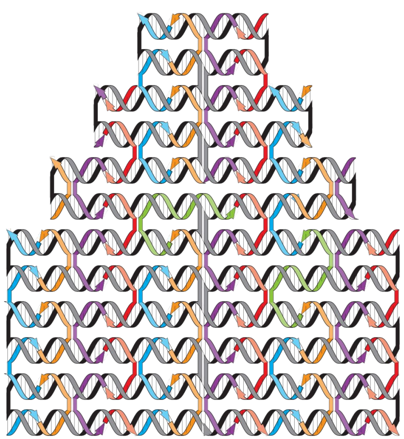
A stochastic and geometrical model for DNA origami self-assembly

Octave Hazard
(PhD 2021-2024)



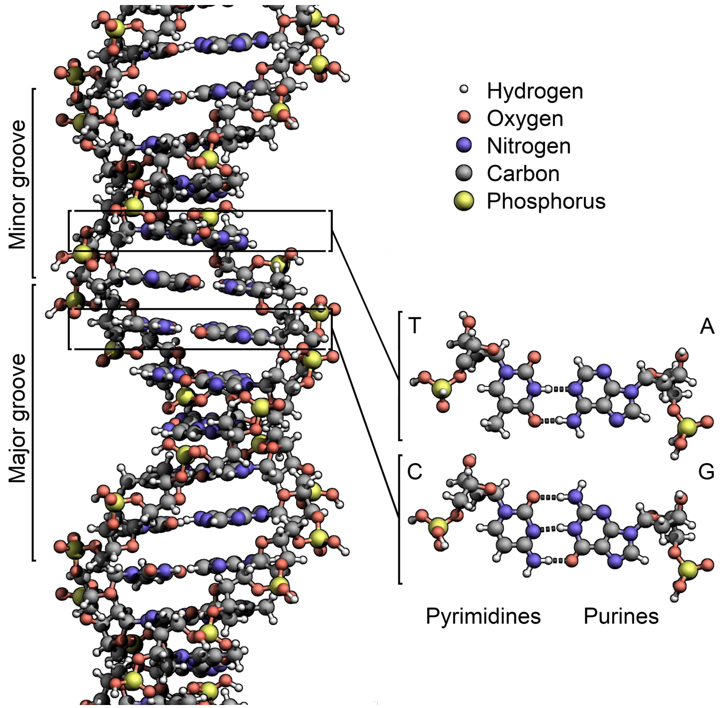
UNIVERSITÉ DE LYON | ENS DE LYON

LIP

Lab PHYS ENS de LYON

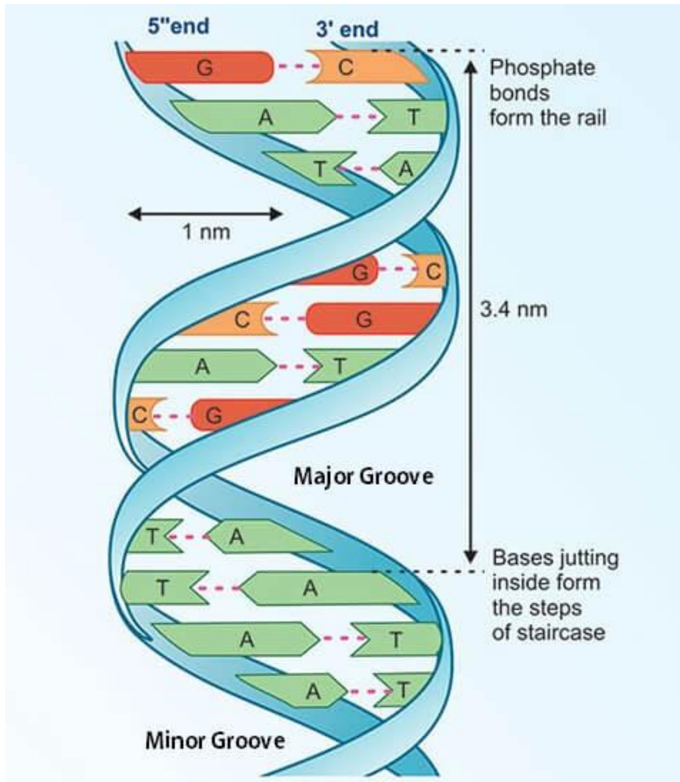
LBMC

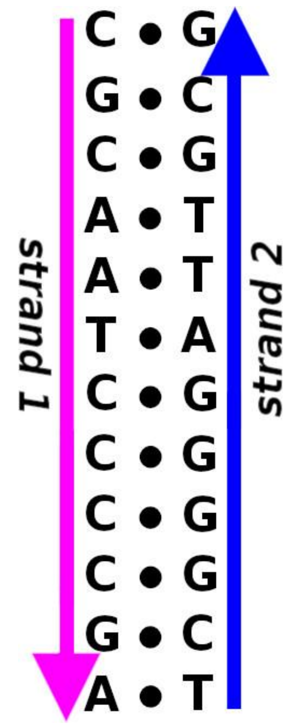
DNA structure



Double helix structure

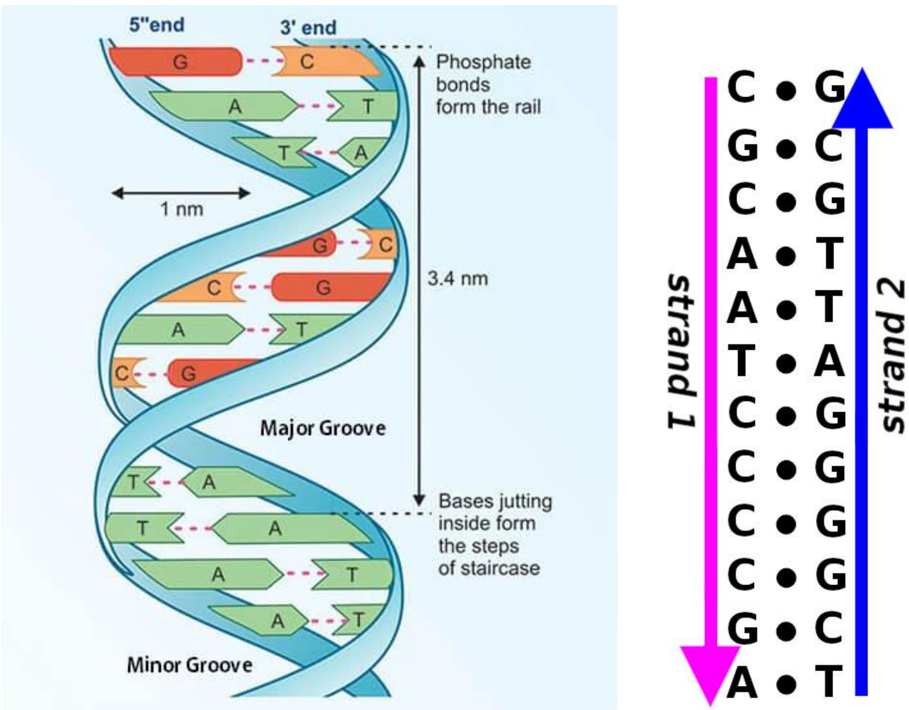
DNA structure



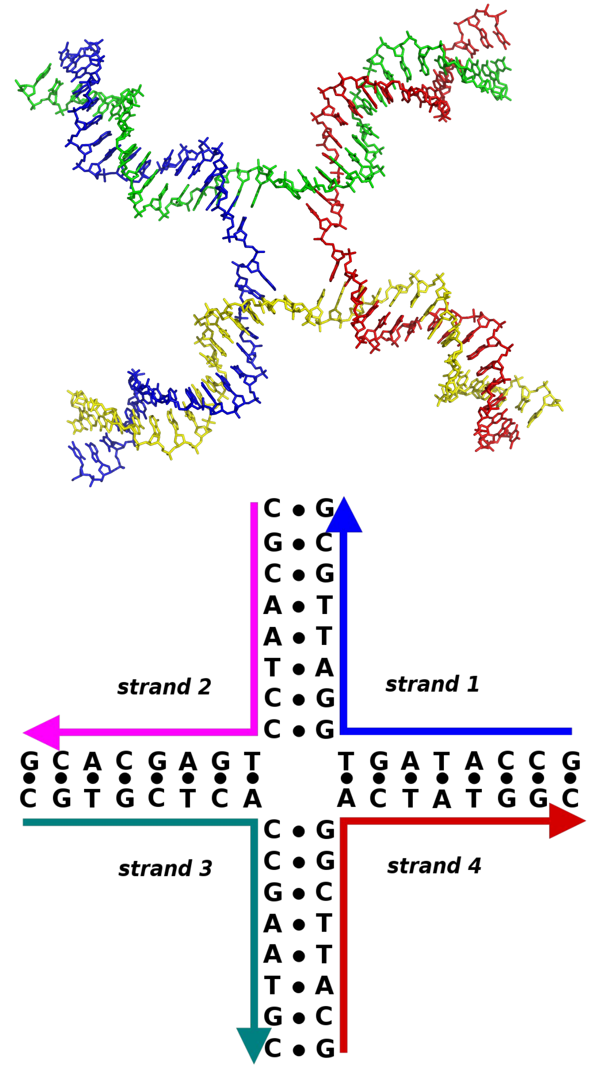


Double helix structure

DNA structure

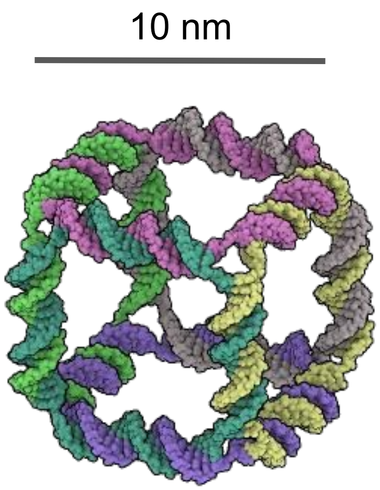


Double helix structure

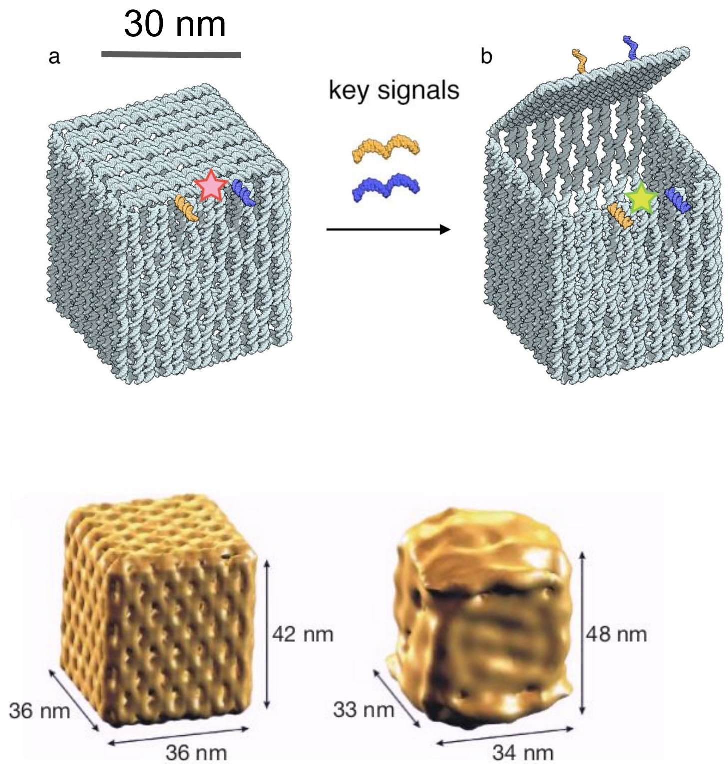


Example of junction (here with 4 arms)

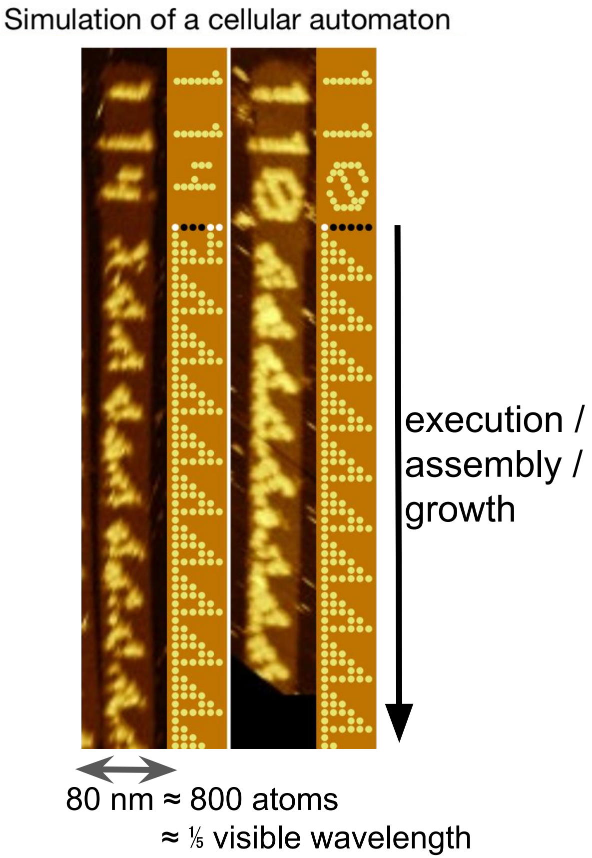
From DNA nanostructures to DNA computing



