A stochastic and geometrical model for DNA origami self-assembly

Octave Hazard
(PhD 2021-2024)
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


Simulation of a cellular automaton

80 nm ≈ 800 atoms
≈ \( \frac{1}{5} \) visible wavelength

execution / assembly / growth
DNA Origami

DNA Origami

Scaffold DNA

Annealing

Folding

Complementary domains

Scaffold

Staples

DNA origami nanostructure
DNA Origami

1) design

2) order staples strands
+ scaffold strand

3) annealing
90°C → 25°C in 12h

4) characterization

AFM

500 nm
Building larger DNA structures

DNA Origami reliable
Short strands assembly less reliable
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

DNA Origami 1

DNA Origami 2

(identical scaffold strand)
Building larger DNA structures
Assembling separately ring and square origamis
Building larger DNA structures

First results: not so great

40nM M13
20nM links
200nM staples
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

Partial annealing step

Final annealing step
Finding key staple strands for scaffold differentiation

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

   ![Fully formed halves](image1)

   ![Link staples](image2)
Finding key staple strands for scaffold differentiation

2. With **50% staples** as seed, **checkerboard pattern**: half with 50% staples remaining staples strands

halves with 50% staples
Refining the choice of seed (energy model)

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:
Further investigating DNA origami formation

**Thermodynamics fundations**

\[
\begin{align*}
5' & \text{C-G-T-T-G-A} \quad 3' \\
3' & \text{G-C-A-A-C-T} \quad 5' \\
\Delta G^\circ_{37 \text{(pred.)}} &= \Delta G^\circ_{\text{CG/GC}} + \Delta G^\circ_{\text{GT/CA}} + \Delta G^\circ_{\text{TT/AA}} \\
&+ \Delta G^\circ_{\text{TG/AC}} + \Delta G^\circ_{\text{GA/CT}} + \Delta G^\circ_{\text{init.}}
\end{align*}
\]

SantaLucia 1998

**Topological considerations**

Dannenberg et al., 2015

Majikes et al., 2020

**Domain level simulation**

Majikes et al., 2020

**Nucleotide level simulation**

Menssen et al., 2021

**Experimental data**

Schneider et al., 2019
Understanding DNA Origami formation: 

**Thermodynamics**

\[
\begin{align*}
5' & \quad \downarrow \downarrow \downarrow \downarrow \downarrow \ C - G - T - T - G - A \quad 3' \\
3' & \quad G - C - A - A - C - T \quad 5'
\end{align*}
\]

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

\[
= -2.17 - 1.44 - 1.00 - 1.45 - 1.30 + 0.98 + 1.03
\]

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

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

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:

- Attachment
- Detachment
- Elongation
- Shortening
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

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

- 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!