Failure mechanisms in DNA self-assembly: barriers to single fold yield

Jacob M. Majikes, Paul N. Patrone, Anthony J. Kearsley, Michael Zwolak, J. Alexander Liddle

Abstract: Understanding the folding process of DNA origami is a critical steppingstone to the broader implementation of nucleic acid nanofabrication technology but is notably non-trivial. Origami are formed by several hundred cooperative hybridization events – folds – between spatially separate domains of a scaffold, derived from a viral genome, and oligomeric staples. Individual events are difficult to detect. Here, we present a real-time probe of the unit operation of origami assembly, a single fold, across the scaffold as a function of hybridization domain separation – fold distance – and staple:scaffold ratio. This novel approach to the folding problem elucidates a predicted, but previously unobserved blocked state that acts as a limit on yield for single folds, and which may manifest as a barrier in whole origami assembly.

DNA origami is an exceptionally successful method of fabricating nanostructures with nearly arbitrary complexity[1]. However, limitations in characterization have precluded a full understanding of the origami folding process and its effect on yield. This in turn has prevented development of the necessary design-structure-property axioms. The path to these axioms can be revealed by studying simplified model systems. Here, we use fluorescence melt curve experiments to probe the *yield of one such model system – a single fold – as a function of folding distance and staple concentration. In doing so we observe a neglected intermediate state which results in a barrier for single folds and may adversely affect yield in DNA origami formation.*

Cooperative effects that occur during assembly are a major contributor to DNA origami's effectiveness as a self-assembly technique, but they also frustrate observation of the details of the folding process. The primary cooperative forces (configurational entropy, inter-staple base stacking, and bimolecular/unimolecular binding) change as each of the 2 to 4 domains in each of the approximately 200 staples bind. When viewed independently of the hundreds of concurrent hybridization events, these primary cooperative effects are conceptually straightforward to understand. It is their interplay during the myriad dynamically-dependent events which is nontrivial[2,3]. Cooperativity can significantly shift the free energy landscape, particularly for binding events occurring in close proximity[2,4]. Additionally, it is unclear how wide, in free energy terms, the funnel of events is, or to what degree it can be disrupted by misfolding [5].

The inherent complexity of the folding process is further compounded by the large design space available for DNA nanostructure fabrication. Design choices include type of structure[1,6,7], scaffold routing[8,9], scaffold sequence permutation[4], and staple motif[10,11]. Often, a given choice will simultaneously limit other choices and perturb multiple cooperative energetic factors. To predict assembly properties, such as yield, it is necessary to understand how the numerous design choices propagate through several hundred interconnected reversible events. The more completely one seeks to describe the relationship between design and cooperativity, the more daunting the task becomes, not unlike viewing the impossible triangle depicted in Fig. 1. Despite this, understanding design-yield relationships is critical to the implementation of DNA nanotechnology. To advance towards this goal, we take a reductionist approach, developing the tools needed to reassemble this complex interconnected system piece by piece.

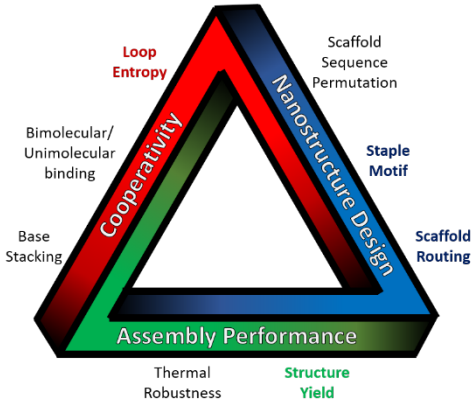

Fig. 1. The impossible triangle of complex self-assembly. We are able to deconvolve the bolded items by examining only a single fold.

We characterize the unit process of DNA origami folding under highly controlled conditions, measuring yield as a function of folding distance and staple:scaffold ratio. By examining a single fold, and thus limiting the number of possible topology states, we enable evaluation of a thermodynamic model for this unit-step system. Additionally, we detect the presence of a previously computationally observed[12] *but experimentally unobserved blocked state which is a potentially important mechanism for yield reduction in single folds, as well as a 'speed bump' for origami assembly. These effects have also been seen in DNA nanoswitch fabrication*[13]*. This blocked state may contribute to the observed "nucleation-and-growth" behavior*[8,14,15] *in whole DNA origami self-assembly, and provides an interesting perspective from which to understand the limits to growth of infinite tile systems. Our results reveal exciting distinctions between the behavior of individual folds and of multiple folds acting in concert, as well as the role of staple motifs. We consider this work a set of essential first steps towards understanding the full complexity of DNA nanostructure self-assembly.*

The paper is organized as follows: we first discuss the state of the art and the need for a more complete picture of the thermodynamics of DNA nanostructure formation. We then present a very brief survey of the strengths and weaknesses of characterization techniques that have been applied to DNA origami. Next, we introduce our model experimental system and a simplified state diagram for use in the later analysis. Following this, we construct an equilibrium model for folded-state yield as a function of staple:scaffold ratio and fold entropy. With this in hand, we discuss the affine transformation and the associated uncertainty quantification that is key to extracting meaningful data from our experiments. After describing the methods, we present compelling evidence for the existence of the blocked state and its dependence on the staple:scaffold ratio. Finally, we conclude with a discussion of how consideration of the blocked state provides insight into the process of origami formation.

Thermodynamics of DNA nanostructure formation:

We begin by considering the nearest-neighbor model of DNA thermodynamics, commonly used by DNA thermodynamics calculators[16]. It is effective for simple oligomer dsDNA melting and annealing, despite

the lack of a consensus regarding the change in heat capacity, ΔC_p, of hybridization[17]. Unfortunately, as noted above, in addition to simple base-paring interactions, DNA nanofabrication systems exhibit significant cooperativity between hybridization events[2,4,18]. While much effort has been devoted to modeling the folding process[2,4,9], the predictive power of the resulting models is limited, as the uncertainty in the magnitude of cooperative effects compounds over hundreds of events. Coarse grained simulations have also been tailored to these systems, and present another approach to handling the complexity of an origami[19]. While they face limitations in available simulation size and timescale, they can still provide valuable insight. In particular, work by Snodin *et al.*, observed that the blocked state was the major obstacle to assembly within simulations at high staple excesses for small origami[12]. Even when folding can be directly probed, the number of folding events presents a significant obstacle to analysis[20].

