Supporting Information for “A coarse-grained model for DNA origami”

Roman V. Reshetnikov¹,²,³,†, Anastasia V. Stolyarova³,⁴,†, Arthur O. Zalevsky³,†, Dmitry Y. Panteleev¹, Galina V. Pavlova¹, Dmitry V. Klinov⁵,⁶, Andrey V. Golovin²,³ and Anna D. Protopopova⁵,⁷

¹Institute of Gene Biology, Russian Academy of Sciences, Vavilova str., 34/5, 119334, Moscow, Russia
²A.N.Belozersky Institute of Physical and Chemical Biology, Lomonosov Moscow State University, Leninskiye gory, 1-40, 119992, Moscow, Russia
³Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, GSP-1, Leninskiye Gory, 1-73, 119234, Moscow, Russia
⁴Skolkovo Institute of Science and Technology, Nobel Street 3, 143026, Moscow, Russia
⁵Federal Research and Clinical Center of Physical-Chemical Medicine, Malaya Pirogovskaya str. 1a, 119435, Moscow, Russia
⁶Moscow Institute of Physics and Technology (State University), 9 Institutskiy per. Dolgoprudny, 141700, Moscow Region, Russia
⁷Department of Cell & Developmental Biology, Perelman School of Medicine, University of Pennsylvania, BRB II/III 421 Curie Boulevard, Philadelphia, PA 19104, USA

*To whom correspondence should be addressed. Email: reshetnikov@fbb.msu.ru

†These authors contributed equally to this work

© 2017 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Figure S1. DNA origami architecture. A: in the honeycomb lattice, a staple oligonucleotide has an opportunity to jump to an adjacent helix with an interval of seven base pairs. Starting from the position when the staple is looking towards the helix 1 (lower right scheme) it makes a turn of 240° within the interval. Thus, after seven base pairs the staple oligonucleotide faces the helix 3; the next possible transition occurs 240° later (480° or 14 base pairs from the start) in the direction of the helix 2. After two full turns of the helix (720°, 21 base pairs), displayed with the fading shades of gray, the cycle repeats. B: in the square lattice a helix has four potential neighbors (lower right scheme), and the cycle takes three turns of the helix, or 32 base pairs. Crossover periodicity is eight base pairs, or 270°. Note that the correspondence is not strict, therefore the helix makes a full turn within 10.67 base pairs instead of 10.5 base pairs in the honeycomb lattice.

Figure S2. Opening of a hexagonal nanocontainer (1). A: the intended trajectory of the predetermined motion; B: different views of the CanDo prediction for the system, according to which the equilibrium shape of the object corresponds to the closed conformation.
Figure S3. **Shape prediction for an icosahedron monomer subunit (2).** A: the initial scheme converted to 3D coordinates; B: different views of the CanDo prediction for the equilibrium form of the system; C-G: major shapes of the monomer predicted by our model with clustering analysis of the MD trajectory. According to the simulation results, the monomer has a flexible structure with the main modes of motion associated with the bending around the crossbar and the movements of the terminal beams. The large number of almost equal-sized clusters indicates that the monomer does not have a single equilibrium shape, appearing as an ensemble of conformations instead.

Figure S4. **Scheme of the distance between the dyes from the initial scheme of the Sq-system, front view on the short arm.** The distance between the central axes of the arm-forming helices is 22 Å. The dyes were placed at the distance of 19 bp from the crossover between the helices, which gives a B-DNA twist angle of 652°. Resulting positions of the dyes are shown with black-shaded circles.
1 COMPUTATIONAL MODEL

Note S1.1 Input data processing

Our tool uses a CaDNAno-generated json file with a scheme of a DNA origami design to generate two run input files for MD simulation in GROMACS (3):

- structure file in pdb format containing 3D coordinates of the design in the COSM representation;
- distance restraints file describing fixed nonbonded interactions between the COSM particles (Note S1.4)

Translation from the CaDNAno scheme to the COSM representation begins from the 5’-end of a scaffold chain. If the scaffold chain is circular, the algorithm breaks the chain in the middle of the first long duplex region (at least 21 bp for the honeycomb lattice and 36 bp for the square lattice).

First, we divide the scheme into a sequence of individual double- and single-stranded regions starting from the 5’-end. The boundaries of the double-stranded regions are either at the beginning of the adjacent single-stranded regions or at crossovers formed by the scaffold chain. In the latter case, the next duplex region follows the scaffold chain and starts on the neighboring strand.

Second, all double- and single-stranded regions are filled with COSM particles as described in the main text of the article. Initial grid coordinates inherited from the json file are translated into Cartesian coordinates assuming that neighboring helices are spaced 24 Å apart and the step between base pairs is 3.4 Å (4) (Figure S5). The obtained Cartesian coordinates of the particles in COSM representation are written in a pdb file.