Chen & Seeman,
Nature 1991

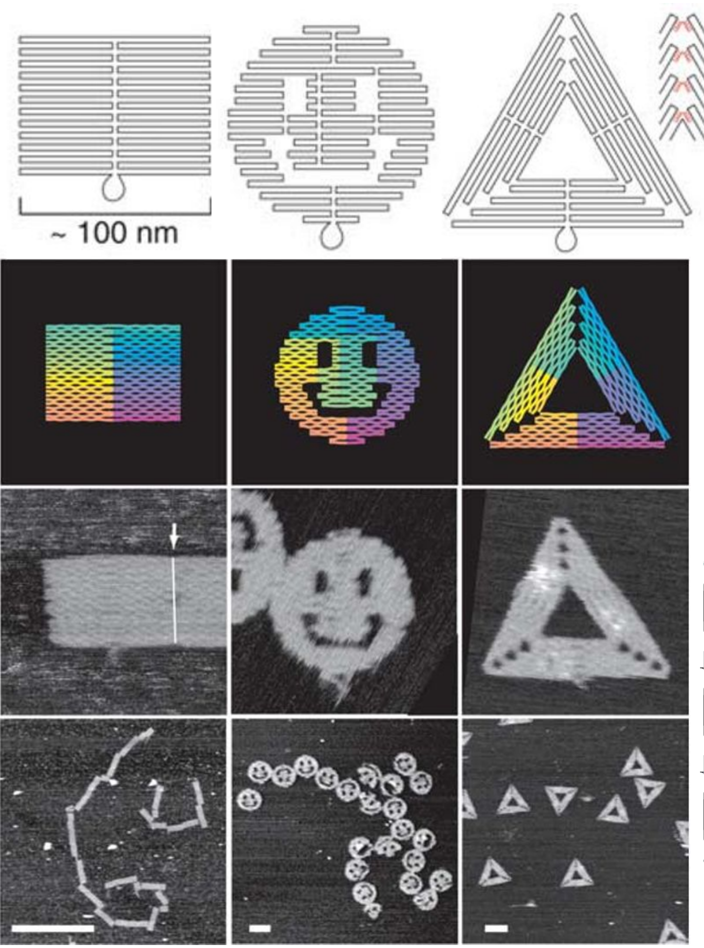


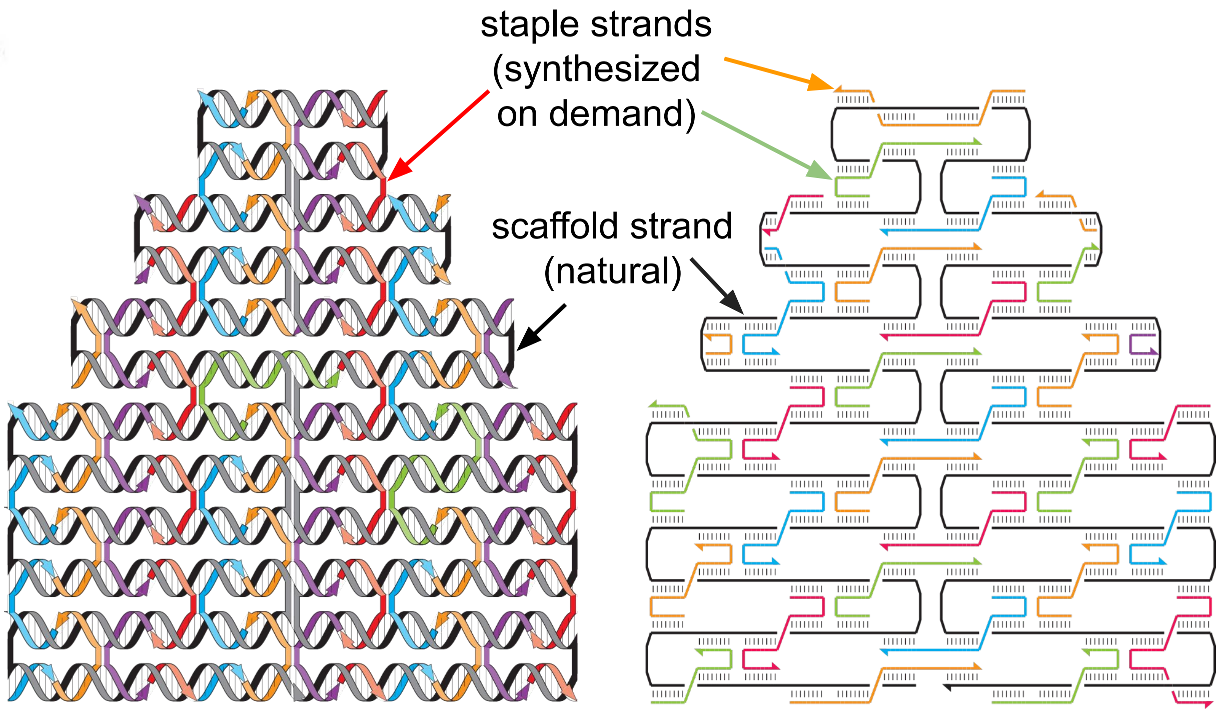
Andersen et al, *Nature* 2009



Woods et al, *Nature* 2019

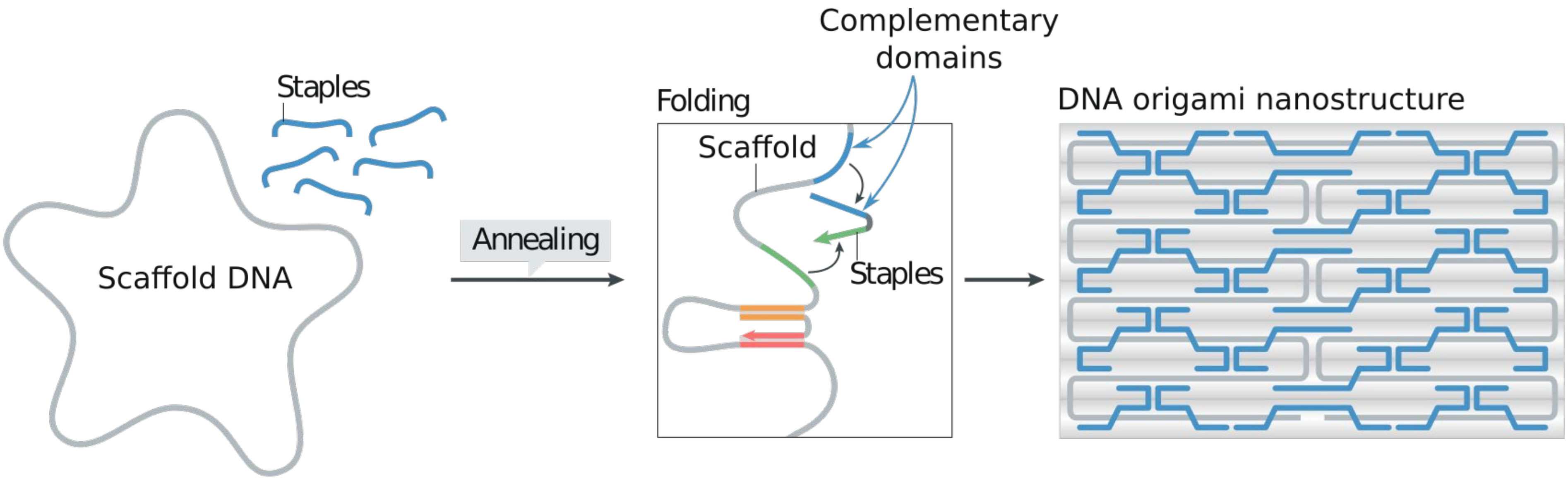
DNA Origami





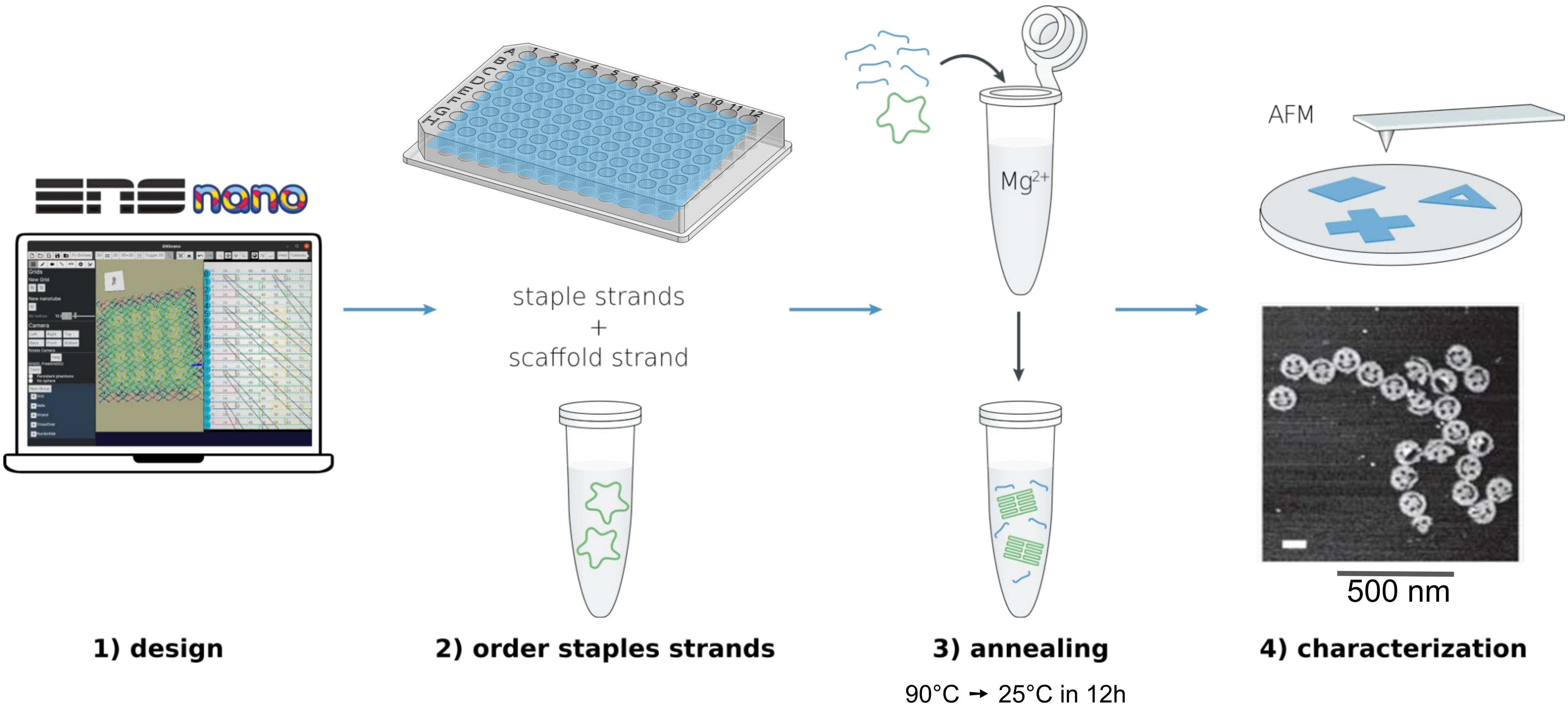
Rothemund, P. “Folding DNA to create nanoscale shapes and patterns.” *Nature* 440, 297–302 (2006).

DNA Origami



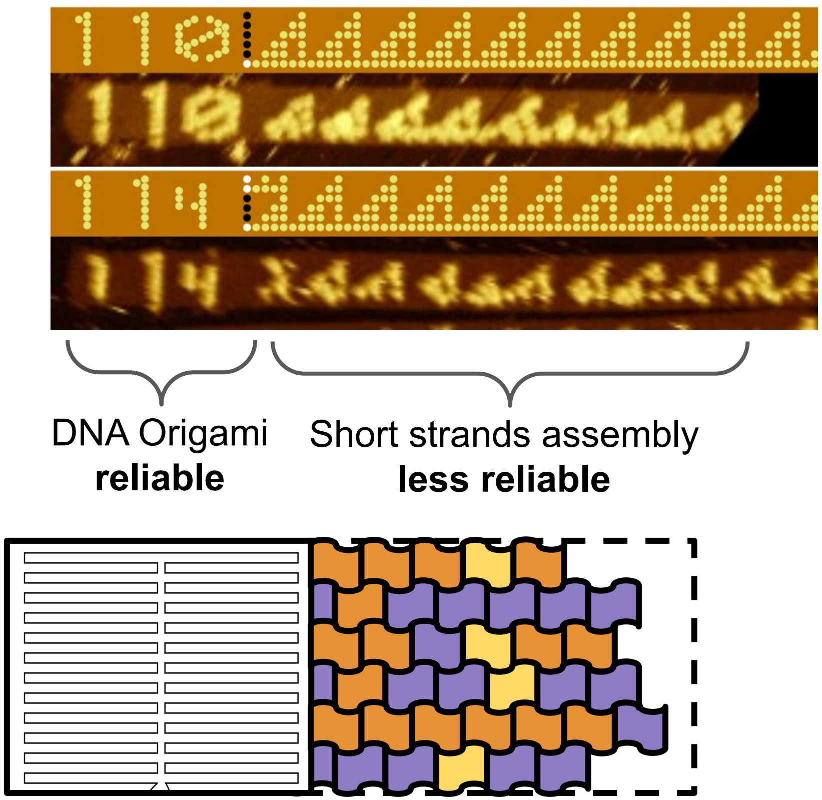
Staples
Scaffold DNA
Annealing
Folding
Scaffold
Complementary domains
Staples
DNA origami nanostructure

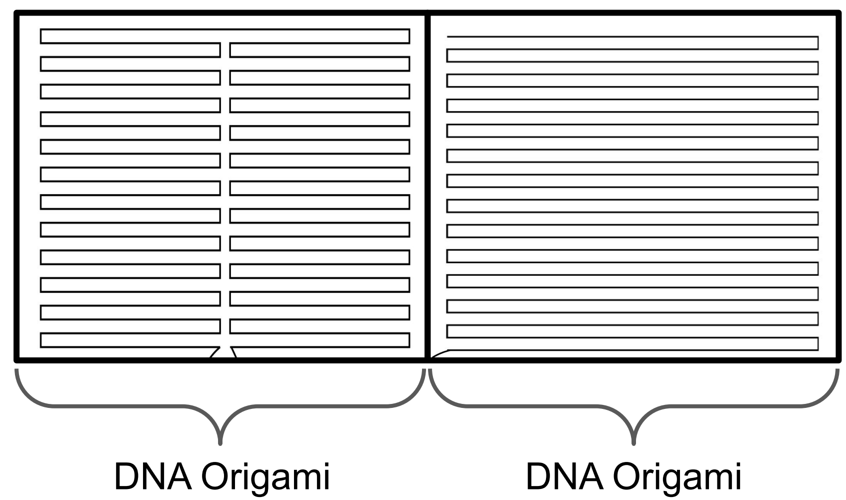
DNA Origami



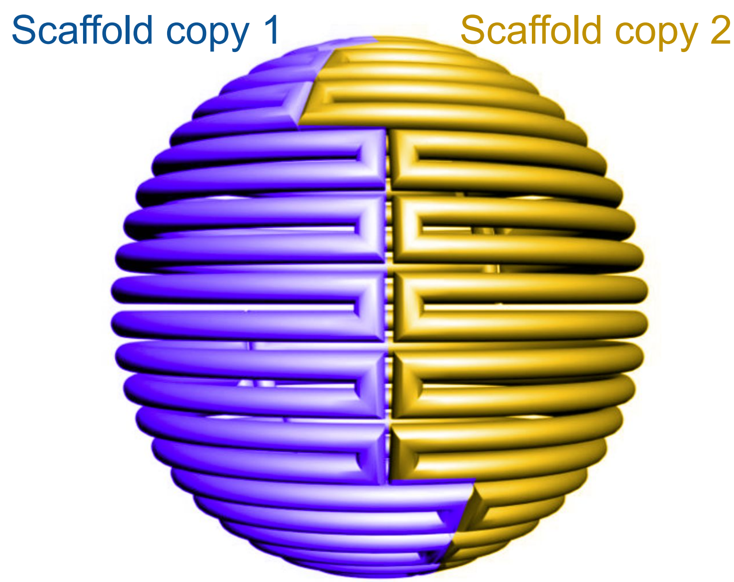
staple strands
+
scaffold strand
Mg2+
AFM
500 nm
1) design
2) order staples strands
3) annealing
90°C → 25°C in 12h
4) characterization

