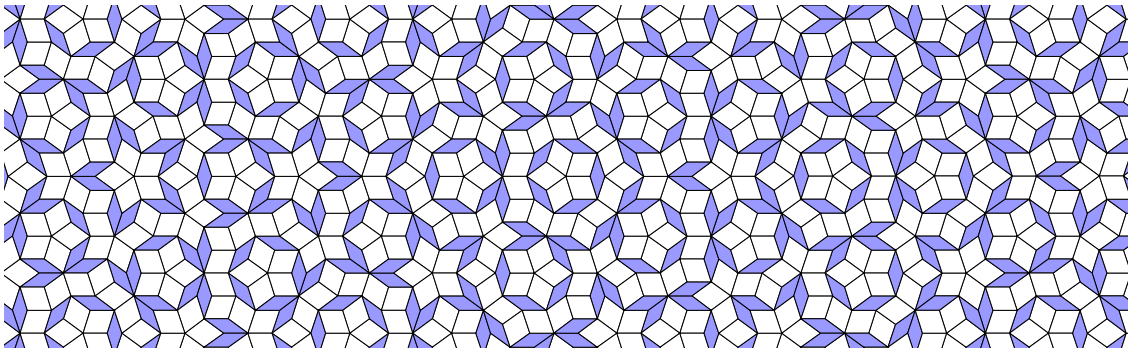


Figure 14.1: Penrose tiling

The notion of *cut-and-project* tiling appeared in 1981 [N G81] as a tool to study Penrose tiling (fig. 14.1). It quickly became central to the study of quasicrystals. Quasicrystals, discovered in 1984 [She+84], contain symmetries that are forbidden in classical crystallographic structures. Indeed, classical crystals have a periodic structure in three directions and can therefore be mathematically described as a repeated pattern along some $\mathbb{R}^3$ lattice. In the case of quasicrystals, the structure is aperiodic (or periodic in only one or two directions) but can be described as the projection of a higher-dimensional periodic structure. This is the *cut-and-project* method.

Our study on this subject is mainly based on Ilya Galanov's work [Gal19], although some choices of notation or definition slightly diverge from it.

Definition 14.1.1 (Cut-and-project)**.** A $n \to d$ *cut-and-project scheme* consists of a linear subspace E of $\mathbb{R}^n$ of dimension d, called *slope*, its orthogonal complement $E^\perp$ and a lattice $\mathcal{L} \subset \mathbb{R}^n$. The projection along $E^\perp$ onto E (resp. along E onto $E^\perp$) is denoted $\pi : \mathbb{R}^n \to E$

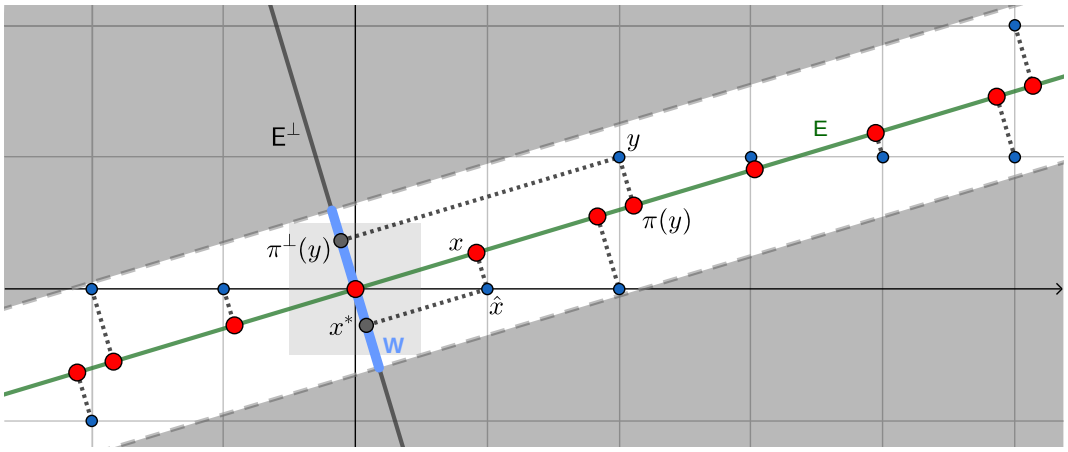


Figure 14.2: Exemple of a canonical $2 \to 1$ cut-and-project construction. The constructed points are in red.

(resp. $\pi^\perp : \mathbb{R}^n \to E^\perp$) $\pi_{|\mathcal{L}}$ must be injective and for $x \in \pi(\mathcal{L})$, $\hat{x}$ (called *lift* of x) denotes the unique element of $\mathcal{L}$ such that $\pi(\hat{x}) = x$. Lastly, the *star map* is the map $* : \pi(\mathcal{L}) \to \pi^\perp(\mathcal{L})$ defined by $x^* = \pi^\perp(\hat{x})$ for all $x \in \pi(\mathcal{L})$.

$$\begin{array}{ccccc} E & \xleftarrow{\pi} & \mathbb{R}^n = E \oplus E^\perp & \xrightarrow{\pi^\perp} & E^\perp \\ \cup & & \cup & & \cup \\ \pi(\mathcal{L}) & \xleftarrow{\pi_{|\mathcal{L}}} & \mathcal{L} & \xrightarrow{\pi^\perp_{|\mathcal{L}}} & \pi^\perp(\mathcal{L}) \end{array} \qquad x \mapsto x^*$$

A bounded subset W of $E^\perp$ with non-empty interior is called a *window* or *acceptance domain*. If the boundary ∂W has zero Lebesgue measure, it is called regular. In the canonical cut-and-project scheme, the window is chosen to be $W = \pi^\perp([-\frac{1}{2}, \frac{1}{2}]^n)$ (the projection on $E^\perp$ of the unit hypercube centered on the origin) and lattice $\mathcal{L}$ to be $\mathbb{Z}^n$.

The window acts as a filter for points which are to be projected: using a cut-and-project scheme, one can define a tiling by considering the projections $\pi(z)$ of lattice points $z \in \mathcal{L}$ whose orthogonal projection $\pi^\perp(z)$ is in the window W. The vertices of the tiling are then:

$$\Sigma(W) = \{x \in \pi(\mathcal{L}) \mid x^* \in W\} = \{\pi(y) \mid y \in (E + W) \cap \mathcal{L}\}$$

In this context, one can construct a *Penrose tiling* (see fig. 14.1) using a $5 \to 2$ canonical cut-and-project scheme, with the slope E being the plane with basis vectors:

$$u = \begin{pmatrix} 1 \\ \cos(2\pi/5) \\ \cos(4\pi/5) \\ \cos(6\pi/5) \\ \cos(8\pi/5) \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 \\ \sin(2\pi/5) \\ \sin(4\pi/5) \\ \sin(6\pi/5) \\ \sin(8\pi/5) \end{pmatrix}$$

We will also study the *Golden tiling* (fig. 14.3). This is a $4 \to 2$ canonical cut-and-project tiling with slope basis vectors:

$$u = \begin{pmatrix} -1 \\ 0 \\ \phi \\ \phi \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 \\ 1 \\ \phi \\ 1 \end{pmatrix} \quad (\text{where } \phi = \frac{\sqrt{5}+1}{2} \text{ denotes the golden ratio})$$

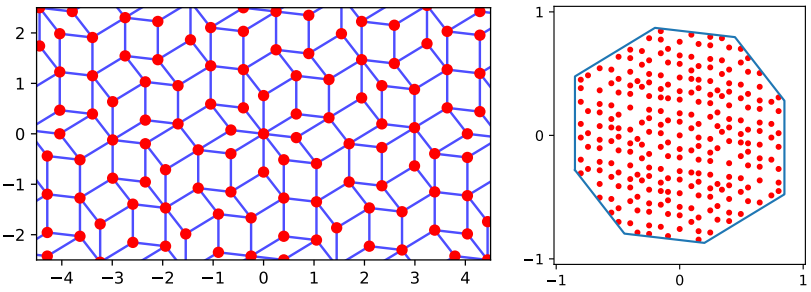


Figure 14.3: $4 \to 2$ **cut-and-projet construction of the Golden tiling.** In red are the tiling vertices on the left and their projection on the window on the right.

14.2 Atlas and window

Due to the linearity of $*$, we can observe that the configuration of the vertices around a certain point x of the tiling only depends on the position of x^* in the window W. Indeed, if $x = \pi(y)$ with $y \in \mathbb{Z}^n$ and $dx = \pi(dy)$ with $dy \in \mathbb{Z}^n$, then :

$$(x + dx) \in \Sigma(W) \Leftrightarrow (x + dx)^* \in W \Leftrightarrow x^* + dx^* \in W \Leftrightarrow x^* \in W - dx^*$$

(Where $A - B$ denotes the set $\{a - b \mid a \in A \text{ and } b \in B\}$ and $A \pm b = A \pm \{b\}$))

Hence, for $dx \in \pi(\mathbb{Z}^n)$, the vertices $x \in \pi(\mathbb{Z}^n)$ such that x and $x + dx$ are part of the tiling are exactly those that verify $x^* \in W \cap (W - dx^*)$.

Definition 14.2.1 (Pattern)**.** We call *pattern* an ordered pair $(P, \overline{P})$ of two disjoint subsets of $\pi(\mathcal{L})$. P represents the set of points that are present in the pattern while $\overline{P}$ is the set of points that are absent in the pattern. A pattern $(P, \overline{P})$ is said to *appear* at some position $x \in \pi(\mathcal{L})$ of the tiling $\Sigma(W)$ if $x + P \subset \Sigma(W)$ and $(x + \overline{P}) \cap \Sigma(W) = \emptyset$.

For $r \in \mathbb{N}$, we call patern of size (or radius) r a pattern of the form $(P, \pi(\mathcal{B}_1(r))\backslash P)$ with $0 \in P$ and $\mathcal{B}_1(r) \subset \mathbb{Z}^n$ (where $\mathcal{B}(r)$ denote the closed ball of radius r centered at 0, using norm $\|.\|_1$).

In the following, a pattern $(P, \overline{P})$ of radius r can be assimilated to P (since $\overline{P} = \pi(\mathcal{B}(r))\backslash P$ is then determined by P and r)

One can observe that a pattern $(P, \overline{P})$ appears in the tiling $\Sigma(W)$ at position x if and only if:

$$x^* \in \left(\bigcap_{dx \in P} W - dx^* \right) \setminus \left(\bigcup_{dx \in \overline{P}} W - dx^* \right)$$

Definition 14.2.2 (Atlas)**.** For a chosen radius $r \in \mathbb{N}$, we call the *radius r atlas* (or *r-atlas*) the set of radius r patterns of the tiling.