Cooperativity confounds the assembly process in three ways. First, inter-staple base stacking between neighboring strands provides an enthalpy bonus, of the same magnitude as that gained by adding a base [21], changing the subdomain melting temperature, T_m, by up to 5 °C. Second, if a staple is already partially bound, its remaining sub-sequences bind unimolecularly. This can increase the remaining domains' T_m by as much as 50 °C. However, this intra-fold cooperativity depends on the stiffness of the rest of the DNA structure. It thus can also decrease the remaining domains' T_m when the unimolecular fold is across a stiff structure. This is a consequence of the third factor: inter-fold cooperativity. The occurrence of any given fold influences the kinetics and thermodynamics of the remaining folds by changing the stiffness and distances between binding domains of those folds. In other words, the configurational entropy associated with binding the scaffold acts as an energetic penalty collectively across all binding events that change the topology[9,18,22].

Whether an event consumes more or less than its average share of the configurational entropy penalty, is determined by the topology state of the scaffold before and after the event. In origami systems, the most relevant measure of this is the fold distance between the two points being joined. The total energetic cost of folding the DNA into a rigid origami raft may be approximated by considering the packing of viral DNA into a T7 bacteriophage capsid. Assuming a linear scaling with polymer length compared to the T7 genome[23], an M13mp18 scaffold would incur a free energy cost at 25 °C, ΔG_{25}, of only 92 kcal.mol^{-1} (385 kJ.mol^{-1}) at room temperature to force the scaffold into a similarly compact configuration. By comparison, the ΔG_{25} of the whole origami assembly, and that of a single staple, as estimated from a random sequence of identical length, are ≈ 10 000 kcal.mol^{-1} (≈ 40 000 kJ.mol^{-1}) and ≈ 45 kcal.mol^{-1} (≈ 188 kJ.mol^{-1}), respectively.

While the absolute magnitude of the entropy penalty may seem inconsequential compared to the overall energetics, the size of the penalty associated with an individual fold can significantly destabilize that fold until the topology state, and thus the penalty, changes[8,9,14]. As with grain boundary energetics in metals, the cooperative entropy penalty is quite small but has a disproportionate control over the material microstructure and thus properties. Unfortunately, measurement of these small but important thermodynamic quantities and the population statistics of the possible topology states is limited by available characterization techniques.

With recent exceptions[20], the resolution of imaging techniques and traditional biological methods has limited our ability to characterize DNA origami[24]. Gel electrophoresis is by far the most convenient technique for characterizing DNA nanostructure yield, but also has the poorest resolution, as it can only

discern differences in electrophoretic mobility[10,11], such resolution is wholly insufficient for structural differences of less than tens of staples. Fluorescence-based techniques have been used to monitor these systems, but the limitations of data analysis associated with the sheer number of possible topologies make meaningful interpretation of the results difficult-to-impossible[20,25]. DNA Point Accumulation for Imaging in Nanoscale Topography (DNA-PAINT) has been used to quantify end-state inclusion yields for individual staples[3]. Finally, atomic force microscopy (AFM) and transmission electron microscopy (TEM) have been used to monitor the folding process, and have most commonly been applied postmortem[9,26,27]. Both of these imaging techniques lack the resolution required to easily image the free and flexible scaffold in order to observe topology changes. An interesting application of imaging relevant to this work is AFM observation of folding as a function of temperature[8,14,15] or through some assembly constraint[9,26]. As discussed above, if cooperativity were not active in DNA origami folding, one might expect a sequence-specific order of assembly favoring longer and GC rich staples, distributed more or less randomly through the origami. These direct imaging studies of folding all indicate whole origami assembly behaves as a sequence-agnostic nucleation-and-growth system, and have been taken as evidence of cooperativity[8,14,15].

Results & Discussion:

Model System:

The model system we employ provides an excellent example of the deceptive simplicity of DNA origami, as illustrated by the state diagram shown in Fig. 2. In this diagram, we neglect larger multi-scaffold polymerization products under the assumption that nanomolar scaffold concentrations are sufficiently dilute to minimize multi-scaffold products. It is worth noting that at less-than-stoichiometric staple:scaffold ratios, i.e., < 1 , this assumption may be less robust when the folding process incurs a large entropic penalty. Our system comprises the scaffold molecule, a folding oligomer staple, and two reporter oligomers labeled with a fluorophore and quencher respectively[22]. The scaffold consists of the

