Mathematical Tools for Self-assembly

Jo Ellis-Monaghan
KdVI

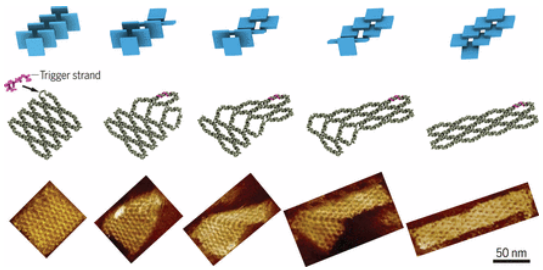
Objective

Not so much about DNA self-assembly, but rather

- How mathematical tools can support self-assembly projects in general
- Some of the range of the tools

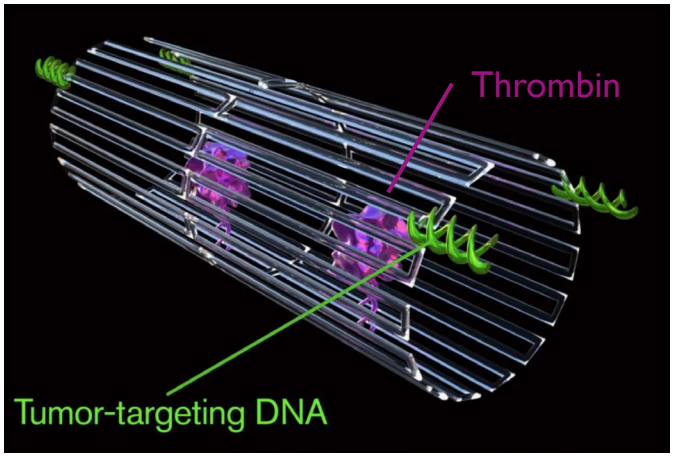
Why self-assembling nanostructures?

Nano-robotics

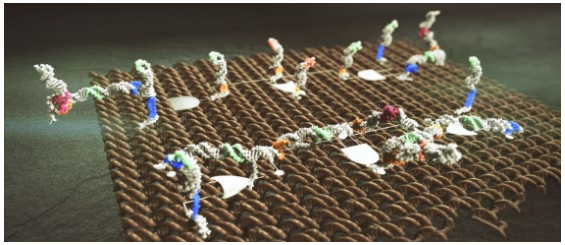


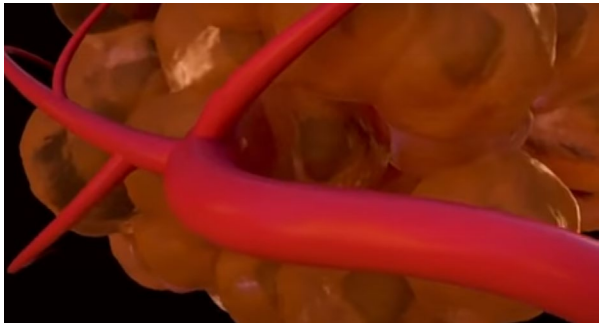
Reconfiguration of DNA molecular arrays driven by information relay. Song, Li, Wang, Mayer, Mao, Ke
Science 28 Jul 2017: Vol. 357, Issue 6349