For a point x of the tiling and some radius r, the position of x^* in the window W entirely determines the pattern of radius r centered at x of $\Sigma(W)$. Viewed from another perspective, to each point w in the window W, one can associate a unique pattern of radius r $(P, \overline{P})$ whose lift $\hat{P} = \{\hat{x} \mid x \in P\}$ corresponds to the set of points $b \in \mathcal{B}(r)$ such that $w + \pi^\perp(b) \in W$. In other words, $P = \{x \in \pi(\mathcal{L}) \mid w + x^* \in W\}$.

This defines a partition of the window. In the case where $\pi^\perp(\mathcal{L})$ is dense in $E^\perp$, the set of patterns thus defined exactly corresponds to the radius r atlas of the tiling. In order to compute the atlas of a cut-and-project tiling, one can compute this partition of the window.

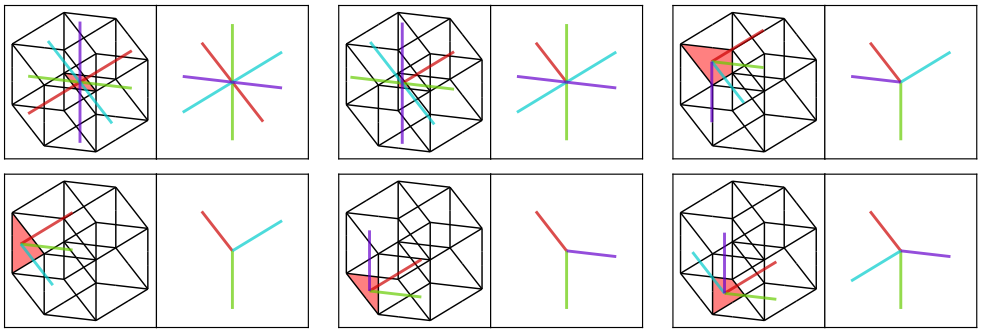


Figure 14.4: Some patterns from the radius 1 atlas of the Golden Tiling, along with their corresponding region and projection in the window. Each color corresponds to a unit vector of $\mathbb{Z}^4$, projected onto E or $E^\perp$.

14.3 Local rules and self-assembly

What we will be focussing on now is to understand how the atlas of a tiling can be used to incrementally construct this tiling using only local rules.

Definition 14.3.1 (Local rules)**.** A tiling obtained by $n \to d$ cut-and-project with slope E has *local rules* if for some radius r, any $n \to d$ cut-and-project tiling whose radius r atlas is included in that of the former has the same slope E.

This definition suggests that for a cut-and-project tiling with local rules, a finite portion X of the tiling and an atlas can constitute sufficient information to deduce an infinite number of points of the tiling. Indeed, assuming the window W of the tiling is convex (which is true for canonical cut-and-project constructions), the portion of the tiling provides partial information about it: since the slope is fixed by the atlas, x^* is also known for all $x \in X$ and the convex hull C of $\{x^* \mid x \in X\}$ then satisfies $C \subset W$. A finer analysis is possible when we know the window is a hypercube of the form $W = \pi^\perp(y + [0,1]^n)$ with $y \in \mathbb{R}^n$. In this case, only y is unknown and can be approximated with increasing precision as the size of the known portion of the tiling grows.

From this reflection, we can hope to assemble a tiling with local rules starting from a portion of the tiling, called a *seed*, and making each addition of a point or tile dependent only on its already constructed local neighborhood. For this, we use the notion of a vertex (or tile) forced by the local configuration.

Definition 14.3.2 (Local configuration)**.** For a portion of the tiling $X \subset \Sigma(W)$ and a vertex $x \in X$, the *local configuration of radius* r around x is the set of relative positions of the vertices of X within r edges (in $\Sigma(W)$) of x. We denote it by $C_X(x,r)$, and it can be expressed as: $C_X(x,r) = \{v \in \pi(\mathcal{B}(r)) \mid x + v \in X\}$.

Definition 14.3.3 (Forced vertex)**.** Given a local configuration $C_X(x,r)$ and an r-atlas $\mathcal{A}(r)$, we consider the set of patterns in $\mathcal{A}(r)$ that contain $C_X(x,r)$. Let $\tilde{\mathcal{A}}$ be the intersection of these patterns. We then define the set of vertices forced by the local configuration by:

$$F_r(C_X(x,r)) := \tilde{\mathcal{A}} \backslash C_X(x,r)$$

We could similarly define the notion of a tile forced by the local configuration, considering a tile as a set of four vertices whose lifts form a unit square of $\mathbb{Z}^n$.

The above concepts are central to the development of an incremental tiling assembly algorithm. For this purpose, we have primarily worked on the Golden tiling, due to its ease of representation (the slope E and its orthogonal $E^\perp$ are 2-dimensional), its atlases that are calculable by partitioning its window ($\pi^\perp(\mathbb{Z}^n)$ is dense in $E^\perp$), and of course, because it possesses local rules [Gal19]. Several simple algorithms are possible (results of these algorithms are shown in figure 14.5):

Algorithm 2: Randomized construction

Input: finite seed $S \subset \pi(\mathbb{Z}^4)$
while `True` **do**
  randomly draw a vertex $x \in S$
  compute forced vertices $F_r(C_S(x, r))$
  add $x + F_r(C_S(x, r))$ to S
end

Algorithm 3: Deterministic Construction

Input: finite seed $S \subset \pi(\mathbb{Z}^4)$
while `True` **do**
  $X \leftarrow \emptyset$
  for $x \in S$ **do**
    compute $F_r(C_S(x, r))$
    add $x + F_r(C_S(x, r))$ to X
  end
  add X to S
end

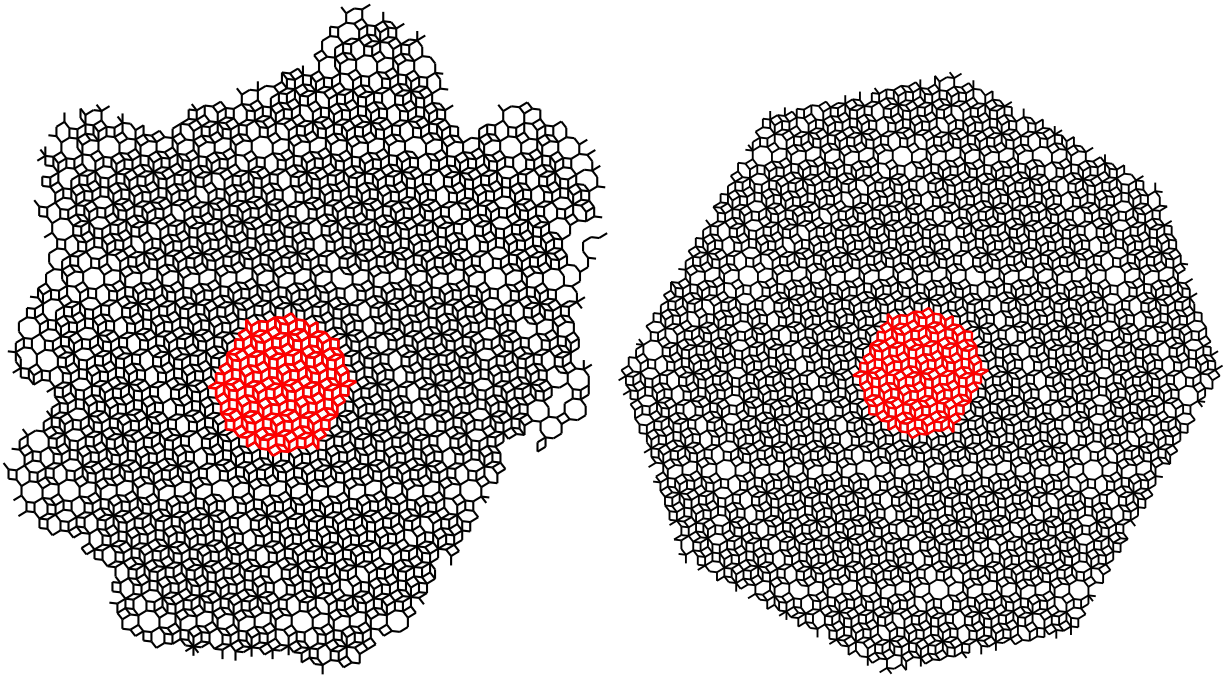


Figure 14.5: Intermediate results of incremental assembly, with algorithm 2 (left) and algorithm 3 (right), both using the atlas of radius 7. The seed is drawn in red.
It can be observed that some points do not seem to be determinable by an incremental construction algorithm. This can be explained by the fact that, although the slope of the cut-and-project tiling is fixed by the atlas, the seed only fixes an approximation of its window: $W' \subsetneq W$. Only the points x with $x^* \in W'$ are then constructible. These images confirm the results of Galanov in his thesis [Gal19].

Recently, Fernique and Galanov reported a proof that such purely local algorithm is capable of constructing a Golden tiling [FG22]. More specifically, they showed in this preprint that for any finite pattern $\mathcal{P}$, there exists a seed $S \supset \mathcal{P}$ and an atlas radius r such that their algorithm builds the *cylinder set* of $\mathcal{P}$ (i.e the intersection of all Golden tilings that have $\mathcal{P}$

as subset).

Many slight variations or optimizations of these algorithms are possible. Our study has remained at a relatively experimental stage, and the most suitable version for modeling the self-assembly of a tiling with DNA strands has yet to be found. In particular, for an atlas radius greater than 1, the information on the local configuration would need a way to be transmitted between non-adjacent (but close enough) tiles.

Chapter 15

A Penrose tiling DNA origami

Before working on the self-assembly of a tiling made of DNA strands, we sought to design a structure representing a chosen portion of the Penrose tiling, notably to test our ability to image this kind of structure using atomic force microscopy. This allows us to first study the feasibility of this project, relying on already proven experimental methods. To this end, we used the DNA origami method.

15.1 The Penrose origami design

First of all, it should be noted that most vertices of a Penrose tiling have an odd degree (i.e. number of adjacent edges). Therefore, a graph represented by a portion of the tiling is generally not Eulerian. To remedy this problem, it is sufficient to double each edge. Here, we have chosen to represent each edge of the tiling with two double helices, one above the other, in hopes of improving its rigidity. We thus aim to design a flat origami with two identical layers.

