Supplementary materials for

Evolutionary Refinement of DNA Nanostructures Using Coarse-Grained Molecular Dynamics Simulations.
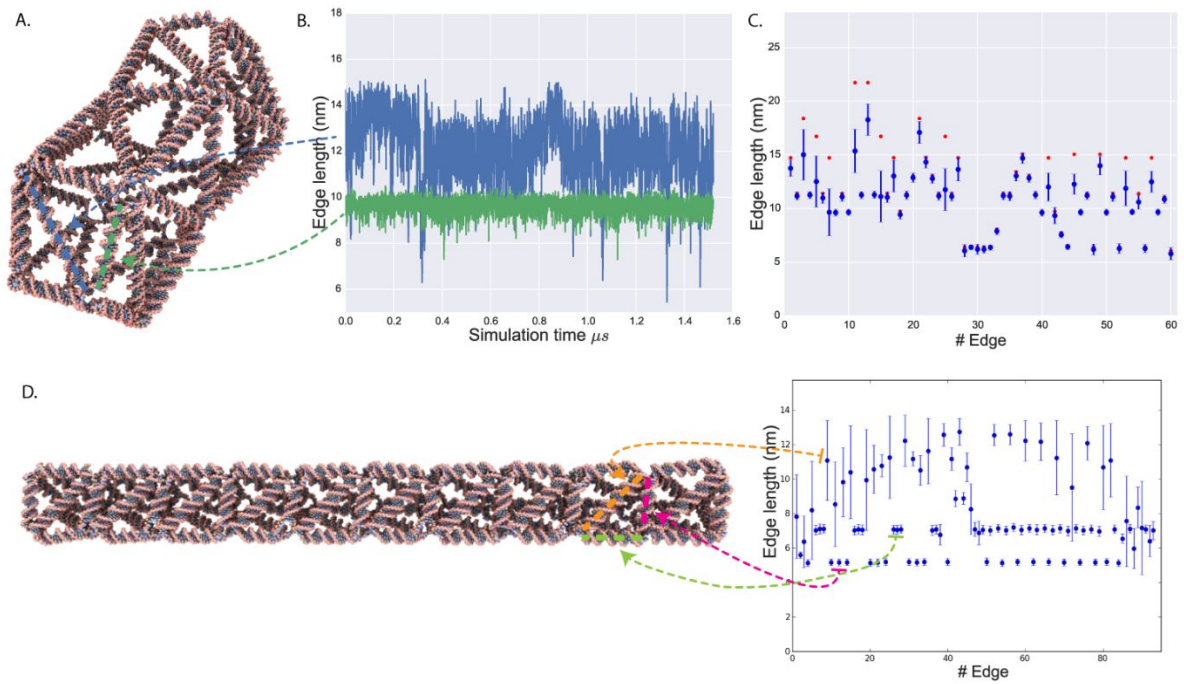
Erik Benson, Marco Lolaico, Yevgen Tarasov, Andreas Gådin and Björn Högberg*