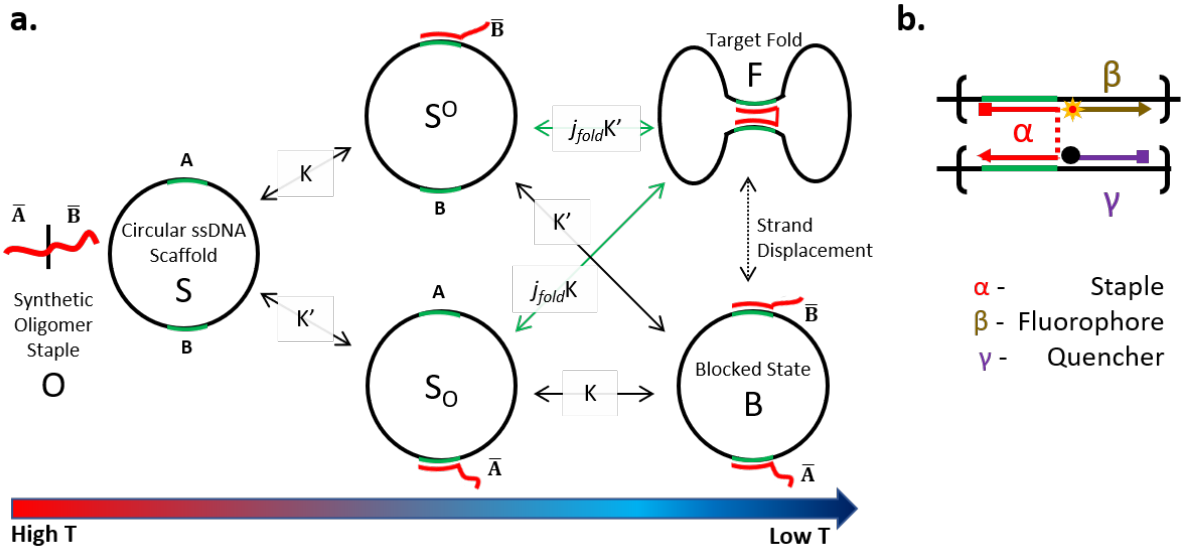

Fig. 2. a) Example state diagram for a two-domain synthetic staple oligomer whose domains, $\bar{A}$ and $\bar{B}$, bind to two disparate complementary positions, A and B (green) on the circular scaffold. K and K' correspond to the equilibrium constants associated with the hybridization energetics for the A and B domains, respectively. b) Fluorescent reporter scheme used in this work.

M13mp18 genome hybridized over most of its length with long oligomers designed to remove secondary structure as discussed in the methods. As designed, the only assembly events which occur below 75 °C are hybridization between scaffold and staple at the two binding domains.

As shown in Fig. 2a, the majority of the reactions are bimolecular (black arrows) and are sensitive to the staple and scaffold concentrations. The two reactions (green arrows) resulting in the target fold are unimolecular and would, with well-separated single domain melting temperatures, be insensitive to staple and scaffold concentration, but since they are coupled to the bimolecular reactions, this is not the case here (see, e.g., Majikes *et al*. [22] for a design that does separate them). Since the hybridized state of each domain is highly favored at low temperature, the final state will consist entirely of the folded or blocked states to the extent possible given the initial concentrations of scaffold and staple, and provided that the quench is not too rapid.

It should be noted that our reporter cannot distinguish between the blocked state and a scaffold without a staple. It is reasonable to assume, below stoichiometry, that the fraction of unquenched reporters is directly correlated to an unfolded state, whereas at high excess that fraction is linked to the blocked state. The system will progress left to right during annealing, as states with the largest number of hybridized bases will be favored at low temperature.

In contrast to our simplified system depicted here, typical staples have at least three binding domains and experience cooperative contributions from many more neighboring states. The diagram for a three-domain staple is more complex than those shown in in Fig. 2. For full origami systems of hundreds of coupled staples, the state diagram becomes exponentially more complex. In such cases, extracting meaningful information about reaction mechanisms can be problematic as uncertainties propagate, and are specific to the assumed mechanisms[28]. By probing a much simpler system, we are better able to determine the characteristics of individual events and are less likely to err by assuming an incorrect mechanism.

As illustrated in Fig. 2b, our reporter scheme enables the staple concentration to be varied while maintaining the fluorophore- and quencher-labeled strands at constant concentrations. We note that this scheme requires a different quencher-labeled oligomer for each folding distance, which may increase concentration variations associated with UV-Vis concentration measurements and pipetting different fold-distance samples. The reporter shown in Fig. 2b signals a change in topology, rather than a change in hybridization. Using our previously developed[22,29] affine transformation technique to improve background correction, together with approximate knowledge of the fluorescence of the quenched system, we estimate the fraction of the scaffold in the folded and blocked states. This simple reporting scheme is only viable due to the reduced nature of the system: a typical staple with three domains would require at least five reporting channels. Using this approach, we examined the effect of staple concentration on the yield of folds of 131 bases, 1187 bases, and 3407 bases with a symmetric staple having binding domains of 15 bases on either side, which we denote as a 15-15 base motif. In addition to probing the effect of staple concentration, we also examined the influence of staple motif by comparing the yield of a 15-15 base motif with a 25-25 base motif, and an asymmetric 10-40 base motif on the 1187 base fold.

Gel electrophoresis has proven effective at detecting and differentiating various supercoiled and knotted DNA structures through changes in rigidity and compactness[30], and can readily distinguish between linear and circular variants of the same strand[30], with pulsed-field methods established for

particularly large genomic DNA[31]. However, gel mobility is unaffected by the transition between the circular and figure-eight forms in our system. This is a consequence of several factors: first the addition of the folding staple changes the mass of the system by ≈ 0.5 %, second, the presence of numerous nicks in the dsDNA means that folding induces no additional changes to the structure, third, the high-salt buffer (12.5 mM Mg^{2+}) means the structure is already relatively compact, fourth, the fold is unlikely to change the rigidity of the structure significantly, and not as a function of fold location. An extensive exploration of both DC and pulsed-field gel methods failed to identify any conditions that could distinguish the circular and figure-eight states.

Equilibrium model for yield:

The equilibrium concentrations versus temperature for the model depicted in Fig. 2a are as follows: to simplify modeling of these reactions, we neglect strand displacement[32] and use *j*-factor notation, in which the effect of folding is treated as a change in local concentration rather than as an explicit entropic and/or enthalpic penalty[33]. The *j*-factor decreases as the entropy penalty increases, as defined by Eq. *1*, where ΔG_{fold} encompasses both entropic and enthalpic penalties, R is the gas constant, and T is the temperature[33]. If only entropic penalties were included, the *j*-factor would be temperature independent.

