Award #: 1740212

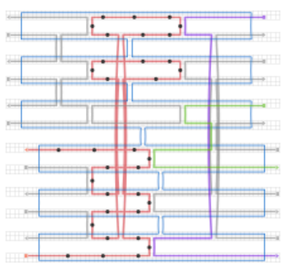
# CSSI Element: SI2:SSE: Collaborative Research: Integrated Tools for DNA Nanostructure Design and Simulation

PI: Alek Aksimentiev, Co-PI: Shawn M. Douglas

Institutions: Univ. of Illinois Urbana Champaign, Univ. of California San Francisco

## A

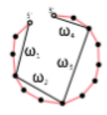
Autobreak finds optimal break solutions by treating the search as a **k-shortest-paths** problem.



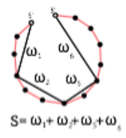
Each staple is encoded as a graph with nodes as possible **breakpoints**.

Edges are added to provide paths from the start to the end of the staple.

A **path** represents a solution whereby the staple is broken at each visited node.



Edge weights (w) are assigned by a scoring function, described below.

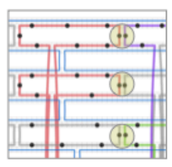



The score for any solution is the sum of its edge weights.

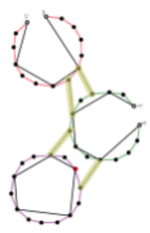
If we choose edge weights ($\omega$) to represent a favorable property of the design, then shorter paths will represent better break solutions.

$$\omega = \ln P_{staple}$$

We assign $\omega$ the log probability that the staple flanked by the two breaks will fold. The calculation of this probability is described below.



Our global approach allows for constraints to **avoid** breaking both halves of a crossover.



After setting up the full graph according to user-specified rules to determine allowed breaks positions, we apply several iterations of the **k-shortest-paths** algorithm to find many possible solutions and then return the best result.

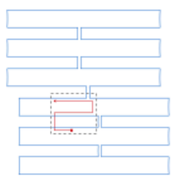
$$P_{origami} = \prod P_{staple}$$

$$\ln P_{origami} = \sum \ln P_{staple}$$

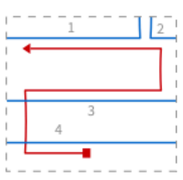
We can score the entire origami solution by multiplying the probabilities of all the staples folding. However, addition operations are is less computionally expensive, so we sum the logs.

## B

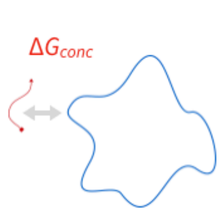
We developed a simple nativistic thermodynamic model for calculating break path edge weights.



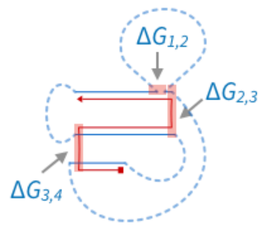
There are many ways to score a staple break, including length, melting temperature ($T_m$), or the presence of long segments above a length threshold (Martin & Dietz 2012). We tested several, and here we summarize our preferred method.



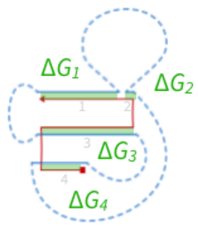

Consider a staple designed to bind four segments of the scaffold. To calculate the edge weight for this staple route, we consider 3 terms:




1. concentration-dependent free energy change assocated with each staple **associating** with the scaffold




2. the energy cost for each scaffold **loop closure** whereby the staple restricts movement of distant scaffold segments




3. free energy gained by **hybridization** of the staple to the scaffold

$$\Delta G_{total} = \Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_4 + \Delta G_{1,2} + \Delta G_{2,3} + \Delta G_{3,4} + \Delta G_{conc}$$

Favorable: SantaLucia & Hicks 2004

Unfavorable: Dunn KE *et al.* 2015

Unfavorable: Dunn KE *et al.* 2015

$$\Delta G_{total} = \sum \Delta G_i + \sum \Delta G_{i,i+1} + \Delta G_{conc}$$

These values are pre-calculated once at the start of the algorithm and used to score each break solution.