Building larger DNA structures



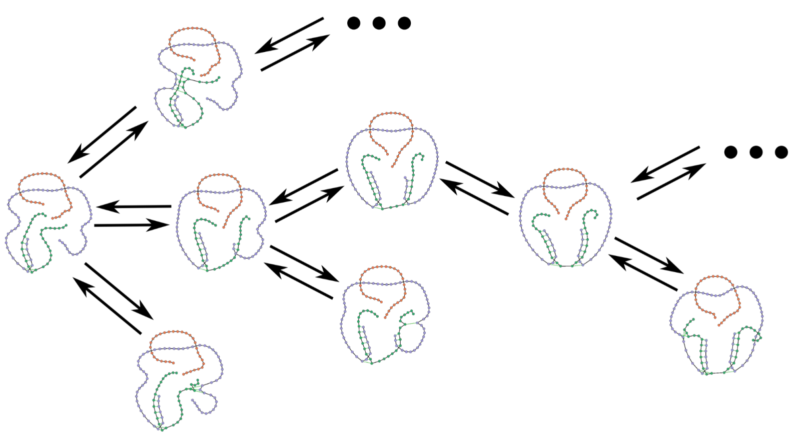


My project and motivations

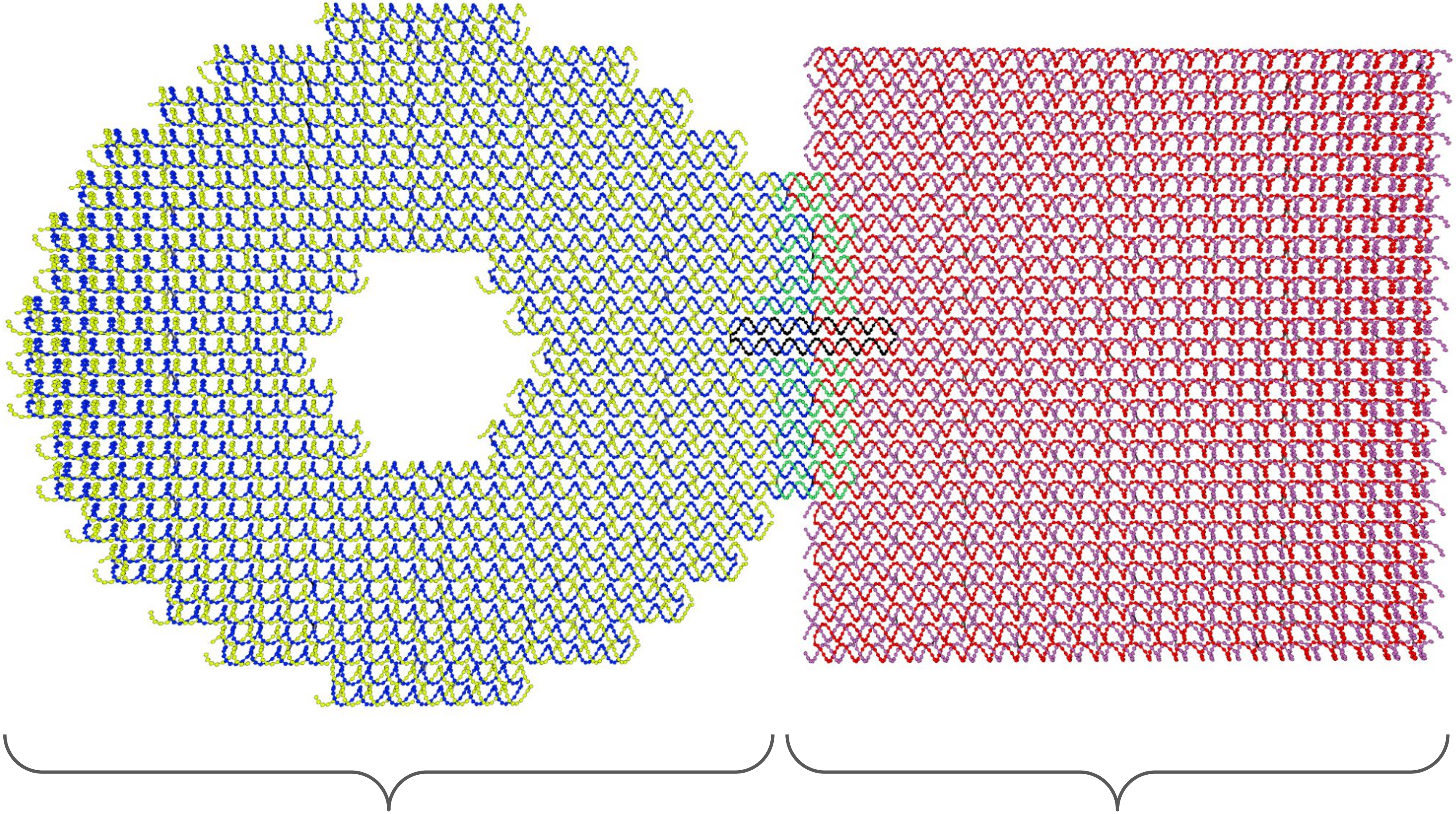


- Building larger DNA origami structures using several (identical) scaffold strands

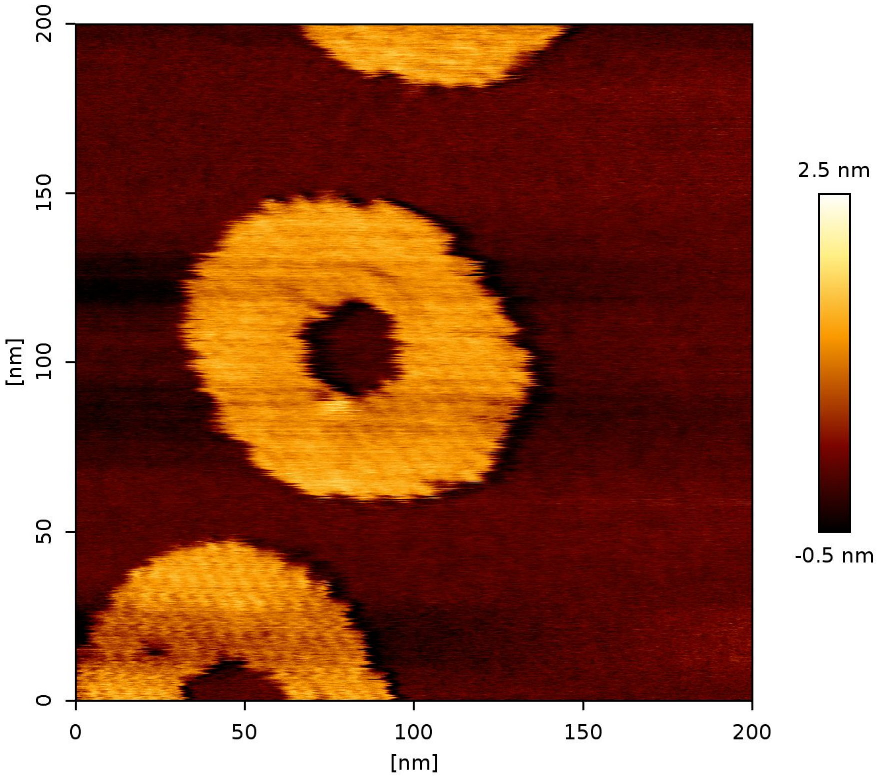
- Better understanding and controlling the folding process

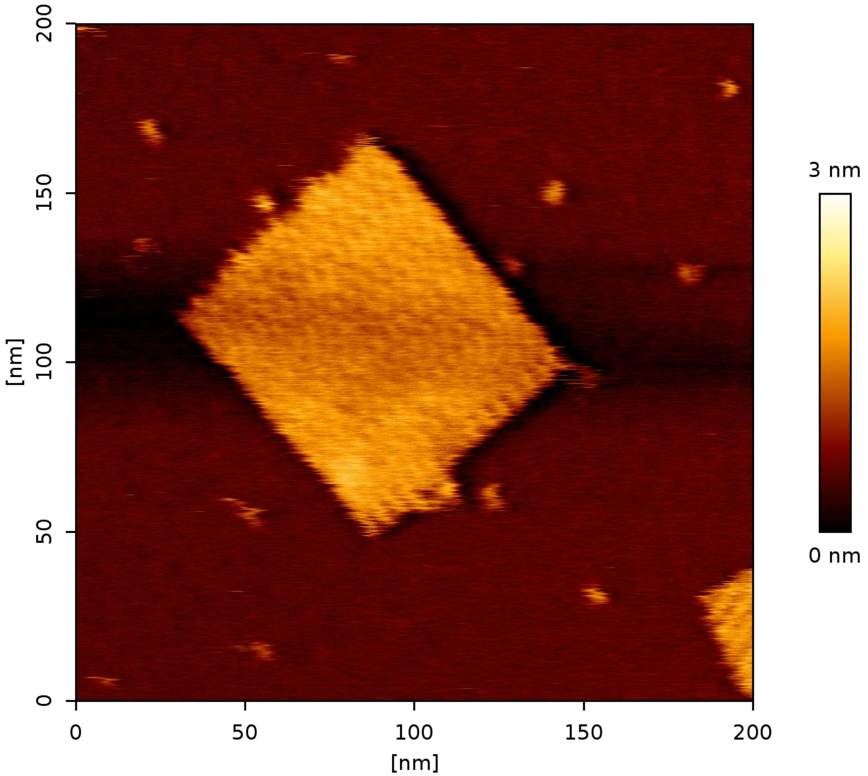


A first test design

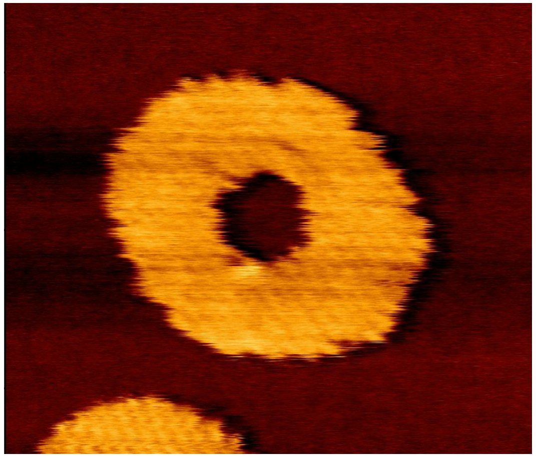


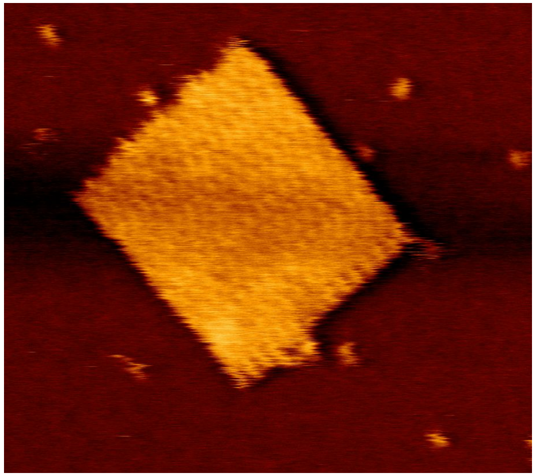
Building larger DNA structures
Assembling separately ring and square origamis



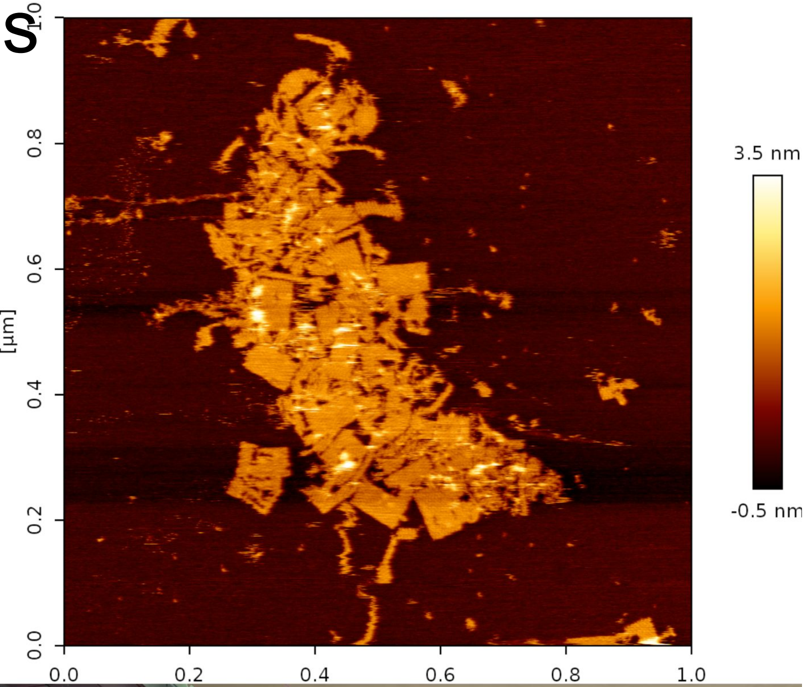


Building larger DNA structures
First results : not so great

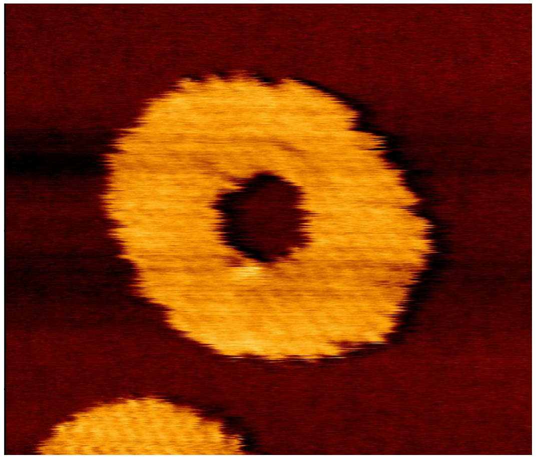


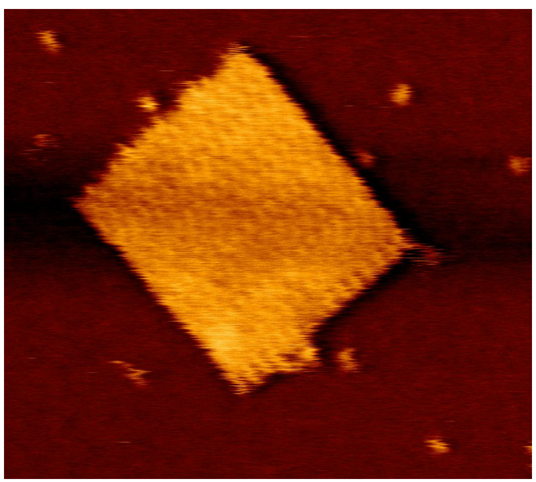


40nM M13
20nM links
200nM stapples
90->25°C, 60h



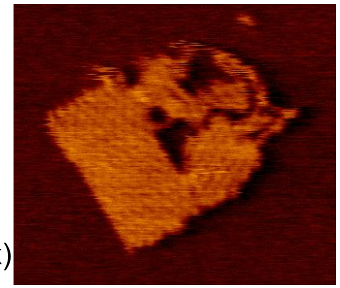
Building larger DNA structures
First results : not so great

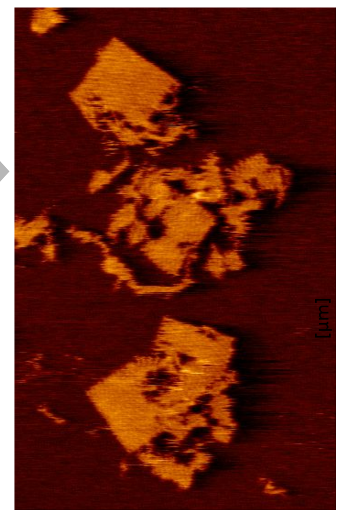


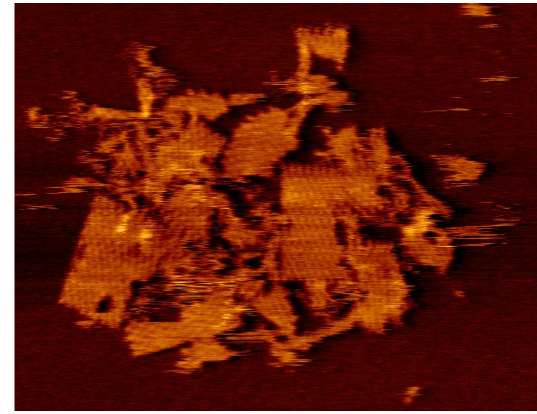


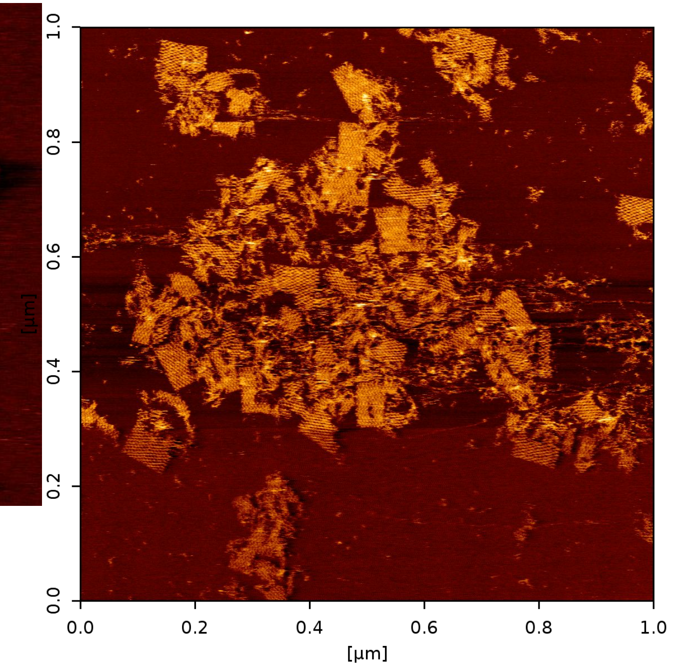
40nM M13 (1x)
20nM black staples (0.5x)
200nM staples (5x)

