






Cite this: *Soft Matter*, 2022, 18, 6716

## Polymorphic self-assembly of helical tubules is kinetically controlled†

Huang Fang, <sup>a</sup> Botond Tyukodi,<sup>ab</sup> W. Benjamin Rogers <sup>a</sup> and Michael F. Hagan <sup>\*a</sup>

In contrast to most self-assembling synthetic materials, which undergo unbounded growth, many biological self-assembly processes are self-limited. That is, the assembled structures have one or more finite dimensions that are much larger than the size scale of the individual monomers. In many such cases, the finite dimension is selected by a preferred curvature of the monomers, which leads to self-closure of the assembly. In this article, we study an example class of self-closing assemblies: cylindrical tubules that assemble from triangular monomers. By combining kinetic Monte Carlo simulations, free energy calculations, and simple theoretical models, we show that a range of programmable size scales can be targeted by controlling the intricate balance between the preferred curvature of the monomers and their interaction strengths. However, their assembly is kinetically controlled—the tubule morphology is essentially fixed shortly after closure, resulting in a distribution of tubule widths that is significantly broader than the equilibrium distribution. We develop a simple kinetic model based on this observation and the underlying free-energy landscape of assembling tubules that quantitatively describes the distributions. Our results are consistent with recent experimental observations of tubule assembly from triangular DNA origami monomers. The modeling framework elucidates design principles for assembling self-limited structures from synthetic components, such as artificial microtubules that have a desired width and chirality.

Received 23rd May 2022,  
Accepted 12th August 2022

DOI: 10.1039/d2sm00679k

[rsc.li/soft-matter-journal](http://rsc.li/soft-matter-journal)

### 1. Introduction

Many biological functions rely upon the assembly of self-limited structures that have well-defined finite sizes, and yet are much larger than the size of the individual building blocks. Examples include the assembly of protein capsomers into viral shells with the appropriate size to encapsulate the viral nucleic acid, assembly of tubulin into microtubules with diameters that confer sufficient rigidity to mechanically support the cell,<sup>1,2</sup> and, within butterfly wings, the organization of chitin into nanostructured domains on the scale of visible light to make the tissue iridescent.<sup>3,4</sup> In contrast, most structures assembled from synthetic building blocks undergo *unlimited* growth into crystals or amorphous materials.<sup>5–7</sup> The biological structures described above are examples of ‘curvature-controlled’ assemblies, in which the building blocks assemble with a preferred curvature that leads the structure to close upon itself in one or more directions.

There has been an intense interest in mimicking such functional biological structures by developing synthetic building blocks that can be pre-programmed to assemble with curvatures leading to self-closure. To this end, researchers have recently used DNA origami (*e.g.*<sup>8,9</sup>) and protein design (*e.g.*<sup>10–12</sup>) to engineer building blocks that assemble into polyhedral capsids or tubules with designed diameters. However, due to thermal fluctuations and kinetic effects, assembled structures typically exhibit polymorphism in the limited dimension rather than a single well-defined diameter.<sup>13–15</sup> Understanding the factors that control this size distribution is essential for achieving functional self-limited assemblies. In this article, we use computer simulations and kinetic models to understand the dynamical pathways of helical tubule assembly, and the resulting polymorphic distribution of assembled tubule structures.

Curvature-controlled assemblies in biology frequently rely on symmetry principles to maximize their ‘economy’ of assembly, meaning the size of the structure that can be assembled for a given number of distinct subunit species.<sup>15</sup> For example, icosahedral symmetry maximizes the number of identical subunits (60) that can be used to assemble a shell, and many viruses assemble icosahedral capsids.<sup>16–19</sup> In this sense, helical tubules are even simpler than icosahedral capsids—there is an infinite family of helical tubules with different diameters and

<sup>a</sup> Martin Fisher School of Physics, Brandeis University, Waltham, Massachusetts 02454, USA. E-mail: [hagan@brandeis.edu](mailto:hagan@brandeis.edu)

<sup>b</sup> Department of Physics, Babes-Bolyai University, 400084 Cluj-Napoca, Romania

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2sm00679k>



itches, each of which can be assembled from a single subunit species with identical conformations throughout the structure. However, because the subunit curvature changes only slightly between different tubule structures with similar geometries within this family, tubule assembly is highly susceptible to polymorphism. That is, when subunits associate with imperfect geometries during assembly, and these imperfections fail to anneal before becoming trapped by further subunit association, the resulting assembled structures deviate from the ground state tubule structure. Consequently, the geometry distribution of tubules assembled in a finite time depends on a competition between kinetic and thermodynamic factors, and can differ significantly from the equilibrium distribution. Identifying these factors from experiments alone is challenging because most intermediate structures are transient and present at concentrations which are too low to experimentally detect or characterize.

Computer simulations can help to understand self-limited size distributions by revealing the dynamical pathways leading to assembly. However, in comparison to the extensive body of theoretical and computational modeling of icosahedral capsids or shells (e.g.<sup>20</sup>), there has been relatively limited study of tubule assembly (e.g.<sup>21–29</sup>). Thus, the mechanisms controlling tubule assembly and closure have yet to be completely explored.

In this article, we perform kinetic Monte Carlo simulations on a model of triangular subunits motivated by recent experiments demonstrating the assembly of DNA origami building blocks into helical tubules.<sup>9</sup> By comparing the distribution of dynamically assembled tubules with equilibrium results, we find that the size and morphology distribution is kinetically controlled. In particular, the structural ensemble is typically quenched shortly after a nascent assemblage first closes upon itself to form a cylindrical tubule. Through a combination of dynamical simulations, free-energy calculations, and simple analytical models, we determine how the resulting size distribution depends on control parameters such as the bending modulus and the pre-programmed target curvature. These results may guide the experimental design of more efficient and accurate self-assembling artificial tubule structures.

The remainder of the article is organized as follows: In Section II A and B, we introduce the kinetic Monte Carlo algorithm that we use to model tubule self-assembly. We then discuss the predicted assembly trajectories and geometry distribution of assembled tubules. In Section II C, we compare simulation outcomes to observations from experiments on tubules self-assembled from DNA origami subunits, and obtain an estimate of the bending rigidity in the experimental system. In Section III A–C, we present calculations of the equilibrium tubule geometry distribution and, through comparison with simulation results, show that the assembled geometry distribution is kinetically controlled. In Section III D–E, we construct a kinetic model that captures these dynamical effects, and use it to predict the assembly behavior as a function of the control parameters. Finally, in Section IV, we discuss implications for future experiments, as well as limitations and possible extensions of the model.

## II. Simulations

### A. Computational model

In our model, monomers are triangular structures composed of three vertices connected by harmonic bonds. We choose a triangle geometry because it closely matches the DNA origami subunits in the motivating experiments,<sup>8,9</sup> and similar models based on triangular monomers have been recently developed for elastic membranes,<sup>30</sup> icosahedral shell self-assembly,<sup>31–35</sup> and geometrically frustrated finite-length tubule-like structures.<sup>36</sup> The model and simulation algorithm can be readily extended to other monomer geometries; for example, Mohajerani *et al.*<sup>37</sup> modeled the assembly of hepatitis B virus (HBV) capsids from protein dimer subunits.

The Hamiltonian is

$$H = \sum_{i \in \text{BoundEdgePairs}} -E_B + \frac{1}{2}B(\theta_i - \theta_{0,i})^2 + \sum_{j \in \text{Edges}} \frac{1}{2}k_S(l_j - l_{0,j})^2, \quad (1)$$

with  $E_B$  as the monomer-monomer binding energy (set as a positive constant);  $\theta_i$  and  $\theta_{0,i}$  as the instantaneous and preferred dihedral angle between two monomers bound at a common edge  $i$ ,  $B$  as the bending modulus;  $l_j$  and  $l_{0,j}$  as respectively the instantaneous and stress-free lengths of an edge  $j$ ; and  $k_S$  as the stretching modulus. The three monomer edges are inequivalent, and setting  $\theta_0$  at each of the monomer edges defines the ground state (target) tubule geometry (Fig. S1, ESI†). For simplicity, we set  $E_B$  and  $l_0$  to be identical among all three edges of a monomer, and we consider only a single monomer species. Moreover, motivated by the material properties of DNA origami subunits and proteins, we focus here on the limit of thin sheets, in which the bending deformations are much lower in energy than stretching. Therefore, we set  $k_S = 200k_B T/l_0^2$  throughout this study so that the monomer edges are nearly fixed in length, and we vary the bending modulus as a control parameter (Fig. 1).

We use Monte Carlo moves to relax the structure, including vertex moves to relax structural degrees of freedom, monomer association and dissociation moves to model assembly and

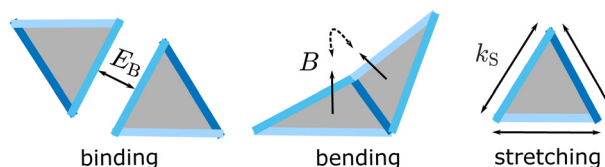


Fig. 1 Schematic of the model. The Hamiltonian includes terms that represent edge stretching, monomer-monomer binding, and bending. Each of the three monomer edges is a different type, and only pairs of edges with the same type can bind. In this work, all edge types have the same binding energy. Each edge type  $i$  has a different preferred ('ideal') dihedral angle,  $\theta_{d,i}^{(m,n)}$ , the set of which determine the target structure ( $m, n$ ). The energetic cost of deviations from preferred edge lengths and the dihedral angles are controlled by the stretching and bending moduli,  $k_S$  and  $B$ .



disassembly, as well as moves to model internal rearrangement events such as the splitting and merging of cracks within a structure. All the moves guarantee detailed balance (see ESI† Section IX and X for details about the algorithm and the moves). Provided that this set of movements represents the transitions that are relevant for actual tubule assembly, with approximately correct relative rates for the different moves, the Monte Carlo trajectories can be qualitatively mapped onto the system dynamics. This mapping can be tested by comparing simulation results against experimental observations of tubule assembly kinetics and the structural ensemble of assembled tubules. The edge fusion and fission moves, which respectively bind two free edges on the structure boundary or split two edges that are already bound, are particularly important for closure/reopening of the tubule structure. We show below that the rate of tubule closure relative to its growth can significantly affect the assembly pathways. Therefore, we define a control parameter – the *edge fusion rate*  $f_{\text{fusion}}$ , as the ratio between the attempt frequency of edge fusion/fission moves and the unit timescale (which is set to the frequency of vertex moves).