Eq. 1: $j_{\text{fold}} = e^{\left(-\Delta G_{fold}/RT\right)}$

The equilibrium constants, *K* and *K'*, are associated with the hybridization energetics for the A and B domains, respectively, and are determined by the free staple oligomer $[\mathcal{S}]$, ssDNA scaffold template $[\mathcal{T}]$, singly-bound scaffolds $[\mathcal{T}^s]$ & $[\mathcal{T}_s]$, the blocked state [*B*], and the folded state [*F*] concentrations. Eq. 2 through Eq. 5 represent the equilibrium constants for the reactions in Fig. 2.

Eq. 2: $K = \frac{[\mathcal{T}^{\mathcal{S}}]}{[\mathcal{T}][\mathcal{S}]} = \frac{[B]}{[\mathcal{T}_{\mathcal{S}}][\mathcal{S}]}$

Eq. 3: $K' = \frac{[\mathcal{T}_{\mathcal{S}}]}{[\mathcal{T}][\mathcal{S}]} = \frac{[B]}{[\mathcal{T}^{\mathcal{S}}][\mathcal{S}]}$

Eq. 4: $j_{\text{fold}} * K = \frac{[F]}{[\mathcal{T}_{\mathcal{S}}]}$

Eq. 5: $j_{\text{fold}} * K' = \frac{[F]}{[\mathcal{T}^{\mathcal{S}}]}$

The total concentration of both the scaffold and staples must be conserved to their initial values ($\mathcal{T}^o$ and $\mathcal{S}^o$), the result of which is shown in

Eq. 6 and Eq. 7.

Eq. 6: $\mathcal{T}^o = [\mathcal{T}] + [\mathcal{T}^{\mathcal{S}}] + [\mathcal{T}_{\mathcal{S}}] + [B] + [F]$

Eq. 7: $\mathcal{S}^o = [\mathcal{S}] + \left[\mathcal{T}^{\mathcal{S}}\right] + [\mathcal{T}_{\mathcal{S}}] + 2[B] + [F]$

Eq. 2 through Eq. 7, can be rearranged to yield a cubic polynomial which can be analytically solved for the folded, [*F*] and blocked, [*B*], states as a function of the *j*-factor, j_{fold}, initial staple $\mathcal{S}^o$ and scaffold $\mathcal{T}^o$ concentrations, and temperature. The analytical solution simplifies at low temperature, where all

hybridization events happen, more details in SI section 6, resulting in the yield, Υ, shown in Eq. 8 and Eq. 9 for $\mathcal{S}^o < \mathcal{T}^o$ and $\mathcal{S}^o > \mathcal{T}^o$, respectively.

Eq. 8: $\Upsilon = \frac{[F]}{\mathcal{T}^o} = \frac{\mathcal{S}^o}{\mathcal{T}^o}$

Eq. 9: $\Upsilon = \frac{[F]}{\mathcal{T}^o} = \frac{1}{2\mathcal{T}^o}\left(-\left(j_{\text{fold}} + \mathcal{S}^o - 2\mathcal{T}^o\right) + \sqrt{\left(j_{fold} + \mathcal{S}^o - 2\mathcal{T}\right)^2 + 4j_{\text{fold}}\mathcal{T}^o}\right.$

This solution was used to model and fit the yield of the folded state at the low-temperature plateau, and the full, temperature-dependent cubic solution to was used to provide context for the T_m, as discussed in the SI.

Origami-relevant approximation:

Examining these equations in the context of whole origami allows for derivation of an interesting rule of thumb. Typically, DNA origami systems are processed with staple:scaffold ratios, $\mathcal{S}^o/\mathcal{T}^o$, ≥ 10× – meaning that it is reasonable to approximate $[\mathcal{S}] \approx \mathcal{S}^o$. As the yield at the end of an anneal is the primary concern of experimentalists, one may also assume sufficiently low-temperature conditions such that only the blocked and folded states are energetically favorable. Using either Eq. 2 and Eq. 4 or using Eq. 3 and Eq. Eq. 5 one may find the relationship between folded and blocked states, Eq. 10.

Eq. 10: $\left.[F]\right/_{[B]} = \left.j_{\text{fold}}\right/_{[\mathcal{S}]}$

From Eq. 10 we can find the yield of the folded state after annealing (i.e., at low temperature where all hybridization events have occurred) as a function of initial staple concentration and *j*-factor, shown in Eq. 11.

Eq. 11: $\Upsilon = \frac{[F]}{\mathcal{T}^o} = \frac{[F]}{[F]+[B]} = \frac{1}{1+\left.[B]\right/_{[F]}} = \frac{1}{1+\left.\mathcal{S}^o\right/_{j_{\text{fold}}}} \to \frac{j_{\text{fold}}}{\mathcal{S}^o}$

Eq. 11 provides a useful tool for the experimentalist, limiting to $j_{\text{fold}}/\mathcal{S}^o$ for small $j_{\text{fold}}/\mathcal{S}^o$ ratios. It indicates that absent cooperative effects, the yield of a single fold in an origami is proportional to the ratio of the initial staple concentration and the *j*-factor. We reserve our discussion of cooperativity until later.

Methods:

For every fold distance and staple:scaffold ratio examined here, 6 replicate measurements were made. For the fluorescence temperature dependence 12 replicates were measured, and for the background signal 96 replicates were measured. The fluorescence was measured during annealing after a one minute denaturation at 95 ˚C followed by a quick ramp to 80 ˚C, the data was then collected as the temperature was reduced by 0.75 ˚C per minute for 99 steps to 5.75 ˚C.

The 7,249 base M13mp18 genome used as the scaffold was partially hybridized by approximately 200 oligomers around the entire molecule, except for the regions which hybridized to the staple or reporter strands. These oligomers were designed to create a repeating pattern of 32 bases of dsDNA and 5 bases of ssDNA around the scaffold. This was done to strike a balance between removing unpredictable secondary structure and creating an overly stiff dsDNA scaffold which would interfere with the folding.