Department of Medical Biochemistry and Biophysics

Karolinska Institutet

17177 Stockholm (Sweden)

*E-mail: bjorn.hogberg@ki.se

Supplementary figure 1. Analysis of wireframe DNA origami structures using oxDNA simulations. A) A barrel shaped DNA origami structure was simulated and the rigidity of its constituent helices was measured by tracking the distance between one nucleotide in the second to firsts and one nucleotide in the second to last base pair of the helix throughout the simulation. B) Time series of the the end to end distance of two adjacent helices in the structures displays drastically different behaviors. C) The mean distance and standard deviation of the values of these time series can be used to describe the flexibility of the helices in a structure (blue markers and bars), and it is apparent that some helices have significantly more flexibility than others, the red markers show the corresponding end to end distances for straight b-DNA helices. D) Analysis of a wireframe rod with a repeating structural motif shows that the helices stratify into three groups. Here the group of longest helices show the largest standard deviations in their time series, meaning that they would contribute the most to the flexibility of the structure as measured by our metric. These three groups correspond to the horizontal, vertical and diagonal helices of the structure, where the diagonal helices show the largest standard deviation.
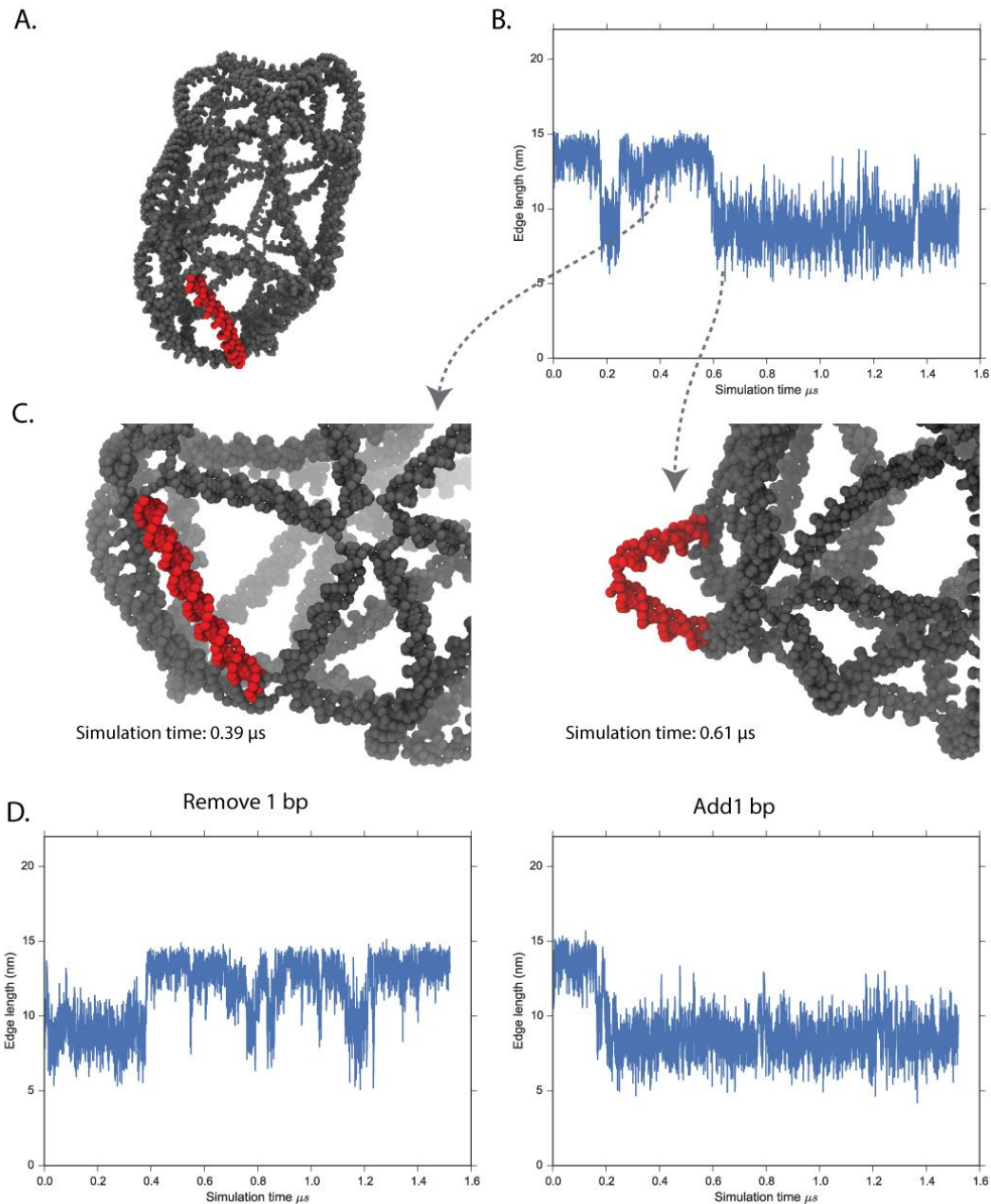
Supplementary figure 2. Example of poorly performing helix, and effects of mutation. A) A frame from a simulation of a barrel shaped DNA origami structure with a poorly performing edge highlighted in red. B) The end-to-end trajectory of the highlighted helix shows a bi-modal behavior of the edge where it alternates between a longer and a shorter state. C) Renders of typical frames from the two states show that the staple break-point "nick" in the middle of helix opens in the shorter state causing major deformation of the helix and a short end-to-end length. D) The helix was mutated by removing or adding one base-pair and the structure was simulated and analyzed again. On this helix, reducing the length by on base pair increased the amount of time spent in the longer non-bent state while increasing the length by one base-pair increase the amount of time spent in the bent state.