To determine well-defined steady-state distributions, we evolve the system in the grand canonical ensemble, in which the assembling structure exchanges monomers with a bath at a fixed chemical potential  $\mu$ . This situation approximately describes tubule distributions at a point in a reaction with a corresponding free monomer concentration  $c_0 = c_{\text{SS}} \exp(\mu/k_B T)$  with  $c_{\text{SS}}$  the standard state concentration. We set  $\mu = -3k_B T$  throughout this study. For the purposes of comparing our results with the DNA origami tubule assembly experiments,<sup>9</sup> we use the same standard state concentration as specified in that work,  $c_{\text{SS}} \approx 10 \mu\text{M}$  (corresponding to approximately 100% monomer volume fraction). This results in a bath concentration of  $c_0 \approx 500 \text{ nM}$ . See Section II C for a discussion of how our simulated system compares to the experiments and our rationale for parameter choices.

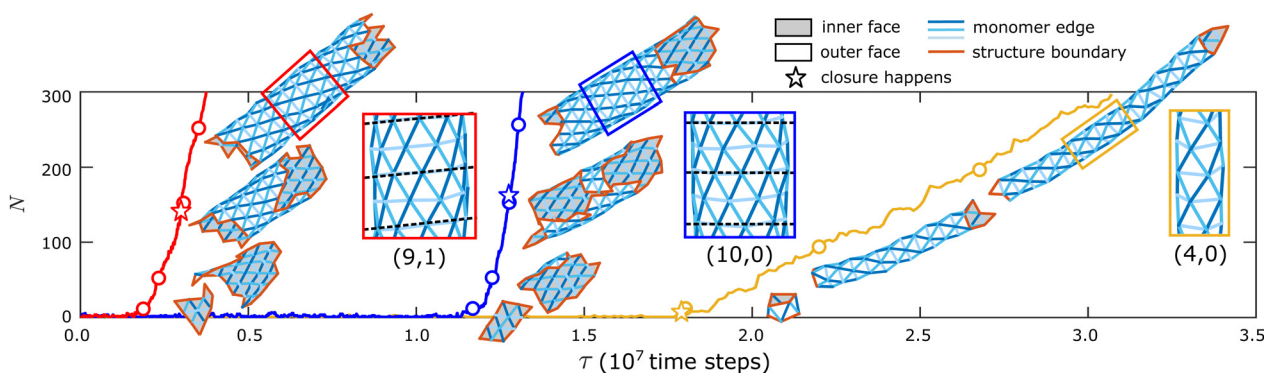
To model assembly from a dilute system of monomers, for which binding between different tubules is negligible,

simulations are restricted to have only one structure within the simulation box. The initial condition for each dynamical assembly simulation is one monomer in the simulation box, which then assembles (and disassembles) through association (and dissociation) of monomers through exchanges with the bath. Since monomers can only associate to an existing structure, and only single monomer association/dissociation is allowed, the system is guaranteed to maintain only one structure. For thermodynamic integration, the system is initialized from, and restricted to, a closed tubule lattice (see Section III A and VII B, ESI†).

## B. Simulation results

**Assembly trajectories.** As expected from generic models of tubule or filament assembly,<sup>15</sup> assembly in our model requires that the subunit concentration and binding affinity exceed threshold values. Below these values, the mixing entropy of free monomers out-competes the interactions stabilizing tubule formation, and equilibrium is dominated by monomers. For the subunit concentration that we focus on (determined by  $\mu = -3k_B T$ ), we find that the critical binding affinity corresponds to  $E_B \approx 4.5k_B T$ . This observation is supported by our thermodynamic integration results (described in Section III A and VII B, ESI†). As shown in ESI† Section VII B, with  $E_B = 4.5k_B T$ , the bulk energy density of an assembled tubule is about  $-3k_B T$ , and thus equal to the bath chemical potential.

Above the threshold binding affinity, the Monte Carlo trajectories exhibit a rich dynamics which proceeds through a series of stages, including nucleation, closure, and growth. Fig. 2 shows snapshots from example simulation trajectories at three different parameter sets. During assembly with a large target tubule diameter (red and blue curves), after an initial period of transient assembly and disassembly, the structure surpasses the critical nucleus size (approximately 5 monomers for these conditions) and grows steadily as a curved two-dimensional sheet. Eventually, the boundary edges at opposite sides of the curling sheet begin to touch, and the edges bind.



**Fig. 2** Simulation trajectories showing the number of monomers  $N$  in a self-assembled structure as a function of elapsed time  $\tau$  (number of Monte Carlo sweeps). Snapshots show the configurations at indicated times. Dashed lines in the zoomed-in snapshots label the direction of the edge that is the most perpendicular to the longitudinal direction. The assembled tubule geometries are: (9,1) for the red trajectory, (10,0) for the blue trajectory, and (4,0) for the yellow trajectory. The red and blue trajectories have the same target tubule geometry of (10,0) and the same binding energy of  $E_B = 6.0k_B T$ . The target tubule geometry for the yellow trajectory is (5,0) and the binding energy is  $E_B = 5.0k_B T$ . Other simulation parameters are the same for all the trajectories:  $B = 20k_B T$ ,  $f_{\text{fusion}} = 10^{-3}$ .



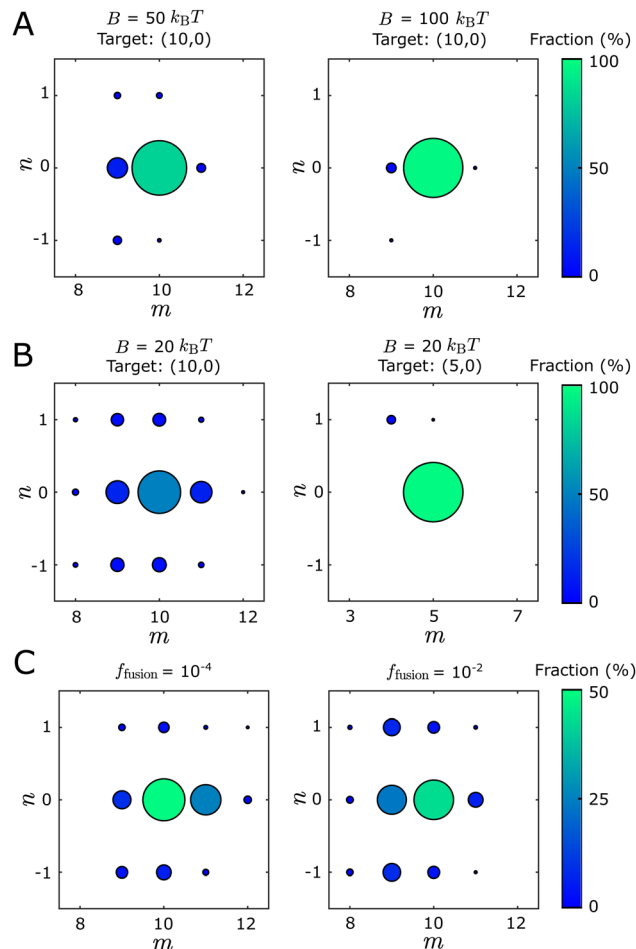
We denote the first such binding event as the point of tubule closure. After closure, the tubule geometry is highly stable and the structure undergoes steady growth from both ends.

Even though the red and blue trajectories have the same target structure, they assemble different tubule geometries, denoted by different pairs of integer numbers  $(m, n)$  based on the convention from carbon nanotubes.<sup>38</sup> Representing a tubule as a curled triangular lattice that closes upon itself, the index  $m$  gives the number of lattice sites on one turn around the helix, while  $n$  gives the number of lattice sites in the orthogonal direction (along the long axis of the tubule) (see Fig. S1 for a schematic of the naming convention and Section I A, ESI† for a detailed description of the notation). In the trajectory with a small target width (yellow line), tubule closure corresponds to the formation of the critical nucleus, after which the tubule undergoes steady growth.

**Tubule geometry distributions.** We performed simulations over a wide range of parameter values to learn how the tubule morphologies arising from dynamical trajectories depend on the relevant physical parameters, such as the bending modulus  $B$ , the diameter of the target tubule geometry  $D_0$ , and the fusion rate  $f_{\text{fusion}}$ , which influences the closure kinetics. We measured the distribution of tubule geometries at the end of each simulation. Simulations were performed until the length of the structure  $L$  grew to approximately three times the tubule circumference, since the geometry distribution is stable by this point (no geometry fluctuations occur beyond this size). We estimated the distributions from 1000 independent trials at each parameter set.

We find that tubule structures with different geometries, as well as structures that fail to close, can assemble in the dynamical simulations under the same set of parameter values. We classify the self-assembly outcomes into three categories: *defect-free* tubules, *defective* tubules, and *open structures*. A tubule is *defective* if part of the structure fails to close or multiple tubule geometries are locally identified within the same structure (see Fig. S3 and Section III A, ESI† for identification details). *Open structures* arise when nonuniform curvature causes opposite boundary edges to ‘miss’ the opportunity to bind to each other, leading to a spiral structure that resembles a toilet paper roll (Fig. S3, ESI†). The fraction of defective tubules increases as the bending modulus  $B$  decreases or the target diameter  $D_0$  increases (Fig. 3). The fraction of defective and open structures increases with the binding energy (Fig. S10, ESI†). To avoid conditions under which defective structures are too prevalent, for all results in the main text we set the binding energy to  $E_B = 6k_B T$  and the chemical potential to  $\mu = -3k_B T$ , so that assembly is sufficiently reversible to allow monomer detachment and annealing.<sup>39–42</sup> With these parameters, the fraction of defective tubes is generally below 30%.