Note S1.2 Bonded interactions

Note S1.2.1 Bond stretching interactions A bond between two particles \(i\) and \(j\) is stretched with a harmonic potential:

\[
V_b(r_{ij}) = \frac{1}{2}k^b_{ij}(r_{ij} - b_{ij})^2
\]

where \(k^b_{ij}\) is a bond stretching constant, \(b_{ij}\) is an equilibrium distance between the particles, and \(r_{ij}\) is an actual distance between them. The values of the bond stretching constants \(k^b_{ij}\) were derived from the Young’s modulus \(Y\) of the DNA:

\[
k^b_{ij} = \frac{YS}{b_{ij}}
\]

Here, \(S\) is a cross-section area of the DNA (3.14 \(nm^2\) for dsDNA and 0.3 \(nm^2\) for ssDNA). The product \(K_0 = YS\), sometimes called the stretch modulus, was measured experimentally using laser tweezers and is equal to 1087 pN for dsDNA and 800 pN for ssDNA (5). Bond lengths in double-stranded regions depend on the number of base pairs between individual particles, assuming 3.4 Å as a distance between two adjacent base pairs (see Table S1). For single-stranded regions, bond length was set to 6.3 Å (6).
Note S1.2.2 Bond angle interactions

Chain stiffness is described with a harmonic potentials for the angle between three neighbouring particles $i$, $j$, and $k$

$$V_a(\theta_{ijk}) = \frac{1}{2} k^{\theta}_{ijk} (\theta_{ijk} - \theta_{ijk}^0)^2$$ (3)

where $k^{\theta}_{ijk}$ is an angle force constant, $\theta_{ijk}^0$ is an equilibrium angle between the particles, and $\theta_{ijk}$ is an actual angle between them.

Angle force constants $k^{\theta}_{ijk}$ were assigned according to the worm-like chain model of DNA (7), where nucleic acid is approximated as an isotropic, continuously flexible rod with a bending energy

$$E = \frac{1}{2} p l k_B T \theta^2$$ (4)

where $k_B T$ is the thermal energy, $\theta^2$ is a mean square angle between the two segments located at the contour distance $l$ from each other, and $p$ is the DNA persistence length (8). We used $p=510$ Å (5) and $p=20$ Å (9) for double- and single-stranded DNA, respectively.

From 3 and 4, the angle force constants are:

$$k^{\theta}_{ijk} = \frac{p}{l_{ijk}} k_B T$$ (5)

where $l_{ijk}$ corresponds to the summarized bond length between the three particles $i$, $j$, and $k$.

Angle force constants are assigned according to the worm-like chain model of DNA (7), based on the DNA persistence length (510 Å for the double-stranded DNA (5), and 20 Å for the single-stranded DNA (9)).

Note S1.3 Non-covalent interactions

Note S1.3.1 Lennard-Jones interactions

Particles $i$ and $j$ interact with each other on a distance $r_{ij}$ via the Lennard-Jones potential:

$$V_{LJ}(r_{ij}) = 4 \epsilon_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6$$ (7)

Here, $\sigma$ corresponds to an effective size of a particle and $\epsilon$ determines the strength of the interaction between particles. To calculate the parameters for a pair of particles $i$ and $j$ the following combination rules are used:

$$\sigma_{ij} = \frac{1}{2} (\sigma_{ii} + \sigma_{jj})$$ (8)

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}}$$ (9)

Effective sizes of the particles $\sigma_{ii}$ in our model correspond to diameters of simulated building blocks increased by 1 Å to simulate a solvent shell of DNA origami (see Table 1). The values of the $\epsilon_{ii}$ parameters were selected by fitting of computational models for best correspondence with experimental data. Strength of non-covalent interactions between all types of particles was set to 1.4 kJ/mol. The only exception was the N particle, which had the value of described with a harmonic potential

$$V_d(\phi_{ijkl}) = k_d (1 + \cos(n\phi - \phi_s))$$ (6)

where $k_d$ is a dihedral angle force constant, $\phi$ is the angle between the planes, with zero corresponding to cis conformation, $n$ is the multiplicity factor and $\phi_s$ is a phase angle. Torsional stiffness of $35 \cdot 10^3$ pN nm/rad (10, 11) was used for treatment of torsional angles.
Figure S6. Selection of the optimal $\epsilon$ parameter for $N$ particles based on simulations of a bent DNA origami structure. A: a COSM model of the bent DNA origami structure (12). Vectors plotted through the red dots illustrate how the bending angle was measured. B: Dependence of the bending angle on the $\epsilon$ parameter. Each box plot shows the distribution of angles obtained from $3 \times 10^5$ integration steps-long trajectories with a time step of 100 fs. The box-plot corresponding to the best $\epsilon$ value is outlined with red.