A.
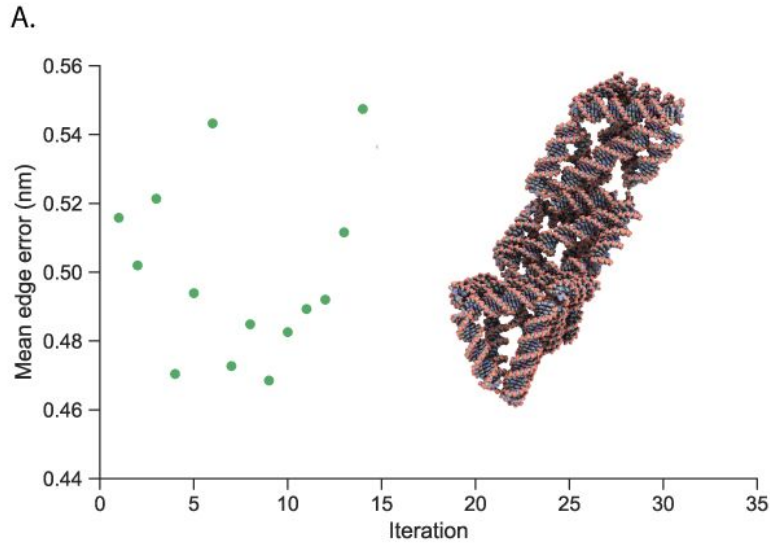
B.

Iteration	Modify helices			
1	6	19	14	10
2	6	27	19	29
3	6	27	19	14
4	6	27	19	29
5	6	19	27	14
6	6	21	26	27
7	6	27	19	29
8	19	14	10	27
9	6	26	21	27
10	6	26	21	27
11	6	26	27	21
12	6	26	27	21
13	6	21	19	27
14	6	27	26	21
15	6	27	21	26
16	6	26	27	21

Supplementary figure 3. Effect of longer simulation time. A very small DNA origami structure consisting of 30 helices and a total of 722 base-pairs was simulated with the constant progression iterative scheme and $5 \cdot 10^8$ time steps. This resulted in an iteration time of around 66 hours. A) The simulation initially showed a positive trend towards lower errors, but this was reversed after 9 iterations. B) Analysis of what helices were modified in the simulation, and which mutations were incorporated, blue helix numbers represent mutations of one extra base-pair being incorporated, and red helix