The geometry distributions of defect-free tubules depend on the control parameters. Fig. 3 shows the distributions of assembled tubules for different bending moduli, target diameters, and fusion rates. The size and color of the circular symbols represent the fraction of different tubule geometries within the defect-free population. Only tubule geometries with populations  $\geq 1\%$  are labeled in the plot. Fig. 3(A) shows tubule geometry



**Fig. 3** Tubule geometry distributions from Monte Carlo assembly trajectories depend on the bending modulus  $B$ , target tubule geometry, and fusion rate  $f_{\text{fusion}}$ . The color and size of each circle indicate the fraction of the corresponding tubule geometry within the defect-free population. (A) Geometry distributions for  $B = 50k_B T$  (left) and  $B = 100k_B T$  (right) with  $E_B = 6k_B T$ ,  $f_{\text{fusion}} = 10^{-3}$ , and target tubule geometry  $(10,0)$ . The fraction of defect-free tubules is about 97% for both cases. (B) Geometry distributions for different target geometries with  $B = 20k_B T$ ,  $E_B = 6k_B T$ , and  $f_{\text{fusion}} = 10^{-3}$ . The fraction of defect-free tubules is about 88% for both cases. (C) Geometry distributions for indicated values of  $f_{\text{fusion}}$  with  $B = 20k_B T$ ,  $E_B = 6k_B T$ , and target tubule geometry  $(10,0)$ . As  $f_{\text{fusion}}$  decreases from  $10^{-2}$  to  $10^{-4}$ , the fraction of defect-free tubules decreases from 88% to 52%, while the fraction of open structures increases from 0 to 45%. Each distribution in (A–C) is estimated from 1000 independent simulation trajectories.

distributions for two bending moduli  $B$ , with other parameters fixed. As  $B$  increases, the fraction of the target tubule  $(10,0)$  increases while the fraction of the off-target tubules decreases. This is consistent with thermodynamics, since the deviations of dihedral angles required for off-target geometries increase in energy with  $B$ . Fig. 3(B) compares the distributions for two different target geometries. As the diameter of the target geometry  $D_0$  increases, the fraction of the target geometry decreases while the fraction and variety of observed off-target geometries increases. This result is also consistent with thermodynamics, since the difference of the ideal dihedral angles between the



target state and neighboring tubule geometries is smaller for larger  $D_0$  (Fig. S2, ESI<sup>†</sup>). Therefore, with the same extent of dihedral angle fluctuation, the number of accessible off-target states increases as  $D_0$  increases. Similar results were described in ref. 43.

As shown in Fig. S10 (ESI<sup>†</sup>), the tubule geometry distribution does not significantly change for binding energies in the range  $E_B \in [5, 6.5] k_B T$ , suggesting that there is at most a weak dependence on binding energies for the regime we focus on. We discuss this result further in Section III H, ESI<sup>†</sup>.

Interestingly, even for the same Hamiltonian (in which  $B$  and  $D_0$  are fixed), changing the edge fusion rate  $f_{\text{fusion}}$  changes the skew of the geometry distribution. Fig. 3(C) shows that as  $f_{\text{fusion}}$  increases from  $10^{-4}$  to  $10^{-2}$ , the geometry distribution changes from skewing above to skewing below the target geometry (10,0). Meanwhile, the proportion of open structures increases from 0 to around 45% as  $f_{\text{fusion}}$  decreases from  $10^{-2}$  to  $10^{-4}$  (Fig. S7, ESI<sup>†</sup>). This observation reflects the fact that decreasing the closure rate increases the chance that the two edges ‘miss’ each other. As the two edges grow past one another, the size of a curvature fluctuation required to enable closure becomes increasingly unfavorable energetically, and thus more rare. The continued growth of the structure boundary then leads to the spiraling structure described above. See Section III D for further discussion.

### C. Comparison of simulations and experiments

We now compare the results of our dynamical assembly simulations to recent experiments. We find that the morphology distribution of simulated tubules semi-quantitatively agrees with those observed in the experiments,<sup>9</sup> suggesting that the model incorporates the essential physics of the experiments.

Hayakawa *et al.*<sup>9</sup> designed triangular monomers from DNA origami that self-assemble into helical tubules (Fig. 4). The monomers interact with each other along their edges through shape-complementary interactions driven by blunt-end DNA base stacking.<sup>8</sup> The interactions are specific—each monomer edge interacts only with the same edge type on a neighboring subunit. The bevel angles of the edges of each monomer  $\{\theta_{\text{id},i}^0\}$  determine the preferred dihedral angles, which, in turn, set the preferred curvatures of the assembly.

The data set against which we compare our simulation results is obtained from an experimental system that resulted in a most probable tubule geometry of (9,4). Since a key unknown parameter from the experiments is the bending modulus  $B$ , we performed simulations with a target geometry of (9,4) at four values of the bending modulus:  $B \in \{5, 10, 15, 20\} k_B T$ . All other parameters were fixed to their default values (see Section II). We found that assembled structures were highly defective for  $B = 5 k_B T$ . For  $B \geq 10 k_B T$  the majority of tubules were well-formed, with distributions peaked around the target geometry of (9,4). The width of the distribution becomes progressively narrower with increasing  $B$ , as described in Section II B.

We found that a value of  $B = 10 k_B T$  resulted in a geometry distribution of assembled tubules that closely resembles the distribution observed in the experiments (Fig. 4(B)). To facilitate

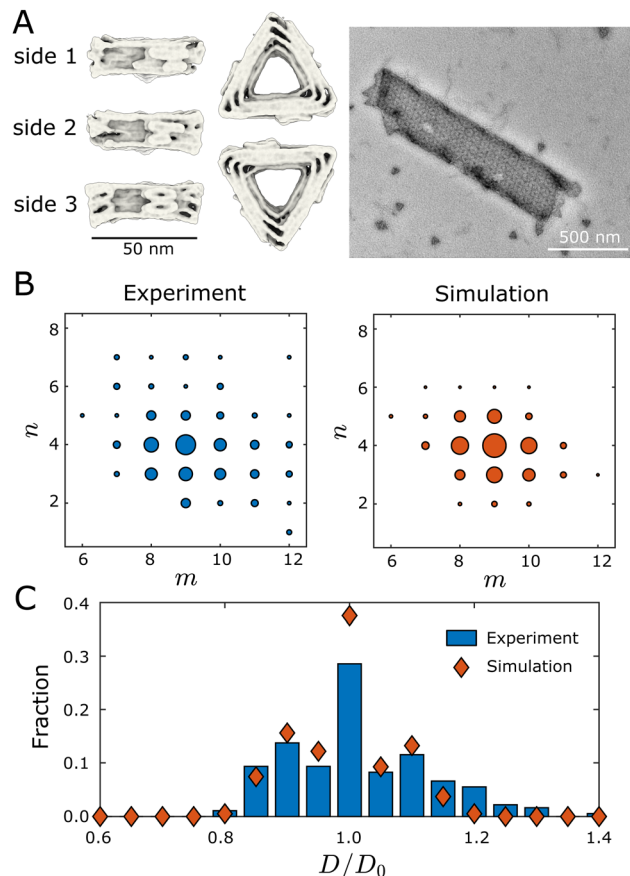


Fig. 4 Comparing the geometry distributions of tubules assembled in simulations and experiments. (A) Cryogenic electron microscopy reconstructions of a DNA origami monomer, and a transmission electron microscopy image of an assembled tubule. The left panel shows the monomer under different views. Two monomers bind along their edges through shape-complementary interactions driven by blunt-end DNA base stacking. The right panel shows an assembled (9,4) tubule. Images in (A) were provided by authors of ref. 9. (B) Tubule geometry distributions measured from experiments<sup>9</sup> and simulations. The size of each circle indicates the fraction of the corresponding tubule geometry within the defect-free population. Each simulation data point is estimated from 1000 independent dynamical MC trajectories that are initialized from a single monomer. (C) Comparing tubule width distributions between experiments (blue bars) and simulations (red symbols).  $D_0$  is the diameter of the ideal (9,4) tubule. Simulation parameters: the target geometry is (9,4),  $E_B = 6.0 k_B T$ ,  $B = 10 k_B T$ , and  $f_{\text{fusion}} = 10^{-3}$ .

comparison between the two distributions, Fig. 4(C) plots the fraction of different tubule geometries against the diameter of the tubules, where  $D_0$  is the diameter of the (9,4) tubule geometry. Although the simulation distribution is slightly narrower than the experiment, we observe that the distributions match fairly closely, especially considering that we have not quantitatively optimized  $B$ . Results for the other simulated values of  $B$  are shown in Fig. S11, ESI<sup>†</sup>.

The comparison between simulations and experimental results in Fig. 4(B) and Fig. S11 (ESI<sup>†</sup>) suggests several important qualitative conclusions: (1) the simple model considered here produces results which are semi-quantitatively consistent with those observed in the experiments. As we will show below,



these results can only be explained through a combination of kinetic and thermodynamic effects, which suggests that the highly simplified dynamics of our model captures the most relevant physics. (2) It was not possible to directly estimate the bending modulus within the experiments. The simulation results suggest a bending modulus on the order of  $10k_{\text{B}}T$ , which is comparable to that of a lipid bilayer membrane ( $B/k_{\text{B}}T \sim 10\text{--}20$ ).<sup>44,45</sup> This computational result could be tested in future experiments that measure the distribution of angular fluctuations between monomers. (3) Given the qualitative agreement between the computational and experimental results, the simulations can provide a predictive guide for future experiments. We note that a definitive comparison of our simulation results to the experiments, and a precise estimate of the experimental bending modulus, will require additional experimental data sets. In the subsequent sections, we use the simulations and simple models to understand the effect of relevant control parameters on the morphology distributions of assembled tubules.