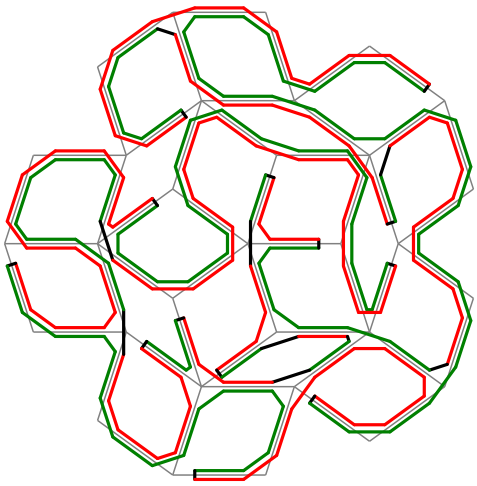


Figure 15.1: Diagram of the chosen path for the scaffold DNA strand, in the case of a radius 4 portion of the Penrose tiling. The portions of the scaffold strand on the upper layer of the origami are in green, and in red for the lower layer.

One naive idea for routing the scaffold strand would be to take any Eulerian circuit on the graph. However, the resulting loop might then describe a non-trivial knot, making the configuration impossible to realize. In order to avoid this, we chose to construct the path from a left-hand depth-first traversal of a spanning tree of the edges of the selected portion of the Penrose tiling. What we call a spanning tree of the edges can be seen as a spanning tree of the line graph. We then minimize the number of layer changes in a greedy manner. Thus, we obtain the routing of the scaffold strand that will serve as the backbone of our origami. An example result of this process is shown in figure 15.1.

The next step involves adding staples along the entire scaffold strand. These staples should measure between 15 and 80 nucleotides and must force and maintain the folding of the scaffold strand into the desired shape. Thus, at each vertex of the tiling, the DNA double helices representing the edges emanating from the vertex must

be connected either by the scaffold strand or by staples. We also want a junction to be made between the two layers of the origami near the middle of each edge in order to ensure alignment.

Several strategies are possible for placing the staples: one can have strands that circle each face of the tiling or strands forming a star around each vertex. We chose the latter option, primarily because it allows easy creation of inter-layer connections at the middle of each edge. One, two or three staples are used to circle around each vertex.

In order to minimize the lengths of connections created by the staples or the scaffold strand between the double helices, we then used the ENSnano software [LS21] and its physics simulation. This allowed us to find the optimal rotations of the double helices as well as the ideal location of the inter-layer connection at the middle of each edge.

Finally, we added a few extra nucleotides at some connections that were still too long between two double helices.

The design resulting from this process used only 6509 of the 7249 nucleotides of the scaffold strand, so we added staples to fold the remaining length into an "L" shape on the side of the origami. This will allow us to know the orientation of the origami in AFM images. An overview of the final design is shown in figure 15.2.

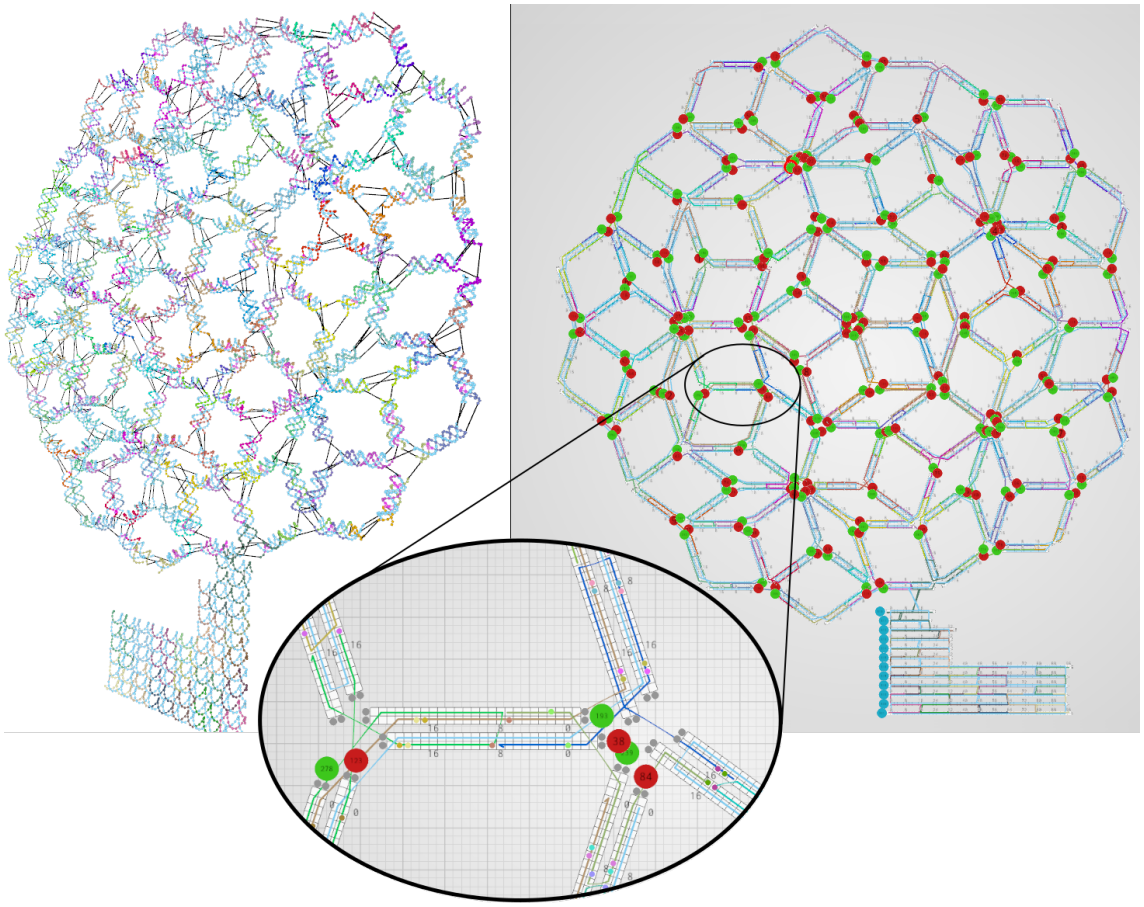
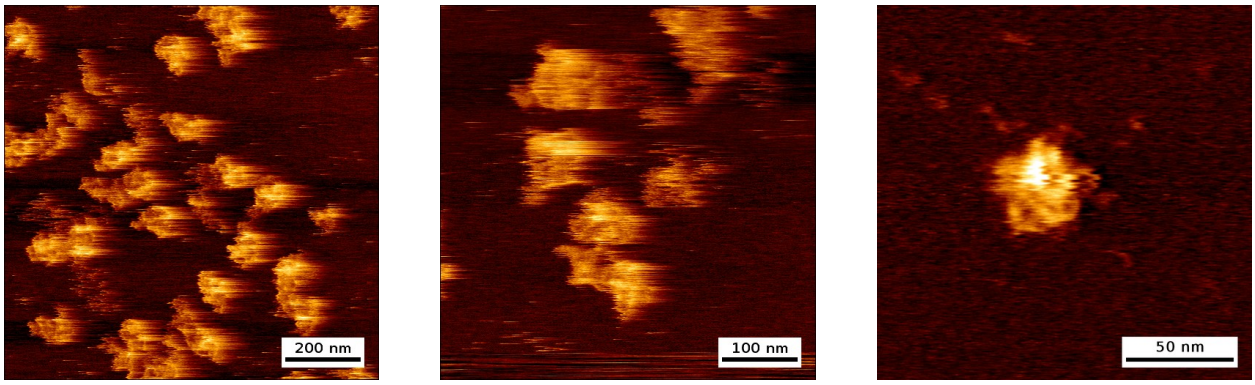


Figure 15.2: Final design of the origami representing the Penrose tiling, visualized in the ENSnano software [LS21]. On the left is a 3D view and on the right a simplified representation.

15.2 Results and perspectives

DNA origamis were prepared in TAE buffer with 12.5 mM $MgCl_2$, using 10 nM of M13mp18 scaffold for gel electrophoresis samples, and 1 nM for direct AFM imaging samples, together with a 10-fold excess of staple strands.



(a) **(b)** **(c)**

Figure 15.3: AFM images of the Penrose DNA Origami designed during the intership. (a, b) First annealing and imaging attempt. Annealing (∼100 minutes): 95°C for 5 minutes then from 95°C to 25°C at the rate of −1°C every 1 min 20 sec. Imaged in TAE buffer, without filtration nor Ni^{2+} addition. (c) Annealing (∼20 hours): 95°C for 5 minutes, from 85°C to 70°C at the rate of −1°C every 7 minutes, from 70°C to 40°C at the rate of −1°C every 28 minutes, from 40°C to 25°C at the rate of −1°C every 15 minutes. Then the sample was filtered using an Amicon filter (0.5 mL), 10 µL were deposited onto freshly cleaved mica together with 10 µL of a 12.5 mM Ni^{2+} solution, and ∼60 µL TAE buffer were added after ∼2 minutes just before AFM imaging.
The structures about 100 nm wide are the origamis, while the smaller ones about twenty nanometers wide in the third image are staples, which are in excess in the solution.
It can be observed that the produced images are particularly blurry. This is due to the sample adhering poorly to the mica disk used as a support. Indeed, we found that the AFM tip moves and deforms the sample during the scan. We therefore accelerated the movements of the tip to reduce this effect, which explains the streaks of several tens of nanometers to the right of each origami.

While some AFM images obtained are encouraging, we were unable to properly image this initial design despite numerous attempts, varying the annealing parameters, staple and M13mp18 scaffold concentrations, or buffer. We have several hypotheses to explain these difficulties, as well as some design improvement possibilities.

First, it quickly became apparent that the origami did not adhere well to the mica disk used as a support for AFM microscopy, despite using a proven protocol. This does not necessarily indicate a design or assembly flaw, but it makes imaging very difficult. We tried several methods found in the literature to address this issue. Adding Nickel Chloride $NiCl_2$ [Wan+19], for example, allowed for better adhesion of the origami to the mica, but also seemed to modify its shape by folding or contracting it excessively (see fig. 15.3c). Some imaging results are shown in figure 15.3.

We also sought to verify the formation of origami by migrating some samples on an agarose gel. This allowed separating fully formed origamis from excess staples and any unpaired scaffolds or other structures formed by the strands present in the solution. The result is shown in figure 15.4. Three bands can be distinguished under wells 2 and 3. The lowest corresponds to the staples (and possibly the small structures they may form among themselves), which is confirmed by the bands at the same height under wells 4 and 5.