numbers represent deletions of one base-pair being incorporated in the next generation. The simulations show a high degree of determinism where the same helices are consistently ranked as the lowest performing.
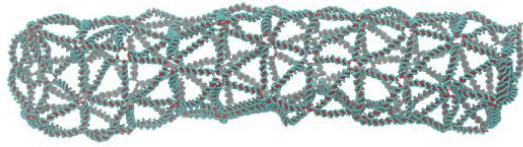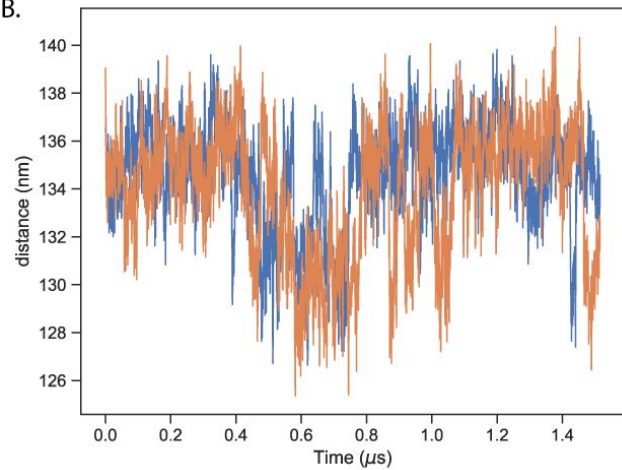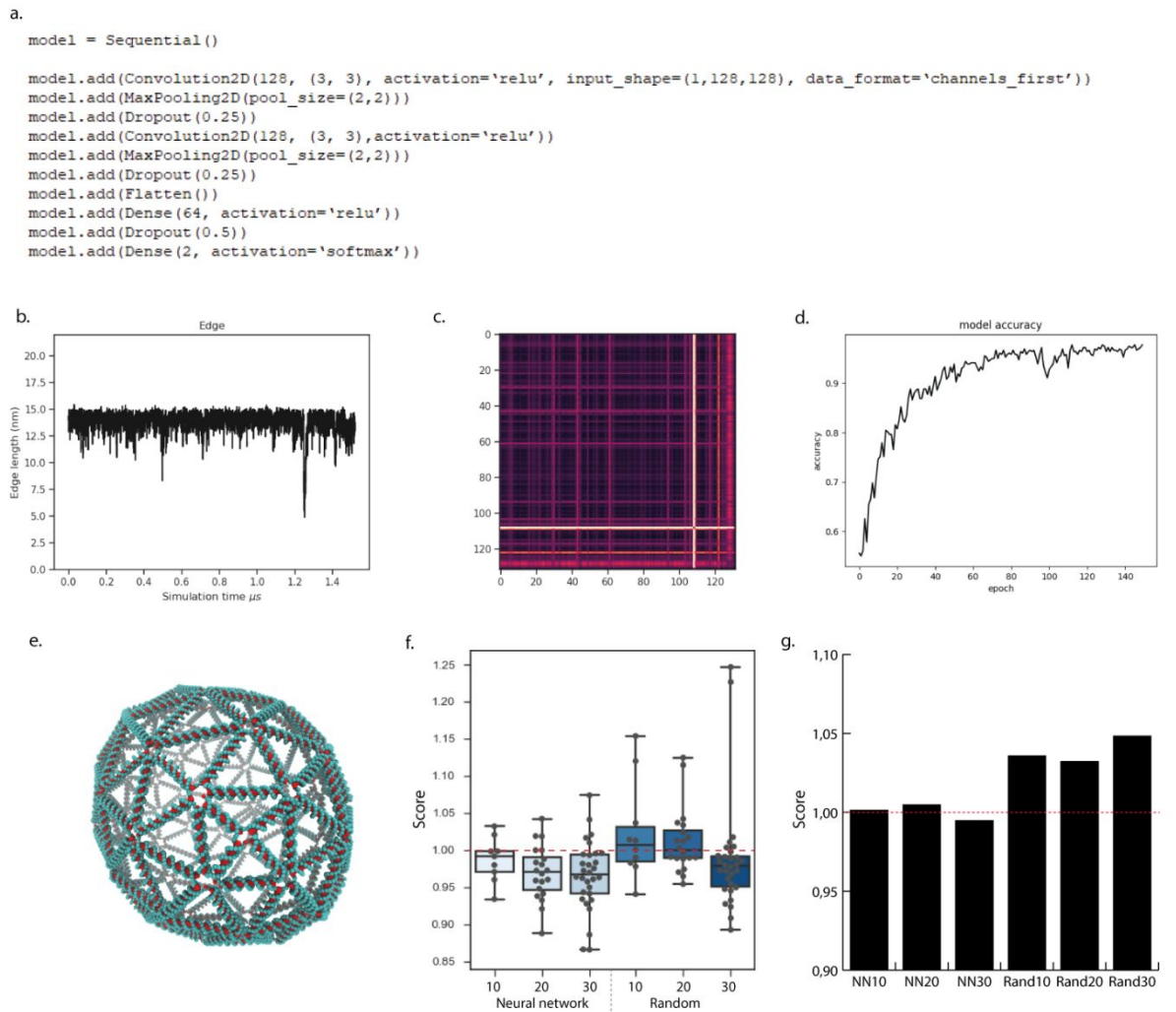
Supplementary figure 4. Additional analysis of multiplexed interative evolution of a wireframe rod with a hexagonal cross-section. A) rendering of a representative state of the simulation of the initial structure. B) End-to-end length of the rod through the simulation of the initial structure (blue) and after 20 generations of refinement (orange). Calculated by generating an average coordinate of the helices that compose the "top" and "bottom" caps of the structure and then calculating the distance between these two coordinates in every saved frame of the trajectory.