**Comparison between simulated systems and the experimental conditions in ref. 9.** We have chosen parameters that ensure the simulations are qualitatively similar to the motivating experiments while maintaining computational tractability. In particular, we focus on values that place the simulations in the same assembly regime as observed in the experiments, while enabling sampling a statistically significant ensemble of assembly trajectories. We summarize these parameters and their correspondence to the experiments here.

In Hayakawa *et al.*,<sup>9</sup> the total monomer concentration is  $c_0 \approx 10$  nM. If we set our standard state volume to that used in Hayakawa *et al.*,  $c_{\text{SS}} = 1/\nu_{\text{monomer}}N_{\text{A}} \approx 10$  M, with  $\nu_{\text{monomer}}$  the volume of a monomer and  $N_{\text{A}}$  Avogadro's number, then our chemical potential value  $\mu = -3k_{\text{B}}T$  corresponds to a bath concentration of  $c_0 \approx 500$  nM.

We use this larger concentration because it allows assembly dynamics to occur on faster timescales compared to the experimental concentration, thus making our dynamical simulations more computationally tractable. The higher concentration causes association to occur on shorter timescales, and, to maintain the same assembly regime as in the experiments, the binding energies can be somewhat smaller, making unbinding occur on shorter timescales. In particular, for  $\mu = -3k_{\text{B}}T$ , the threshold binding energy for assembly is  $E_{\text{B}} \approx 4.5k_{\text{B}}T$  (see Section II B and VII B, ESI†). For our dynamical simulations to be in the experimentally relevant regime, they require binding energies that are somewhat larger than the threshold energy, so that nucleation occurs within computationally accessible timescales (and similarly experimentally accessible timescales), which corresponds to the value we focus on,  $E_{\text{B}} = 6.0k_{\text{B}}T$ . At this binding energy, the probability for a pair of fused edges to undergo fission becomes sufficiently small that the closure event is nearly irreversible, but assembly still involves a significant nucleation barrier. Both conditions are consistent with experimental observations. Recall also that our simulation algorithm enforces the dilute assembly regime by construction, consistent with the low experimental concentration. Moreover, the net bulk free energy density of assembled tubules in the simulations,  $\Delta\epsilon \approx -3E_{\text{B}}/2 - \mu = -6k_{\text{B}}T$  is close to

the value at early stages of the experiments (before free monomers are significantly depleted): with  $c_0 = 10$  M the initial chemical potential is  $\mu \approx -7k_{\text{B}}T$ , while the binding energy was estimated as  $E_{\text{B}} \approx 9k_{\text{B}}T$ , resulting in  $\Delta\epsilon \approx -6.5k_{\text{B}}T$ .

Other differences between the simulations and experiments are as follows. The simulations are performed in the grand canonical ensemble and thus have a constant bulk monomer concentration. In contrast, the experiments are in the canonical ensemble (fixed total monomer concentration). Thus, the bulk monomer concentration is depleted as assembly occurs, so tubules that nucleate at later times in the experiment assemble at lower chemical potentials. Tubule morphology distributions are measured over all times, and thus average over these differences in bulk concentration. However, based on the observation that the width distribution of well-formed tubules does not depend sensitively on binding energy in this parameter regime (Fig. S10, ESI†), we expect that it is also insensitive to the chemical potential. In addition, the tubules in the experiments have an exponential distribution of lengths (as expected for the canonical ensemble). The experiments are performed for one week, by which time the longest tubule structure is about 2  $\mu\text{m}$ , which is roughly 10 times its diameter. In the simulations, we analyze all tubules after they have grown to a length of 10 times the tubule diameter. However, this difference does not affect the results, since, as we show in this work, the morphology distribution is set at the time of closure and is thus roughly independent of tubule length.

**Controlling the monomer–monomer binding affinity.** The experiments show that the monomer–monomer binding affinity increases with increasing salt concentration,<sup>8,9</sup> likely due to screening of electrostatic repulsions between monomers. Our binding affinity parameter  $E_{\text{B}}$  thus qualitatively maps to the salt concentration used in the experiments. However, the affinity increases much more strongly with magnesium compared to sodium than would be predicted by simple screening, suggesting specific interactions with magnesium. Therefore, we have not attempted a quantitative mapping between  $E_{\text{B}}$  and experimental parameters.

### III. Theoretical models

To determine whether kinetic effects influence the observed geometry distributions, we compute the equilibrium tubule geometry distribution and compare it against those observed in simulations. We first perform the calculation accounting for the discrete tubule geometries allowed by the finite monomer size, and then we simplify the calculation by adopting the continuum limit.

#### A. Discrete equilibrium model

Motivated by the high rigidity of DNA origami subunits,<sup>8,9</sup> we focus on the regime of high stretching modulus in this work, so that the relative edge length fluctuations are small,  $k_{\text{B}}T/k_{\text{s}}l_0^2 \ll 1$ . Thus, in the following calculation we assume that the curvature within a tubule is uniform, and for an assembled geometry ( $m, n$ )



all the dihedral angles are approximately equal to their ideal value  $\theta_{\text{id},i}^{(m,n)}$ , with  $i \in 1, 2, 3$  as the indices of the three sides of a subunit. We denote the ideal angles for the target geometry as  $\theta_{\text{id},i}^0$ . The free energy per monomer  $g_L^{(m,n)}$  in a tubule geometry  $(m, n)$  with length  $L$  is then approximately given by

$$g_L^{(m,n)} = \frac{2a_0\gamma}{L} - \left(\frac{3E_B}{2} + Ts\right) + \frac{1}{4}B \sum_{i=1,2,3} \left(\theta_{\text{id},i}^{(m,n)} - \theta_{\text{id},i}^0\right)^2 \quad (2)$$

in which  $a_0$  is the area of a monomer,  $\gamma$  is the line tension accounting for unsatisfied interactions at the two tubule boundaries,  $E_B$  is the binding energy,  $B$  is the bending modulus,  $T$  is the temperature, and  $s$  is the per-monomer entropy. The bending energy term is reduced by a factor of two compared with eqn (1) because it is shared between two neighboring monomers. The equilibrium probability  $P^{(m,n)}$  to assemble the tubule geometry  $(m, n)$  with length  $L$  is given by

$$P_L^{(m,n)} \propto \exp[-\beta N(g_L^{(m,n)} - \mu)], \quad (3)$$

where  $N$  is the number of monomers in the structure and  $\mu$  is the chemical potential. In the grand canonical ensemble  $\mu$  is equal to the bath chemical potential, while in the canonical ensemble (conserved total monomer concentration)  $\mu = k_B T \ln(c_0/c_{\text{SS}})$  with  $c_0$  the concentration of free monomers, with  $c_{\text{SS}}$  the standard state concentration.

We consider the large  $L$  limit, in which the contribution from the line tension can be ignored, so the free energy per monomer becomes independent of length and will be denoted as  $g^{(m,n)}$ . Further, at equilibrium, the free energy per monomer of the geometry that minimizes the free energy (in this case the target geometry) is approximately equal to the chemical potential  $\mu$ .<sup>15,46</sup> Since the bending energy of the target geometry is zero, the equilibrium chemical potential is given by  $\mu_{\text{eq}} \approx -\left(\frac{3E_B}{2} + Ts^*\right)$ ,

where  $s^*$  is the entropy per monomer of the target structure. Assuming the entropy is roughly independent of geometry, the probability distribution is then dominated by the bending energy, resulting in

$$P^{(m,n)} \propto \exp\left[-\frac{1}{4}\beta NB \sum_{i=1,2,3} \left(\theta_{\text{id},i}^{(m,n)} - \theta_{\text{id},i}^0\right)^2\right], \quad (4)$$

which shows that the probability to assemble an off-target structure decreases exponentially as the bending modulus  $B$  increases.

To test this analysis, we used an adapted thermodynamic integration algorithm to compute the free energy for different tubule geometries  $g_L^{(m,n)}$ . In brief, the algorithm evaluates the free energy change for each geometry along a thermodynamic pathway that gradually transforms the Hamiltonian of the system from a reference state (an Einstein solid with the same number of vertices) to our computation model (eqn (1)). We find that the measured free energy difference between different tubule geometries closely agrees with the bending energy difference, confirming the validity of the simplifications described above. See Fig. S17 and Section VII, ESI† for details

about free energy computations and the comparison to the bending energy.

## B. Continuum equilibrium model

To obtain an approximate analytical expression for the tubule width distribution, we adopt the continuum limit and neglect the presence of defects. In this limit, the bending energy as a function of tubule diameter  $D$  is given by the Helfrich energy<sup>47</sup>

$$g^D = \frac{2a_0\gamma}{L} - \left(\frac{3E_B}{2} + Ts\right) + 2\tilde{B}a_0\left(\frac{1}{D} - \frac{1}{D_0}\right)^2, \quad (5)$$

with  $D_0$  as the diameter of the target structure and  $\tilde{B}$  as the effective bending modulus in the continuum limit. The continuum bending modulus is related to the bending modulus  $B$  of the discrete model by  $\tilde{B} = (\sqrt{3}/2)B$ .<sup>30</sup>

We evaluate the equilibrium width fluctuations of the closed tubules ( $\Delta D \equiv \sqrt{\langle(D - \langle D \rangle)^2}\rangle$ ) as a function of their length  $L$ . By performing analogous simplifications to the discrete model (see Section II, ESI†), we obtain:

$$\Delta D \cong \sqrt{\frac{\sqrt{3}D_0^3 k_B T}{6\pi B L}}. \quad (6)$$

Eqn (6) shows that the relative equilibrium width fluctuations decrease with bending modulus, but increase with target diameter as  $\Delta D/D_0 \sim D_0^{1/2}$ , as found for spherical curvature-controlled capsids.<sup>15</sup> However, a key difference for tubules is that the equilibrium fluctuations become negligible for tubules with large aspect ratios  $L \gg D$ . Thus, the observations from simulations and experiments of appreciable width fluctuations in large-aspect ratio tubules indicate that kinetic effects are important in determining the polymorphism.