1.2 kJ/mol for better correspondence with experimental data (Figure S6).

Note S1.3.2 Coulomb interactions Particles $i$ and $j$ with charges $q_i$ and $q_j$ interact with each other via the Coulomb potential:

$$ V_C(r_{ij}) = \frac{1}{4\pi \epsilon_0 \epsilon_r r_{ij}} \frac{q_i q_j}{r_{ij}} $$

(10)

where $r_{ij}$ is the distance between the particles, $\epsilon_0$ is the vacuum permittivity and $\epsilon_r$ is a relative dielectric constant (default value is 1).

Initial estimates of the particle charges were set as a combination of component nucleotides charges,

$$ q_0 = \sum_{n=1}^{N} q_n $$

(11)

where $N$ is the number of the nucleotides simulated by the particle, $q_n = -1$ is the charge of an individual nucleotide. Then, all charges were reduced by a factor of square root of $\epsilon_r = 80$ (dielectric constant of water) to simulate water environment. Finally, the charges were corrected with regard to the effect of salt concentration on the B-DNA electrostatics. There are two terms that determine the magnitude of the effect:

1. Neutralizing term (see Supplementary Note S1.5.1): cations surround a negatively charged DNA, providing an electrostatic shielding. The ion environment modulates DNA interactions with itself and other
charged molecules. For example, decreasing ionic strength from 260 to 110 mM increases the effective diameter of rod-like viruses’ filaments from 9.2 to 10.5 nm (13). A similar behaviour was observed for 6-helix DNA origami (14);

2. **Repulsing term** (see Supplementary Note S1.5.2): high salt concentrations reduce dielectric constant of the solution, enhancing electrostatic interactions between the nucleotides.

Default charge values $q'$ in our model were empirically optimized for the ionic strength of 30 mM representing a standard DNA origami buffer salt composition (5 mM NaCl, 10 mM MgCl$_2$). The best correspondence between the experimental and simulated data was achieved with an additional reduction of the charges by 89%:

$$q' = 0.11 \frac{q_0}{\sqrt{\epsilon_\omega}} \quad (12)$$

Charge shielding of single-stranded DNAs by counterions is less significant (15), therefore there was no extra charge correction for the S particles in our model. Final charges of the particles are specified in Table 1. Optionally, custom ionic strength representing, for example, different concentration of MgCl$_2$ could be taken into account as described in Note S1.5.

**Note S1.3.3 Nonbonded interactions cutoff** To speed up the calculations and minimize background noise the nonbonded interaction terms are used in their shifted form, where both energy and force become zero beyond a cutoff distance of 20 Å; nonbonded interactions between the nearest neighbors along the coarse-grained strand were also excluded.

### Note S1.4 Distance restraints

For a successful relaxation of an origami structure we use several types of distance restraints (Table S1, Figure S7):

1. Restraints rejoining the ends of the circular scaffold strand in case if it was artificially broken (Figure S7A, distance $d_e$);

2. Restraints modeling crossovers formed by the scaffold chain (Figure S7A, distance $d_{sc}$);

3. Restraints describing staple-formed crossovers:
   - Distance between adjacent helices (Figure S7A, distance $d_c$);
   - Diagonal distance between a particle participating in crossover formation and a particle of type H or PT neighboring to its counterpart (Figure S7A, distance $d_{cg}$);

4. Lattice type-specific restraints that control the angle between the adjacent strands, 120° for the honeycomb lattice and 180° for the square lattice (Figure S7B, distance $d_{lg}$).

For each type of distance restraints, we use a piecewise linear/harmonic potential with three threshold values for a distance between particles $r_{ij}$:

$$V_{dr}(r_{ij}) = \begin{cases} 
\frac{1}{2} k_{dr} (r_{ij} - r_0)^2 & \text{for } r_{ij} < r_0 \\
0 & \text{for } r_0 \leq r_{ij} < r_1 \\
\frac{1}{2} k_{dr} (r_{ij} - r_1)^2 & \text{for } r_1 \leq r_{ij} < r_2 \\
\frac{1}{2} k_{dr} (r_2 - r_1)(2r_{ij} - r_1 - r_2) & \text{for } r_2 \leq r_{ij}
\end{cases} \quad (13)$$
Table S1. Parameters of distance restraints

| Type     | \( r_0, \ \text{Å} \) | \( r_1, \ \text{Å} \) | \( r_2, \ \text{Å} \) | \( k_{dr}, \ \text{kJ/mol nm}^2 \) |
|----------|----------------|----------------|----------------|-----------------------------------|
| \( d_e \) | 3.30           | 3.40           | 3.70           | 1000                              |
| \( d_{sc} \) | 3.30         | 3.40           | 3.70           | 1000                              |
| \( d_c \) | 20.04          | 21.45          | 22.86          | 1000                              |
| \( d_{cg}^* \) | 31.11       | 32.04          | 33.00          | 100                               |
| \( d_{cg}^{**} \) | 33.78      | 34.64          | 35.53          | 100                               |
| \( d_{lg}^* \) | 36.80       | 37.50          | 37.70          | 100                               |
| \( d_{lg}^{**} \) | 48.00       | 52.00          | 52.30          | 100                               |