90->25°C, 60h



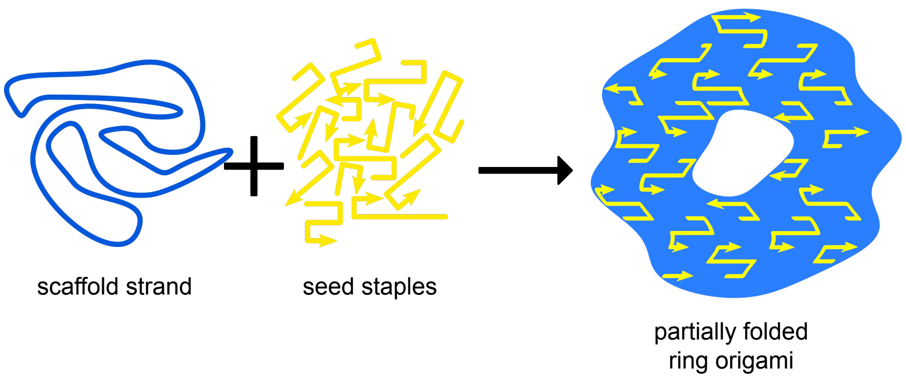




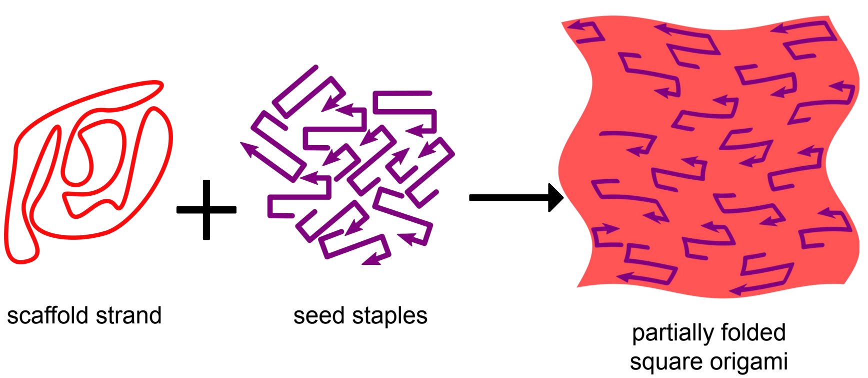


Finding key staple strands for scaffold differentiation

Adding a partial folding step

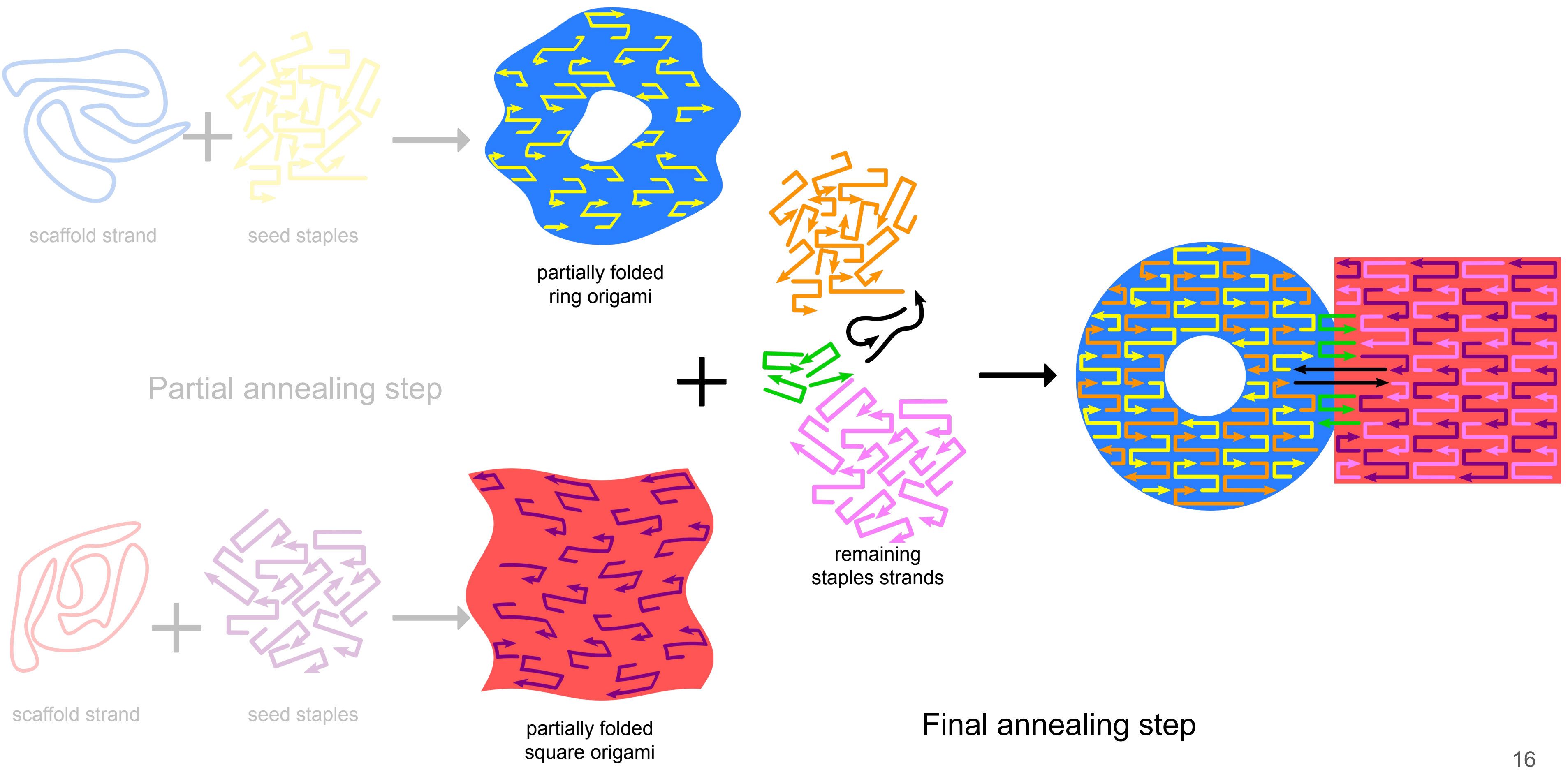


Partial annealing step



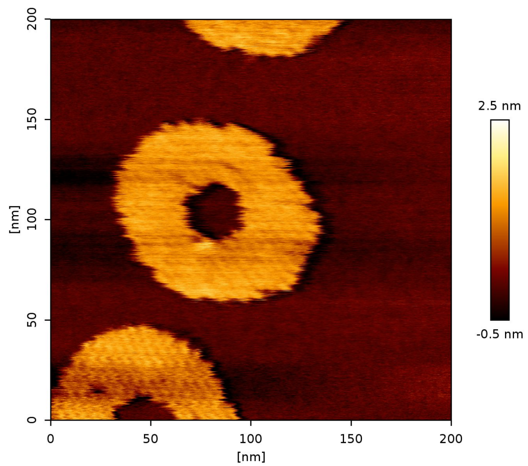
Finding key staple strands for scaffold differentiation

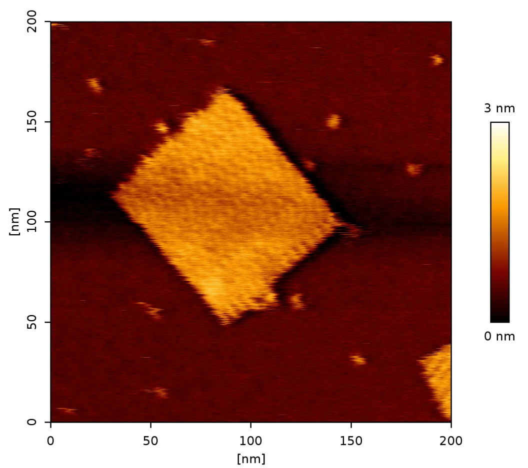
Adding a partial folding step



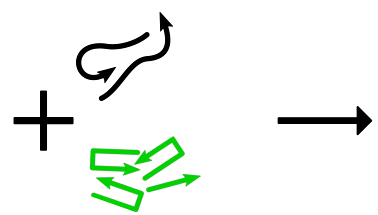
Finding key staple strands for scaffold differentiation

1. With **all staples** as seeds :

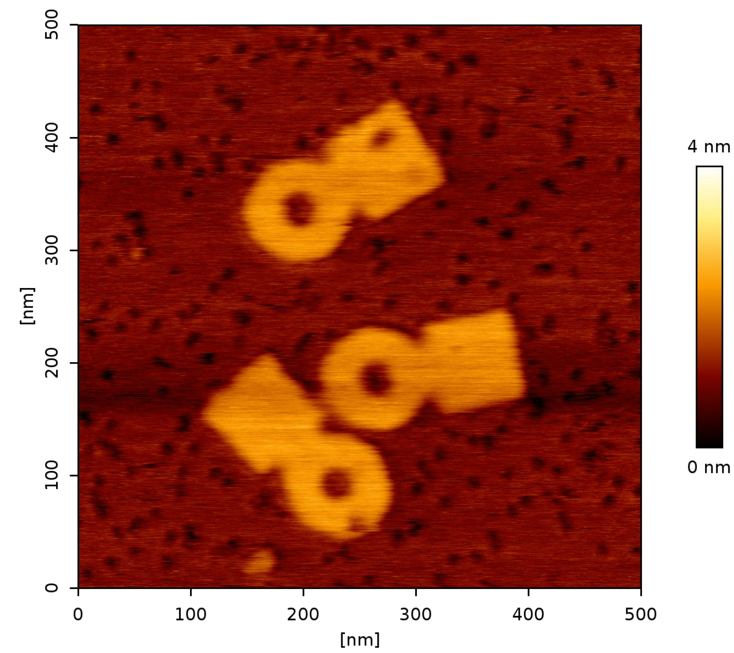




Fully formed halves

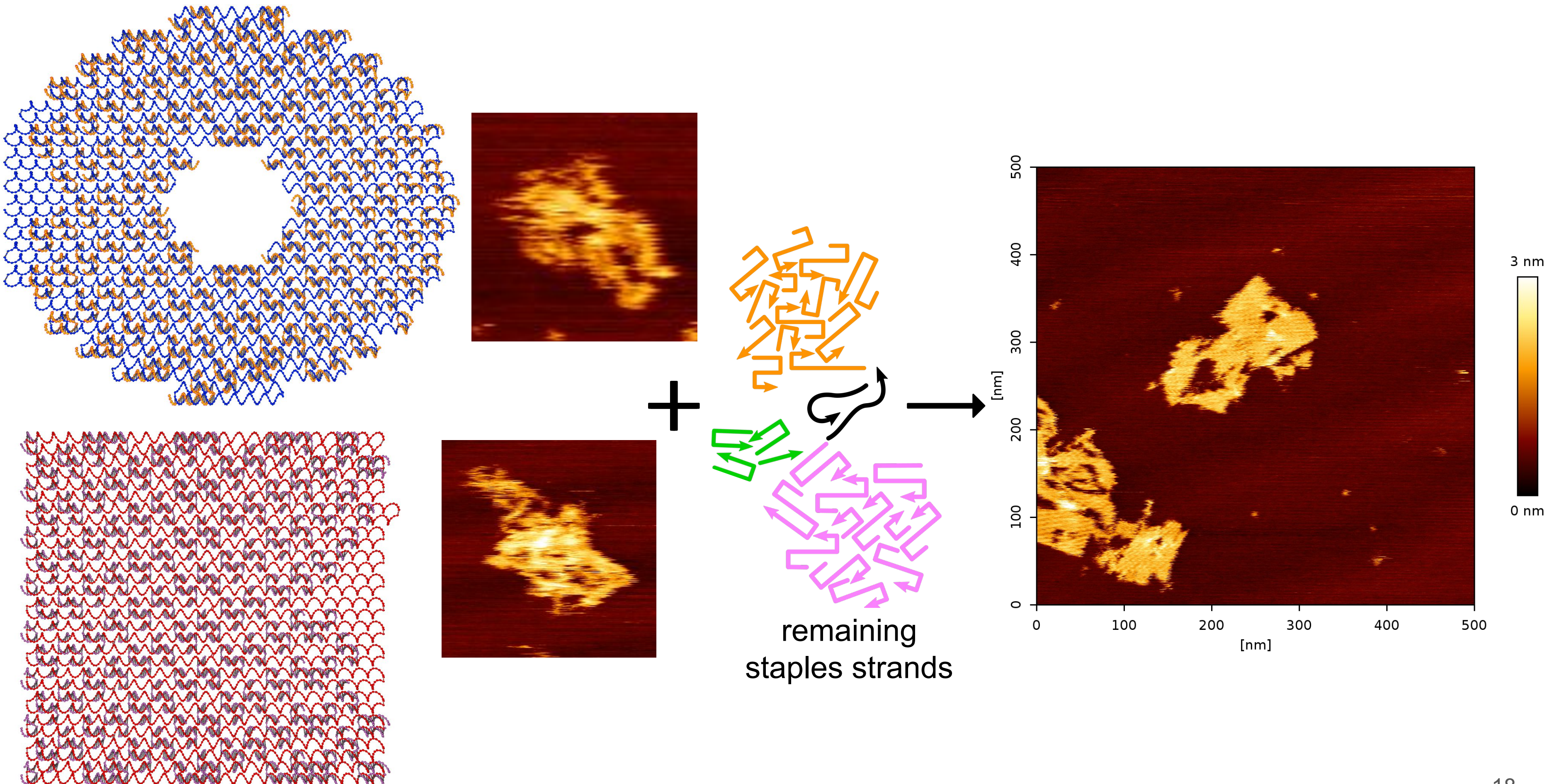


Link staples



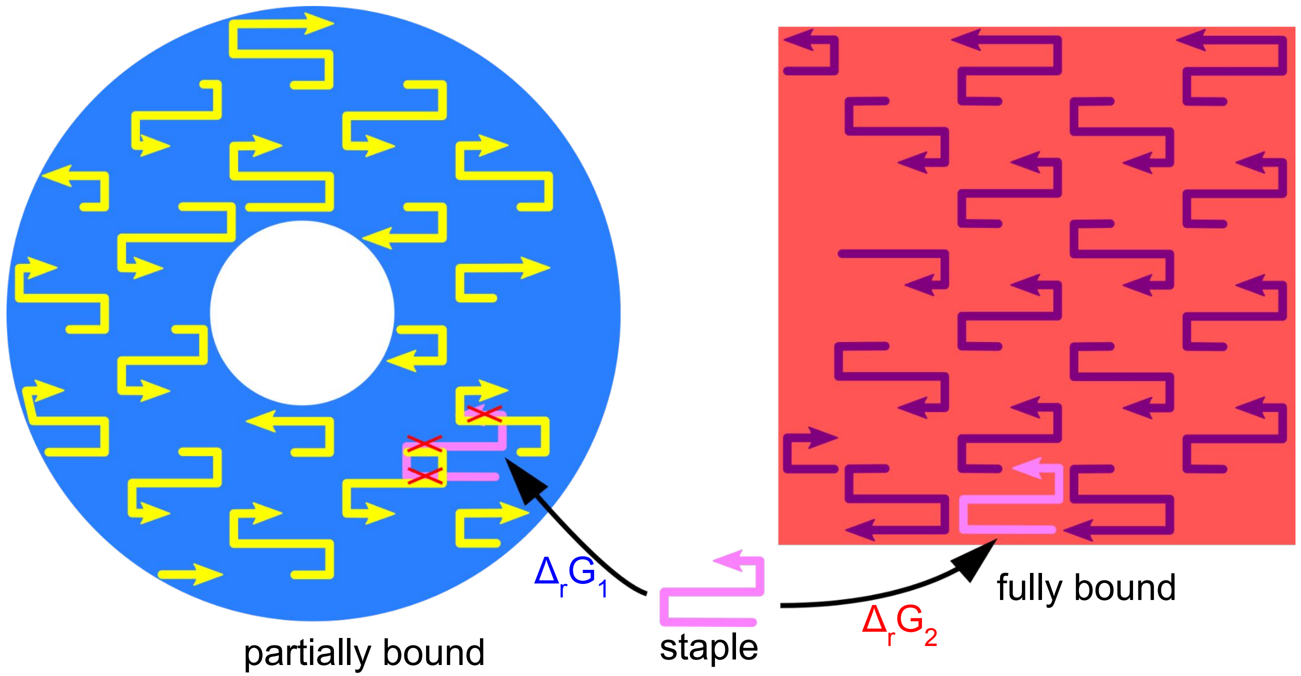
Finding key staple strands for scaffold differentiation

2. With **50% staples** as seed, **checkerboard pattern**:



halves with 50% staples

Refining the choice of seed (energy model)



$$|\Delta_r G_1| < |\Delta_r G_2|$$

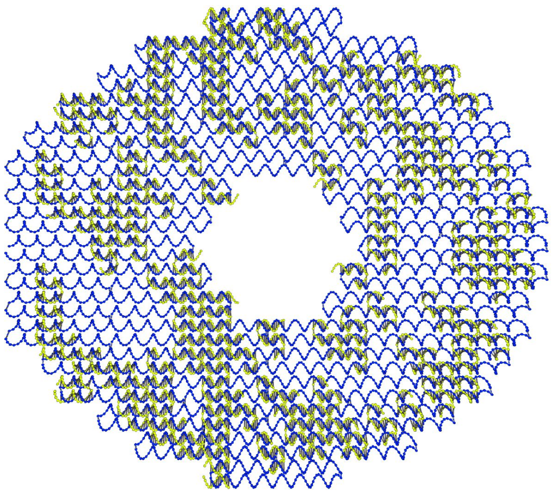
Goal: Minimize the number of **seed staples**.