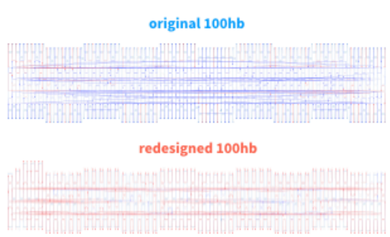
$$P_{staple} = \frac{e^{-\Delta G_{total}/RT}}{1 + e^{-\Delta G_{total}/RT}}$$

We can use the Boltzmann distribution to express the $\Delta G$ for each staple as a probability of the staple folding, and use the values as edge weights for the k-shortest-paths algorithm described above.

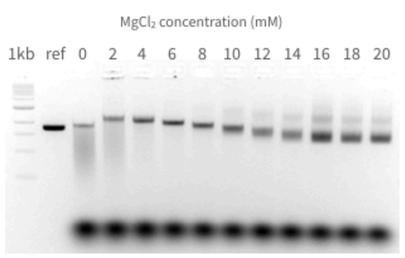
## C

Redesigning published structures with our new Autobreak tool

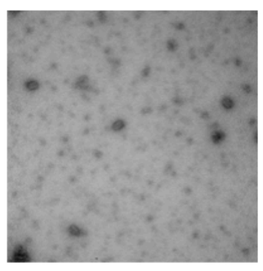
To demonstrate the utility of our approach, we redesigned several published structures. We were able to improve the quality and yield of the 100-helix-bundle by Martin and Dietz, which is a particularly challening design. We used our model to score the original design for comparison.



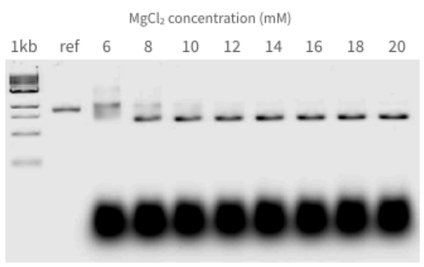

Strand diagrams with staple colors scaled by predicted probability of folding at 50°C (**blue** = lower, **red** = higher)



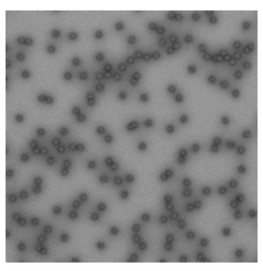

Agarose gel analysis of **original 100hb** (Martin & Dietz 2012)



Neg. stain TEM of **original 100hb** 12-day fold at 20 mM $MgCl_2$ (Martin & Dietz 2012)

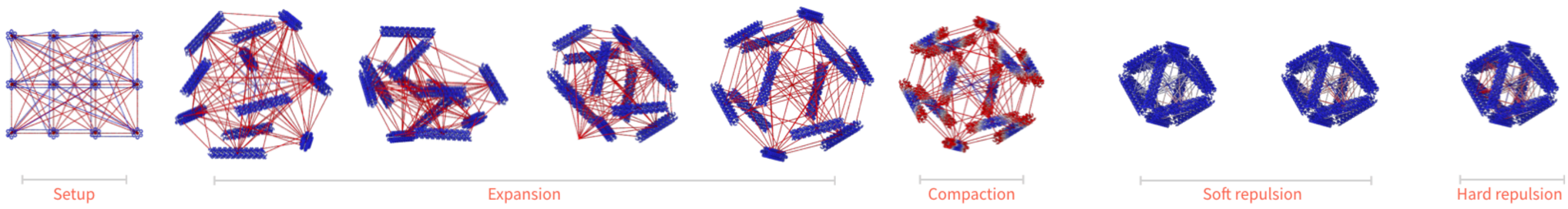



Agarose gel analysis of our **redesigned 100hb**



Neg. stain TEM of our **redesigned 100hb** 3-day fold at 20 mM $MgCl_2$.