## C. Comparing simulation results against equilibrium width distributions

By comparing the simulation and equilibrium computation, we find that tubule geometry distributions from the dynamical simulations have larger variances than predicted by the equilibrium models. Fig. 5 compares the width fluctuations  $\Delta D$  measured in the simulations to the scaling law (eqn (6)) from the equilibrium computation for different values of the bending modulus and target geometry. In general, we see that the distributions observed in simulations have larger variances than the equilibrium results. Importantly, the observed  $\Delta D$  collapse to the equilibrium scaling with respect to the target diameter  $D_0$  and bending modulus  $B$ , but not at the tubule length at which the geometry measurements are performed ( $L_{\text{end}} \sim 3\pi D_0$ , black dashed line in Fig. 5). The measured diameter fluctuations are much larger than the equilibrium value. Instead, the fluctuations are roughly consistent with the equilibrium prediction for the smaller value of  $L_{\text{close}} \sim 1.5 D_0$  at which the tubules closed. Indeed, the results match the equilibrium prediction with  $L = L_{\text{close}}$  for all parameter values except  $\Delta D/l_0 \lesssim 0.5$ ; below this threshold the fluctuations are smaller than the discrete monomer size and the continuum approximation breaks down.



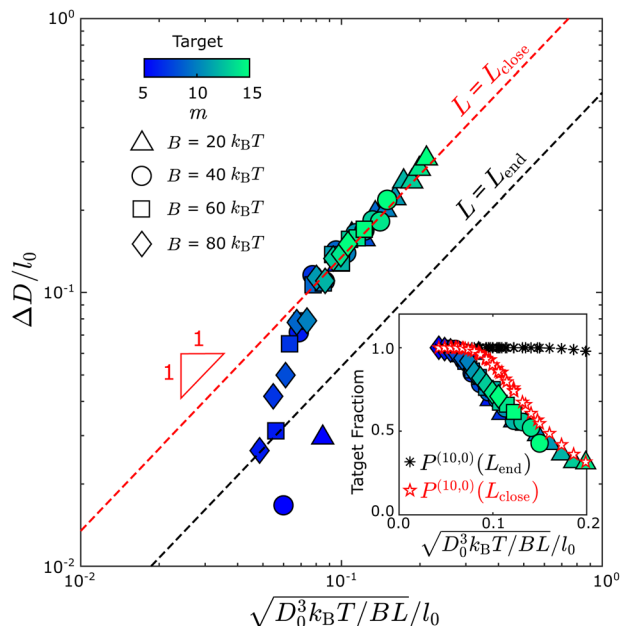


Fig. 5 Comparing the tubule geometry distribution from Monte Carlo assembly trajectories with the equilibrium theory shows that tubule closure fixes tubule geometries out of equilibrium. Tubule width fluctuations  $\Delta D$  measured from simulations at different parameter sets, plotted according to the scaling from equilibrium theory (eqn (6)). The dashed black line shows the equilibrium result for the tubule length at the end of the simulation (with  $L = L_{\text{end}}$  in eqn (6)), while the dashed red line shows the expected result if the geometry is quenched at the point of closure (with  $L = L_{\text{close}}$  in eqn (6)). Different symbols represent different bending modulus values  $B$ , and the color shows the first lattice number  $m$  of the target tubule geometry  $(m, n)$ ; all structures in this dataset have  $n = 0$ . The inset shows an analogous comparison for the fraction of tubules within the defect-free population that have the target geometry. The black asterisk symbols show the discrete model prediction (eqn (4)) with  $L = L_{\text{end}}$  and the red pentagon symbols show the discrete model prediction with  $L = L_{\text{close}}$ . Other simulation parameters:  $E_B = 6k_B T$  and  $f_{\text{fusion}} = 10^{-3}$ .

Note that similar results were observed for the same computational model in ref. 43 and were shown to be consistent with experiments on DNA origami subunits assembling the tubules in ref. 9. A detailed description of how we measure the tubule diameter and the closure size is given in the ESI† (Fig. S4 and S5).

The fact that the fluctuations are consistent with the equilibrium prediction, but at the smaller length  $L_{\text{close}}$ , indicates that the geometry distribution is *kinetically* controlled. This conclusion is consistent with the observation from simulations that the geometry rarely changes once a tubule closes, which can be understood from the fact that, after closure, all monomers have their maximum number of bonds except those at the two tubule ends. Rearrangement of the tubule geometry requires breaking a significant number of bonds, and thus overcoming a large free energy barrier. Note that the substitution of  $L_{\text{close}}$  into eqn (6) amounts to a quasi-equilibrium assumption: because assembly occurs near equilibrium for the parameters considered in Fig. 5, the tubule geometry distribution at the time of closure is nearly consistent with the equilibrium distribution

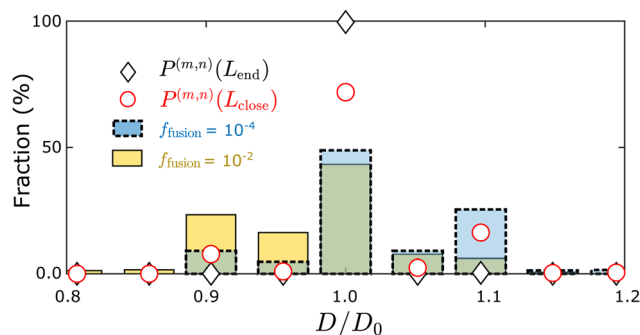


Fig. 6 Comparing measured tubule width distributions to the discrete equilibrium model (eqn (4)). Bars are the simulation results with indicated values of  $f_{\text{fusion}}$ , while symbols represent the equilibrium results, with tubule length at the simulation endpoint or at closure respectively. Other simulation parameters:  $B = 20k_B T$ ,  $E_B = 6k_B T$ , and the target tubule geometry is  $(10, 0)$ .

at the corresponding tubule length  $L_{\text{close}}$ . However, this condition breaks down for larger values of the edge fusion rate  $f_{\text{fusion}}$  as discussed next.

We also compared the fraction of each tubule geometry  $(m, n)$  predicted by the discrete model against the simulation results, which indicated a similar trend as for the continuum model: The distribution computed using  $L_{\text{close}}$  is much closer to the simulation results as compared with using  $L_{\text{end}}$  (Fig. 6), in terms of the width distribution and the yield of the target geometry (Inset of Fig. 5). Here, we replot the distribution in Fig. 3(C) against the diameter of the assembled tubule geometries (bars in Fig. 6). However, as  $f_{\text{fusion}}$  increases from  $10^{-4}$  to  $10^{-2}$ , the skewness of the tubule width distribution changes from below to above  $D_0$ . The equilibrium computation at  $L_{\text{close}}$  does not predict the change in skewness resulting from the change in the assembly kinetics.

This result shows that the simple picture based on a quasi-equilibrium morphology distribution at  $L_{\text{close}}$  does not capture all kinetic effects that control the tubule morphology distribution. In Section III D we develop a model that accounts for these additional dynamical influences.

#### D. Kinetic model for tubule geometry distributions

The results shown thus far suggest that factors affecting the size and geometry of the tubule at the moment of closure are the key determinants of the observed steady-state geometry distribution. In this section, we develop a discrete model that incorporates both the kinetics and thermodynamics of the system and we show that it semi-quantitatively describes the simulation results. We present an analogous continuum model in Section V C, ESI†

To simplify the analysis, we focus on parameters for which the critical nucleus size is small compared to the closure size, which covers most of the parameter space that we consider in this work (see Fig. 2). Therefore, in the model we assume that nucleation occurs well before closure, and thus the two processes are independent. In a future work, we will extend the



model to account for the case when the closure and critical nucleus sizes are comparable, and thus the two processes are coupled.

We consider the structure before closure as a circular disk that is bent to have the stress-free curvature of the target tubule (see Fig. S12, ESI† for a schematic of the model). The size  $N$  of the disc grows with a rate  $k_{\text{grow}}$  that is proportional to its boundary length:

$$k_{\text{grow}}(N) = k_{\text{grow}}^0 \sqrt{N}, \quad (7)$$

where  $k_{\text{grow}}^0$  is a factor that depends on  $B$  and  $E_B$  (see Fig. S8 and Section III F, ESI† for details about growth rate measurements). To simplify the model, we ignore the stochasticity in subunit association by assuming that  $N$  increases by one subunit at regular time intervals given by  $\Delta t = 1/k_{\text{grow}}$ .

While remaining at a size  $N$ , the open structure also attempts to close with a rate  $k_{\text{close}}$ . Once the structure closes, we assume that it does not reopen. Allowing for a finite reopening probability is straightforward, but has a negligible effect on the results for the parameters that we focus on because reopening is rare and/or transient. We assume  $k_{\text{close}}$  decreases exponentially with the free energy barrier to closure  $\Delta G_{\text{close}}$ , which arises primarily from the bending elastic energy due to the difference in the curvature of the closed structure and the stress-free structure. Section III G, ESI† presents estimates and measurements of  $\Delta G_{\text{close}}$ .