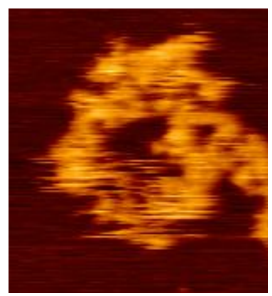
Constraints: $|\Delta_r G_1|$ low for every **pink staple**

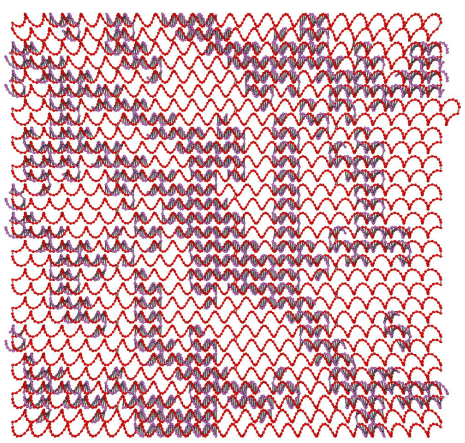
→ We can use **linear programming** (assuming $\Delta_r G_1$ is linear)

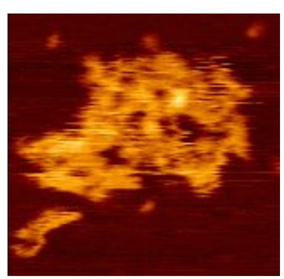
Finding key staple strands for scaffold differentiation

3. With **48% staples** as seed, **linear optimization problem**:

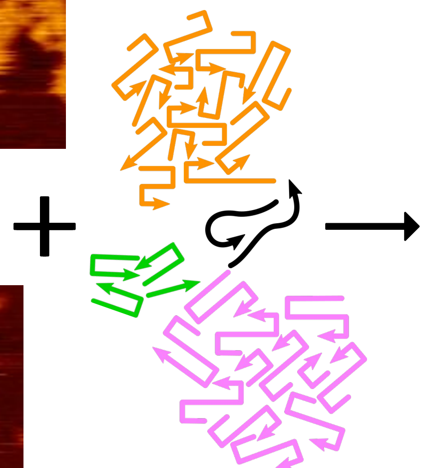




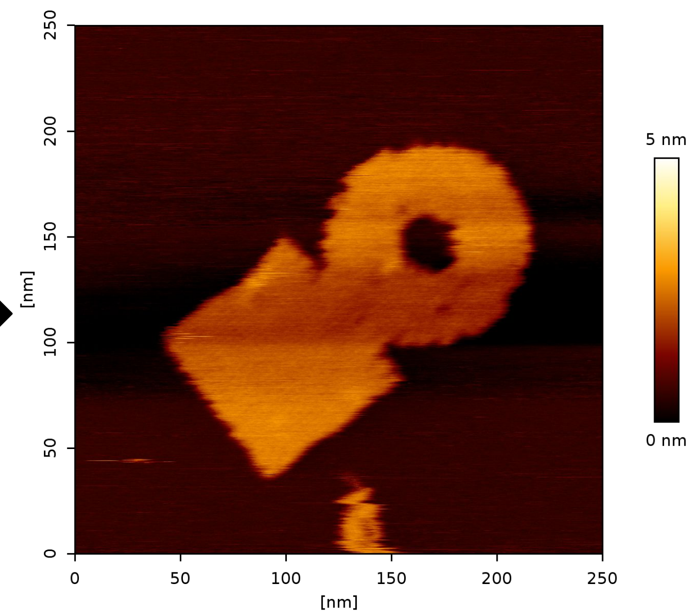




halves with 48% staples



remaining staples strands



Further investigating DNA origami formation

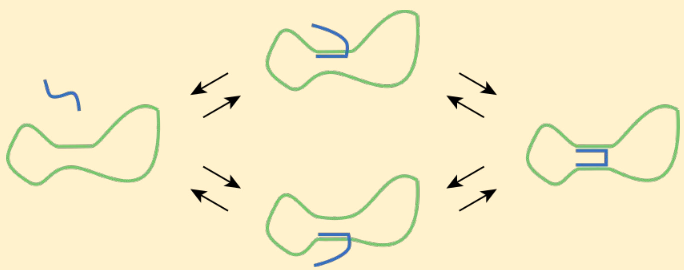
Thermodynamics fundations

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

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

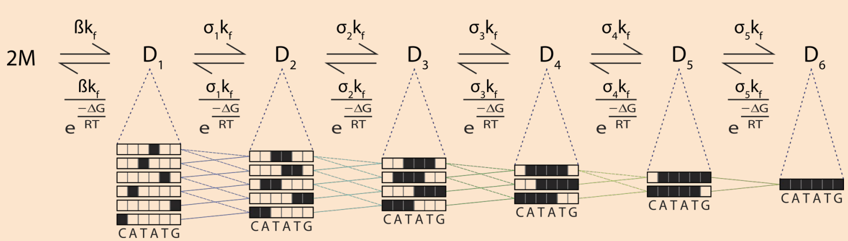
SantaLucia 1998

Domain level simulation



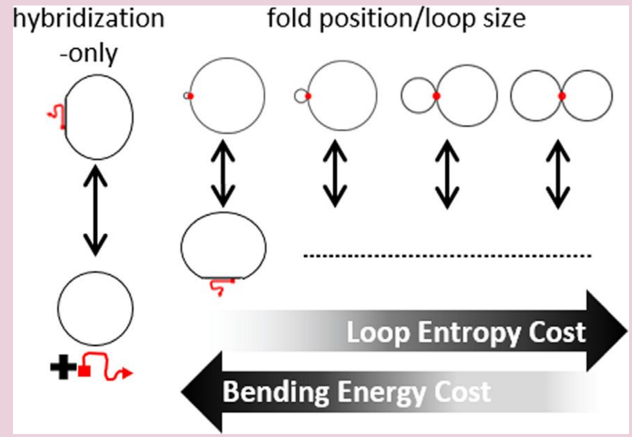
Dannenberg et al., 2015

Nucleotide level simulation



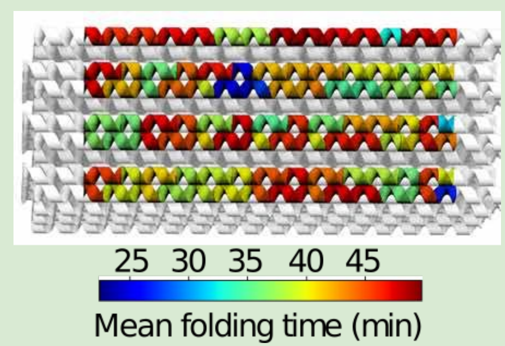
Menssen et al., 2021

Topological considerations



Majikes et al., 2020

Experimental data



Schneider et al., 2019

Understanding DNA Origami formation: *Thermodynamics*

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

$$\Delta G^\circ_{37}(\text{prediction}) = \Delta G^\circ(\text{CG/GC}) + \Delta G^\circ(\text{GT/CA}) + \Delta G^\circ(\text{TT/AA}) + \Delta G^\circ(\text{TG/AC}) + \Delta G^\circ(\text{GA/CT}) + \Delta G^\circ(\text{init.})$$

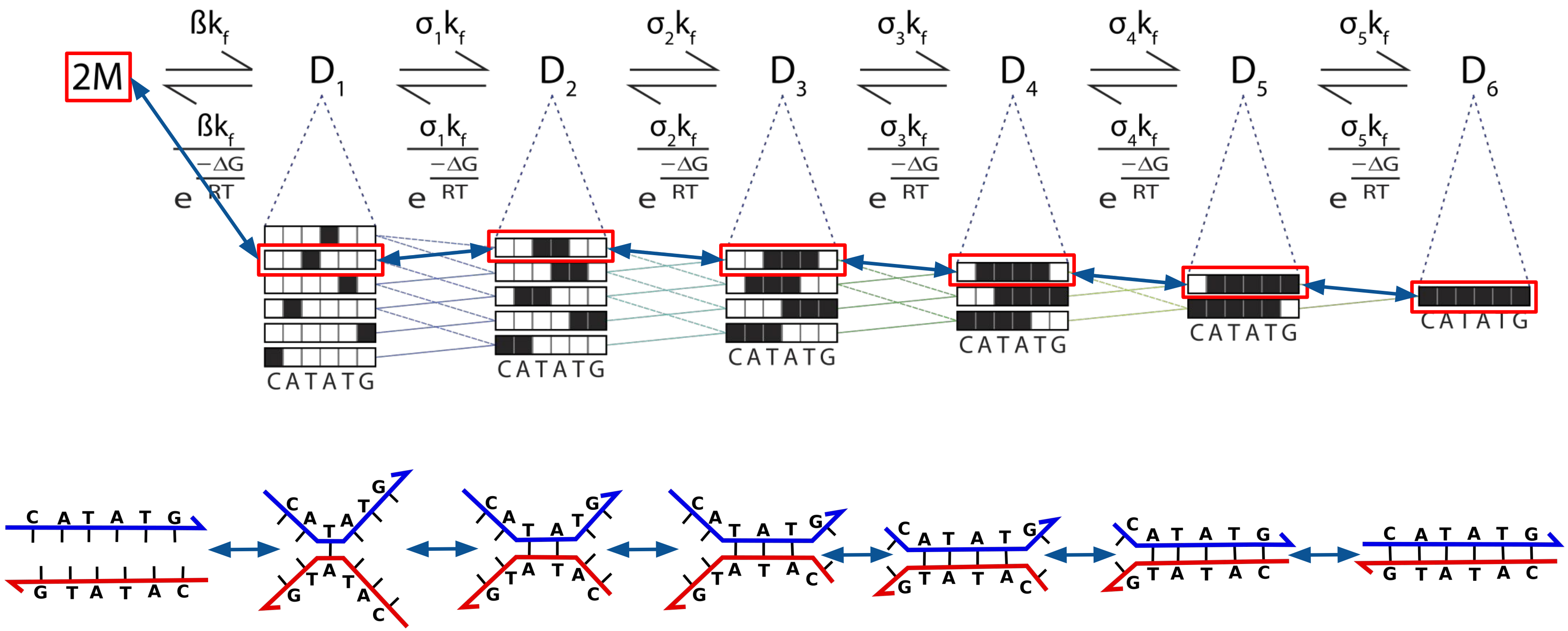
$$= -2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.98 + 1.03$$

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

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

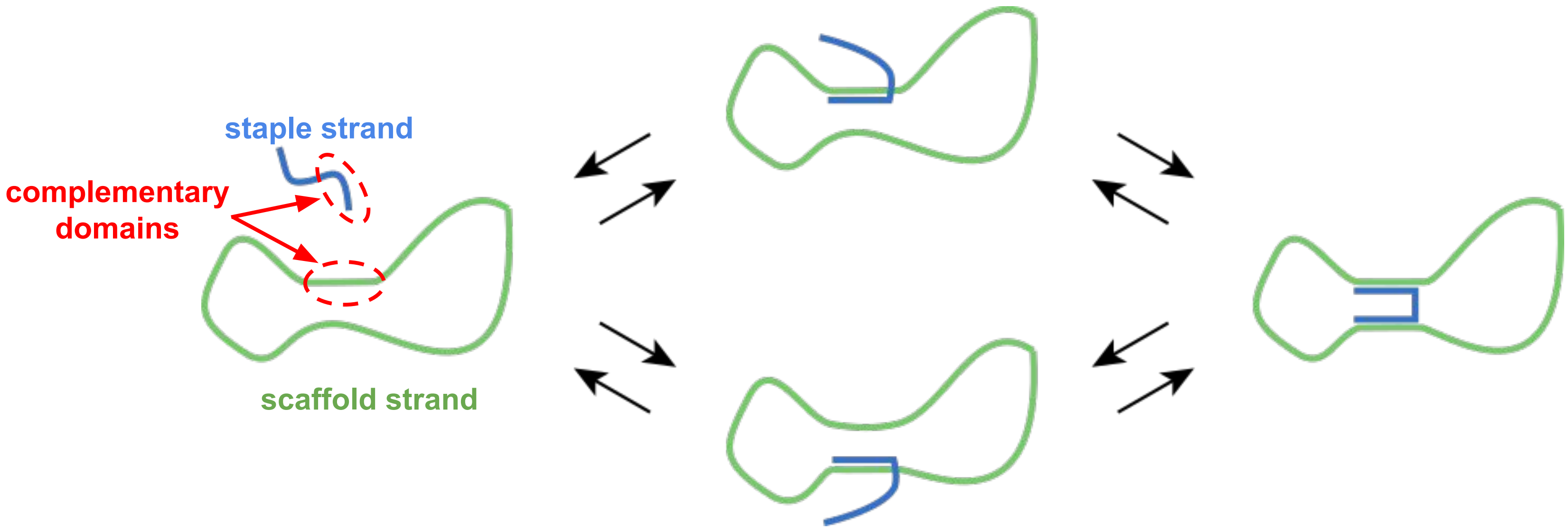
SantaLucia, « A Unified View of Polymer, Dumbbell, and Oligonucleotide DNA Nearest-Neighbor Thermodynamics » (1998)

Understanding DNA Origami formation:
kinetic Monte-Carlo model



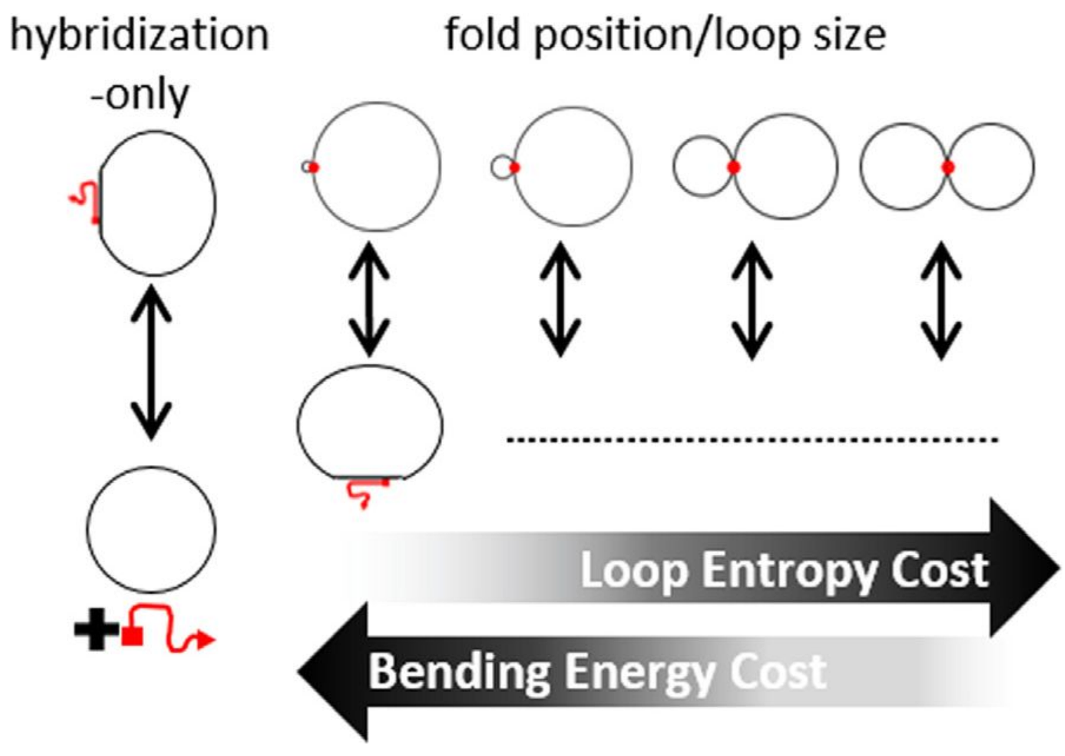
Menssen, Kimmel, et Tokmakoff, « Investigation into the mechanism and dynamics of DNA association and dissociation utilizing kinetic Monte Carlo simulations » (2021).