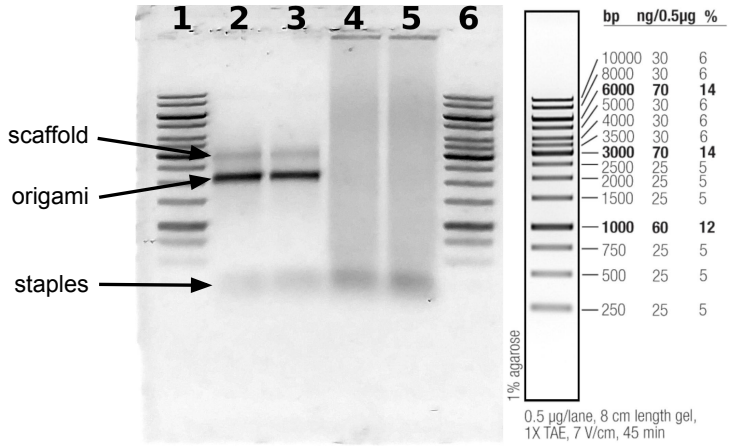


Figure 15.4: Agarose gel electrophoresis of the origami (1% agarose). Wells 2-3: Origami, wells 4-5: staples alone, wells 1 and 6: GeneRuler (linear double-stranded nucleotide length correspondence given on the right). Annealing: 95°C for 5 minutes then from 95°C to 25°C at the rate of -1°C every 1 min 20 sec.

We believe the highest band corresponds to the remaining unpaired scaffolds, and the one just below to the correctly formed origamis. The migration speeds here are related to the compactness of the structure more than its total nucleotide count, which explains why the origami migrates faster than the lone scaffold.

The sharpness and intensity of the band corresponding to the origami suggest a certain robustness of the assembly: there does not seem to be a continuum of partial or defective assemblies, which would have resulted in a broader, smeary and lighter band. However, we were unable to properly image the origamis we extracted from this band, preventing us from confirming this reasoning with certainty.

Following these observations, we identified several limitations and possible improvements to the design.

Adding more free nucleotides at certain junctions could allow the arrangement of double helices around the vertices of the tiling to be closer to the desired configuration. Indeed, although this is not clearly visible in the AFM images, this shortcoming may explain the observed deformation of the origamis compared to the initial design. Additionally, the physical simulations we conducted with the ENSnano software (using the rigid helices simulator) [LS21] seem to confirm this defect. To address it, we can draw on similar proven designs [Wan+19].

The choice to make the origami in two layers poses several problems: it is difficult to find rotations of the double helices that allow all the connections between double helices to be sufficiently short in the ENSnano design. Also, we suspected that these two layers could be one of the reasons why the origami adheres poorly or is deformed on the mica.

Finally, the choice of 21 nucleotides per edge is quite practical for designing the origami but is probably too small for several reasons. First, the ratio between the length and width of the edges is too large for the origami to have the desired tiling shape. Additionally, this likely poses folding robustness issues. Finally, this small mesh size makes it particularly difficult to image. Most similar origamis we have seen in the literature have edges that are two to five times longer [Wan+19].

Chapter 16

Leads for designing self-assembling aperiodic tilings

We also began working on different designs of elementary DNA tiles for the self-assembly of a cut-and-project tiling.

The literature currently offers several DNA tile systems capable of transmitting information during their progressive self-assembly, thus performing more or less complex calculations. For example, one can mention the double crossover tile system of Rothemund et al. [RPW04] used to generate Sierpinski triangles, or the SST (for *single-stranded tiles*) systems, which use tiles of about forty nucleotides to perform calculations by extending a DNA nanotube [Woo+19].

However, these solutions mostly use a set of tiles of the same shape: while we can draw inspiration from them, they do not directly adapt to our project.

16.1 Design approaches for implementing local rules

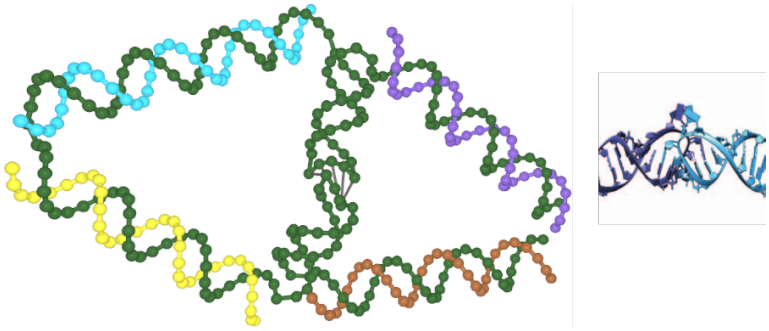
We have considered several designs for elementary tiles for the self-assembly of a cut-and-project tiling with local rules. For this, two types of elementary pieces are possible: star-shaped structures representing the vertices of the tiling, or rhombus-shaped structures representing its tiles. Since we have not yet determined which option the theory of cut-and-project tilings may favor, we have begun exploring both possibilities.

First, the simplest design would be to use, for each tile of the tiling, a single DNA strand that folds into a rhombus, attaching itself to the strands of adjacent tiles. However, this could pose robustness issues for the tiles. Indeed, the angle of each rhombus in the tiling would be difficult to control, which could lead to flat rhombuses in some places.

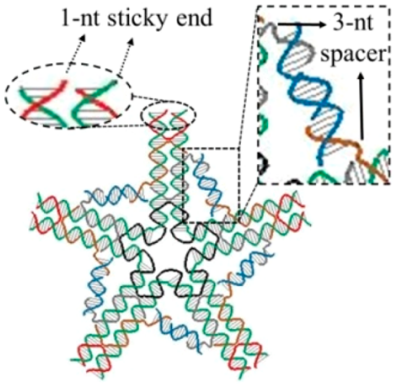
A solution to this problem is to slightly modify the design of a tile to form a rigid bond along its minor diagonal. To do this, the strand is extended to form two double helices that meet in the center of the tile, where they attach to each other over a few nucleotides. This process, called a *kissing loop*, was initially introduced for RNA origami construction [GRA14]. A visualization of this design is shown in figure 16.1a.

Another solution is a star-shaped elementary piece design, representing a vertex of the tiling. An example is given in figure 16.1b.

Due to the need to use multiple DNA strands per elementary piece, this solution is a bit less elegant and more time-consuming to implement. However, it is more versatile and has the advantage of having already been tested with several variants for simpler tilings. The design proposed in Liu et al.'s article [Liu+19] is not directly usable here due to the



(a) Tile design using a *kissing loop* we made in ENSnano. On the left, the green strand is the tile of interest while the other strands are segments of adjacent tiles. On the right, a detailed representation of the configuration of an RNA kissing loop (from [GRA14]).



(b) Example design for a star-shaped tiling piece, using a large number of strands (from [Liu+19]).

Figure 16.1: Examples of tile design for implementing local rules.

low specificity of the bonds between the pieces, but it seems easily adaptable by slightly modifying the functioning of these bonds.

16.2 First hard-coded version and sequence design

An initial step towards the self-assembly of a tiling is to test a tile design by hard-coding the self-assembly of a portion of the tiling. Instead of using a finite set of tiles to assemble the entire tiling, we first considered assembling a portion of the Penrose tiling with a set of entirely distinct tiles. For this, we chose to test the simplest solution first: each DNA strand describes the outline of a face of the tiling. A representation of this type of assembly for a small portion of the Penrose tiling is given in figure 16.2.

Once the shape design is chosen, the next step is to work on selecting the sequences to use. There are several constraints to respect. To generate them, we use a relatively simple algorithm that draws random sequences and then filters them according to the criteria of interest. In particular, each strand must be unable to fold onto itself or bind to a strand if it is not intended by the design. For this, the set of binding energies between domains that should bind must be higher and intersect as little as possible with the set of binding energies of domain pairs that are normally separated.

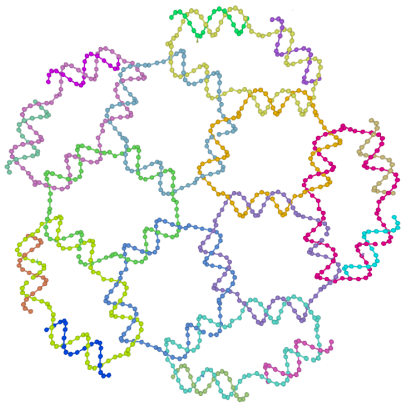


Figure 16.2: Proposed design for the hard-coded self-assembly of the Penrose tiling (shown here with few tiles for readability). Our hope is that it could take the desired shape through electrostatic repulsions, despite the lack of rigidity of the tiles.

In the end, the time constraints of the internship prevented us from finishing designing and ordering those strands. The poor experimental results on the Penrose tiling origami discouraged us from further investigating self-assembly of cut-and-project tilings in DNA, and we decided to start exploring the design methods for multi-scaffold DNA origami.

Conclusion. This work marks only the initial step towards the self-assembly of DNA-based cut-and-project quasicrystals. Adjustments to various experimental elements are needed, and a better understanding of how to adapt the local assembly algorithm to model the self-assembly of DNA bricks is required. Then, finalizing both the design of these DNA bricks and the corresponding experimental protocol will be needed in order to successfully synthesize an aperiodic DNA hydrogel.

Conclusion

I was first introduced to the domain of DNA nanotechnologies and DNA computing during my Master's internship, where I explored theoretical and experimental approaches for the algorithmic assembly of quasi-crystals using DNA strands. I implemented a new greedy algorithm for the local growth of Golden Octagonal tilings, that allowed us to reproduce and support findings in Ilya Galanov's PhD research. Additionally, I explored some experimental leads on how to grow cut-and-project quasicrystals using DNA. While I couldn't achieve all my initial objectives during those few months, this internship was an extensive introduction to several aspects of DNA computing.

This led me to further explore this fascinating and interdisciplinary domain for three more years, by tackling the largely unresolved problem of controlling the growth of multi-scaffold DNA origami. Initially motivated by lifting the size limitation of the single-pot DNA origami technique, this study allowed us to better understand origami folding mechanisms and scaffold differentiation control.

Our first chimeric assemblies discouraged the use of nucleation-based differentiation methods (such as presented by Evans et al. [Eva+24]) for multi-scaffold DNA origami. But, inspired by their concept of a seed of selectively chosen staple strands to facilitate self-assembly, we implemented a different approach where the seed acted as a barrier to prevent undesired conformations rather than as a nucleation source. We proposed several algorithmic selection methods for such seed of staple strands and experimentally evaluated those methods. I used a neural network-based method for computing the yield of each sample, using YOLOv5 for object detection on AFM images to count correctly folded bi-origamis. This allowed us to demonstrate the efficiency of our final seed-selection algorithm, which translated thermodynamic constraints into an integer programming problem.