Each sample was nominally 50 µL in volume with constituents in absolute and relative concentrations described in table 1. Pipetting of samples was performed via a programmable automatic pipettor with a 0.1 µL to 10 µL volume range. This necessitated four different concentration stocks of the staple strand at approximately (0.1, 1, 10, and 100) µmol.L^{-1}. These stocks were created by dilution and their concentration measured by UV-Vis spectroscopy, except for the 0.1 µmol.L^{-1} which did not have sufficient absorbance. For this sample, the average pipetting uncertainty was added in quadrature to the concentration uncertainty in the 1 µmol.L^{-1} stock from which it was diluted.

	Absolute Concentration	*Ratio to Scaffold*
Scaffold	27.3 ± 1.4 nmol.L^{-1}	--
Staple	7.0 ± 0.3 to 2820 ± 41 nmol.L^{-1}	0.26 ± 0.02 × to 103 ± 6 ×
Secondary Structure Strands	165.7 ± 9.0 nmol.L^{-1}	5.2 ± 0.6 ×
Quencher	282.0 ± 18 nmol.L^{-1}	10.3 ± 0.8 ×
Fluorophore	25.8 ± 1.7 nmol.L^{-1}	0.9 ± 0.08 ×

Table 1. Concentration of DNA strands in annealing experiments, absolute and relative to the scaffold concentration. Uncertainties in staple concentration are the standard deviation of at least four UV-vis measurements. The uncertainty in concentration is negligible for the 100 µMol/L stock, 0.5 % for the 10 µmol.L^{-1} stock, 5 % for the 1 µmol.L^{-1} stock, and 5.4 % for the 0.1 µmol.L^{-1} stock. The absorbance of the 0.1 µmol.L^{-1} stock was too low to measure by UV-Vis and its concentration uncertainty was taken to be the addition in quadrature of the uncertainty for 1 µmol.L^{-1} stock from which it was diluted and pipetting uncertainty, taken as a maximum possible value of 2 %. These values were then corrected for systematic underestimation of concentration in UV-vis due to dsDNA content.

The uncertainties due to pipetting and UV-vis measurement precision discussed above can potentially underestimate the uncertainty in the final concentration. As shown in the SI, the scaffold and at least two of the staple strands can exhibit weak secondary structure: since dsDNA has a lower absorbance than ssDNA this would result in a systematic underestimation of concentration. This systematic uncertainty can be as much as 20 %[34]. While the predicted secondary structure shown in the SI would indicate a much lower systematic bias, we chose to measure the systematic underestimation in this system by comparing the absorbance in buffer and in 50 % Dimethylsulfoxide (DMSO) with appropriate blank/background samples[34]. The difference between these absorbances was treated as the systematic underestimation of concentration. The underestimation of secondary structure strands, quencher and fluorophore were assumed to be equal to that of the staple as the experiment is likely to be insensitive to small changes in concentration for these strands.

Further, it is unlikely that every quencher strand is perfectly functionalized with the correct molecule. Although less likely, it is possible that the quenching itself is not complete. In either of these cases the yield and signals reported here would need to be rescaled.

Another potential source of uncertainty is the way in which the state diagram, Fig. 2, would change as a result of competing secondary structure in the scaffold or staple molecules. This would then have difficult-to-model effects on yield and T_m.

Affine Transformation and Uncertainty Quantification:

In this work, we use a variation of the hierarchical affine transformation data analysis technique developed previously[22,29]. This technique assumes that every measured data set is generated by a "true" signal, σ(T), that is distorted by an unknown amount of temperature-dependent fluorescence and background effects. For example, we model the DNA measurements presented here, using Eq. 12, where $m_i(T)$ is the i^{th} measurement, R(T) is a temperature-dependent fluorescence efficiency, B(T) is a background signal, and the unknown affine parameters a_i, b_i, c_i characterize the relative contributions of each source. These parameters should encompass variations in concentration due to pipetting error, variations in optical alignment along the 96-well plate, as well as dark-current in the detector for a particular sample, i.

Eq. 12: $m_i(T) = a_i\sigma(T)R(T) + b_iB(T) + c_i$

The goal of our analysis is to estimate the affine parameters, the figure of merit for which is the collapse of all replicates, or realizations $\mathrm{m_i(T)}$, onto the true signal, σ(T).

As R(T) and B(T), are initially unknown, similar models are also proposed for them, shown in Eq. 13 and Eq. 14. The true background B(T), is found by collapsing 96 realizations, $B_i(T)$, with scaling (d_i) and offset (e_i) affine parameters for each realization i. Using this true background 12 realizations of the fluorescence temperature dependence, $R_i(T)$, are collapsed into the true R(T) by the scaling (f_i), background scaling (g_i), and offset (h_i) affine transformations.

Eq. 13: $B(T) = d_iB_i(T) + e_i$

Eq. 14: $R(T) = f_iR_i(T) + g_iB(T) + h_i$

In this analysis, we formulate and solve a constrained optimization problem that requires estimates of the true signal to collapse onto one another in a physically meaningful way. This achieves effective baseline and background subtraction without fitting empirical models and uses the feasibility of collapse (which is not always possible) as an internal consistency check. If the system cannot collapse, it may be an indication that the model has left out some physical source of signal variation, or that the predetermined threshold for variation after collapse was too stringent. This is an improvement over traditional baseline fitting and subtraction as it provides a warning that the user may be, unknowingly, manipulating numerous discretionary choices to over-fit the data in order to meet prior expectations.

While the uncertainty quantification (UQ) for affine transformations of melt curves is addressed in depth in our previous work[22,29], the experiment here presents a significant complication. In particular, the data collapse process only determines the signals up to a multiplicative factor and constant offset common to them all. Thus, two additional equations are required to set the scale of the transformed curves. In Majikes *et al*. [22], it was known that the low and high temperature data represented fully folded and unfolded states, respectively, providing the necessary information. In the present work, however, that is not true. In fact, the yield of the folded state at low temperature is the unknown quantity to be measured. In the raw melt curves, this corresponds to unknown asymptotic fluorescence behavior of the signal at low temperature. *Informally, we may view this problem as not knowing what value of fluorescence reported by the instrument corresponds to zero.*

We address this problem by performing measurements on a collection of wells with no fluorophores. This provides a mean background signal and offset. A non-trivial statistical analysis is then required to determine the fraction of constant offset due to background sources of fluorescence versus electronics.