Understanding DNA Origami formation:
Domain level simulation



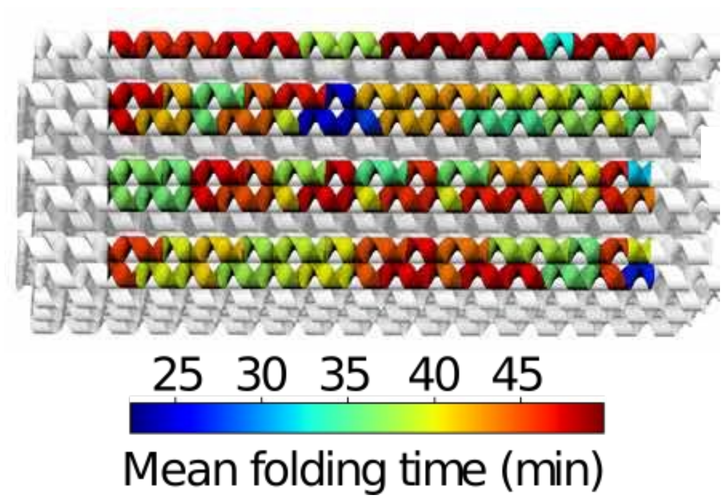
Dannenberg et al., « Modelling DNA origami self-assembly at the domain level » (2015)

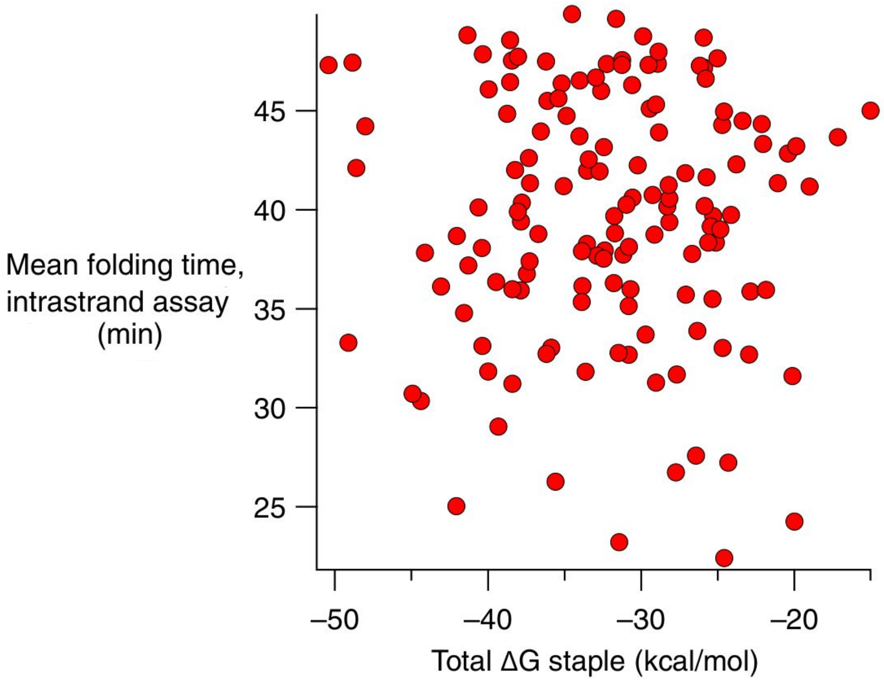
Understanding DNA Origami formation:
Geometrical/topological considerations



Majikes et al., « Revealing thermodynamics of DNA origami folding via affine transformations » (2020),

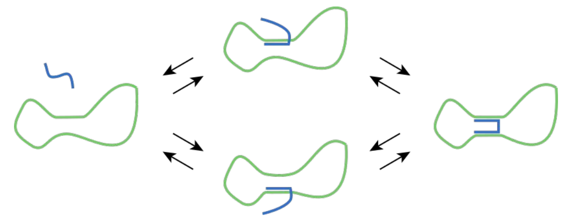
Understanding DNA Origami formation:
Measuring staple attachment delay



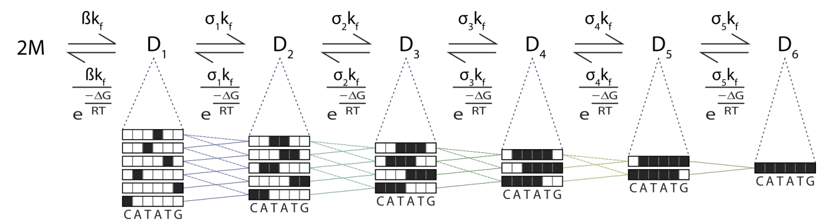


Schneider, Möritz, et Dietz, « The sequence of events during folding of a DNA origami » (2019),

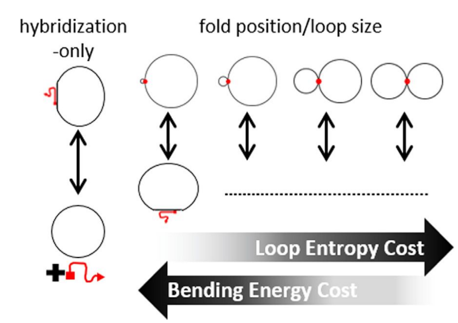
My work: a stochastic model that incorporates all these approaches



- Origamis are “big” (~7000 nts)
  → **domain approach**



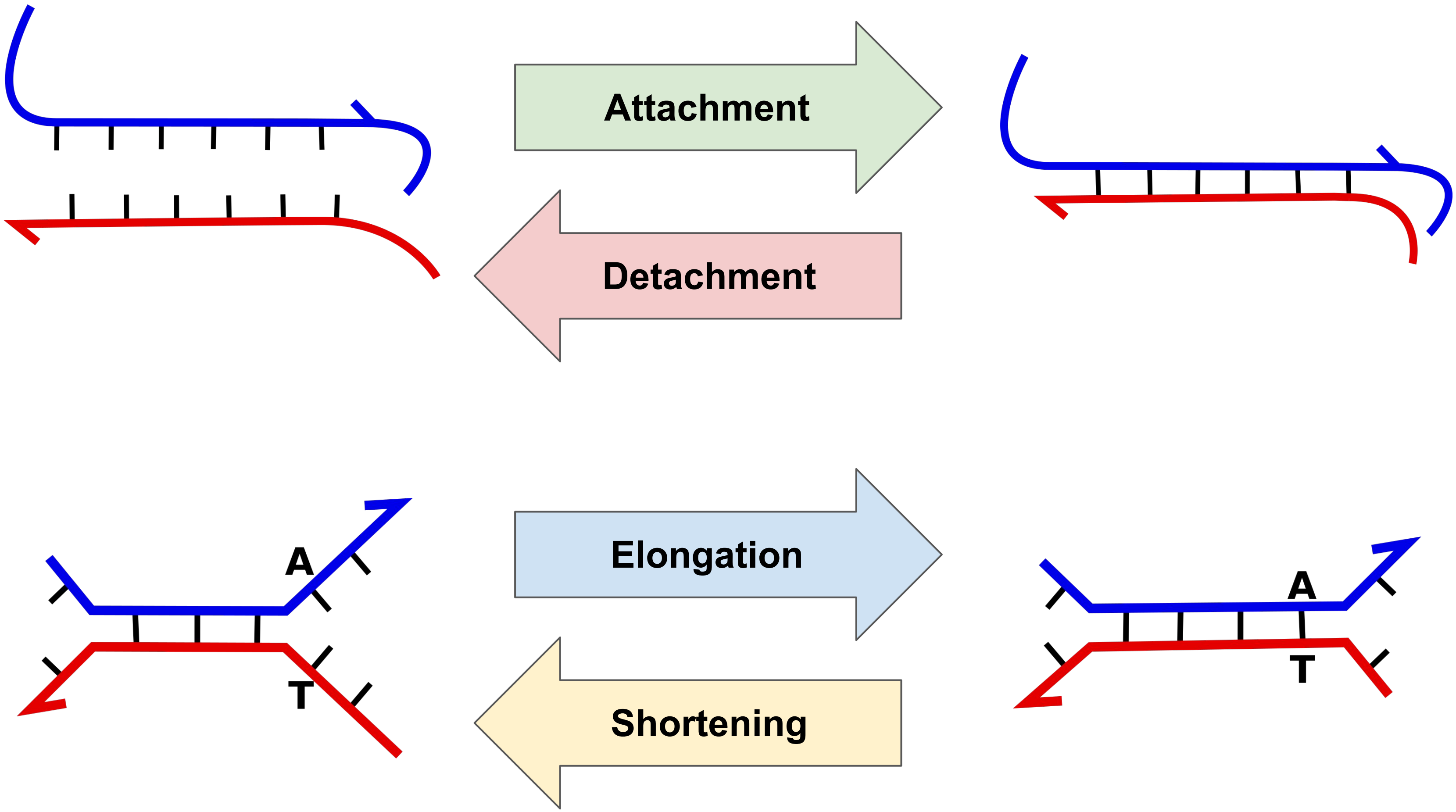
- Detecting unpredicted formations
  → **nucleotide level events**



- Complex shapes / scaffold routing
  → **topological and geometrical considerations are important**

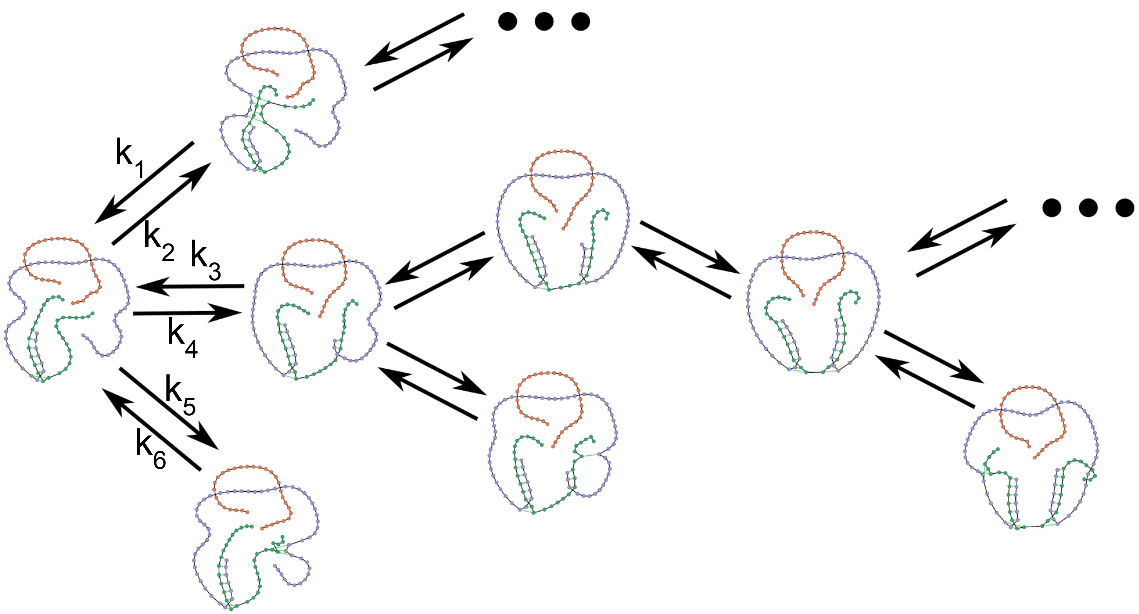
Our model: Authorized state transitions

Simulation with 4 types of transitions:

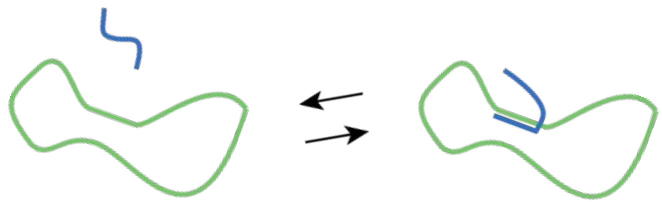


Our model: Kinetic Monte-Carlo simulation

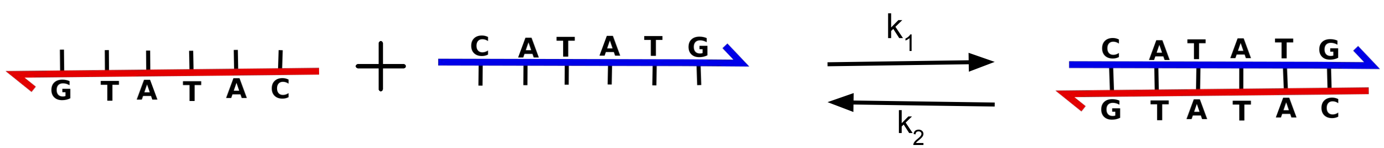
- **Initial state**: a bunch of unattached strands
- **Possible transitions**: Attachments, Detachments, Elongations and Shortenings when possible at the current state.
- **Transition rate**: proportionate to the probability of occurring as the next transition



Our model: Computing transition rates
Bimolecular reactions



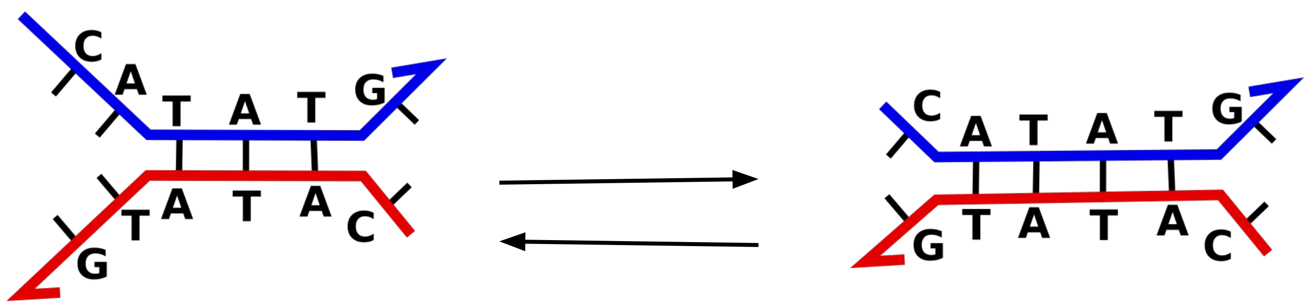
Bimolecular domain Attachment / Detachment = simple chemical reaction



$$\frac{k_1}{k_2} = e^{-\frac{\Delta G_{\text{attach}}}{RT}}$$ where ΔG_{attach} is computed from the sequence and condition parameters (temperature and salt concentrations).