The progress made in synthesizing custom-sequence scaffold strands has advanced rapidly [Eng+19], with enhanced cost-efficiency and their very recent availability in commercial catalogs such as Tilibit. While this development partially mitigates our initial goal of constructing larger DNA origami with identical scaffold strands, our approach remains valuable for studying and controlling the folding pathways of plural-outcome DNA assemblies. Future work could extend our method to more sophisticated designs, including 3D assemblies or constructs requiring more than two scaffolds.

We also used our two-step seed-based folding method to investigate the assembly of a frustrated shape in DNA origami, and explored an alternative method for scaffold differentiation, using an additional origami as a mold. Although we couldn't fully complete and experimentally validate these two projects, it allowed me to learn transmission electron microscopy, and it was an adequate opportunity to develop a simple protocol for producing a short DNA scaffold from a double-stranded plasmid (which will be useful to our group in the future).

Those two projects can surely benefit from further investigations, for example by redesign-

ing our frustrated shape bi-origami to reduce dependency on experimental parameters, and reattempting to fold our shape-sorter-inspired design using newly ordered DNA strands and freashly prepared buffer solutions.

Lastly, I also conducted complementary studies on modeling and simulating DNA origami self-assembly. I devised a new multiscale yet simplistic model aiming at efficiently simulating DNA origami self-assembly, showing promising first results by capturing several interesting mechanisms, such as nucleation zipper and rapid oscillations around a metastable state until reaching a more stable one. Addressing some of its limitations could further enhance the model's applicability in experimental validation and refinement for DNA nanotechnology applications – a potential avenue for a dedicated PhD thesis.

Appendix A

Materials and methods

DNA strands

All staple strands were purchased from Integrated DNA Technologies (IDT) in IDTE Buffer (pH 8.0, 10 mM Tris-HCl/0.1 mM EDTA) in 96-well plates at a concentration of 200 µM each. M13mp18 scaffold strands were purchased from Tilibit Nanosystems, at a 100 nM (224 µg/mL) concentration. Dissolved in buffer containing 10 mM TRIS-BASE, 1 mM EDTA.

Sample preparation

Unless otherwise stated, DNA origamis were prepared using this reference mix (valid for the first annealing reaction in the case of a multistep folding):

- buffer: 40 mM of Tris-acetate and 1 mM of EDTA (TAE), with 12.5 mM $MgCl_2$,
- DNA strands: 10 nM of M13mp18 scaffold strand and a 5-fold excess of staple strands (the strand concentrations were generally higher for origamis using the sspUC19 scaffold and were specified individually, for each experiment).

Atomic force microscopy

Unless otherwise stated, AFM imaging was carried out as follows: a small volume of sample (0.5 to 10 µL, depending on its concentration) was deposited onto freshly cleaved mica for a total quantity of 20 ng to 60 ng of DNA origami, and 30 µL of TAE buffer with 12.5 mM Mg^{2+} were added. After at least 5 minutes, 2 µL of 25 mM Ni^{2+} solution and 30 µL of TAE/Mg^{2+} buffer were added. AFM images were taken in AC Mode (Fast Imaging) on a JPK NanoWizard 4 microscope with a Fastscan module, using NanoWorld Ultra-Short probes (USC-F0.3-k0.3-10).

Bibliography

[Alb08] Bruce Alberts. *Molecular Biology of the Cell*. Garland Science, 2008. 1734 pp.

[AMM44] Oswald T. Avery, Colin M. MacLeod, and Maclyn McCarty. "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types". In: *The Journal of Experimental Medicine* 79.2 (1944), pp. 137–158.

[And+08] Ebbe S. Andersen, Mingdong Dong, Morten M. Nielsen, Kasper Jahn, Allan Lind-Thomsen, Wael Mamdouh, Kurt V. Gothelf, Flemming Besenbacher, and Jørgen Kjems. "DNA Origami Design of Dolphin-Shaped Structures with Flexible Tails". In: *ACS Nano* 2.6 (2008), pp. 1213–1218. DOI: 10.1021/nn800215j.

[And+09] Ebbe S. Andersen, Mingdong Dong, Morten M. Nielsen, Kasper Jahn, Ramesh Subramani, Wael Mamdouh, Monika M. Golas, Bjoern Sander, Holger Stark, Cristiano L. P. Oliveira, Jan Skov Pedersen, Victoria Birkedal, Flemming Besenbacher, Kurt V. Gothelf, and Jørgen Kjems. "Self-assembly of a nanoscale DNA box with a controllable lid". In: *Nature* 459.7243 (2009), pp. 73–76. DOI: 10.1038/nature07971.

[Ben+15] Erik Benson, Abdulmelik Mohammed, Johan Gardell, Sergej Masich, Eugen Czeizler, Pekka Orponen, and Björn Högberg. "DNA rendering of polyhedral meshes at the nanoscale". In: *Nature* 523.7561 (2015), pp. 441–444. DOI: 10.1038/nature14586.

[Bet+22] J. Gordon Betts, Kelly A. Young, James A. Wise, Eddie Johnson, Brandon Poe, Dean H. Kruse, Oksana Korol, Jody E. Johnson, Mark Womble, and Peter DeSaix. *3.3 The Nucleus and DNA Replication - Anatomy and Physiology 2e | OpenStax*. 2022. URL: https://openstax.org/books/anatomy-and-physiology-2e/pages/3-3-the-nucleus-and-dna-replication (visited on 09/24/2024).

[BQG86] G. Binnig, C. F. Quate, and Ch. Gerber. "Atomic Force Microscope". In: *Physical Review Letters* 56.9 (1986), pp. 930–933. DOI: 10.1103/PhysRevLett.56.930.

[CCD71] Mary E. Craig, Donald M. Crothers, and Paul Doty. "Relaxation kinetics of dimer formation by self complementary oligonucleotides". In: *Journal of Molecular Biology* 62.2 (1971), pp. 383–401. DOI: 10.1016/0022-2836(71)90434-7.

[CD83] D. F. Calef and J. M. Deutch. "Diffusion-Controlled Reactions". In: *Annual Review of Physical Chemistry* 34 (1983), pp. 493–524. DOI: 10.1146/annurev.pc.34.100183.002425.

[CFR22] Alexander Cumberworth, Daan Frenkel, and Aleks Reinhardt. "Simulations of DNA-Origami Self-Assembly Reveal Design-Dependent Nucleation Barriers". In: *Nano Letters* 22.17 (2022), pp. 6916–6922. DOI: 10.1021/acs.nanolett.2c01372.

[Chi+22] Matthew Chiriboga, Christopher M. Green, David A. Hastman, Divita Mathur, Qi Wei, Sebastían A. Díaz, Igor L. Medintz, and Remi Veneziano. "Rapid DNA origami nanostructure detection and classification using the YOLOv5 deep convolutional neural network". In: *Scientific Reports* 12.1 (2022), p. 3871. DOI: 10.1038/s41598-022-07759-3.

[CRF18] Alexander Cumberworth, Aleks Reinhardt, and Daan Frenkel. "Lattice models and Monte Carlo methods for simulating DNA origami self-assembly". In: *The Journal of Chemical Physics* 149.23 (2018), p. 234905. DOI: 10.1063/1.5051835.

[CS91] Junghuei Chen and Nadrian C. Seeman. "Synthesis from DNA of a molecule with the connectivity of a cube". In: *Nature* 350.6319 (1991), pp. 631–633. DOI: 10.1038/350631a0.

[Dan+15] Frits Dannenberg, Katherine E. Dunn, Jonathan Bath, Marta Kwiatkowska, Andrew J. Turberfield, and Thomas E. Ouldridge. "Modelling DNA origami self-assembly at the domain level". In: *The Journal of Chemical Physics* 143.16 (2015), p. 165102. DOI: 10.1063/1.4933426.

[DBC12] Shawn M. Douglas, Ido Bachelet, and George M. Church. "A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads". In: *Science* 335.6070 (2012), pp. 831–834. DOI: 10.1126/science.1214081.

[DDS09] Hendrik Dietz, Shawn M. Douglas, and William M. Shih. "Folding DNA into twisted and curved nanoscale shapes". In: *Science (New York, N.Y.)* 325.5941 (2009), pp. 725–730. DOI: 10.1126/science.1174251.

[DeL+24] Marcello DeLuca, Daniel Duke, Tao Ye, Michael Poirier, Yonggang Ke, Carlos Castro, and Gaurav Arya. "Mechanism of DNA origami folding elucidated by mesoscopic simulations". In: *Nature Communications* 15.1 (2024), p. 3015. DOI: 10.1038/s41467-024-46998-y.

[Der+12] N. D. Derr, B. S. Goodman, R. Jungmann, A. E. Leschziner, W. M. Shih, and S. L. Reck-Peterson. "Tug-of-War in Motor Protein Ensembles Revealed with a Programmable DNA Origami Scaffold". In: *Science* 338.6107 (2012), pp. 662–665. DOI: 10.1126/science.1226734.

[Dey+21] Swarup Dey, Chunhai Fan, Kurt V. Gothelf, Jiang Li, Chenxiang Lin, Longfei Liu, Na Liu, Minke A. D. Nijenhuis, Barbara Saccà, Friedrich C. Simmel, Hao Yan, and Pengfei Zhan. "DNA origami". In: *Nature Reviews Methods Primers* 1.1 (2021), pp. 1–24. DOI: 10.1038/s43586-020-00009-8.

[DLS20] David Doty, Benjamin L. Lee, and Tristan Stérin. "scadnano: A Browser-Based, Scriptable Tool for Designing DNA Nanostructures". In: *26th International Conference on DNA Computing and Molecular Programming (DNA 26) (2020).* 26th International Conference on DNA Computing and Molecular Programming (DNA 26) (2020). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020. DOI: 10.4230/LIPIcs.DNA.2020.9.

[Dou+09a] Shawn M. Douglas, Hendrik Dietz, Tim Liedl, Björn Högberg, Franziska Graf, and William M. Shih. "Self-assembly of DNA into nanoscale three-dimensional shapes". In: *Nature* 459.7245 (2009), pp. 414–418. DOI: 10.1038/nature08016.

[Dou+09b] Shawn M. Douglas, Adam H. Marblestone, Surat Teerapittayanon, Alejandro Vazquez, George M. Church, and William M. Shih. "Rapid prototyping of 3D DNA-origami shapes with caDNAno". In: *Nucleic Acids Research* 37.15 (2009), pp. 5001–5006. DOI: 10.1093/nar/gkp436.