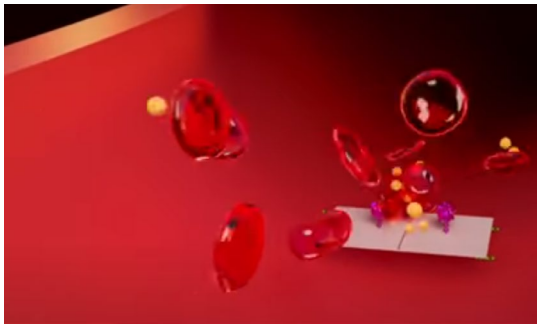
Cancer treatments



Nano-circuitry



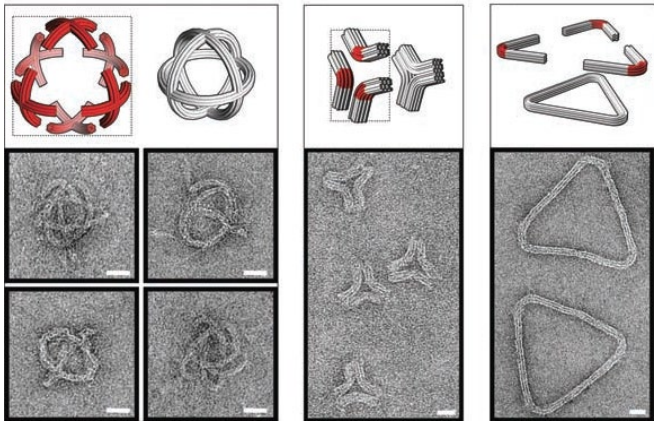




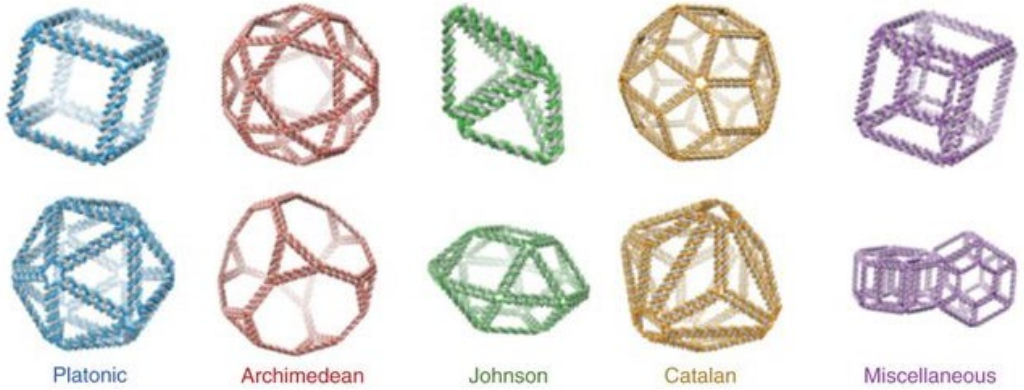
https://www.microsoft.com/en-us/research/blog/researchers-build-nanoscale-computational-circuit-boards-dna/

https://www-nature-com.proxy.uba.uva.nl/articles/nbt.4071.pdf
https://scitechdaily.com/fully-autonomous-cancer-fighting-nanorobots-seek-and-destroy-tumors/

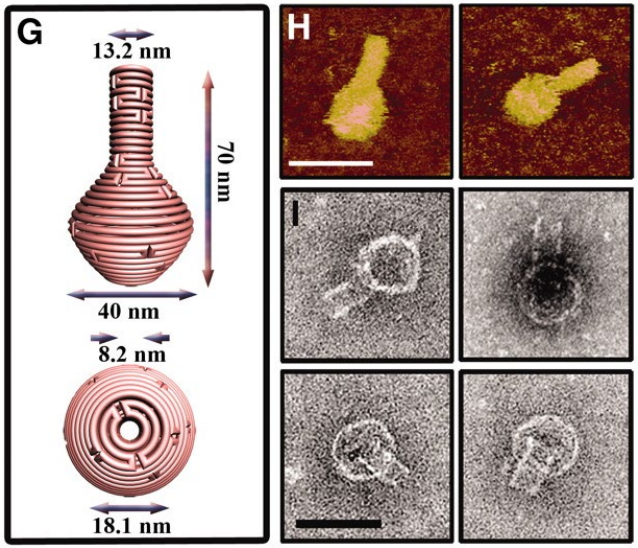
Some (3D) DNA nano-objects



Dietz, Douglas and Shih, Science, 325, 725



Linko, Kostiainen. Nature Biotechnology volume34, pages826–827 (2016)



Han, Pal, Nangreave, Deng, Liu, and Yan, Science, 333, 342

Benson, Mohammed, Gardell, Masich, Czeizler, Orponen , Högberg. Nature 523, 441–444 (23 July 2015)

Objectives

Create new mathematical and computation tools for laboratories producing self-assembled DNA nanostructures.

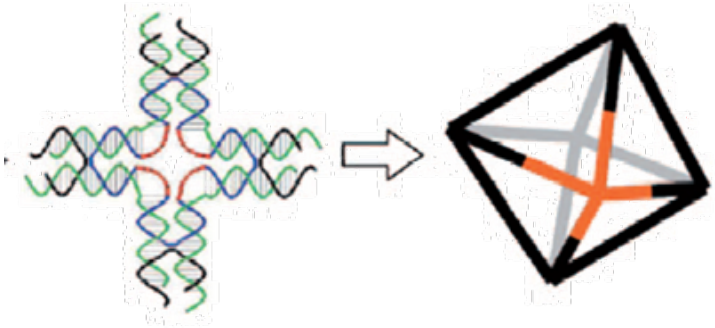
- Several assembly paradigms:

  Tile-based assembly ◆ DNA origami ◆ Reporter strands

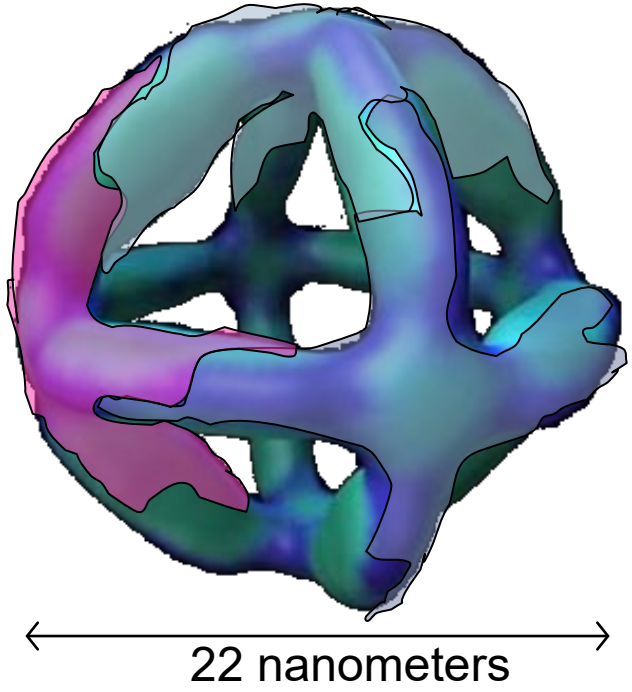
- For each assembly method, usually get:
  - Problem formulation and mathematical formalism
  - Proofs that design strategies are NP-Hard
  - Pragmatic approaches—algorithms, special cases, etc.
  - New mathematical directions arising from the problem