* Honeycomb lattice

** Square lattice

where \( k_{dr} \) is a distance restraint force constant and \( r_0, r_1, r_2 \) are the threshold values. The values of the parameters are specified in Table S1.
Figure S7. Types of distance restraints. A: an origami design scheme in COSM model. Gray lines mark staple-formed crossover sites; B: side view on the design from the panel A. Distance restraints types are named and colored.

Figure S8. Simulation of the shape of three-bundle DNA origami system: effect of salt concentration. A: 2D plot of the system. Black: scaffold chain, gray: staple strands; B: idealized COSM model of the system, face view. Black: scaffold chain, gray: staple crossovers; C: final shape of the system at different ionic strength, top view.

Note S1.5 Salt conditions consideration

Custom salt concentration could be taken into account as a correction coefficient for the Coulomb potential, specified as the \( \epsilon_r \) parameter in the MD configuration file of the standalone version. Nominally, the parameter assigns the relative dielectric constant for a simulation (the default value is 1), but it can be used for multicomponent influence on the electrostatic interactions. To determine the value of the \( \epsilon_r \) parameter, the two salt concentration effect terms should be considered.

Note S1.5.1 Neutralizing term From the Manning equation (16), there is a direct dependence between Debye screening length and the axial charge spacing of the polyion. Consequently, the charge spacing is proportional to a square
root of the ionic strength expressed in mol/L. Electrostatic parameters of our model were optimized for ionic strength of 30 mM, therefore the charges of the particles should be corrected in accordance with the desired ionic strength:

\[ q'' = q' \sqrt{\frac{0.03}{I''}} \]  

(14)

where \( q' \) is the charge of a particle at the ionic strength of 30 mmol (see 12), \( q'' \) is the charge at the ionic strength \( I'' \).

**Note S1.5.2 Repulsing term** According to Hasted and coauthors (17), there is a linear relationship between the dielectric constant and the salt concentration of the solution:

\[ \epsilon = \epsilon_\omega - \delta c \]  

(15)

where \( \epsilon_\omega \) is the dielectric constant of water, \( c \) is the salt concentration and \( \delta \) is an ion-specific parameter (17). Note that the dependence is valid only for salt concentrations below 2 M. At higher salt concentrations significant deviations from linearity were observed and the dielectric decrement was observed to saturate (17).

**Note S1.5.3 Combined effect** Since we include both water and salt environment in the particles’ charge values (see Supplementary Note S1.3.2), the Coulomb potential correction should be provided as a ratio between the default and custom parameters. From 10, 12, 14 and 15,

\[ \epsilon - r = \frac{1}{0.03} + \frac{\epsilon}{\epsilon_\omega} \]  

(16)

**Note S1.5.4 Validation of the model** In our model the distances between the crossover-connected helices are fixed with rigid restraints (see Supplementary Note S1.4), therefore we could not simulate the salt effect on the effective diameter of the 6-helix filaments (14). However, the model allows to observe the interactions between the filaments; to demonstrate that, we designed a system consisting from the three 6-helix loosely connected bundles (Figure S8, A-B).

For evaluation of the ionic strength effect on the origami shape, we monitored the average distance between the centers of mass of the bundles (Table S2).

Increasing ionic strength from 30 mM to 110 mM and then to 260 mM decreased the distance from 10.34±2.53 nm to 7.84±1.56 nm and to 6.88±0.56 nm. The \( d_{260}/d_{110} \) ratio was 0.88, which is consistent with the filament effective diameter ratio of 0.88 (13).

Thus, our model allows to consider the influence of ionic strength. However, because of the model’s limitations and lack of the data for the direct validation, further research is needed in this direction.

**Note S1.6 Workflow**

Our tool takes a file with a caDNAno scheme of an origami object as an input. It generates starting Cartesian coordinates of the structure along with a 2D map that contains the correspondence between the initial plot and our CG model.