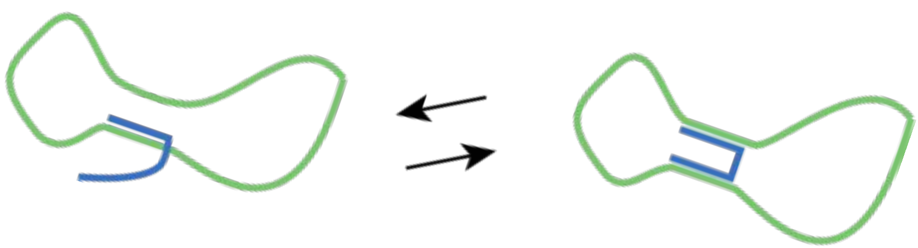
Our model: Computing transition rates
Elongation/Shortening

- Elongation / Shortening: similar dependance on sequence and condition parameters



Our model: Computing transition rates
Unimolecular Attachment/Detachment

- Unimolecular domain Attachment/Detachment:
  - depends on current geometry/topology
  - rate can change due to non-local state modifications
  - sometime impossible (ex: when starting and ending domains are already attached)

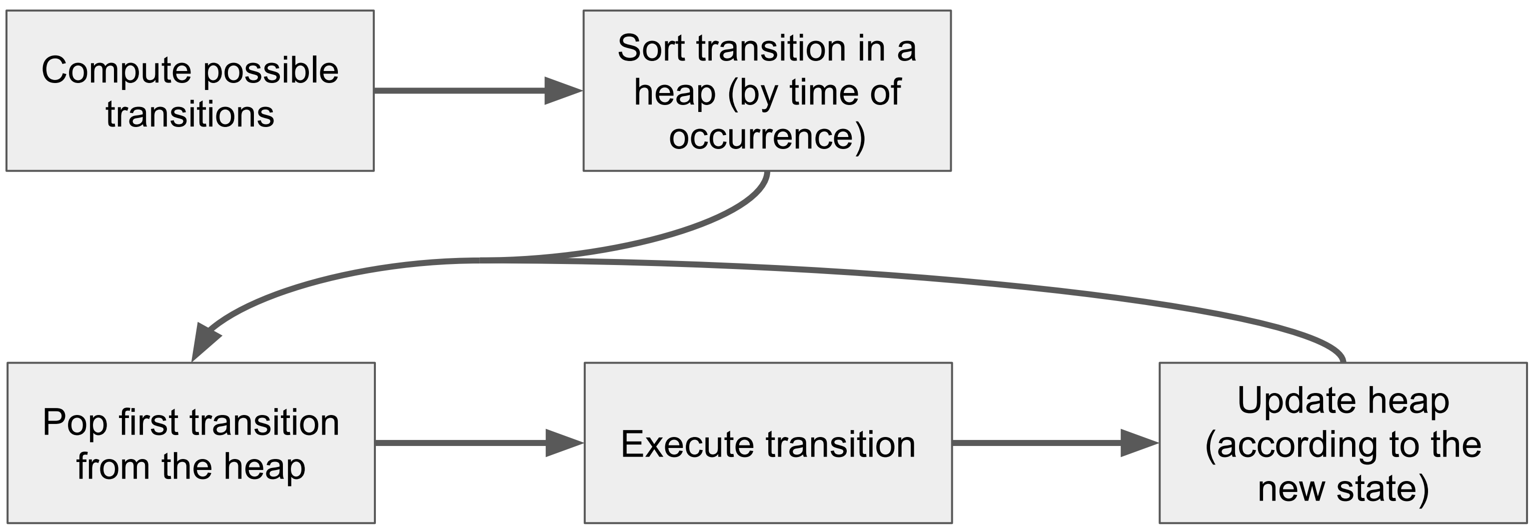


Simulation loop

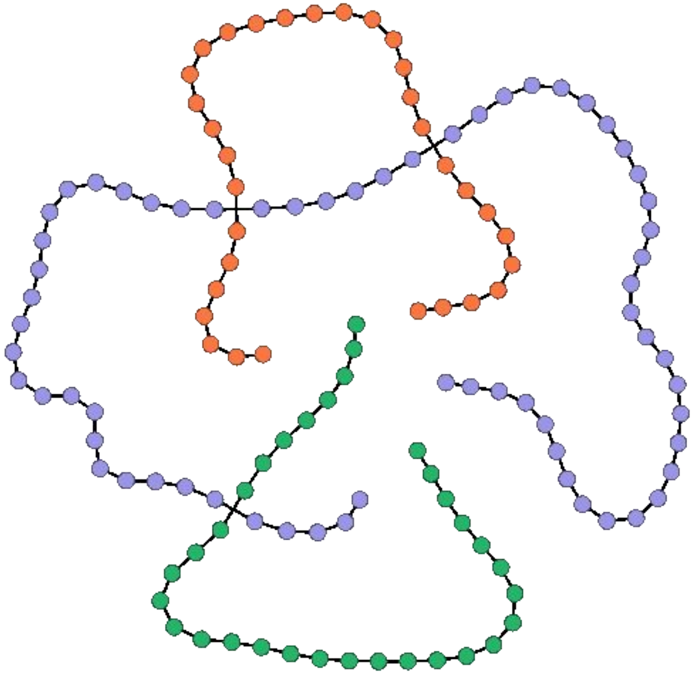
Input: strand sequences (ex: [“ATCCGT”, “AATTAT”, “ATGGCGTGCAGT”, …])

Output: sequence of states

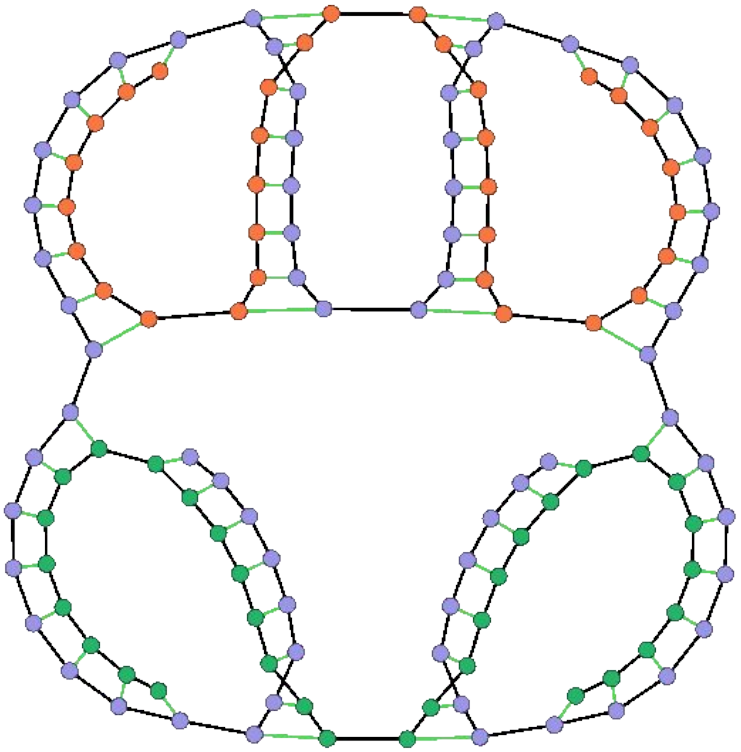
Initial state: all strands unattached



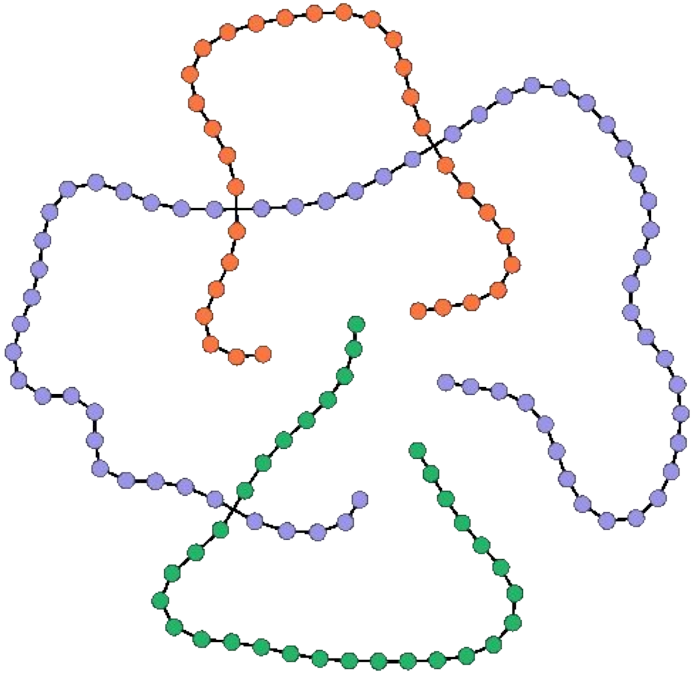
Example of execution (simplified Origami)



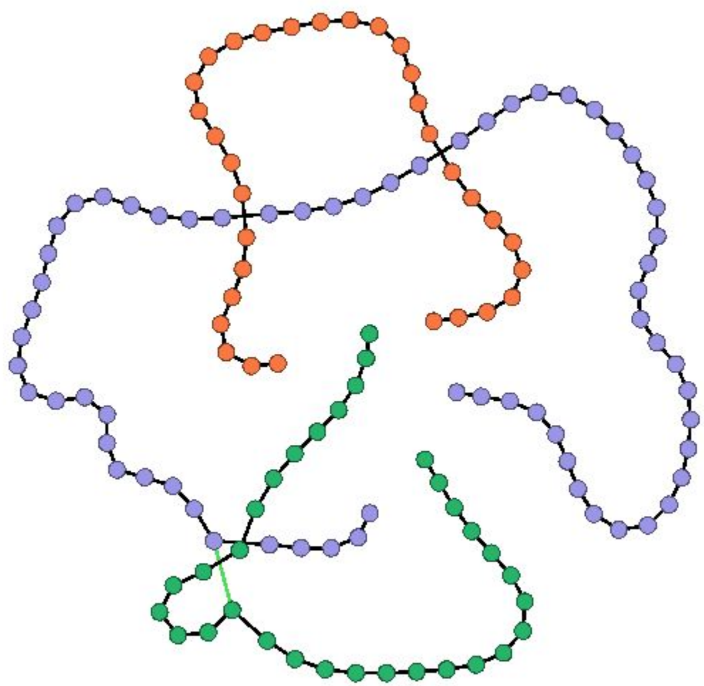
Example of execution (simplified Origami)



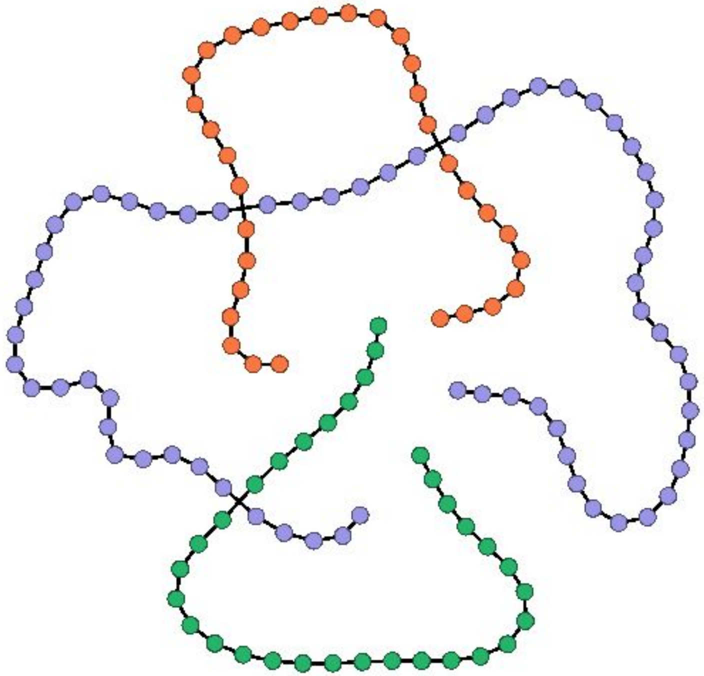
Example of step-by-step execution (simplified Origami)



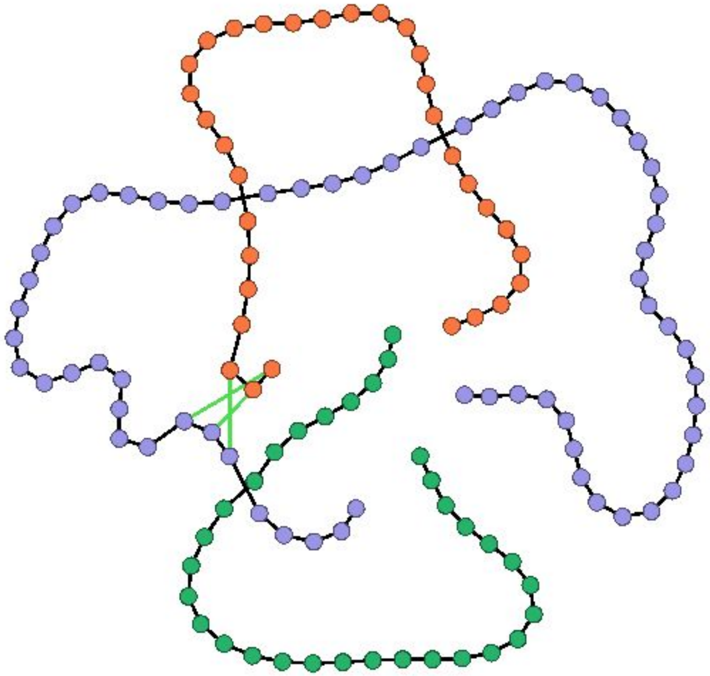
Example of step-by-step execution (simplified Origami)



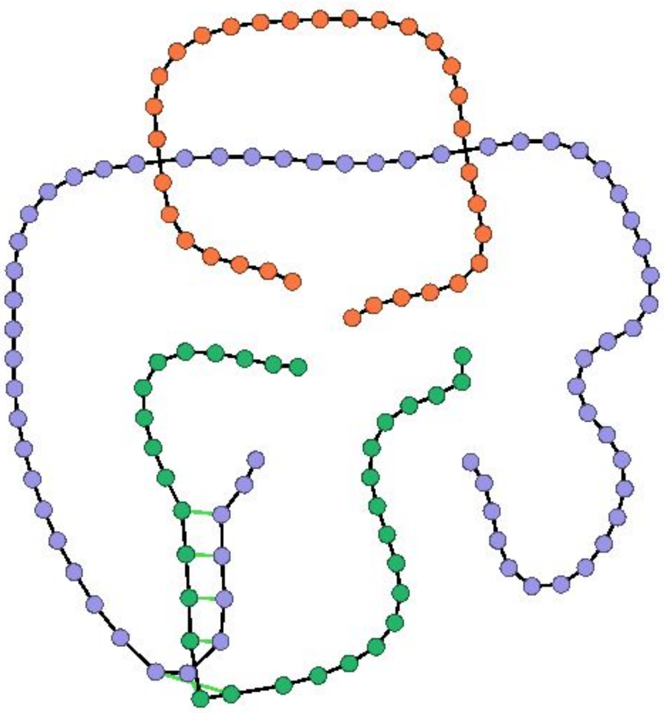
Example of step-by-step execution (simplified Origami)