a.

```
model = Sequential()

model.add(Convolution2D(128, (3, 3), activation='relu', input_shape=(1,128,128), data_format='channels_first'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))
model.add(Convolution2D(128, (3, 3),activation='relu'))
model.add(MaxPooling2D(pool_size=(2,2)))
model.add(Dropout(0.25))
model.add(Flatten())
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(2, activation='softmax'))
```


Supplementary figure 5. Training a neural network to identify edges that can benefit from mutations. From the multiplexed simulations of the hexagonal rod, a total of 528 simulations were performed on variants of the structure with multiplexed mutations. In total 4238 mutations were evaluated and from that, 389 mutation were classified as beneficial. The beneficial mutations were split into a training set of 358 and a test set of 31. a.) A convolutional neural network was constructed in Keras. b.) From the beneficial mutations, the end to end distance through the simulations (before the mutation was introduced) were extracted as a time-series. c.) The time series were converted to 2D recurrence plots and downscaled to 128×128 pixels. d.) The recurrence plots were used to train the networks to classify mutations as beneficial from addition or beneficial from deletion of a base pair. e.) A spherical structure with 198 edges was designed and simulated. The end-to-end times series from of each edge from the simulation was evaluated by the trained neural network which outputted a strength of association to the classes: addition or removal of base pairs. After this, the edges were ranked by how strongly they were classified to either class. f.) The 10, 20 or 30 edges that had the strongest association with a class were modified according to their class, and as a reference 10, 20 or 30 random edges received random mutations, this resulted in six new structures. The new structures were simulated, and the edges were evaluated by their local effect, were a score below 1.0 is positive. Box plots indicate quartiles of data, and whisker plots show full range of data points. g.) The average score of all edges was evaluated for the structures with neural-network directed and random modifications, also here a score below 1.0 is positive.
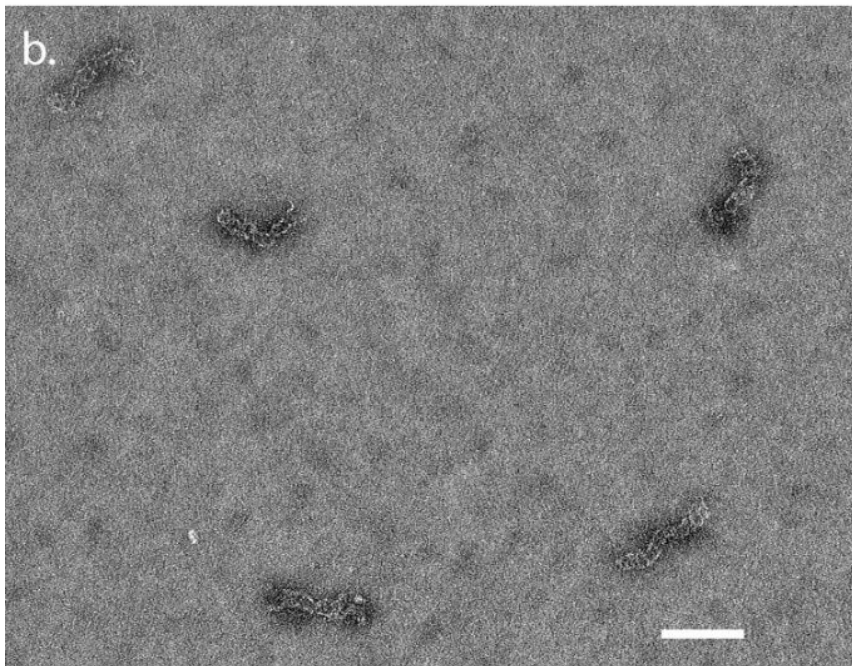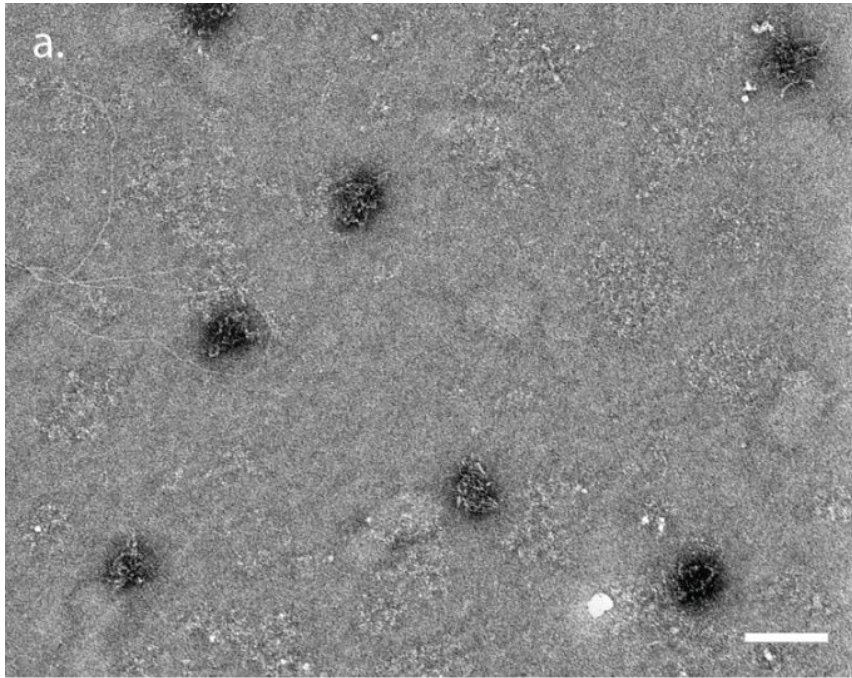
Supplementary figure 6. Experimental assembly of an internally supported square rod before (a.) and after (b.) optimization. Structures were assembled and gel purified as described before (Benson, et al. Nature 523.7561 (2015): 441.). Purified samples were applied on TEM grids and stained with 2 % Uranyl Formate and imaged in a FEI Morgagni 268 at 22 000× magnification. 100 nm scale bars.