A set of lattice- and design-specific distance restrictions

| Ionic strength, mM | \( \epsilon - r \) | distance, nm |
|-------------------|----------------|-------------|
| 30                | 1.00           | 10.34±2.53  |
| 110               | 3.61           | 7.84±1.56   |
| 260               | 8.37           | 6.88±0.56   |
is generated. Then the two steps of geometry optimization are performed using (i) quasi-Newtonian limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (18) and (ii) steepest descent algorithm. The following step is an MD simulation for $9 \times 10^5$ integration steps with a time step of 100 fs, which proved to be an optimal compromise between the processing time and resultant information on the system (Figure S8).

First $10^5$ steps of the trajectory are intended for geometry relaxation, while the remaining $8 \times 10^5$ steps are used for conformational analysis of the structure. These time parameters could be adjusted using a standalone version of the tool. To perform a conformational analysis, the trajectory is cut into individual frames with an interval of 320 integration steps, which then are clustered with affinity propagation algorithm (19) to determine the major conformations (Supplementary Note S1.7).

**Note S1.7 Clustering of output conformations**

Analysis of DNA origami conformational landscape was performed for snapshots of the MD trajectory generated with an interval of 100 ps, starting from $10^5$th iteration step of the trajectory (to skip the geometry relaxation stage) and up to $9 \times 10^5$ integration steps. These parameters have been found to be optimal in terms of processing time and provided sampling, regardless of the system’s size. At this time mark the number of clusters achieves a plateau, while the wall time continues to grow linearly (Figure S9).

For the clustering procedure we used a homemade package **affbio** based on Affinity Propagation (AP) clustering algorithm (19) with the preference parameter value of 120. Basic version from the SciKit package (20) was modified to be (i) applicable to structures of biomolecules and (ii) portable and scalable in different environments, from web-services to supercomputers.

We used root mean square deviation (RMSD) as a similarity measure between the conformations. To efficiently calculate pairwise distances we used pyRMSD package (21) with added MPI parallelization (22, 23) and adaptable local caching scheme which allows better multinode scaling in supercomputing environments. The computed values in a form of an upper triangle similarity HDF5 matrix (24) served as a run input for the AP clustering.

To calculate the median of the similarity matrix, which is a very memory-demanding step, we have written the optimized version of the Livestats package based on the P-square algorithm for dynamic calculation of quantiles and histograms (25). It allows to calculate quantiles in real-time with an error of less than 1% for matrices consisting of $10^4$ elements and more. Our version (with Cython insertions) outperforms the source version by more than two orders in terms of wall time. The clustering step was also rewritten to optimize performance, parallelization and reduce memory footprint.

For an analysis of clustering results we have developed several CLI friendly flexible modules for text and graphical outputs. 3D rendering for multiplots of representative structures with specified cluster sizes is based on PyMOL viewer code.

**Affbio** package is available in Python PIP repository, source code is available under GPLv3 license (26) at [http://vsb.fbb.msu.ru/rhodecode/software/aff_cluster/](http://vsb.fbb.msu.ru/rhodecode/software/aff_cluster/).
Figure S9. Relation between trajectory length, processing (wall) time and conformational sampling (expressed as number of clusters). A: for the Sq-system (size of the system: 72 particles); B: for the hexagonal nanocontainer structure (size of the system: 2068 particles).

2 VALIDATION OF THE COSM MODEL

Note S2.1 Benchmark systems: design and sequences

Scaffold DNA sequence for the benchmark systems:

5'-CCACCATGTCACCAGATAAAACAAATGGCAGTG
CTTCTAGAAGAGACGGAATCGGCAGGCTGCAGC
TGCCAACCCAGAGAATTCCAGAGGAAAAGGTCGGA
GAGGCGAGGGGCAAACACGGGGTTGTGTCTTTA
ACTGATGAAACCAAGGAGGAACTGATTTTTAGGTA
CTGCAGCGGCTCTTGCGATGCAGCTGAGACAACGTA
CGACAAAATATGGAAAAACTTATCCAGAAATAGAA
GGCTGGTGAGTGACAAAGTAGGGCAGGCATGTTGC
AGACCCATCGCCTTTGATGATGACCTGTCGTTTTTA
GATGATAACCTGGTTTACCATATTCTAAGAAAGCAT
TCCGCTAAAAGGTGTGGATGTATCTGGGATCC-3'

Staple sequences for the benchmark systems (Figure S10) are listed in Table S3.
Figure S10. Schemes of the (A) Hc-system and (B) Sq-system, produced in cadnano (http://cadnano.org/). Staple oligonucleotides are labeled.