At a given size  $N$ , the rate  $k_{\text{close}}^{(m,n)}$  of closing into a structure  $(m, n)$  is then approximated by

$$k_{\text{close}}^{(m,n)}(N) = k_{\text{close}}^0 \exp\left(-\frac{\Delta G_{\text{close}}^{(m,n)}(N)}{k_B T}\right) I_{\text{close}}^{(m,n)}(N). \quad (8)$$

Here  $k_{\text{close}}^0$  is the closure attempt rate (*i.e.* the rate in the absence of a barrier), and  $I_{\text{close}}^{(m,n)}$  is a function that indicates whether a particular structure  $(m, n)$  is geometrically compatible with closure at size  $N$ :  $I_{\text{close}}^{(m,n)} = 1$  if it is compatible and  $I_{\text{close}}^{(m,n)} = 0$  if it is incompatible (see Section V B, ESI† for details about the determination of  $I_{\text{close}}^{(m,n)}(N)$ ). Assuming that shape fluctuations are fast in comparison to the net growth timescale, the net closure rate  $\tilde{k}_{\text{close}}$  for a disk with size  $N$  is then given by a sum over all accessible geometries as

$$\tilde{k}_{\text{close}}(N) = \sum_{(m,n)} k_{\text{close}}^{(m,n)}(N). \quad (9)$$

Finally, we evaluate the closure probability as a function of time. To simplify the calculation, we assume that the structure is larger than the critical nucleus size, and that closure is a rare event in comparison to growth. In the absence of closure, the time at which a structure first grows to size  $N$  is thus  $t_N = \sum_{i=1}^{N-1} 1/k_{\text{grow}}(i)$ , and the probability that such a structure stays open for an additional time  $\delta t < t_{N+1} - t_N$  is

$$P_{\text{open}}(t + \delta t, N) = P_{\text{open}}(t_N, N) \exp[-\tilde{k}_{\text{close}}(N) \delta t]. \quad (10)$$

By summing over smaller sizes, we can compute the probability that a structure has stayed open until size  $N$  as

$$P_{\text{open}}(t_N, N) = \prod_{i=1}^{N-1} \exp\left(-\frac{\tilde{k}_{\text{close}}(i)}{k_{\text{grow}}(i)}\right). \quad (11)$$

The probability for the structure to close at size  $N$  is then given by

$$P_{\text{close}}(N) = P_{\text{open}}(t_N, N) \left[1 - \exp\left(-\frac{\tilde{k}_{\text{close}}(N)}{k_{\text{grow}}(N)}\right)\right] \quad (12)$$

The probability to assemble a geometry  $(m, n)$  is then computed by summing over all sizes  $N$  that can close to  $(m, n)$

$$P_{\text{close}}^{(m,n)} = \sum_{N=1}^{N_{\text{max}}} P_{\text{close}}(N) \frac{k_{\text{close}}^{(m,n)}(N)}{\tilde{k}_{\text{close}}(N)}, \quad (13)$$

where the second term on the right-hand side is the conditional probability for assembling the geometry  $(m, n)$ , given that the structure closes at size  $N$ . Eqn (13) shows that the ratio of growth to closure rates, which is a kinetic factor, can significantly affect the tubule geometry distribution. Next, we will compare these predictions against the dynamical simulation results from Section II B.

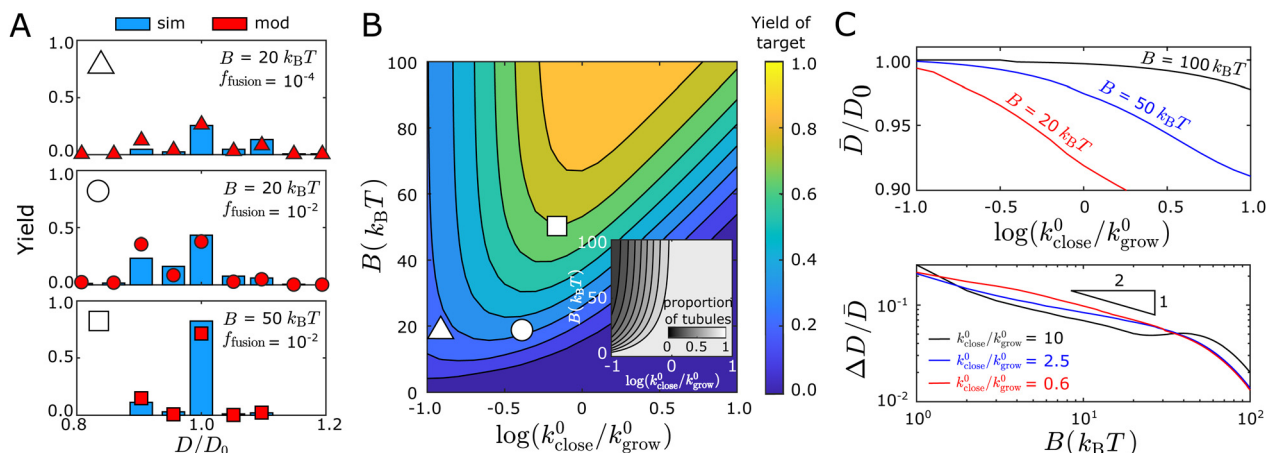
### E. Testing the kinetic model predictions

The simple kinetic model predicts the tubule geometry distribution and *yield* of the target geometry over a wide range of parameter space of the bending modulus  $B$  and the effective closure rate (normalized by the net growth rate,  $\log(k_{\text{close}}^0/k_{\text{grow}}^0)$ ). We define the yield as the fraction of a specific defect-free tubule geometry  $(m, n)$  within the entire population (including the unclosed structures and the defective tubules), and  $k_{\text{close}}^0$  is defined as the rate for an isotropic open structure to close and form the target geometry. In the simulations, the effective closure rate  $\log(k_{\text{close}}^0/k_{\text{grow}}^0)$  is controlled by the parameter  $f_{\text{fusion}}$ , and we measured  $\log(k_{\text{close}}^0/k_{\text{grow}}^0)$  from simulation trajectories as described in Section III F and G, ESI†.

The kinetic model accurately predicts the detailed distribution of defect-free tubule geometries, as well as the fraction of structures that fail to close in the dynamical simulations. Comparisons between the kinetic model and the simulation results are shown for three representative parameter sets in Fig. 7(A). Starting from the top panel, we reduce  $f_{\text{fusion}}$  by  $100 \times$  at fixed  $B$  (middle panel), which does not significantly change the spread of the distribution of closed tubules, but changes the skew from wider than targeted to narrower than targeted. More significantly, the yield of the target geometry decreases from 40% to 28% while the fraction of the target geometry within the defect-free population does not significantly change. This observation is because the proportion of unclosed structures increases from 0 to  $\sim 45\%$  within the entire population (Fig. S7, ESI†). The kinetic model captures this trend.

In the bottom panel, we increase  $B$  at fixed  $f_{\text{fusion}}$  (relative to the middle panel), which narrows the distribution considerably and shifts the mean toward the target diameter. In particular, a





**Fig. 7** The kinetic model captures the geometry distributions observed in simulations. (A) Comparison of the tubule width distribution between the kinetic model (red symbols) and simulation results (blue bars) for three representative parameter sets. The triangle, circle, or square symbol at the top left of each panel indicates its corresponding location in the parameter space for the plot shown in (B). The yield is defined as the fraction of a tubule geometry assembled within the *entire* population of the assembled structures, including the structures that do not close. (B) Color map showing the yield of the target tubule geometry ( $P_{\text{close}}^{(10,0)}$  defined in eqn (13)) predicted by the kinetic model as a function of the bending modulus  $B$  and the normalized closure rate  $k_{\text{close}}^0/k_{\text{grow}}^0$  (shown on a log scale). The inset shows the fraction of closed tubules predicted by the model ( $\sum_i P_{\text{close}}(i)$ ,  $P_{\text{close}}(i)$  is defined in eqn (12)). (C) Kinetic model predictions for the mean ( $\bar{D}$ , top panel) and coefficient of variation ( $\Delta D/\bar{D}$ , bottom panel) with respect to the normalized closure rate and bending modulus. In all cases the target geometry is (10,0).

significant fraction of tubules with sizes  $< D_0$  at  $B = 20k_B T$  shifts to the target geometry with  $D_0$  at  $B = 50k_B T$ . This trend reflects a combination of thermodynamic and kinetic effects. Increasing  $B$  increases the thermodynamic stability of the target geometry relative to competing structures. It also decreases the net growth rate  $k_{\text{grow}}^0$  because monomer association incurs a greater entropy penalty (since fewer configurations are accessible for binding at higher  $B$ ), which increases the effective closure rate (see Fig. 7(B)) and thus favors smaller structures. However, the thermodynamic effect dominates in this case and shifts the distribution upward toward  $D_0$  (see Fig. S15, ESI† for details).

Fig. 7(B) shows that the yield of the target geometry also depends on a combination of thermodynamic and kinetic factors. The value of the bending modulus  $B$  sets an upper limit on the yield, while the yield itself changes nonmonotonically with respect to the normalized closure rate at fixed  $B$ . These trends reflect the fact that the bending rigidity determines the spread of the distribution, while the closure rate mostly influences the mean of the distribution. As noted above, a higher closure rate leads to structures that close earlier and thus shifts the mean toward smaller structures.

The bending rigidity controls both the mean  $\bar{D}$  and the width fluctuation  $\Delta D$  of the distribution, while the effective closure rate mostly influences  $\bar{D}$ . Fig. 7(C) shows the mean of the distribution  $\bar{D}$  (top panel) and the width fluctuation  $\Delta D$  (bottom panel) as functions of the closure rate and the bending modulus. We see that the mean width monotonically decreases with increasing normalized closure rate or decreasing  $B$ . In contrast, the fluctuations  $\Delta D/\bar{D}$  decrease with  $B$  but depend only weakly on the closure rate. The latter trend is consistent with the qualitative results from the quasi-equilibrium model (Section III C) based on the equilibrium geometry distribution

at the time of closure. In particular, the scaling result  $\Delta D \sim B^{-1/2}$  still applies. However, the results for the mean width reflect the fact that the probability for an open structure with size  $N$  is determined by both the effective closure rate at that size and the time for the structure to remain at size  $N$ . As shown above, the closure rate increases with  $k_{\text{close}}^0$  and decreases with  $B$ , while the time for the structure to remain at a given size decreases with increasing  $k_{\text{grow}}^0$ . The irreversible nature of tubule closure plays a key role in this trend, since the smaller structures always have the opportunity to close before larger sizes. Thus, increasing the closure rate or extending the time at a given size will cause the entire distribution to shift toward smaller widths.