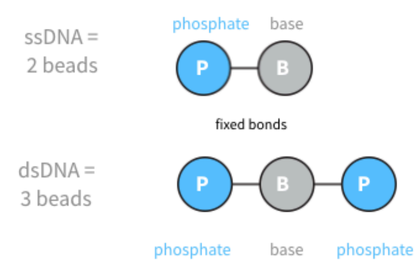
## D



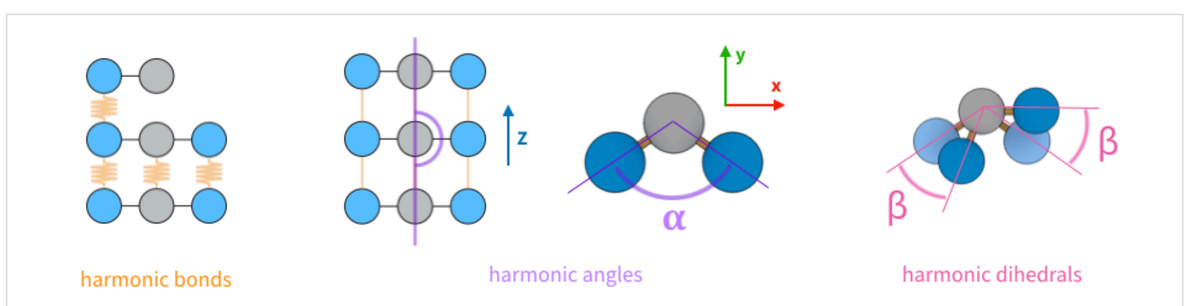

## E


## F

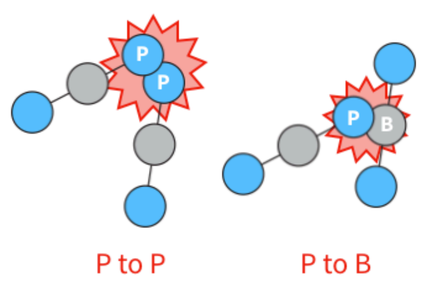
**Coarse Grain** 3-bead (CG3B) model:




Nucleotides



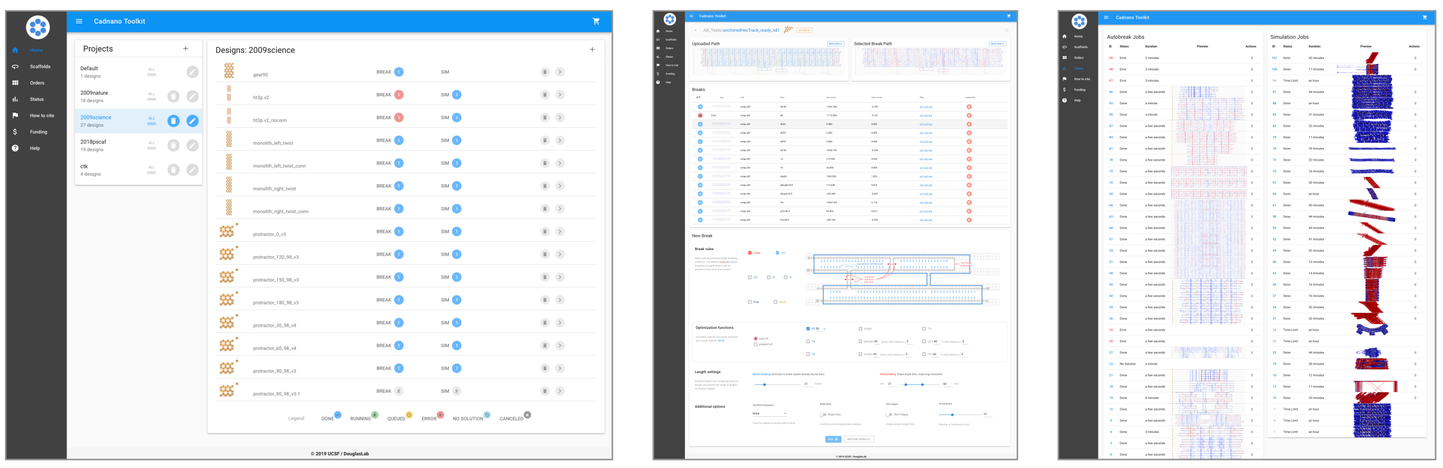

Potentials




Repulsions

## G

We have synthesized and tested about 50 designs using our new **Autobreak** and **Simulation** tools. All data collection is complete, and we are drafting the manuscript. The only pending item before we publish is to complete our web-based software tool (**Cadnano Toolkit**) to make the tools available to the community. This should take about 6 weeks.




Screenshots of the frontend Cadnano Toolkit (CTK ) interface, built with VueJS. The backend API is built with Python, Flask, and nginx.