Branched Junction Molecules

A self-assembled octahedron



http://onlinelibrary.wiley.com/doi/10.1002/anie.200904513/pdf



http://www.scripps.edu/news/press/2004/021104.html

A visualization model

https://www.youtube.com/watch?v=X-8MP7g8XOE

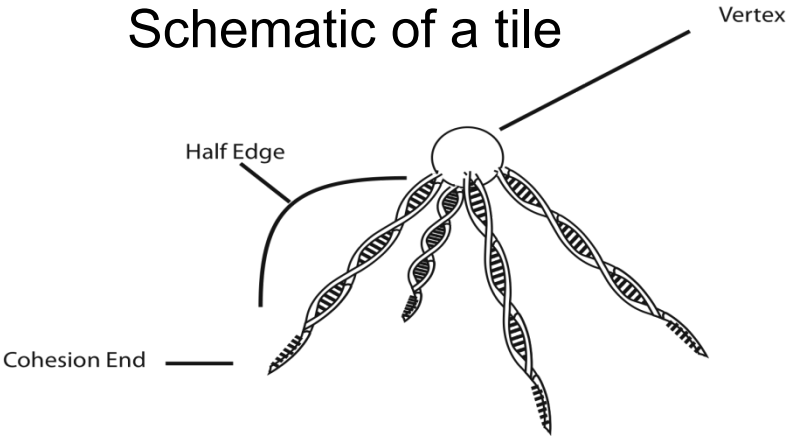
Short and long

https://gulfspecimen.org/specimen/echinodermata/brittle-stars-and-serpent-stars/

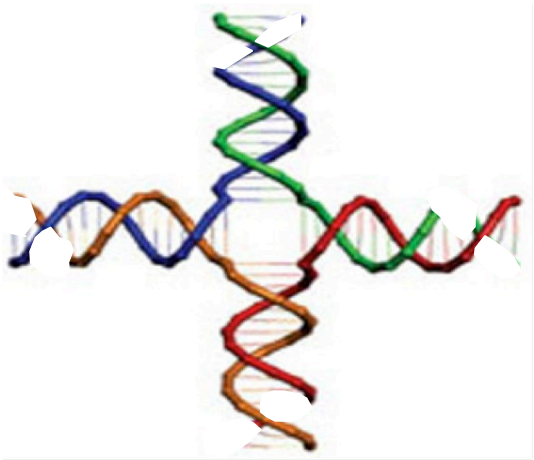
https://www.flickr.com/photos/edbierman/7383798192/

What is a tile?

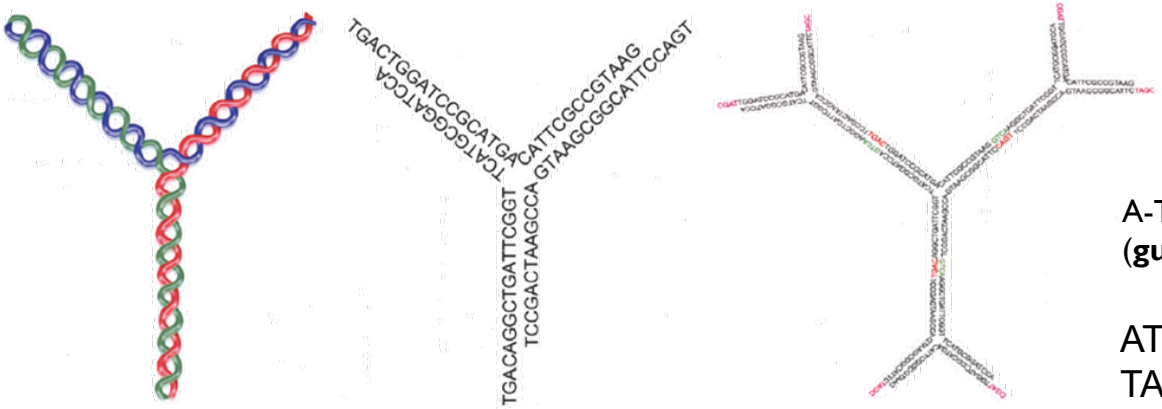
Schematic of a tile



The branched junction molecule



With the bases specified



A-T (**adenine - thymine**) and G-C (**guanine - cytosine**).

ATTCG GGTAACATTCG
TAAGCCCATTG TAAGC

Y-shaped DNA. Schematic diagrams of the structure (left) and sequence (middle) of Y-DNA, and dendrimer-like DNA (right).

D. Luo, “The road from biology to materials,” Materials Today, **6** (2003), 38-43

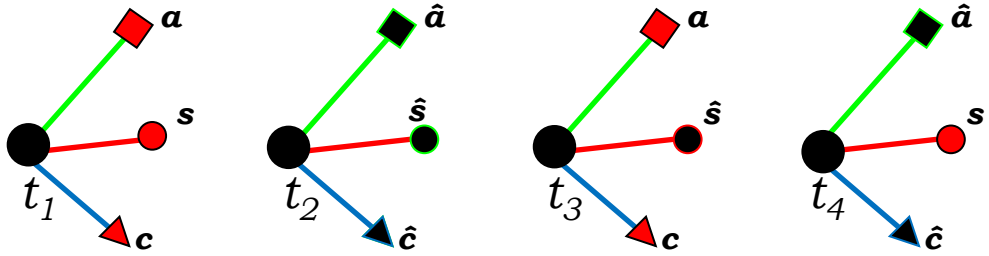
a $\hat{a}$

ATTCG GGTAACATTCG
TAAGCCCATTG TAAGC

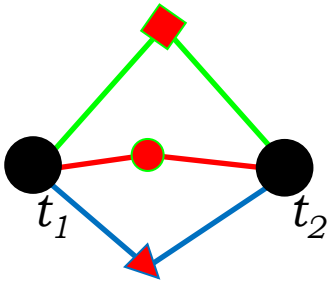
Combinatorial formulation

Both complete complexes and incomplete complexes can be constructed from the this pot P with 4 tiles:

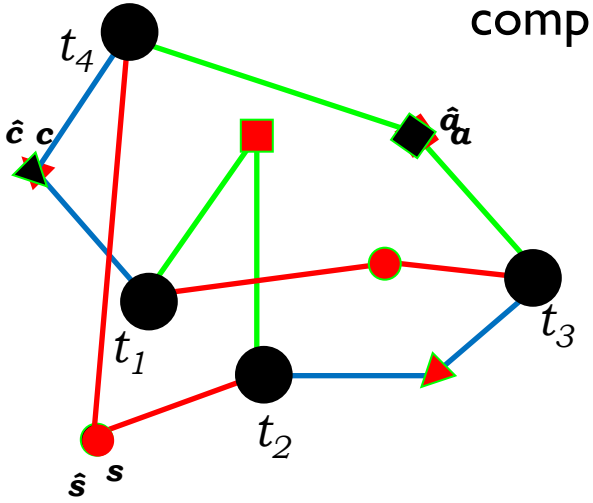
A *pot* P:



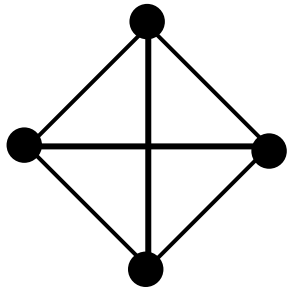
The labs generally want complete complexes



Complete complex



Incomplete complex

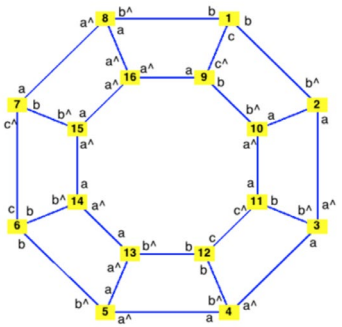


But optimal pots are hard

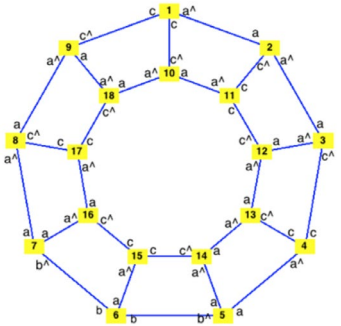
- Computational results with ramifications for the fundamental objectives:
  - If you have a pot *P*, in general you can't efficiently determine what graphs it will construct.
  - If you have a graph *G*, in general you can't efficiently determine a pot that will assemble *G*, but not anything smaller.
- This stinks for the labs, as they would really prefer a fast, general purpose algorithm.
- Changes the focus to: complexity for special classes, pragmatic solutions, approximation algorithms, explicit minimal pots for high-utility graphs, etc.

Pragmatic code

- We can use existing integer linear programing tools for small examples.
- We have specialized code that will handle 2-3 degrees of freedom and can determine if a graph is the unique smallest construct assembled by a pot.
- Simulator with novel use of configuration model.
- Specific designs for sporadic ad hoc cases, but quite difficult to prove optimality.



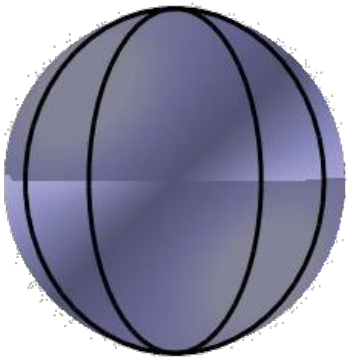
$P = \{\hat{a}^3, ab\hat{c}, b^2c, \hat{b}a\hat{a}\}$



$P = \{\hat{a}c^2, ab^2, \hat{a}a\hat{c}, a^2\hat{b}\}$

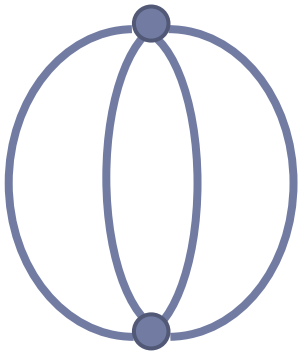
- More algorithms are needed– for special classes, in restricted settings, and for approximate solutions.