[Doy+13] Jonathan P. K. Doye, Thomas E. Ouldridge, Ard A. Louis, Flavio Romano, Petr Šulc, Christian Matek, Benedict E. K. Snodin, Lorenzo Rovigatti, John S. Schreck, Ryan M. Harrison, and William P. J. Smith. "Coarse-graining DNA for simulations of DNA nanotechnology". In: *Physical Chemistry Chemical Physics* 15.47 (2013), pp. 20395–20414. DOI: 10.1039/C3CP53545B.

[DP04] Robert M. Dirks and Niles A. Pierce. "Triggered amplification by hybridization chain reaction". In: *Proceedings of the National Academy of Sciences* 101.43 (2004), pp. 15275–15278. DOI: 10.1073/pnas.0407024101.

[Dun+15] Katherine E. Dunn, Frits Dannenberg, Thomas E. Ouldridge, Marta Kwiatkowska, Andrew J. Turberfield, and Jonathan Bath. "Guiding the folding pathway of DNA origami". In: *Nature* 525.7567 (2015), pp. 82–86. DOI: 10.1038/nature14860.

[Edw+13] Thomas G. W. Edwardson, Karina M. M. Carneiro, Christopher K. McLaughlin, Christopher J. Serpell, and Hanadi F. Sleiman. "Site-specific positioning of dendritic alkyl chains on DNA cages enables their geometry-dependent self-assembly". In: *Nature Chemistry* 5.10 (2013), pp. 868–875. DOI: 10.1038/nchem.1745.

[Eng+19] Floris A. S. Engelhardt, Florian Praetorius, Christian H. Wachauf, Gereon Brüggenthies, Fabian Kohler, Benjamin Kick, Karoline L. Kadletz, Phuong Nhi Pham, Karl L. Behler, Thomas Gerling, and Hendrik Dietz. "Custom-Size, Functional, and Durable DNA Origami with Design-Specific Scaffolds". In: *ACS Nano* 13.5 (2019), pp. 5015–5027. DOI: 10.1021/acsnano.9b01025.

[Erk+14] Michael Erkelenz, Dennis M. Bauer, Rebecca Meyer, Christos Gatsogiannis, Stefan Raunser, Barbara Saccà, and Christof M. Niemeyer. "A Facile Method for Preparation of Tailored Scaffolds for DNA-Origami". In: *Small* 10.1 (2014), pp. 73–77. DOI: 10.1002/smll.201300701.

[Eva+22] Constantine Glen Evans, Jackson O'Brien, Erik Winfree, and Arvind Murugan. *Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly.* 2022.

[Eva+24] Constantine Glen Evans, Jackson O'Brien, Erik Winfree, and Arvind Murugan. "Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly". In: *Nature* 625.7995 (2024), pp. 500–507. DOI: 10.1038/s41586-023-06890-z.

[Fen+03] Liping Feng, Sung Ha Park, John H. Reif, and Hao Yan. "A Two-State DNA Lattice Switched by DNA Nanoactuator". In: *Angewandte Chemie International Edition* 42.36 (2003), pp. 4342–4346. DOI: 10.1002/anie.200351818.

[FG22] Thomas Fernique and Ilya Galanov. *Purely local growth of a quasicrystal.* 2022. DOI: 10.48550/arXiv.2201.00789.

[Fon+18] Pedro Fonseca, Flavio Romano, John S. Schreck, Thomas E. Ouldridge, Jonathan P. K. Doye, and Ard A. Louis. "Multi-scale coarse-graining for the study of assembly pathways in DNA-brick self-assembly". In: *The Journal of Chemical Physics* 148.13 (2018), p. 134910. DOI: 10.1063/1.5019344.

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

[FS93] Tsu Ju Fu and Nadrian C. Seeman. "DNA double-crossover molecules". In: *Biochemistry* 32.13 (1993), pp. 3211–3220. DOI: 10.1021/bi00064a003.

[Fu+16] Jinglin Fu, Yuhe Renee Yang, Soma Dhakal, Zhao Zhao, Minghui Liu, Ting Zhang, Nils G. Walter, and Hao Yan. "Assembly of multienzyme complexes on DNA nanostructures". In: *Nature Protocols* 11.11 (2016), pp. 2243–2273. DOI: 10.1038/nprot.2016.139.

[Fu+22] Daniel Fu, Raghu Pradeep Narayanan, Abhay Prasad, Fei Zhang, Dewight Williams, John S. Schreck, Hao Yan, and John Reif. "Automated design of 3D DNA origami with non-rasterized 2D curvature". In: *Science Advances* 8.51 (2022), eade4455. DOI: 10.1126/sciadv.ade4455.

[Gal19] Ilya Galanov. "On self-assembly of planar octagonal tilings of finite type". PhD thesis. Université Paris-Nord - Paris XIII, 2019.

[Ger+15] Thomas Gerling, Klaus F. Wagenbauer, Andrea M. Neuner, and Hendrik Dietz. "Dynamic DNA devices and assemblies formed by shape-complementary, non–base pairing 3D components". In: *Science* 347.6229 (2015), pp. 1446–1452. DOI: 10.1126/science.aaa5372.

[Gil77] Daniel T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361. DOI: 10.1021/j100540a008.

[GRA14] Cody Geary, Paul W. K. Rothemund, and Ebbe S. Andersen. "A single-stranded architecture for cotranscriptional folding of RNA nanostructures". In: *Science* 345.6198 (2014), pp. 799–804. DOI: 10.1126/science.1253920.

[Gru+08] Andreas R. Gruber, Ronny Lorenz, Stephan H. Bernhart, Richard Neuböck, and Ivo L. Hofacker. "The Vienna RNA Websuite". In: *Nucleic Acids Research* 36 (2008), W70–W74. DOI: 10.1093/nar/gkn188.

[Han+11] Dongran Han, Suchetan Pal, Jeanette Nangreave, Zhengtao Deng, Yan Liu, and Hao Yan. "DNA Origami with Complex Curvatures in Three-Dimensional Space". In: *Science* 332.6027 (2011), pp. 342–346. DOI: 10.1126/science.1202998.

[Han+17] Dongran Han, Xiaodong Qi, Cameron Myhrvold, Bei Wang, Mingjie Dai, Shuoxing Jiang, Maxwell Bates, Yan Liu, Byoungkwon An, Fei Zhang, Hao Yan, and Peng Yin. "Single-stranded DNA and RNA origami". In: *Science* 358.6369 (2017), eaao2648. DOI: 10.1126/science.aao2648.

[He+20] Kai He, Zhe Li, Longfei Liu, Mengxi Zheng, and Chengde Mao. "Assembly of a DNA Origami Chinese Knot by Only 15% of the Staple Strands". In: *ChemBioChem* 21.15 (2020), pp. 2132–2136. DOI: 10.1002/cbic.202000106.

[Hol64] Robin Holliday. "A mechanism for gene conversion in fungi". In: *Genetics Research* 5.2 (1964), pp. 282–304. DOI: 10.1017/S0016672300001233.

[Hua+21] Chao-Min Huang, Anjelica Kucinic, Joshua A. Johnson, Hai-Jun Su, and Carlos E. Castro. "Integrated computer-aided engineering and design for DNA assemblies". In: *Nature Materials* 20.9 (2021), pp. 1264–1271. DOI: 10.1038/s41563-021-00978-5.

[JKC21] Joshua A. Johnson, Vasiliki Kolliopoulos, and Carlos E. Castro. "Co-self-assembly of multiple DNA origami nanostructures in a single pot". In: *Chemical Communications* 57.39 (2021), pp. 4795–4798. DOI: 10.1039/D1CC00049G.

[Jun+10] Ralf Jungmann, Christian Steinhauer, Max Scheible, Anton Kuzyk, Philip Tinnefeld, and Friedrich C. Simmel. "Single-Molecule Kinetics and Super-Resolution Microscopy by Fluorescence Imaging of Transient Binding on DNA Origami". In: *Nano Letters* 10.11 (2010), pp. 4756–4761. DOI: 10.1021/nl103427w.

[Jun+14] Ralf Jungmann, Maier S. Avendaño, Johannes B. Woehrstein, Mingjie Dai, William M. Shih, and Peng Yin. "Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT". In: *Nature Methods* 11.3 (2014), pp. 313–318. DOI: 10.1038/nmeth.2835.

[Ke+12] Yonggang Ke, Luvena L. Ong, William M. Shih, and Peng Yin. "Three-Dimensional Structures Self-Assembled from DNA Bricks". In: *Science* 338.6111 (2012), pp. 1177–1183. DOI: 10.1126/science.1227268.

[Kim+12] Do-Nyun Kim, Fabian Kilchherr, Hendrik Dietz, and Mark Bathe. "Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures". In: *Nucleic Acids Research* 40.7 (2012), pp. 2862–2868. DOI: 10.1093/nar/gkr1173.

[KMS83] Neville R. Kallenbach, Rong-Ine Ma, and Nadrian C. Seeman. "An immobile nucleic acid junction constructed from oligonucleotides". In: *Nature* 305.5937 (1983), pp. 829–831. DOI: 10.1038/305829a0.

[Kuz+12] Anton Kuzyk, Robert Schreiber, Zhiyuan Fan, Günther Pardatscher, Eva-Maria Roller, Alexander Högele, Friedrich C. Simmel, Alexander O. Govorov, and Tim Liedl. "DNA-based self-assembly of chiral plasmonic nanostructures with tailored optical response". In: *Nature* 483.7389 (2012), pp. 311–314. DOI: 10.1038/nature10889.

[Lan+12] Martin Langecker, Vera Arnaut, Thomas G. Martin, Jonathan List, Stephan Renner, Michael Mayer, Hendrik Dietz, and Friedrich C. Simmel. "Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures". In: *Science* 338.6109 (2012), pp. 932–936. DOI: 10.1126/science.1225624.

[Lee+12] Hyukjin Lee, Abigail K. R. Lytton-Jean, Yi Chen, Kevin T. Love, Angela I. Park, Emmanouil D. Karagiannis, Alfica Sehgal, William Querbes, Christopher S. Zurenko, Muthusamy Jayaraman, Chang G. Peng, Klaus Charisse, Anna Borodovsky, Muthiah Manoharan, Jessica S. Donahoe, Jessica Truelove, Matthias Nahrendorf, Robert Langer, and Daniel G. Anderson. "Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery". In: *Nature Nanotechnology* 7.6 (2012), pp. 389–393. DOI: 10.1038/nnano.2012.73.