Interestingly, the closure rate does not significantly influence  $\Delta D/\bar{D}$ . Although the kinetic effects discussed above change the tubule closure size  $N_{\text{close}}$ , they do not significantly change the relative prevalence of different tubule geometries at a given  $N$ . Thus, as long as the shift of the distribution away from  $D_0$  is not too large, the density of states around the preferred geometry at size  $N_{\text{close}}$  remains comparable to that around the target geometry. This distribution is then essentially fixed once closure occurs.

## IV. Conclusions

In summary, we have used kinetic Monte Carlo simulations and free energy calculations to understand the dynamical assembly of helical tubules. Our simulations reveal how assembly pathways and the resulting tubule morphologies depend on control parameters. The geometry distribution of assembled tubules predicted by the simulations semi-quantitatively matches the distribution observed in experiments on tubules assembled



from DNA origami monomers,<sup>9</sup> suggesting that the model captures the key physics of the experimental system. Further, we show that the simulations provide a useful tool to obtain a first-order estimate of the physical parameters of the experimental system, and can serve as a predictive guide for future experiments.

Comparison of the simulation results with an equilibrium calculation shows that the geometry distribution of assembled tubules depends on a balance between thermodynamic and kinetic effects. While the observed magnitudes of the fluctuations in the tubule width  $\Delta D$  match the equilibrium scaling ( $\Delta D/D_0 \sim \sqrt{D_0/BL_{\text{end}}}$ ) with respect to bending modulus  $B$  and preferred diameter  $D_0$ , the distribution of assembled tubules is significantly broader and *independent* of length ( $L_{\text{end}}$ ). This behavior can be explained by the fact that the tubule geometry becomes fixed shortly after an assembling proto-tubule closes upon itself. Closure stabilizes monomer interactions except for those at the two open ends of the tubule, and the topology rearrangements required to significantly change the tubule structure would incur a large free energy barrier. For this reason, the observed geometry distribution fluctuations tend to scale with the tubule length at closure, as  $L_{\text{close}}^{-1/2}$ . For systems in which closure rates are slow in comparison to growth time-scales, the resulting tubule morphology distribution is approximately given by the equilibrium distribution at  $L_{\text{close}}$ . However, for systems with faster closure rates (relative to assembly), additional kinetic effects shift the geometry distribution further out of equilibrium. We developed a simple kinetic model which captures these additional effects.

While the computational model used in this work is based on triangular monomers motivated by the DNA origami experiments,<sup>8,9</sup> experimental and computational systems that assemble tubules from other monomer geometries, such as DNA tiles<sup>14,48</sup> and wedge-shaped monomers,<sup>23</sup> exhibit similar polymorphic behaviors as in the DNA origami system. Moreover, the simple kinetic model derived in Section III D does not assume a particular monomer geometry. The high level of agreement between predictions of this model and the computational results suggests that kinetic control of morphology applies generically to helical tubule self-assembly systems in which growth rates are significantly faster than reopening, regardless of the monomer shape.

### Model limitations and outlook

While our kinetic model closely reproduces the computational results over a wide range of parameter space, it is limited to regimes in which tubule closure occurs above the critical nucleus size. In particular, the model assumes a positive net growth of the assembling tubule and thus is limited to the forward-biased growth phase that occurs beyond the critical nucleus size. In a future work, we plan to study the nucleation behavior in detail, and how the assembly kinetics and geometry distribution change when closure occurs before nucleation. In this study we have primarily focused on parameters that lead to well-formed tubules, with a low fraction of defective tubules.

However, the simulations provide insights into the factors and mechanisms controlling defect formation. For example, analysis of our simulation trajectories suggests that defective tubules frequently arise when closure happens locally and independently at two or more sites on the boundaries, with geometries that are incompatible with the overall tubule geometry. This mechanism results in a local crack between binding sites, which is unable to anneal unless one of the bound edge-pairs breaks. We expect that the probability of observing defects through this mechanism increases as assembly and closure become less reversible, *via* increasing the binding energy  $E_B$  or the intrinsic closure rate constant  $k_{\text{close}}^0$ , or decreasing the bending modulus  $B$ . On one hand, the frequency of multiple independent closure events, and thus the probability of observing cracks by this mechanism, increase with the effective closure rate. As described in Section III D, the effective closure rate increases linearly with  $k_{\text{close}}^0$  (a material property) and exponentially with the free energy barrier to closure  $\Delta G_{\text{close}}$  (which depends both on material properties such as the bending modulus and the geometry of the tubule). On the other hand, the probability of such a crack annealing decreases with  $E_B$ . These observations are consistent with the general principles established from other self-assembly reactions and crystallization (*e.g.*<sup>40,42,49</sup>). When growth rates are sufficiently fast that monomers that associate with strained interactions cannot anneal before additional subunits assemble, defects become locked into the growing structure.

While some potential mechanisms of defect formation are disallowed by the simplifications of our model and simulations, these mechanisms can be neglected in the DNA origami experiments that motivate our work. In particular, the algorithm does not allow for binding between multiple partially assembled structures, but these events are negligible under the dilute assembly conditions with a substantial nucleation barrier that tend to lead to productive assembly.<sup>40,42,50</sup> Similarly, we do not consider binding of subunits along non-complementary edges because in the experiments,<sup>9</sup> monomer-monomer interactions were made highly specific using shape-complementary interactions based on blunt-end DNA base stacking, and there is no evidence of significant binding between non-complementary edges in the experiments. Note that it would be straightforward to extend the model to eliminate these simplifications to describe other systems for which these mechanisms are not negligible.

We also note that a kinetic Monte Carlo algorithm can only be reliably mapped to real dynamics if the move set accounts for all relevant transitions that occur in a given system, with approximately correct relative rates for each move. In this respect it is encouraging that the simulated tubule geometry distribution compares well with experimental observations. However, further comparison against additional data will be required to stringently test the simulated dynamics, and to refine relative rates. In particular, the simulations described here suggest that the rate at which free edges within an assembled tubule bind to each other is an important parameter controlling the closure rate and defect formation.



With the availability of additional experimental data, some of these unknown coarse-grained parameters could be directly estimated from experiments. At the same time, these measurements would provide estimates of unknown experimental parameter values. For example, by optimizing simulation tubule geometry and width distributions against experiments performed at different parameter values (*e.g.* target geometry, and monomer concentration), we could estimate the bending rigidity, as well as the closure and growth rates. Additional experimental techniques could enable directly estimating some coarse-grained parameters in the model. For example, growth rates could be estimated from dynamic light scattering experiments of tubule assembly, while dimerization rates and free energies for specific monomer–monomer edge interactions could be estimated from static light scattering experiments of subunits which each has only a single edge activated for binding.<sup>8,9,51</sup> Angular fluctuations of dimers measured using atomic force microscopy (AFM)<sup>52,53</sup> or estimated from electron density in cryo-electron microscopy experiments<sup>8,9,54</sup> would provide an independent means of estimating the bending rigidity.

Through combination with such experimental techniques, our computational and theoretical study could be used to improve the design of existing experimental platforms for tubule assembly. Further, analysis of simulation trajectories for a validated model will provide insights into mechanisms underlying assembly of tubules in these systems, and potentially other related systems with helical geometries such as microtubules,<sup>21–26,55–57</sup> filamentous viruses,<sup>58–63</sup> and diverse other helical assemblies found in biological systems.<sup>64</sup> The model and computational algorithms described in this work are broadly generalizable. They can be readily adapted to other monomer shapes, such as rectangular DNA tiles,<sup>14,48</sup> wedge-shaped monomers,<sup>23</sup> or biomolecules such as the tubulin dimers that form microtubules,<sup>21–26,55–57</sup> and other assembly symmetries, such as icosahedral capsids<sup>31–35,37</sup> and geometrically frustrated structures.<sup>36</sup> Such modifications require changing the graph structure of the triangulated sheet and the relationships among the monomer size, monomer–monomer interaction angles, and assembly curvature; for example, ref.37 derived interactions for a model of protein dimer subunits from atomistic simulations of HBV capsids. Thus, this modeling framework can be used to provide similar insights into other assembly geometries with different symmetries and mechanisms of self-limitation.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

We acknowledge support from National Institute Of General Medical Sciences R01GM108021 (HF, BT, MFH), NSF DMR-1710112 (HF, WBR), the Brandeis MRSEC NSF DMR-2011846 (HF, BT, WBR, MFH), the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project no. PN-III-P1-1.1-PD-2019-0236,

within PNCDI III (BT), and the Smith Family Foundation (WBR). We also acknowledge computational support from NSF XSEDE computing resources allocation TG-MCB090163 and the Brandeis HPCC which is partially supported by the NSF through DMR-MRSEC 2011846 and OAC-1920147.