Basic design process for DNA origami

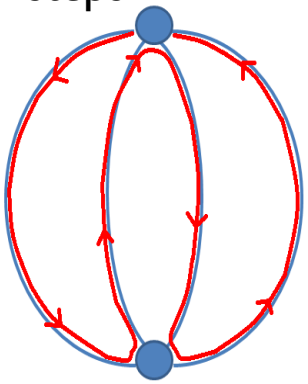


The target

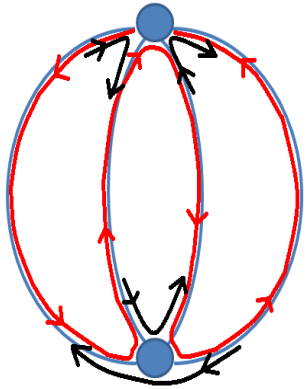
The design steps



Represent the target as a graph



Find an optimal route for the scaffolding strand



Locate the staple strands

MANY constraints:

- Just an Eulerian circuit if the target happens to be an Eulerian graph.
- For non-Eulerian graphs, must adapt either the graph or the circuit to enable the construction.
- Identify structurally appropriate augmenting edges.
- Constraints on turnings
- No interwoven strands
- Symmetry preferred
- Commensurable edge lengths (full turns of DNA)
- Overall length of strand

https://www.youtube.com/watch?v=5yH5LTXxFzk

The generic problem from the lab

- We want to build a molecule with the structure of some graph in three space using DNA origami
- We want the scaffolding strand to follow faces as much as possible, but if it can't, we want it at least not to cross-over or interweave.

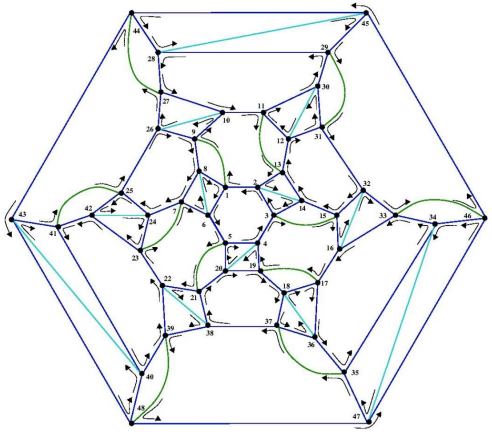
Please provide the best possible route for the scaffolding strand through the graph.

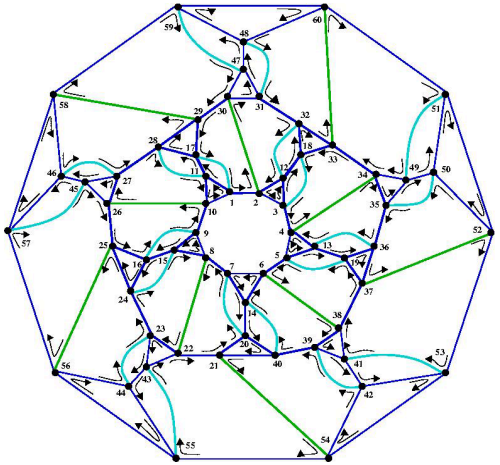
Results

- Existence NP-complete for A-trails in plane graphs (Andersen, Fleischner 95, Bent, Manber 87)
- Polynomial time for non-crossing, planar 4 regular graphs (Kotzig, 2002)
- We give a polynomial time reduction to 3-SAT, which is NP-Complete to show still hard for maxdeg =8.
- Maxdeg = 6 is open, and seems quite difficult.

Basically, proved this design problem is also provably hard.

Often can provide specific designs– or prove they are impossible for given constraints





Reformulation as TSP

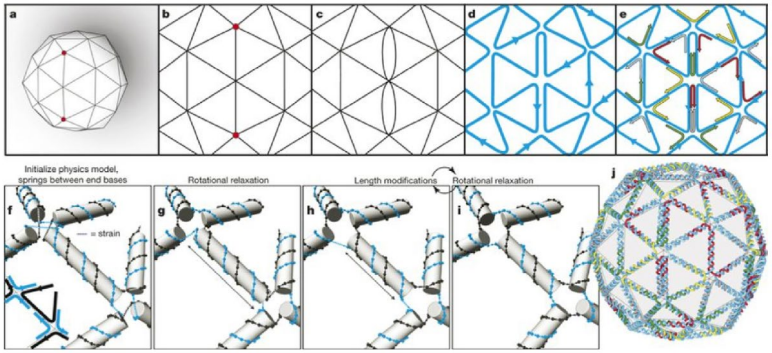
Theorem:
Given an Eulerian graph *G* with (pair consistent) turning costs, there is an associated edge weighted graph L(G_2) that may be constructed in polynomial time so that applying the TSP to L(G_2) yields a minimum cost Eulerian circuit for *G*.

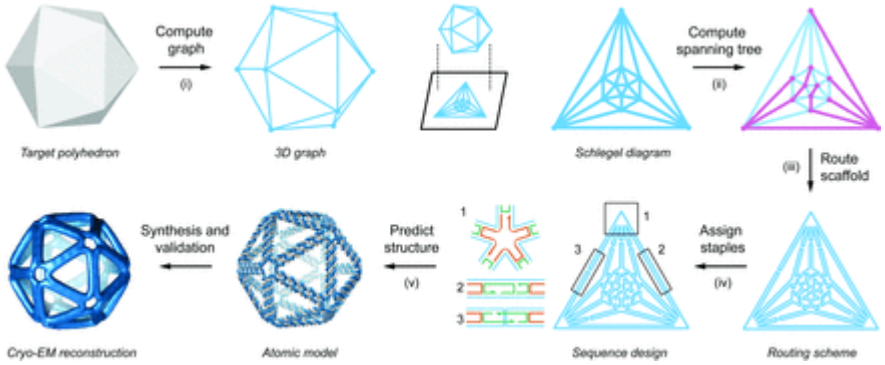
This means that existing TSP programs can be used for quite reasonably sized instances of DNA origami. We have written prototype software for this.