[Lee+16] Jonathan Lee Tin Wah, Christophe David, Sergii Rudiuk, Damien Baigl, and André Estevez-Torres. "Observing and Controlling the Folding Pathway of DNA Origami at the Nanoscale". In: *ACS Nano* 10.2 (2016), pp. 1978–1987. DOI: 10.1021/acsnano.5b05972.

[Lev23] Nicolas Levy. "Conception de nanostructures ADN par des méthodes géométriques : Modèles, logiciel et expériences". PhD thesis. Ecole normale supérieure de lyon - ENS LYON, 2023.

[Li+96] Xiaojun Li, Xiaoping Yang, Jing Qi, and Nadrian C. Seeman. "Antiparallel DNA Double Crossover Molecules As Components for Nanoconstruction". In: *Journal of the American Chemical Society* 118.26 (1996), pp. 6131–6140. DOI: 10.1021/ja960162o.

[Liu+11] Wenyan Liu, Hong Zhong, Risheng Wang, and Nadrian C. Seeman. "Crystalline Two-Dimensional DNA-Origami Arrays". In: *Angewandte Chemie International Edition* 50.1 (2011), pp. 264–267. DOI: 10.1002/anie.201005911.

[Liu+19] Longfei Liu, Zhe Li, Yulin Li, and Chengde Mao. "Rational Design and Self-Assembly of Two-Dimensional, Dodecagonal DNA Quasicrystals". In: *Journal of the American Chemical Society* 141.10 (2019), pp. 4248–4251. DOI: 10.1021/jacs.9b00843.

[Lla+20] Elisa de Llano, Haichao Miao, Yasaman Ahmadi, Amanda J Wilson, Morgan Beeby, Ivan Viola, and Ivan Barisic. "Adenita: interactive 3D modelling and visualization of DNA nanostructures". In: *Nucleic Acids Research* 48.15 (2020), pp. 8269–8275. DOI: 10.1093/nar/gkaa593.

[LS21] Nicolas Levy and Nicolas Schabanel. "ENSnano: A 3D Modeling Software for DNA Nanostructures". In: *27th International Conference on DNA Computing and Molecular Programming (DNA 27) (2021)*. 27th International Conference on DNA Computing and Molecular Programming (DNA 27) (2021). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. DOI: 10.4230/LIPIcs.DNA.27.5.

[MA20] Christopher Maffeo and Aleksei Aksimentiev. "MrDNA: a multi-resolution model for predicting the structure and dynamics of DNA systems". In: *Nucleic Acids Research* 48.9 (2020), pp. 5135–5146. DOI: 10.1093/nar/gkaa200.

[Mar+16] A. E. Marras, L. Zhou, V. Kolliopoulos, H.-J. Su, and C. E. Castro. "Directing folding pathways for multi-component DNA origami nanostructures with complex topology". In: *New Journal of Physics* 18.5 (2016), p. 055005. DOI: 10.1088/1367-2630/18/5/055005.

[MKT21] Ryan J. Menssen, Gregory J. Kimmel, and Andrei Tokmakoff. "Investigation into the mechanism and dynamics of DNA association and dissociation utilizing kinetic Monte Carlo simulations". In: *The Journal of Chemical Physics* 154.4 (2021), p. 045101. DOI: 10.1063/5.0035187.

[Mon+22] Alba Monferrer, Jessica A. Kretzmann, Christian Sigl, Pia Sapelza, Anna Liedl, Barbara Wittmann, and Hendrik Dietz. "Broad-Spectrum Virus Trapping with Heparan Sulfate-Modified DNA Origami Shells". In: *ACS Nano* 16.12 (2022), pp. 20002–20009. DOI: 10.1021/acsnano.1c11328.

[Mon+23] Alba Monferrer, Fabian Kohler, Christian Sigl, Michael Schachtner, David Peterhoff, Benedikt Asbach, Ralf Wagner, and Hendrik Dietz. "DNA origami traps for large viruses". In: *Cell Reports Physical Science* 4.1 (2023), p. 101237. DOI: 10.1016/j.xcrp.2022.101237.

[MYA16] Christopher Maffeo, Jejoong Yoo, and Aleksei Aksimentiev. "De novo reconstruction of DNA origami structures through atomistic molecular dynamics simulation". In: *Nucleic Acids Research* 44.7 (2016), pp. 3013–3019. DOI: 10.1093/nar/gkw155.

[N G81] De N. G. Bruijn. "Algebraic theory of Penrose's non-periodic tilings of the plane. I". In: *Indagationes Mathematicae (Proceedings)* 84.1 (1981), pp. 39–52. DOI: 10.1016/1385-7258(81)90016-0.

[Par+16] Nina Parker, Mark Schneegurt, Anh-Hue Thi Tu, Philip Lister, and Brian M. Forster. *10.2 Structure and Function of DNA - Microbiology | OpenStax*. 2016. URL: https://openstax.org/books/microbiology/pages/10-2-structure-and-function-of-dna (visited on 09/23/2024).

[Pos+24] Gregor Posnjak, Xin Yin, Paul Butler, Oliver Bienek, Mihir Dass, Seungwoo Lee, Ian D. Sharp, and Tim Liedl. "Diamond-lattice photonic crystals assembled from DNA origami". In: *Science* 384.6697 (2024), pp. 781–785. DOI: 10.1126/science.adl2733.

[Raj+11] Arivazhagan Rajendran, Masayuki Endo, Yousuke Katsuda, Kumi Hidaka, and Hiroshi Sugiyama. "Programmed Two-Dimensional Self-Assembly of Multiple DNA Origami Jigsaw Pieces". In: *ACS Nano* 5.1 (2011), pp. 665–671. DOI: 10.1021/nn1031627.

[Ram+19] Saminathan Ramakrishnan, Leonard Schärfen, Kristin Hunold, Sebastian Fricke, Guido Grundmeier, Michael Schlierf, Adrian Keller, and Georg Krainer. "Enhancing the stability of DNA origami nanostructures: staple strand redesign versus enzymatic ligation". In: *Nanoscale* 11.35 (2019), pp. 16270–16276. DOI: 10.1039/C9NR04460D.

[Red+16] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

[Rot06] Paul W. K. Rothemund. "Folding DNA to create nanoscale shapes and patterns". In: *Nature* 440.7082 (2006), pp. 297–302. DOI: 10.1038/nature04586.

[RPW04] Paul W. K. Rothemund, Nick Papadakis, and Erik Winfree. "Algorithmic Self-Assembly of DNA Sierpinski Triangles". In: *PLOS Biology* 2.12 (2004), e424. DOI: 10.1371/journal.pbio.0020424.

[Sae84] Wolfram Saenger. *Principles of Nucleic Acid Structure*. Red. by Charles R. Cantor. Springer Advanced Texts in Chemistry. New York, NY: Springer, 1984. DOI: 10.1007/978-1-4612-5190-3.

[San98] John SantaLucia. "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics". In: *Proceedings of the National Academy of Sciences* 95.4 (1998), pp. 1460–1465. DOI: 10.1073/pnas.95.4.1460.

[See82] Nadrian C. Seeman. "Nucleic acid junctions and lattices". In: *Journal of Theoretical Biology* 99.2 (1982), pp. 237–247. DOI: 10.1016/0022-5193(82)90002-9.

[SH04] John SantaLucia and Donald Hicks. "The Thermodynamics of DNA Structural Motifs". In: *Annual Review of Biophysics and Biomolecular Structure* 33.1 (2004), pp. 415–440. DOI: 10.1146/annurev.biophys.32.110601.141800.

[She+84] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn. "Metallic Phase with Long-Range Orientational Order and No Translational Symmetry". In: *Physical Review Letters* 53.20 (1984), pp. 1951–1953. DOI: 10.1103/PhysRevLett.53.1951.

[SMD19] Fabian Schneider, Natalie Möritz, and Hendrik Dietz. "The sequence of events during folding of a DNA origami". In: *Science Advances* 5.5 (2019), eaaw1412. DOI: 10.1126/sciadv.aaw1412.

[Sno+16] Benedict E. K. Snodin, Flavio Romano, Lorenzo Rovigatti, Thomas E. Ouldridge, Ard A. Louis, and Jonathan P. K. Doye. "Direct Simulation of the Self-Assembly of a Small DNA Origami". In: *ACS Nano* 10.2 (2016), pp. 1724–1737. DOI: 10.1021/acsnano.5b05865.

[SS04] William B. Sherman and Nadrian C. Seeman. "A Precisely Controlled DNA Biped Walking Device". In: *Nano Letters* 4.7 (2004), pp. 1203–1207. DOI: 10.1021/nl049527q.

[SS17] Nadrian C. Seeman and Hanadi F. Sleiman. "DNA nanotechnology". In: *Nature Reviews Materials* 3.1 (2017), pp. 1–23. DOI: 10.1038/natrevmats.2017.68.

[ST19] Paul J. Sanstead and Andrei Tokmakoff. "A lattice model for the interpretation of oligonucleotide hybridization experiments". In: *The Journal of Chemical Physics* 150.18 (2019), p. 185104. DOI: 10.1063/1.5092526.

[STW15] Joseph Malcolm Schaeffer, Chris Thachuk, and Erik Winfree. "Stochastic Simulation of the Kinetics of Multiple Interacting Nucleic Acid Strands". In: *DNA Computing and Molecular Programming*. Ed. by Andrew Phillips and Peng Yin. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 194–211. DOI: 10.1007/978-3-319-21999-8_13.

[Sun+14] Wei Sun, Etienne Boulais, Yera Hakobyan, Wei Li Wang, Amy Guan, Mark Bathe, and Peng Yin. "Casting inorganic structures with DNA molds". In: *Science* 346.6210 (2014), p. 1258361. DOI: 10.1126/science.1258361.

[Sun+19] Noah R. Sundah, Nicholas R. Y. Ho, Geok Soon Lim, Auginia Natalia, Xianguang Ding, Yu Liu, Ju Ee Seet, Ching Wan Chan, Tze Ping Loh, and Huilin Shao. "Barcoded DNA nanostructures for the multiplexed profiling of subcellular protein distribution". In: *Nature Biomedical Engineering* 3.9 (2019), pp. 684–694. DOI: 10.1038/s41551-019-0417-0.

[Thu+17] Anupama J. Thubagere, Wei Li, Robert F. Johnson, Zibo Chen, Shayan Doroudi, Yae Lim Lee, Gregory Izatt, Sarah Wittman, Niranjan Srinivas, Damien Woods, Erik Winfree, and Lulu Qian. "A cargo-sorting DNA robot". In: *Science* 357.6356 (2017). DOI: 10.1126/science.aan6558.