Incorrectly estimating these contributions can artificially increase uncertainties in the final determination of yield. While details are reserved for the SI, these estimates are also used to inform regularization terms and constraints that test for feasibility of data collapse subject to the mean effects. Datasets for which this collapse could not be achieved were rejected. This is particularly noticeable for systems having low yield, which have a lower signal-to-noise ratio.

The final uncertainty analysis for yield leverages the observation that after subtracting the mean background and offsets, variation in the low-temperature asymptotic behavior of a melt curve reflects the unbiased uncertainty in zero fluorescence. Thus, by using each realization in turn as a reference and mapping the remaining curves to it, we generated a distribution of yield estimates, from which we computed a mean and standard deviation, more details are in the SI.

Finally, two sources of uncertainty in this experiment are not accounted for in the affine transformations. First, the quantum yield of the fluorophore-quencher pair might not be zero. This is unlikely given the < 1 nm spacing between fluorophore-labeled position and quencher-labeled position. The fluorophore and quencher were located on bases which could otherwise be engaged in a crossover junction, Fig. 2c, so at a minimum the fluorophores are exiting their respective dsDNA helices facing one another. It is likely that they may be in close enough contact to exhibit static quenching[35]. Second, are the uncertainties discussed in the methods section, particularly in measuring the staple concentrations.

Confirmation of the blocked state:

Typical DNA origami have been annealed successfully at staple concentrations ranging from 5× to 100× the scaffold concentration. To probe the blocked state, we examined the yield of the blocked and folded states as a function of staple concentration in this range. We explored a small number of concentrations below 1×, as under these conditions the blocked state should be vanishingly rare, and the fraction of the folded state should scale linearly with the staple concentration.

The results of the fluorescence measurements, for a fold distance of 3407 bases (approximately the mid-point of the scaffold), are shown in Fig. 3. The uncertainty envelopes were inferred through the affine

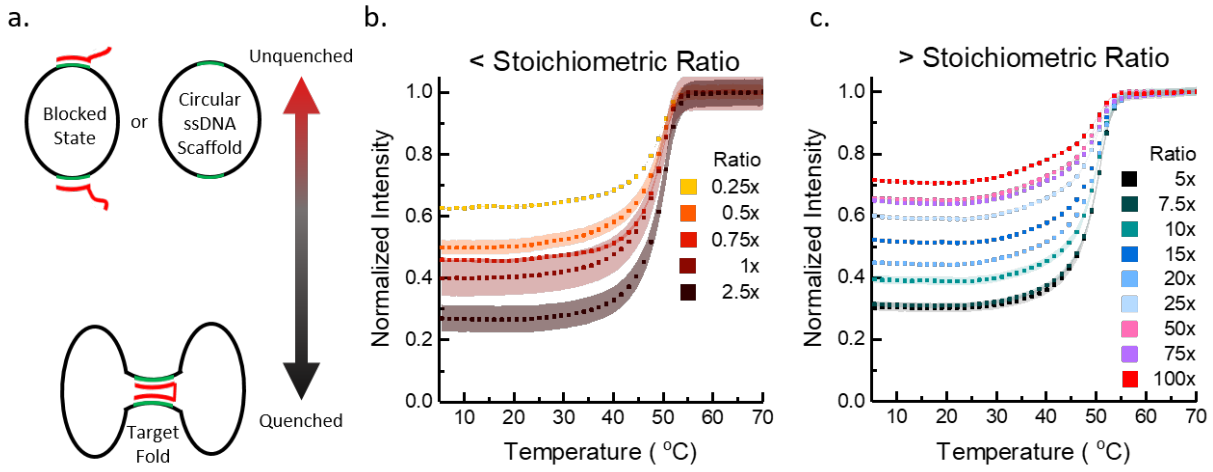

Fig. 3. a) schematic of unquenched, or fluorescing, versus quenched fluorophore states, b) & c) Anneal curves of normalized reporter fluorescence intensity at b) staple:scaffold ratios < 1 and c) staple:scaffold ratios > 1 for a fold distance of 3407 bases and a staple with symmetric 15 base binding domains. The shaded regions around each curve represent the uncertainty envelope at one standard deviation.

transformation analysis described above. We take the normalized intensity to be equivalent to the fraction of molecules in the unfolded state and use [1 - Normalized Intensity] to define the single-fold yields. As discussed in the uncertainty section, less-than-perfect quenching of the fluorophore would result in a proportional rescaling of Fig. 3 to a lower normalized intensity. However, as noted above, we do not anticipate this as a likely possibility.

The normalized intensities in Fig. 3 indicate a successful probe of the state diagram in Fig. 2b. From Fig. 3b and Fig. 3c, one can infer that the yield of the folded state increases as a function of staple concentration up to about 5× the scaffold concentration. While this is expected from literature on whole origami folding[1], the decrease in yield, a result of the blocked state becoming increasingly favored, as staple concentration increases past this level, provides unique insight in the context of whole origami. Moreover, yield should theoretically be maximum at the stoichiometric staple:scaffold ratio. Thus, the observed offset provides some evidence of the interplay of kinetic and quench effects. However, due to the flatness of the theoretical yield curves for moderate excess, as well as measurement uncertainties and sparse data, it is difficult to determine the exact process occurring in this region.

While there is an anecdotal consensus that whole origami yield is improved with > 10× staple relative to the scaffold, this is not the case for a single fold. For whole origami, the higher initial staple:scaffold ratio may 'jump-start' the events on the left side of the state diagram in Fig. 2a, giving cooperative effects a broader range of time/temperature in which to push the system toward the target structure via strand displacement.