E-M, McDowell, Moffatt, Pangborn
Blakely, E-M, Falcigno, Pangborn

Other pragmatic approaches

- A-trail algorithm, restricted to spherical meshes and with some thickened edges –Benson, Mohammed, Gardell, Masich, Czeizler, Orponen, Hogberg, and approximation algorithms for triangulation on higher genus surfaces –Mohammed, Hajij
- Fast algorithm if all edges thickened —Veneziano, Ratanalert, Zhang, Zhang, Yan, Chiu, Bathe



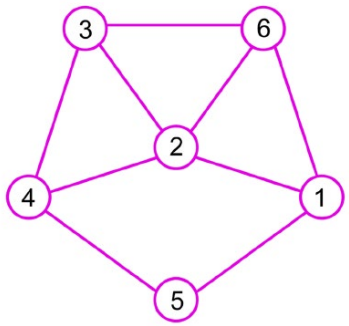


The routing problem on the backend

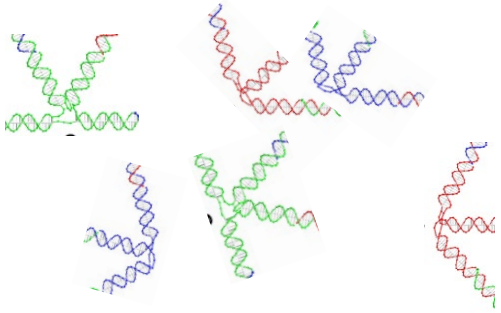
- Tile-based assembly and strand routing are examples of 'front end' design problems.
- But also need good walks through the target structures, called ***reporter strands,*** for reading the output from biomolecular computing or other experiments.

Biomolecular computing of 3-coloring

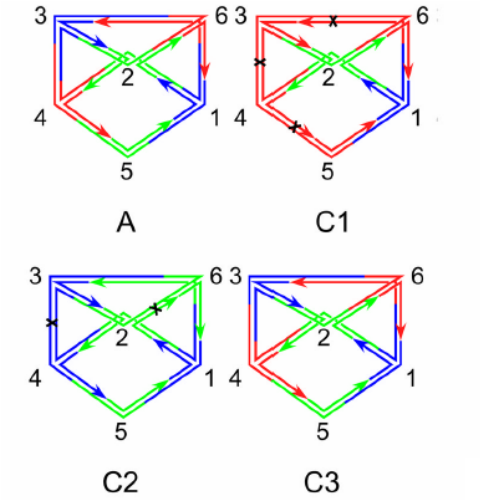
The graph to test for 3 colorability

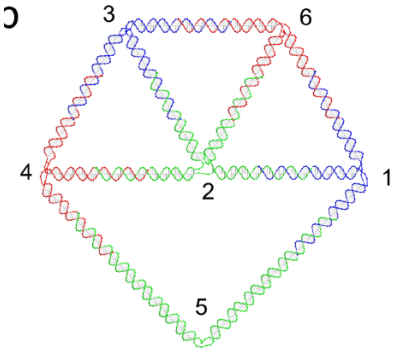


1. Make tiles that assemble the graph in three different `colors', but only unlike tiles bond properly.

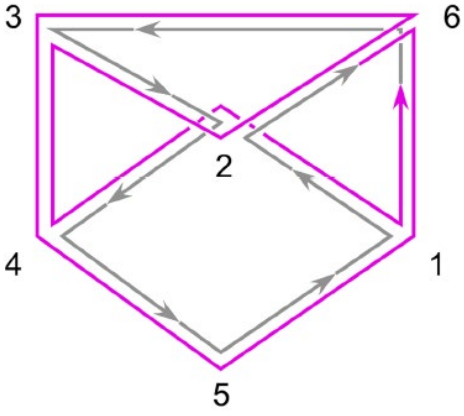


2. Self-assemble, noting 'nicks' between like colors.





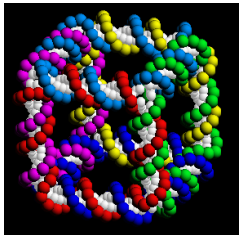
3. Strip off 'reporter strand' to read and confirm solution.



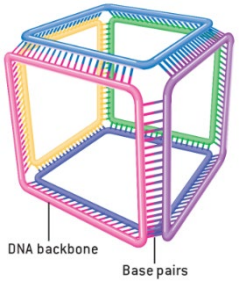
Construction of a DNA Nano-Object Directly Demonstrates Computation, G. Wu, N. Jonoska, and N. C. Seeman *Biosystems*. 2009 November ; 98(2): 80–84. (Experimental results, not just theory.)

Topological graph theory connection

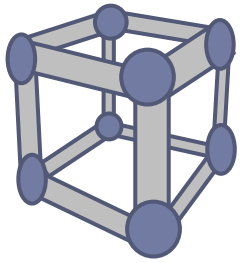
http://seemanlab4.chem.nyu.edu/



DNA cube



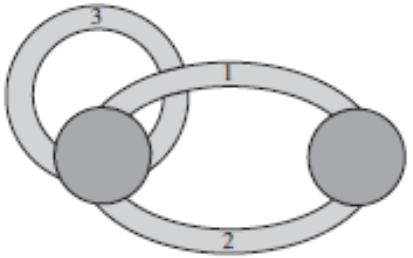
Schematic of cube (twisting of DNA strands not shown).