Table S3. Staple oligonucleotides

| ID | Sequence |          |
|----|----------|----------|
|    | Sq system |          |
| st1 | TCTATTTCTGGATAAGTTTTTCCGATGGGTC |          |
| st2 | TCTATTTCTGGATAAGTCTTGGATTTTCCGATGGGTC |          |
| st3 | AAATCAGCTTGCAGCCCTGCGGATT |          |
| st4 | GATCCCCAGATACATACACACCTTGTTTATCTGGTACATGGT |          |
| st5 | GAATGCTTTCTCTAGAAATAT |          |
| st6 | AAACGACAGGTCATCATCAAAGGAATATTTTG |          |
| st7 | TCGTACTCGTTCAGCTGACATCATCTCTAA |          |
| st8 | CCGCTTTTCTGGCCCTCTCGG |          |
| st9 | CGCTCTCTTCTAGGAAGCACTGCCATTTTACGG |          |
| st10 | GTAAACCGAGTTGCAAGCGCCCTGACACCTTA |          |
| st11 | TCAACCATGCGCCCTGCACTTTCTTTGTCACCGCC |          |
| st12 | CCGCTCTCGTTCATCTTAAAATGTAATTCAGTTAAGACA |          |
| st13 | GGAATCTCTGCGGTTGCAAGCTCCCTCCTTGTT |          |
|    | Hc system |          |
| oli1 | TCGACAGCGTCCAGTTGCTCTTGGTTTCATAGACACCTA |          |
| oli2 | AAACGACAGTCCAGCTCTTTCTTTTCTCTTAATATGCTT |          |
| oli3 | ATCCAGCTACATTCCACGGAATGCTTTTCTGGTCAT |          |
| oli4 | TTCTGGATAGTGCTTTACAGCTATTTTCTTGGTGGCATATCTACAT |          |
| oli5 | CCGGGGTITGTGAAGCTGCGCCG |          |
| oli6 | GTCACTCAGCAAGAATATGCTT |          |
| oli7 | AACGACATTCTGACATGTGCTT |          |
| oli8 | TTCCGCTTCAAGACCAAC |          |
| oli9 | TCACTGTTCATTTTTCACG |          |
| oli10 | AAGCACCATTACCTTGGGCTACATGCTT |          |
| oli11 | ATGTAAGTCCAGTCTTTCAGG |          |
| oli12 | CCTCTCCGACTGACATAGCTT |          |
| oli13 | AAATCAGTCCAGAAGGCGCCTGACGTGTC |          |
Figure S11. All-atom models of the benchmark systems. A: initial (left) and final (right) conformations of the Hc-system, front and side views; B: initial (left) and final (right) conformations of the Sq-system, side and top views.

Note S2.2 All-atom MD simulation

For an additional validation of our DNA origami model we performed all-atom MD simulations of the benchmark systems. We assembled the systems using 3DNA v.2.3 (27) for creating of the constituent DNA strands and PyMOL v. 1.8.6.0 (28) for arrangement of the strands in space according to the corresponding CaDNAno scheme (Figure S10). The coordinates of the short arm of the Sq-system were altered in such a way that the system acquired the T-shape. Even though we manually optimized the relative positions of benchmark systems’ helices to match the lattice geometry and locations of crossover sites, initial coordinates of the systems were far from the equilibrium state. Therefore, large time scale simulations are needed to get an appropriate sampling for the analysis of structure and dynamics of the all-atom models.

We used the SIRAH coarse-grained solvent model (29) to get 100 ns-long MD trajectories for the systems in a reasonable time. We tried different coarse-grained solvent models, namely Martini (30) and SIRAH (29), adding corresponding entries to the parmbsc0 (\(\chi\) OL4) force field (31). The SIRAH model demonstrated the better performance (data is not shown, but available on request).

It has been previously shown that the SIRAH solvent model allows to reliably describe a conformational landscape of a protein-DNA complex (32) in a coarse-grained representation. Our results suggest that the model is compatible with the
Figure S12. Radius of gyration of all-atom and COSM models of the Hc-system through the first 5 ns of corresponding MD trajectories. Black: all-atom model (parmbsc0 (χ OL4) force field), red: COSM model.

1. parting of helices with formation of the chickenwire pattern;
2. slight bending of the "legs" accompanied by twisting of the "legs" along the longitudinal axes;
3. spreading of the "legs" in both in- and out-of-the-body-plane directions. The out-of-body-plane movement led to a slight lateral twist of the "body".

Twisting and bending motions, along with formation of the chickenwire pattern, were the reasons of the most noticeable alterations of the Sq-system's structure (Figure S11B). Unfortunately, the alterations were too insignificant to analyze the details.