Yield as a function of staple:scaffold ratio and fold distance:

Single-fold yields are shown as a function of staple:scaffold ratio and fold distance in Fig. 4a. All three folding distances have the expected turnover in yield between 1× to 10× staple concentration and all fall below 50 % yield at high staple:scaffold ratios. The turnover for 138, 1187, and 3407 base folds occur at between 5 × and 1×, although the breadth of the yield turnover for the broader peaks creates additional uncertainty in identifying the true peak position. A turnover at a ratio higher than 1× for any single fold is not seen in the model, but the broad peaks, which make experimental determination of the optimum staple concentration uncertain, do match our model. We do not anticipate this to be a kinetic effect as anneals performed two times slower replicated the trend (SI Fig S4). Nor do we anticipate it to be due to off-product sequences in the fold strands themselves as they were ordered PAGE purified. A possible cause could be unprogrammed interaction between fold strands and those removing secondary

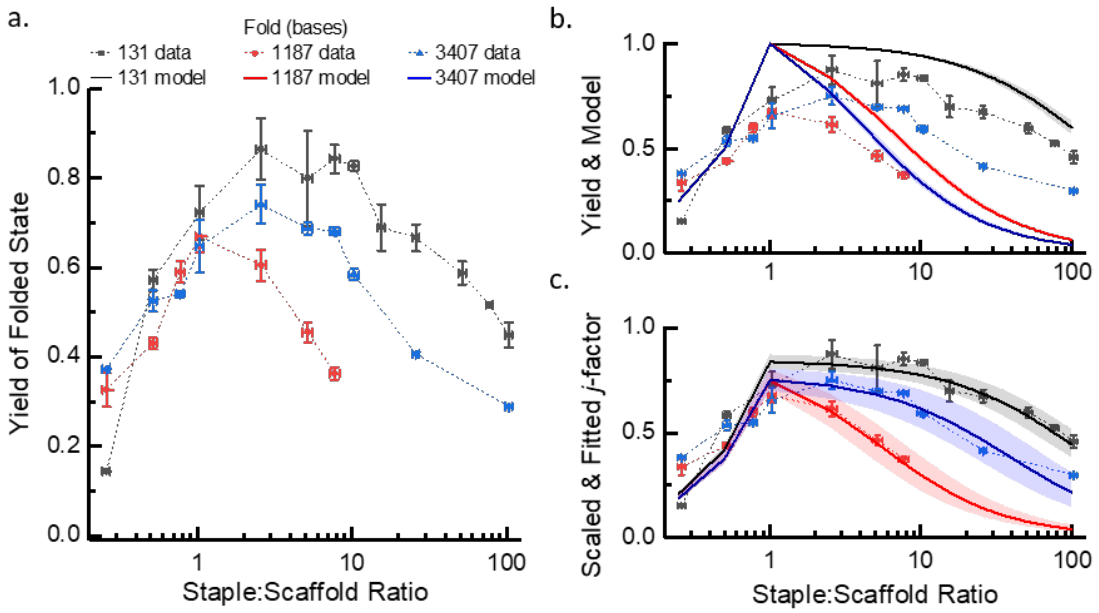

Fig. 4. a) Measured yield of the folded state as a function of staple:scaffold ratio for 131, 1187, and 3407 base folds. The yield is maximum at intermediate staple:scaffold ratios and decreases in both the small and large staple concentration regimes. In the former case, there are not enough staples to bind to all the scaffolds and form the folded state, while in the latter, it is the formation of the blocked state that limits yield. b) Model prediction using literature j-factor values, and c) with y-axis scaling and fitted j-factor. The x-axis uncertainty bars represent a standard deviation of UV-vis quantification. The y-axis uncertainty presented represents a standard deviation as determined by the affine transformation. The model uncertainty envelopes were calculated by the upper and lower uncertainty bounds, as represented by a standard deviation above and below the j-factor extracted in previous work.

structure. The trend of increasing then decreasing yield, modulated by the fold-distance generally matches the predicted yields, shown in Fig. 4b. The model yield was plotted for both b) *j*-factors derived from our previous work[22]. The model was then scaled, and its *j*-factors fitted, c), to show the ability of the model to capture the trends in the data, particularly when the data is rescaled, which assumes a 'leaky' FRET reporter.

	j-factor from[22] *(mol.L^{-1})*	*fitted j-factor (mol.L^{-1})*	*scale factor (a.u.)*
131 base fold	3.6×10^{-6} $^{+4.5 \times 10^{-7}}_{-5.2 \times 10^{-7}}$	2.7×10^{-6} $^{+5.9 \times 10^{-7}}_{-7.5 \times 10^{-7}}$	0.89 ± 0.03
1187 base fold	1.7×10^{-7} $^{+1.3 \times 10^{-8}}_{-1.4 \times 10^{-8}}$	1.3×10^{-7} $^{+4.0 \times 10^{-7}}_{-5.6 \times 10^{-7}}$	0.75 ± 0.05
3407 base fold	1.0×10^{-7} $^{+9.6 \times 10^{-9}}_{-1.1 \times 10^{-8}}$	9.5×10^{-7} $^{+3.4 \times 10^{-7}}_{-5.4 \times 10^{-7}}$	0.75 ± 0.05

Table 2. Anticipated and fitted j-factors for the model in Fig. 4. Uncertainties represent one standard deviation, and for j-factor are reported separately for upper and lower values.

The *j*-factors from in Fig. 4b and c are given in Table 2, and indicate that the fitted *j*-factors are comparable in magnitude to those fitted in previous work. The *j*-factors from previous work were directly re-fitted using the melting temperature, T_m, data: a short discussion of this choice and its alternatives is presented in the SI.

One unanticipated facet of this data is the lower yield for the 1187 base fold as compared to the 3407 base fold. As both the folds are displaced from what might be expected, it is difficult to say whether one, the other, or both are anomalous, or to ascribe a mechanism to this observation.