Ribbon graph model, with ribbon edge boundaries corresponding to DNA backbones

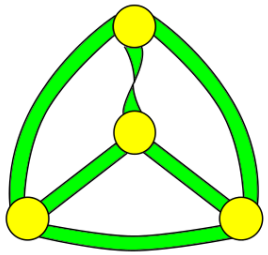
(a) A cellularly embedded graph G.

(b) G as a band decomposition.



(c) G as a ribbon graph.

Non-cellularly embedded graph



K_4 in Projective Plane

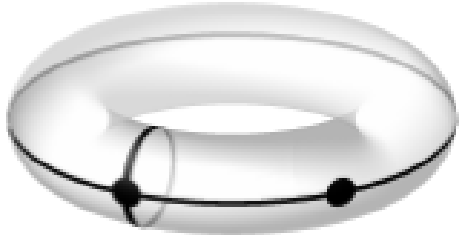
Sophisticated mathematical tools here

- For reporter strands in non-Eulerian graphs, we want a closed walk that:
  1. Traces the whole graph
  2. Repeats as few edges as possible
  3. If it repeats any edge, it doesn't double back, and it retraces the edge in the reverse direction (since single stranded DNA is directed, and two strands have to be going in opposite directions if they bind to a double helix).

- Equivalent to:

  Given a graph, cellularly embed it in some surface so that there is a special 'outer face' that all the edges are incident to.

- Moreover, find such an embedding with the smallest possible size outer face (fewest edges)



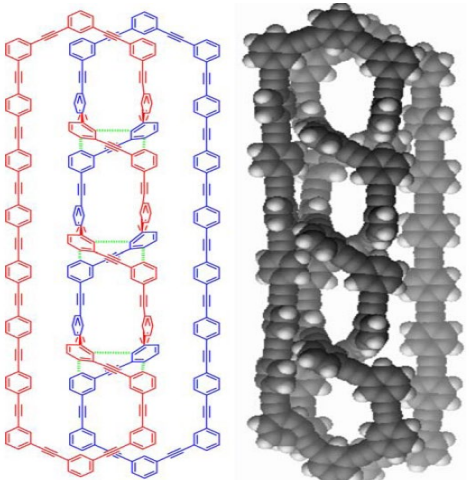
Critical Questions

- Does such an edge-outer embedding always exist? Yes— Jonoska, Seeman, Wu and Ellingham, E-M
- Is there a fast algorithm for finding a solution? Yes—Very fast! Ellingham, E-M
- Is there a fast algorithm for finding a best possible solution?

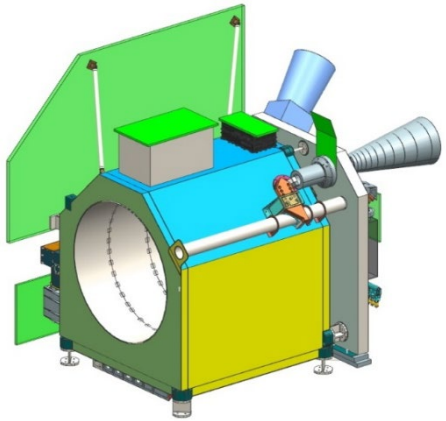
No….This is provably hard, even for 3-regular, cubic, plane graphs. Ellingham, E-M

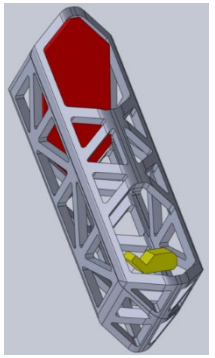
But now working on some very nice special cases, for which we have found fast algorithms.

Not just DNA… hydrocarbons, space deployment, statistical mechanics, computer chip design, crystals



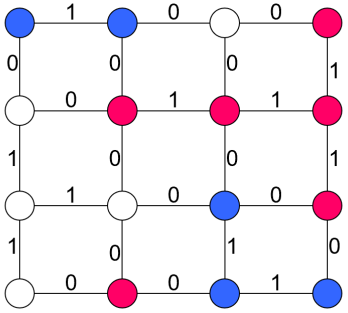
M. Cox, J. Ellis-Monaghan, T. Hughes, K. Mondanaro, "Hydrocarbon links in an octet truss





NASA ATLAS project

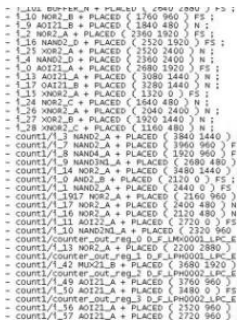
J. Coolidge, K. O'Brien (SMC 2011), "Utilizing the octet truss in the design of lateral transfer retroreflectors.