Note S2.2.1 Sampling efficiency of the COSM model Highly simplified model of DNA origami allowed us to significantly speed up the scanning of a conformational landscape for the benchmark systems at the expense of atomistic and even single-nucleotide details. For example, first 5 ns of the COSM trajectory for the Hc-system contained about 20 large conformational changes, while the all-atom model did not move far from the initial configuration (Figure S12). While we do not assert that our model allows to locate a global minimum of the system, it nevertheless makes it possible for the system to evolve through more than one local minimum. As an illustration, let us consider the results of the clustering analysis of MD trajectory snapshots (see Figure S9). For both systems, the number of clusters reached a plateau at the middle of the trajectory. This means that a stable number of cluster centers (local minima) appeared, between which transitions occurred during the rest of the trajectory. Otherwise, we would...
Figure S13. Coarse-grained oxDNA model of the Hc-system: initial conformation after geometry optimization step, front view. The scaffold chain is colored rainbow, staple strands are colored gray; B: three major conformations of the Hc-system, side (top) and front (bottom) views

observe a gradual increase in the number of clusters with time, reflecting the increase of different conformations in the system. Thus, our model avoids convergence at the nearest local minimum, which makes it a suitable tool for geometry optimization and scanning of a conformational landscape of DNA origami designs.

Note S2.3 oxDNA MD simulation

Since the all-atom MD simulations of the benchmark systems did not provide a sufficient conformational sampling for the AFM data interpretation, especially for the Hc-system, we also carried out a coarse-grained modeling of the system. For that, we used the oxDNA model which provides a reliable physical representation of the thermodynamic and mechanical properties of single- and double-stranded DNA (34), including DNA origami (35).

We generated the initial coordinates of the Hc-system with the cadnano-interface.py script from the oxDNA v.2.2.2 package. After that, we performed a $10^4$ steps-long MD simulation of the Hc-system with the following parameters:

- `newtonian-steps = 2`
- `diff-coeff = 2.5`
- `thermostat = john`
- `dt = 0.005`

The final conformation from this step was used as starting coordinates for a $5 \times 10^7$ steps-long trajectory with the simulation parameters:

- `newtonian-steps = 103`
- `diff-coeff = 2.5`
- `thermostat = brownian`
- `dt = 0.005`
- `T = 220K`

The last $3 \times 10^7$ steps of the trajectory were used for the conformational analysis (as described in Supplementary Note S1.7), which revealed three main shapes of the system (Figure S13). The main mobility of the Hc-system was...
associated with out-of-body-plane movements of the legs, which was consistent with the prediction made by our model (see Figure 4B). Just like in our model, the legs’ swinging movements were accompanied by a twisting of the body (Figure S13B). According to the γ angle values, oxDNA slightly overestimated the flexibility of the system (γ_{oxDNA} = 84.8 ± 20.5 degrees, γ_{COSM} = 64.4 ± 15.4 degrees, γ_{AFM} = 64.7 ± 25.6 degrees, p < 0.001). We associate it with the temperature of the simulation set too high. In the preliminary simulations we set the parameter value equal to 300, 280, 260 and 240K and observed the melting of the system in all the cases listed. 220K was the first temperature value that produced a stable trajectory, but it is possible that the kinetic energy of the system was still artificially high at that temperature. Note that Snodin and coauthors reported a simulation of DNA-origami self-assembly at 338K (35); it probably implies that the correspondence between physical and simulated temperature in the oxDNA model is not strict and depends on the target system, which is a common issue with coarse-grained models (see, for example, the study of Jiang and Hansmann (36)).

In summary, modes of the Hc-system’s motion predicted by our model were consistent with the oxDNA MD simulation results. Both models described highly dynamic behavior of the system expressed in swinging motions of the legs and body twisting.

**Note S2.4  FRET experimental details**

From the spectral data a FRET efficiency $E$ for the systems was calculated using the method of Gordon et al. (37). A distance $r$ between the fluorophores was derived from equation

$$E = \frac{R_0^6}{R_0^6 + r^6}$$  \hspace{1cm} (17)

where $R_0$ is a Forster distance for a given pair of fluorophores,

$$R_0 = 0.2108 \sqrt[6]{\frac{k^2}{k^2 \Phi_0 n - 4^2 J}}$$  \hspace{1cm} (18)

Here, $k$ is the orientation factor, $k^2 = 2/3$. $\Phi_0$ is a quantum yield of a donor, which is 0.9 for FAM. A refractive index of a medium $n$ is 1.353 for TE buffer at 20 °C. $J$ is a parameter characterizing an overlap between the donor emission spectrum and the acceptor excitation spectrum, it was calculated with ale 2.2 software (UV-Vis-IR Spectral Software 2.2, FluorTools, www.fluortools.com) and was equal to 2.517 * 10^{15}. Therefore, $R_0$ for our systems was equal to 58.4 Å.