## References

- 1 R. C. Weisenberg, *Science*, 1972, **177**, 1104–1105.
- 2 A. Desai and T. J. Mitchison, *Annu. Rev. Cell Dev. Biol.*, 1997, **13**, 83–117.
- 3 J. Han, H. Su, D. Zhang, J. Chen and Z. Chen, *J. Mater. Chem.*, 2009, **19**, 8741–8746.
- 4 V. Saranathan, C. O. Osuji, S. G. J. Mochrie, H. Noh, S. Narayanan, A. Sandy, E. R. Dufresne and R. O. Prum, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 11676–11681.
- 5 V. N. Manoharan, *Science*, 2015, **349**, 1253751.
- 6 M. A. Boles, M. Engel and D. V. Talapin, *Chem. Rev.*, 2016, **116**, 11220–11289.
- 7 K. Lee, W. Wagermaier, A. Masic, K. P. Kommareddy, M. Bennet, I. Manjubala, S.-W. Lee, S. B. Park, H. Cölfen and P. Fratzl, *Nat. Commun.*, 2012, **3**, 725.
- 8 C. Sigl, E. M. Willner, W. Engelen, J. A. Kretzmann, K. Sachenbacher, A. Liedl, F. Kolbe, F. Wilsch, S. A. Aghvami, U. Protzer, M. F. Hagan, S. Fraden and H. Dietz, *Nat. Mater.*, 2021, **20**, 1281–1289.
- 9 D. Hayakawa, T. E. Videbæk, D. M. Hall, H. Fang, C. Sigl, E. Feigl, H. Dietz, S. Fraden, M. F. Hagan, G. M. Grason and W. B. Rogers, arXiv:2203.01421 [cond-mat], 2022.
- 10 J. B. Bale, S. Gonen, Y. Liu, W. Sheffler, D. Ellis, C. Thomas, D. Cascio, T. O. Yeates, T. Gonen, N. P. King and D. Baker, *Science*, 2016, **353**, 389–394.
- 11 E. Golub, R. H. Subramanian, J. Esselborn, R. G. Alberstein, J. B. Bailey, J. A. Chiong, X. Yan, T. Booth, T. S. Baker and F. A. Tezcan, *Nature*, 2020, **578**, 172–176.
- 12 G. L. Butterfield, M. J. Lajoie, H. H. Gustafson, D. L. Sellers, U. Nattermann, D. Ellis, J. B. Bale, S. Ke, G. H. Lenz, A. Yehdego, R. Ravichandran, S. H. Pun, N. P. King and D. Baker, *Nature*, 2017, **552**, 415–420.
- 13 L. Lavelle, M. Gingery, M. Phillips, W. M. Gelbart, C. M. Knobler, R. D. Cadena-Nava, J. R. Vega-Acosta, L. A. Pinedo-Torres and J. Ruiz-Garcia, *J. Phys. Chem. B*, 2009, **113**, 3813–3819.
- 14 A. M. Mohammed and R. Schulman, *Nano Lett.*, 2013, **13**, 4006–4013.
- 15 M. F. Hagan and G. M. Grason, *Rev. Mod. Phys.*, 2021, **93**, 025008.
- 16 D. L. D. Caspar and A. Klug, *Cold Spring Harbor Symp. Quant. Biol.*, 1962, **27**, 1–24.
- 17 J. E. Johnson and J. A. Speir, *J. Mol. Biol.*, 1997, **269**, 665–675.
- 18 T. Dokland, *Struct. Fold. Design*, 2000, **8**, R157–R162.
- 19 P. Natarajan, G. C. Lander, C. M. Shepherd, V. S. Reddy, C. L. Brooks and J. E. Johnson, *Nat. Rev. Microbiol.*, 2005, **3**, 809–817.
- 20 M. F. Hagan and R. Zandi, *Curr. Opin. Virol.*, 2016, **18**, 36–43.



- 21 S. Cheng, A. Aggarwal and M. J. Stevens, *Soft Matter*, 2012, **8**, 5666–5678.
- 22 F. Cheng, I. B. Tsvetkova, Y.-L. Khuong, A. W. Moore, R. J. Arnold, N. L. Goicochea, B. Dragnea and S. Mukhopadhyay, *Mol. Pharm.*, 2013, **10**, 51–58.
- 23 S. Cheng and M. J. Stevens, *Soft Matter*, 2013, **10**, 510–518.
- 24 J. A. Bollinger and M. J. Stevens, *Soft Matter*, 2018, **14**, 1748–1752.
- 25 J. A. Bollinger and M. J. Stevens, *Soft Matter*, 2019, **15**, 8137–8146.
- 26 M. J. Stevens, *J. Chem. Phys.*, 2017, **147**, 044902.
- 27 D. J. Kraft, W. K. Kegel and P. van der Schoot, *Biophys. J.*, 2012, **102**, 2845–2855.
- 28 M. Marchetti, D. Kamsma, E. Cazares Vargas, A. Hernandez García, P. van der Schoot, R. de Vries, G. J. L. Wuite and W. H. Roos, *Nano Lett.*, 2019, **19**, 5746–5753.
- 29 A. Hernandez-Garcia, D. J. Kraft, A. F. J. Janssen, P. H. H. Bomans, N. A. J. M. Sommerdijk, D. M. E. Thies-Weesie, M. E. Favretto, R. Brock, F. A. de Wolf, M. W. T. Werten, P. van der Schoot, M. C. Stuart and R. de Vries, *Nat. Nanotechnol.*, 2014, **9**, 698–702.
- 30 H. S. Seung and D. R. Nelson, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 1005–1018.
- 31 G. M. Rotskoff and P. L. Geissler, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 6341–6346.
- 32 S. Li, P. Roy, A. Travasset and R. Zandi, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 10971–10976.
- 33 J. Wagner, G. Erdemci-Tandogan and R. Zandi, *J. Phys.: Condens. Matter*, 2015, **27**, 495101.
- 34 S. Panahandeh, S. Li, L. Marichal, R. Leite Rubim, G. Tresset and R. Zandi, *ACS Nano*, 2020, **14**, 3170–3180.
- 35 S. Panahandeh, S. Li and R. Zandi, *Nanoscale*, 2018, **10**, 22802–22809.
- 36 B. Tyukodi, F. Mohajerani, D. M. Hall, G. M. Grason and M. F. Hagan, *ACS Nano*, 2022, **16**, 9077–9085.
- 37 F. Mohajerani, B. Tyukodi, C. Schlicksup, J. Hadden, A. Zlotnick and M. F. Hagan, *BioRxiv*, 2022, **v**, 1–18.
- 38 R. K. F. Lee, B. J. Cox and J. M. Hill, *Nanoscale*, 2010, **2**, 859–872.
- 39 M. F. Hagan and D. Chandler, *Biophys. J.*, 2006, **91**, 42–54.
- 40 M. F. Hagan, *Adv. Chem. Phys.*, 2014, **155**, 1–68.
- 41 H. Fang, M. F. Hagan and W. B. Rogers, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 27927–27933.
- 42 S. Whitelam and R. L. Jack, *Ann. Rev. Phys. Chem.*, 2015, **66**, 143–163.
- 43 T. E. Videbaek, H. Fang, D. Hayakawa, B. Tyukodi, M. F. Hagan and W. B. Rogers, *J. Phys.: Condens. Matter*, 2022, **34**, 134003.
- 44 W. K. den Otter and W. J. Briels, *J. Chem. Phys.*, 2003, **118**, 4712–4720.
- 45 R. Dimova, *Adv. Colloid Interface Sci.*, 2014, **208**, 225–234.
- 46 Note that this approximation is still reasonable in the dynamical simulations, although we set  $\mu > \mu_{eq}$  so that the simulated tubule can undergo steady-state growth in length  $L$ , provided that the system is sufficiently close to equilibrium ( $\mu \gtrsim \mu_{eq}$ ).
- 47 W. Helfrich, *J. Phys.*, 1986, **47**, 321–329.
- 48 P. W. K. Rothmund, A. Ekani-Nkodo, N. Papadakis, A. Kumar, D. K. Fygenson and E. Winfree, *J. Am. Chem. Soc.*, 2004, **126**, 16344–16352.
- 49 V. Agarwal and B. Peters, *Advances in Chemical Physics*, John Wiley & Sons, Ltd, 2014, vol. 155, pp. 97–160.
- 50 A. Zlotnick, *Virology*, 2003, **315**, 269–274.
- 51 B. J. Berne and R. Pecora, *Dynamic Light Scattering: With Applications to Chemistry, Biology, and Physics*, Courier Corporation, 2000.
- 52 C. Roduit, S. Sekatski, G. Dietler, S. Catsicas, F. Lafont and S. Kasas, *Biophys. J.*, 2009, **97**, 674–677.
- 53 G. Thomas, N. A. Burnham, T. A. Camesano and Q. Wen, *J. Visualized Exp.*, 2013, 50497.
- 54 T. Nakane, D. Kimanius, E. Lindahl and S. H. Scheres, *eLife*, 2018, **7**, e36861.
- 55 G. M. Alushin, G. C. Lander, E. H. Kellogg, R. Zhang, D. Baker and E. Nogales, *Cell*, 2014, **157**, 1117–1129.
- 56 E. Nogales, *Annu. Rev. Biochem.*, 2000, **69**, 277–302.
- 57 D. Tong and G. A. Voth, *Biophys. J.*, 2020, **118**, 2938–2951.
- 58 T. A. Bharat, T. Noda, J. D. Riches, V. Kraehling, L. Kolesnikova, S. Becker, Y. Kawaoka and J. A. Briggs, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4275–4280.
- 59 L. J. Calder, S. Wasilewski, J. A. Berriman and P. B. Rosenthal, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10685–10690.
- 60 M. Nakano, Y. Sugita, N. Kodera, S. Miyamoto, Y. Muramoto, M. Wolf and T. Noda, *Commun. Biol.*, 2021, **4**, 858.
- 61 L. Selzer, Z. Su, G. D. Pintilie, W. Chiu and K. Kirkegaard, *PLoS Biol.*, 2020, **18**, e3000827.
- 62 Z. Su, C. Wu, L. Shi, P. Luthra, G. D. Pintilie, B. Johnson, J. R. Porter, P. Ge, M. Chen, G. Liu, T. E. Frederick, J. M. Binning, G. R. Bowman, Z. H. Zhou, C. F. Basler, M. L. Gross, D. W. Leung, W. Chiu and G. K. Amarasinghe, *Cell*, 2018, **172**, 966–978.e12.
- 63 C. Xu, N. Katyal, T. Nesterova and J. R. Perilla, *J. Chem. Phys.*, 2020, **153**, 155102.
- 64 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Garland Science, 4th edn, 2002.