[TPQ17] Grigory Tikhomirov, Philip Petersen, and Lulu Qian. "Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns". In: *Nature* 552.7683 (2017), pp. 67–71. DOI: 10.1038/nature24655.

[Wan+19] Wen Wang, Silian Chen, Byoungkwon An, Kai Huang, Tanxi Bai, Mengyuan Xu, Gaëtan Bellot, Yonggang Ke, Ye Xiang, and Bryan Wei. "Complex wireframe DNA nanostructures from simple building blocks". In: *Nature Communications* 10.1 (2019), p. 1067. DOI: 10.1038/s41467-019-08647-7.

[Wan+22] Jianhua Wang, Yuhui Wei, Ping Zhang, Yue Wang, Qinglin Xia, Xiaoguo Liu, Shihua Luo, Jiye Shi, Jun Hu, Chunhai Fan, Bin Li, Lihua Wang, Xingfei Zhou, and Jiang Li. "Probing Heterogeneous Folding Pathways of DNA Origami Self-Assembly at the Molecular Level with Atomic Force Microscopy". In: *Nano Letters* 22.17 (2022), pp. 7173–7179. DOI: 10.1021/acs.nanolett.2c02447.

[WB04] Erik Winfree and Renat Bekbolatov. "Proofreading Tile Sets: Error Correction for Algorithmic Self-Assembly". In: *DNA Computing*. Ed. by Junghuei Chen and John Reif. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004, pp. 126–144. DOI: 10.1007/978-3-540-24628-2_13.

[WC53] J. D. Watson and F. H. C. Crick. "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". In: *Nature* 171.4356 (1953), pp. 737–738. DOI: 10.1038/171737a0.

[WDY12] Bryan Wei, Mingjie Dai, and Peng Yin. "Complex shapes self-assembled from single-stranded DNA tiles". In: *Nature* 485.7400 (2012), pp. 623–626. DOI: 10.1038/nature11075.

[Wil+09a] Sean Williams, Kyle Lund, Chenxiang Lin, Peter Wonka, Stuart Lindsay, and Hao Yan. "Tiamat: A Three-Dimensional Editing Tool for Complex DNA Structures". In: *DNA Computing*. Ed. by Ashish Goel, Friedrich C. Simmel, and Petr Sosík. Berlin, Heidelberg: Springer, 2009, pp. 90–101. DOI: 10.1007/978-3-642-03076-5_8.

[Wil+09b] Ofer I. Wilner, Yossi Weizmann, Ron Gill, Oleg Lioubashevski, Ronit Freeman, and Itamar Willner. "Enzyme cascades activated on topologically programmed DNA scaffolds". In: *Nature Nanotechnology* 4.4 (2009), pp. 249–254. DOI: 10.1038/nnano.2009.50.

[Win+23] Christopher M. Wintersinger, Dionis Minev, Anastasia Ershova, Hiroshi M. Sasaki, Gokul Gowri, Jonathan F. Berengut, F. Eduardo Corea-Dilbert, Peng Yin, and William M. Shih. "Multi-micron crisscross structures grown from DNA-origami slats". In: *Nature Nanotechnology* 18.3 (2023), pp. 281–289. DOI: 10.1038/s41565-022-01283-1.

[Win+98] Erik Winfree, Furong Liu, Lisa A. Wenzler, and Nadrian C. Seeman. "Design and self-assembly of two-dimensional DNA crystals". In: *Nature* 394.6693 (1998), pp. 539–544. DOI: 10.1038/28998.

[Woo+19] Damien Woods, David Doty, Cameron Myhrvold, Joy Hui, Felix Zhou, Peng Yin, and Erik Winfree. "Diverse and robust molecular algorithms using reprogrammable DNA self-assembly". In: *Nature* 567.7748 (2019), pp. 366–372. DOI: 10.1038/s41586-019-1014-9.

[WR11] Sungwook Woo and Paul W. K. Rothemund. "Programmable molecular recognition based on the geometry of DNA nanostructures". In: *Nature Chemistry* 3.8 (2011), pp. 620–627. DOI: 10.1038/nchem.1070.

[WSD17] Klaus F. Wagenbauer, Christian Sigl, and Hendrik Dietz. "Gigadalton-scale shape-programmable DNA assemblies". In: *Nature* 552.7683 (2017), pp. 78–83. DOI: 10.1038/nature24651.

[Yin+08] Peng Yin, Rizal F. Hariadi, Sudheer Sahu, Harry M. T. Choi, Sung Ha Park, Thomas H. LaBean, and John H. Reif. "Programming DNA Tube Circumferences". In: *Science* 321.5890 (2008), pp. 824–826. DOI: 10.1126/science.1157312.

[Yur+00] Bernard Yurke, Andrew J. Turberfield, Allen P. Mills, Friedrich C. Simmel, and Jennifer L. Neumann. "A DNA-fuelled molecular machine made of DNA". In: *Nature* 406.6796 (2000), pp. 605–608. DOI: 10.1038/35020524.

[Zad+11] Joseph N. Zadeh, Conrad D. Steenberg, Justin S. Bois, Brian R. Wolfe, Marshall B. Pierce, Asif R. Khan, Robert M. Dirks, and Niles A. Pierce. "NUPACK: Analysis and design of nucleic acid systems". In: *Journal of Computational Chemistry* 32.1 (2011), pp. 170–173. DOI: 10.1002/jcc.21596.

[Zha+15] Fei Zhang, Shuoxing Jiang, Siyu Wu, Yulin Li, Chengde Mao, Yan Liu, and Hao Yan. "Complex wireframe DNA origami nanostructures with multi-arm junction vertices". In: *Nature Nanotechnology* 10.9 (2015), pp. 779–784. DOI: 10.1038/nnano.2015.162.

[Zha+18] Tao Zhang, Caroline Hartl, Kilian Frank, Amelie Heuer-Jungemann, Stefan Fischer, Philipp C. Nickels, Bert Nickel, and Tim Liedl. "3D DNA Origami Crystals". In: *Advanced Materials* 30.28 (2018), p. 1800273. DOI: 10.1002/adma.201800273.

[Zha+22] Pengfei Zhan, Kevin Jahnke, Na Liu, and Kerstin Göpfrich. "Functional DNA-based cytoskeletons for synthetic cells". In: *Nature Chemistry* 14.8 (2022), pp. 958–963. DOI: 10.1038/s41557-022-00945-w.

List of Figures

1 Représentation schématique d'un origami ADN vii

1.1 Structure of the DNA molecule 3
1.2 DNA replication fork 4
1.3 Holliday Junction 5
1.4 Woodcut *Depth* 5
1.5 DNA self-assembly into a regular lattice 6
1.6 DNA tile assembly 7
1.7 Strand displacement and nanomechanical devices 8
1.8 Biomedical and imaging applications of DNA nanotechnologies 9
1.9 A timeline of the field of DNA nanotechnology 10

2.1 DNA origami principle 11
2.2 Some of the first examples of DNA origami 12
2.3 ENSnano 13
2.4 DNA origami assembly steps 14
2.5 3D DNA origami structures 15
2.6 Multicomponent DNA origami 16

3.1 Repetitive scaffold origami design 20
3.2 Co-self-assembly of multiple DNA origami nanostructures in a single pot . . 21
3.3 Controlling the nucleation of SST structures 22

4.1 Principle of the two twin-domain binding staples 23
4.2 Binding two identical double-layer square origamis 24
4.3 Agarose gel electrophoresis of the paired scaffolds and symetrical bi-origami . 25

5.1 Ring-and-Square bi-origami design 27
5.2 Separately folded ◎ and □ DNA origamis 29
5.3 Chimeras obtained when co-folding the ◎□-origami 30

6.1 2-steps cofolding annealing 36
6.2 Illustration of the penalty p 36
6.3 AFM result of the greedy algorithm 38

7.1 Gel electrophoresis of the different stages of assembly 41

8.1 Neural network training and origami detection 44
8.2 Surface computation for folded DNA and reference ◎□-origamis 45

9.1 Origami scaffold preparation 50
9.2 Origami scaffold preparation 50

9.3 Gel electrophoresis at each step of the production of sspUC19 51
9.4 A DNA origami using sspUC19 52

10.1 Paper fans bi-origami design sketch 54
10.2 Corrugami design 56
10.3 AFM images of the corrugami design 57
10.4 TEM image of the 1-corrugami and 2D class average 58

11.1 A shape-sorter toy 61
11.2 Mold, mini-ring and mini-square origami design 62
11.3 AFM images of each component of the design 63

12.1 Application of the unified nearest neighbor model 68
12.2 Kinetic model reaction scheme 69
12.3 Transitions in the domain level model from [Dan+15] 70
12.4 DNA strand complexes, with and without pseudoknot 71
12.5 Experimental data on individual staple folding time from [SMD19] 72

13.1 State transitions 74
13.2 KMC scheme 76
13.3 Example of simulation execution 79

14.1 Penrose tiling 83
14.2 Exemple of a canonical $2 \to 1$ cut-and-project construction 84
14.3 $4 \to 2$ cut-and-projet construction of the Golden tiling 85
14.4 Some patterns from the radius 1 atlas of the Golden Tiling 86
14.5 Intermediate results of incremental assembly 87

15.1 Diagram of the chosen path for the scaffold DNA strand 89
15.2 Final design of the origami representing the Penrose tiling 90
15.3 AFM images of the Penrose DNA Origami designed during the intership . . 91
15.4 Agarose gel electrophoresis of the origami (1% agarose) 92

16.1 Examples of tile design for implementing local rules 94
16.2 Proposed design for the hard-coded self-assembly of the Penrose tiling 94

List of Tables

5.1 Some of the one-pot folding attempts for the ring-and-square bi-origami design 31

7.1 Summary of the experimental results 42

8.1 Computed yield results for each seed algorithm 45

List of Algorithms

1 KMC algorithm 77

2 Randomized construction 87
3 Deterministic Construction 87

Index

AFM, **14**

Bi-origami, **19**, 23, 27
    ⊙□-origami, **27**
    Corrugami, **53**

Cut-and-project (tiling), **83**

DNA, **3**
DNA Origami, **11**

ENSnano, **12**

Gel electrophoresis, **13**

ILP (Integer Linear Program), 39

KMC (Kinetic Monte Carlo), 69, **76**

M13mp18, **11**, 99

Nt.BspQI, **49**

pUC19, **49**
    sspUC19, **49**

Seed (staple strands), **35**
SST, **6**

T7 exonuclease, **50**
TEM, **14**