**REFERENCES**

1. Douglas, S. M., Bachelet, I., and Church, G. M. (feb, 2012) A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science, 335*(6070), 831–834.
2. Douglas, S. M., Dietz, H., Liedl, T., Hogberg, B., Graf, F., and Shih, W. M. (may, 2009) Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature, 459*(7245), 414–418.
3. Hess, B., K. C. v. d. S. D. and Lindahl, E. (July, 2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput., 4*, 435–447.
4. Leslie, A. G., Arnott, S., Chandrasekaran, R., and Ratliff, R. L. (Oct, 1980) Polymorphism of DNA double helices. *J. Mol. Biol., 143*(1), 49–72.
5. Smith, S. B., Cui, Y., and Bustamante, C. (1996) Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules. *Science, 271*(February), 795–799.
6. Murphy, M. C., Rasnik, I., Cheng, W., Lohman, T. M., and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, **86**(4), 2530–2537.

7. Kratky, O. and Porod, G. (1949) Röntgenuntersuchung gelöster Fadenmoleküle. *Recl. Trav. Chim. Pays-Bas*, **68**(12), 1106–1122.

8. Bustamante, C., Bryant, Z., and Smith, S. B. (Jan, 2003) Ten years of tension: single-molecule DNA mechanics. *Nature*, **421**(6921), 423–427.

9. Chen, H., Meisburger, S. P., Pabit, S. a., Sutton, J. L., Webb, W. W., and Pollack, L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **109**(3), 799–804.

10. Klaue, D. and Seidel, R. (Jan, 2009) Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys. Rev. Lett.*, **102**(2), 028302.

11. Lipfert, J., Kerssemakers, J. W., Jager, T., and Dekker, N. H. (Dec, 2010) Magnetic torque tweezers: measuring torsional stiffness in DNA and RecA-DNA filaments. *Nat. Methods*, **7**(12), 977–980.

12. Yoo, J. and Aksimentiev, A. (2013) In situ structure and dynamics of DNA origami determined through molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.*, **110**(50), 20099–104.

13. Onsager, L. (1949) THE EFFECTS OF SHAPE ON THE INTERACTION OF COLLOIDAL PARTICLES. *Ann. N. Y. Acad. Sci.*, **51**(4), 627–659.

14. Siavashpouri, M., Wachauf, C. H., Zakhary, M. J., Praetorius, F., Dietz, H., and Dogic, Z. (Aug, 2017) Molecular engineering of chiral colloidal liquid crystals using DNA origami. *Nat. Mater.*, **16**(8), 849–856.

15. Ray, S. G., Cohen, H., Naaman, R., and Rabin, Y. (dec, 2005) Where is the Sodium in Self-Assembled Monolayers of Single-Stranded DNA?. *J. Am. Chem. Soc.*, **127**(49), 17138–17139.

16. Manning, G. S. (1981) Limiting laws and counterion condensation in polyelectrolyte solutions. 7. Electrophoretic mobility and conductance. *J. Phys. Chem.*, **85**(11), 1506–1515.

17. Hasted, J. B. and Ritson, D. M. and Collie, C. H. (1948) Dielectric Properties of Aqueous Ionic Solutions. Parts I and II. *J. Chem. Phys.*, **16**(1), 1–21.

18. Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (sep, 1995) A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM J. Sci. Comput.*, **16**(5), 1190–1208.
2015) Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *J. Chem. Theory Comput.*, 11(1), 260–275.

31. Krepl, M., Zgarbová, M., Stadlbauer, P., Otyepka, M., Banáš, P., Koča, J., Cheatham, T. E., Jurečka, P., and Šponer, J. (2012) Reference Simulations of Noncanonical Nucleic Acids with Different \( \chi \) Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.*, 8(7), 2506–2520 PMID: 23197943.

32. Machado, M. R. and Pantano, S. (2015) Exploring LacI-DNA Dynamics by Multiscale Simulations Using the SIRAH Force Field. *J. Chem. Theory Comput.*, 11(10), 5012–5023 PMID: 26574286.

33. Mark, P. and Nilsson, L. (2001) Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*, 105(43), 9954–9960.

34. Šulc, P., Romano, F., Ouldridge, T. E., Rovigatti, L., Doye, J. P. K., and Louis, A. A. (2012) Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.*, 137(13), 135101.

35. Snodin, B. E. K., Romano, F., Rovigatti, L., Ouldridge, T. E., Louis, A. A., and Doye, J. P. K. (feb, 2016) Direct Simulation of the Self-Assembly of a Small DNA Origami. *ACS Nano*, 10(2), 1724–1737.

36. Jiang, P. and Hansmann, U. H. (Jun, 2012) Modeling Structural Flexibility of Proteins with Go-Models. *J Chem Theory Comput.*, 8(6), 2127–2133.

37. Gordon, G. W., Berry, G., Liang, X. H., Levine, B., and Herman, B. (may, 1998) Quantitative fluorescence resonance energy transfer measurements using fluorescence microscopy. *Biophys. J.*, 74(5), 2702–2713.