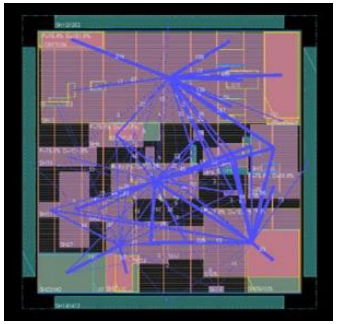


Potts model

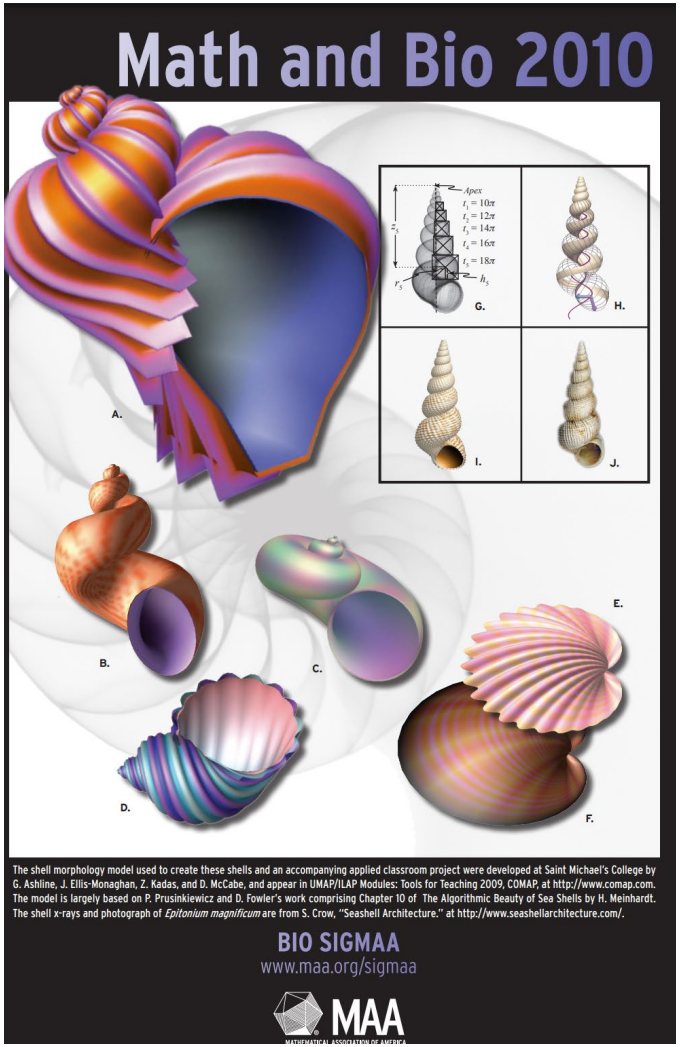
VLSI

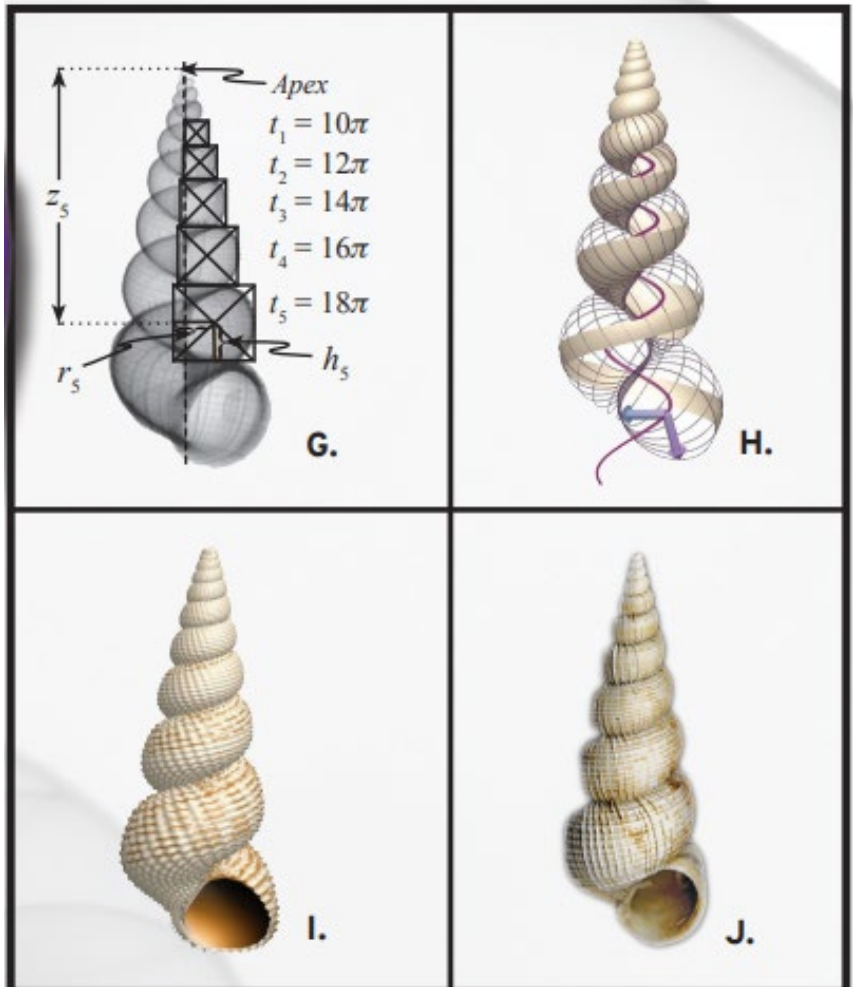
Graph Drawing





Just for fun





Epitonium magnificum

Alas…

- Math is not magic