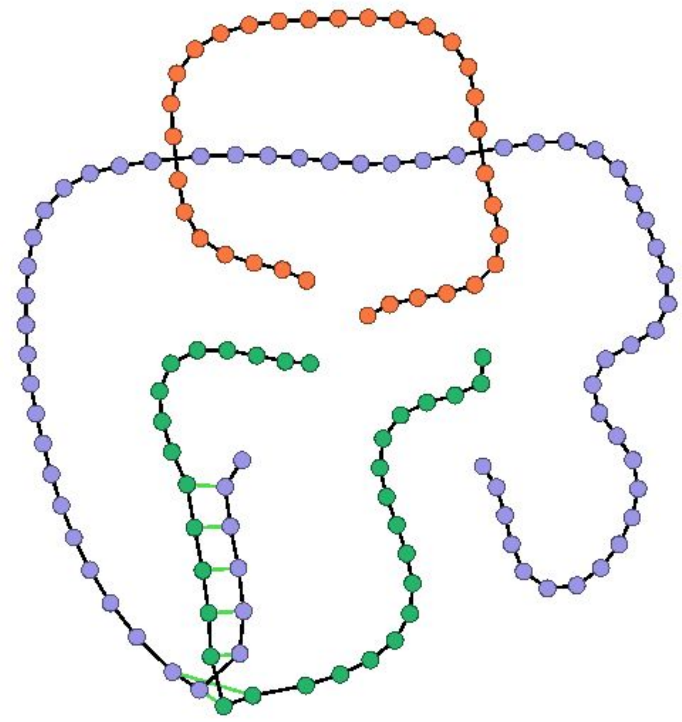
Example of step-by-step execution (simplified Origami)



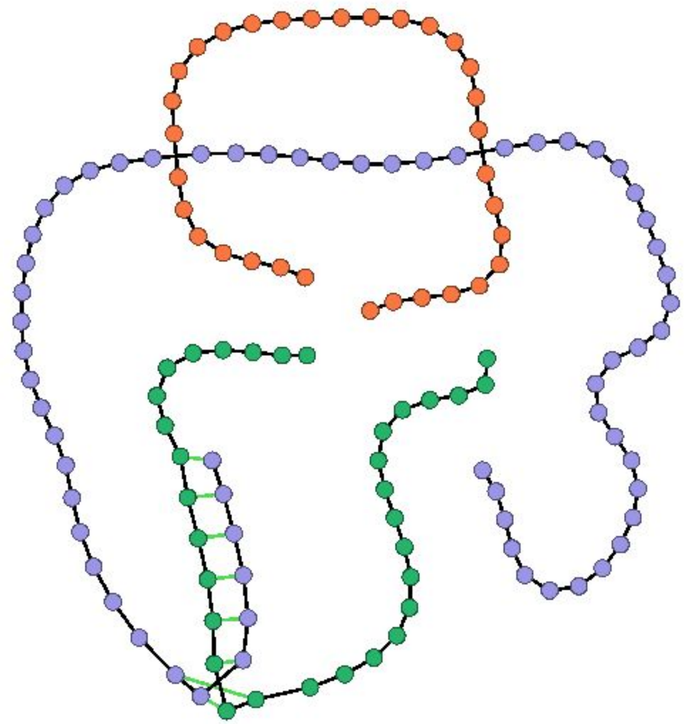
Example of step-by-step execution (simplified Origami)



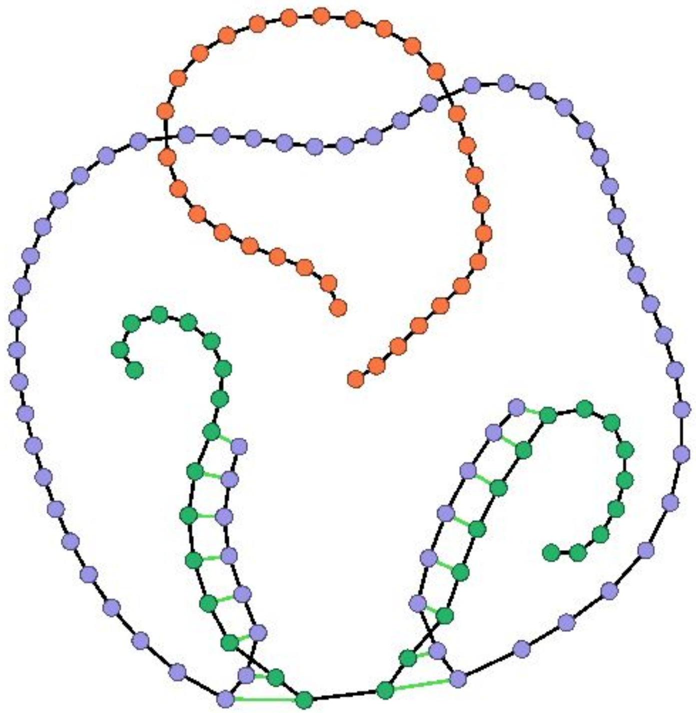
Example of step-by-step execution (simplified Origami)



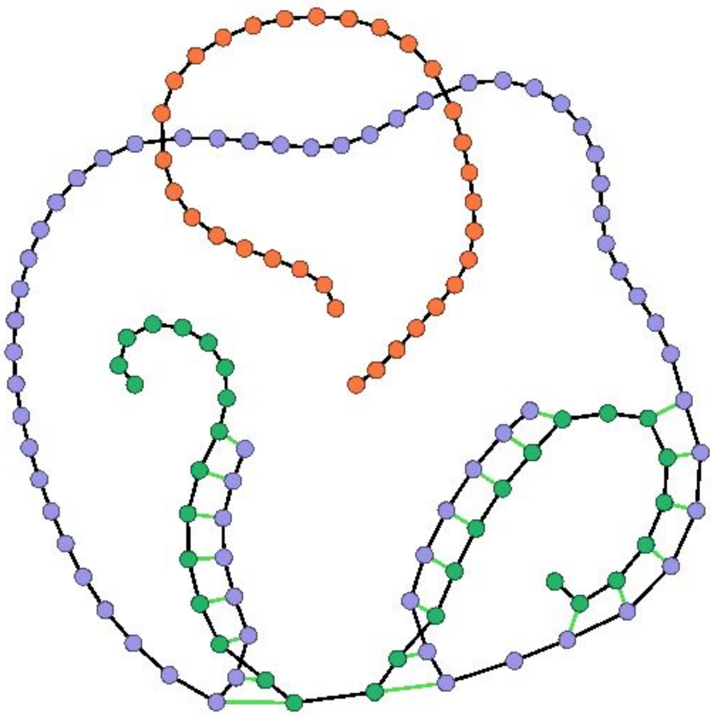
Example of step-by-step execution (simplified Origami)



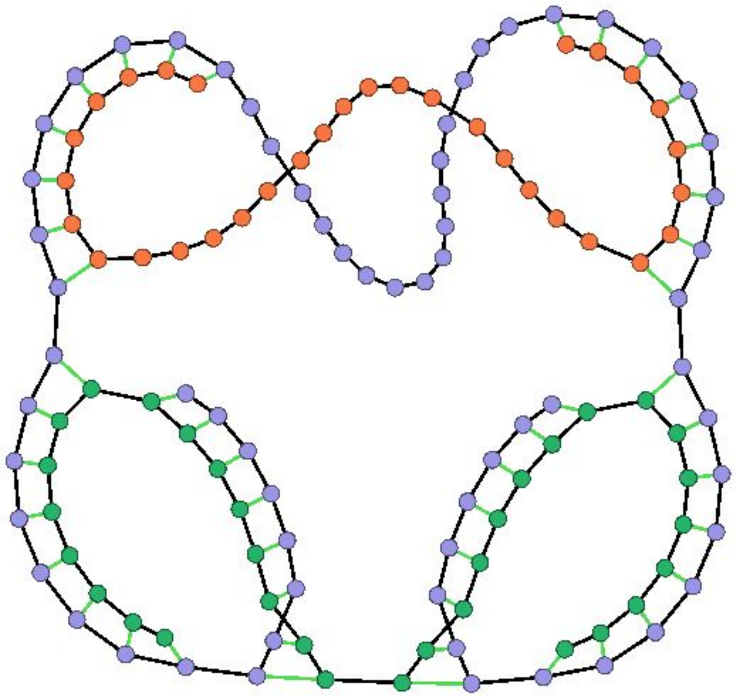
Example of step-by-step execution (simplified Origami)



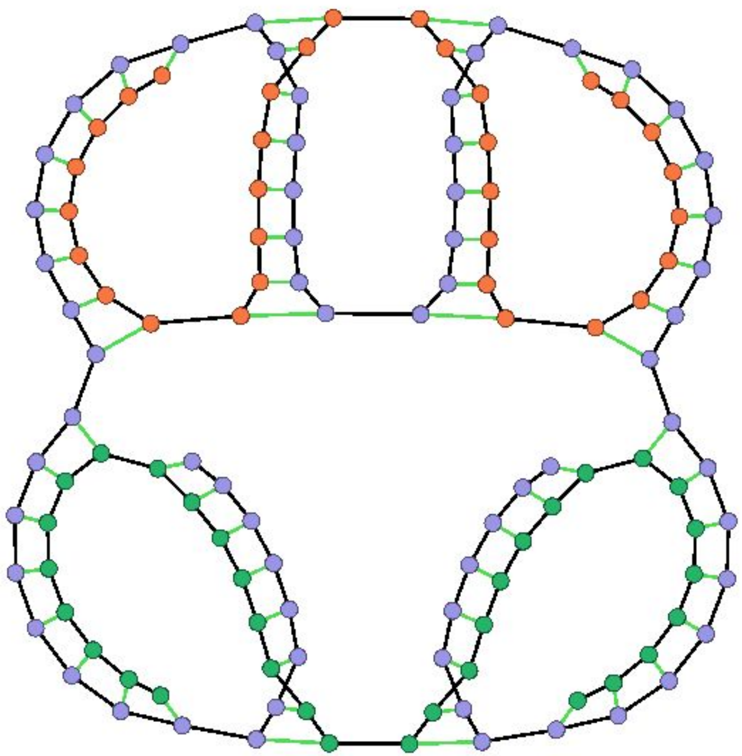
Example of step-by-step execution (simplified Origami)



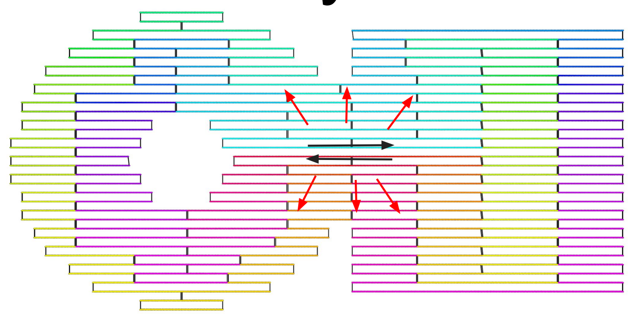
Example of step-by-step execution (simplified Origami)

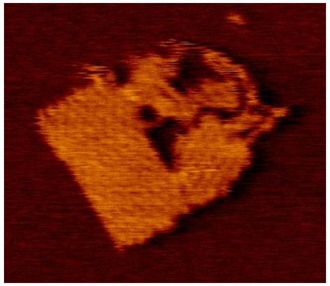


Example of step-by-step execution (simplified Origami)

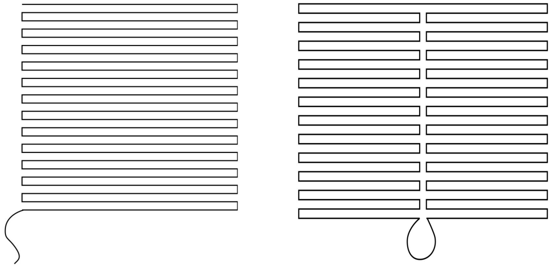


The model should allow us to study :





- Nucleation phenomenons,
- Chimeric or ill-formed origami,
- Influence of design choices :
  - scaffold routing,
  - stapling method.



- Influence of experimental parameters :
  - strand concentrations,
  - salt concentrations,
  - temperature,
  - temperature curve.

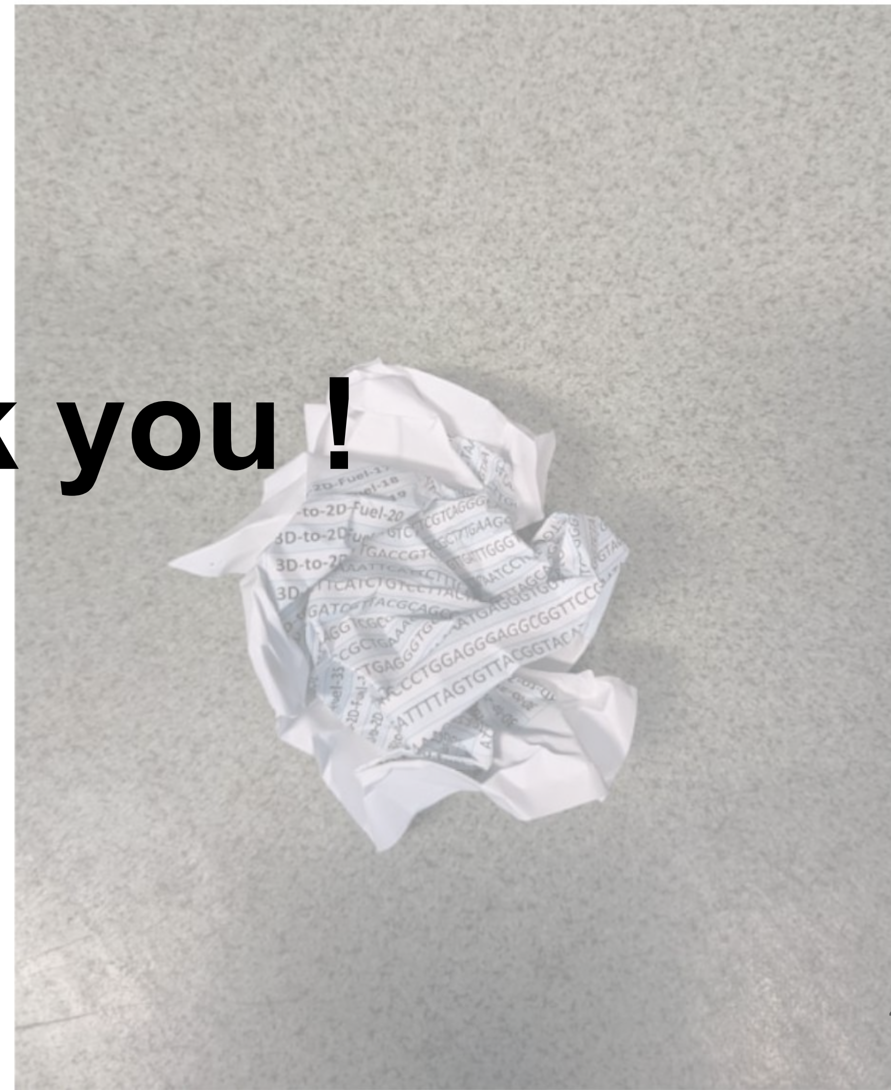
How to improve the model (help wanted)

- Finer use of **topology** and **geometry** of the system:
  - **topology:** implementation of distance-dependant rates
  - a mathematical model for **loop entropy cost**
  - **How much** I need to know about the **2D / 3D positioning** of the origami components during simulation ?

- **Model simplifications**:
  - which transitions are **impactless** ? (e.g. short bimolecular attachments ? self-attachment ?)
  - **Shortcut** fast sequences of events (e.g. random walks)

My PhD

Expectation **Reality**



Thank you !