Tentatively, we observe that the 3407 fold has a gentler, or stretched, decay in yield at high staple:scaffold ratio than one might expect. We therefore make the assumption that the 3407 base fold is anomalous. Moreover, due to presence of some intrastaple and binding domain complementarity, strand displacement may be playing a role in the apparent the *j*-factor increase. This led us to examine the T_m as a consistency check, discussed in detail in SI section 3. In the T_m data, we see the same inversion between expected and observed behavior for the 1187 base fold and 3407 base fold. Numerical solution of Fig. 4 through Eq. 7, was used to model a T_m trends based on the fitted *j*-factors in Table 2. These predicted values were comparable to the observed T_m, indicating that the observed T_m and yield data are self-consistent.

Our comparison between differing staple motifs, which bridge the same fold distance but with different length domains on either side of the fold, are shown in Fig. 5. These varying motifs were tested at the 1187 base fold distance. The motifs themselves consisted of a 30 base staple with two 15 base domains (the default motif used above), a 50 base staple with two 25 base domains, and a 50 base asymmetric staple with one 10 base domain and one 40 base domain.

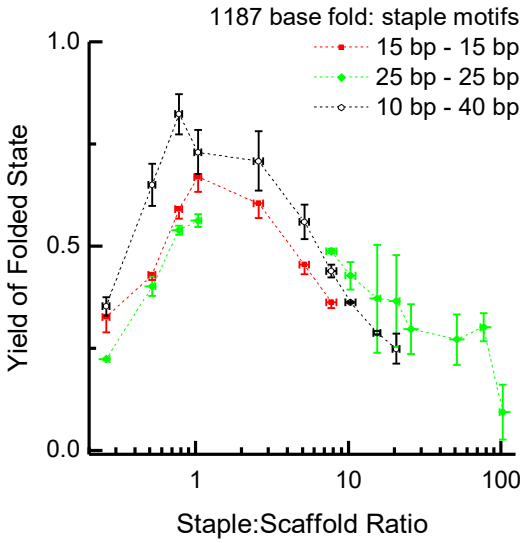

Fig. 5. Yield of folded state as a function of staple excess for different staple motifs bridging the same fold distance. X-axis uncertainty presented as maximum bound of possible UV-vis underestimation of concentration due to dsDNA content. The Y-axis uncertainty presented represents a standard deviation as determined by the affine transformation.

While the asymmetric staple deviates slightly from the two symmetric staples at low staple:scaffold ratio this deviation is on the order of one to two standard deviations and is far smaller than for a difference in fold distance. Given the uncertainties in the data, it is reasonable to concluded that these systems had the same yield of the folded state. This is consistent with the notion that the energetic difference between the blocked and folded states is independent of the size of staple domains, since, in both folded and blocked states, both binding domains have hybridization of the same sequences.

Fig. 5 highlights a relevant feature of the affine transformations present, but less visible, in Fig. 4. The affine transformations do dramatically improve data quality and reduce potential user error in analysis, however if the transformations fail to collapse replicates of a sample, there is little recourse other than to reject that sample or loosen uncertainty constraints. In Fig. 4, the rejected samples were exclusively low-signal. However, for the 25 bp - 25 bp 1187 base fold, numerous samples of moderate scaffold:staple ratio failed to collapse. Therefore, we have confined our claims to those supported by the data presented.

The reporter used here clearly shows the presence of a blocked state, however it does not provide enough information to fully interrogate the complex behavior of these single folds. Future work

implementing a reporter system for multiple states in the state diagram will allow us to better probe the unexpected yield and T_m values, provide a more comprehensive dataset for kinetics experiments, and enable interpretation of multiple simultaneous folds designed for origami-style cooperativity.

The blocked state in whole origami & tiles:

We have demonstrated that single-fold yield decreases as the staple:scaffold ratio increases above one. However, when going from a single fold to the many folds that occur in a typical DNA origami assembly process, other processes need to be considered. Specifically, cooperative effects, such as the systematic reduction in fold distances as the assembly processes proceeds, can clearly – based on the tradeoff between *j* and $\mathcal{S}^o$ discussed above – diminish the occurrence and impact of the blocked state. This state will occur less because of enhanced *j*-factors after partial assembly and the impact can also be reduced, due to strand displacement, if blocking does occur (with the rate of displacement enhanced by shorter distances). We note, however, that cooperativity can go both ways – stiffer structures due to partially assembled regions can disfavor certain folds and thus favor blocked states.

Given this, it is clear that different, full origami structures may display different propensities for blocking. Bringing the techniques developed in this work to bear on this more complex problem will thus require incorporating additional aspects into the experimental design. However, as a first step towards this larger goal, we examine the assembly of the notched rectangle at intermediate-to-large staple excess. This structure, as seen in Fig. 6, still forms at high fidelity even at 100x staple excess. This suggests that cooperative effects, whether by enhanced *j*-factors or strand displacement (or both), are at play, and may underlie the nucleation-and-growth behavior seen in studies directly imaging the folding process[8,14,15].

Additionally, we have observed folding of 100 × staple:scaffold ratio systems when the system is simply taken from the thermocycler at 95 °C and immediately imaged. The folding at this high ratio and rapid quench is somewhat surprising as both the results presented here and those of Snodin *et al.* would lead one to expect dramatically misfolded origami[12]. It is also worth emphasizing that our single fold results and the work by Snodin *et al.* regard purified staples (experimentally and computationally, respectively), where those used to fold the notched rectangle in Fig. 6 are not purified. Assuming the manufacturers' per-base DNA synthesis yield of 99.5 %, 15 % of the strands in the notched rectangle staple pool are truncated and could poison a binding site until removed by a correct copy via strand displacement. We anticipate the folding under these conditions is aided by the notched rectangle origami's 8 base – 16

base – 8 base staple motif, which may mitigate the blocked state. In future work we hope to identify design/anneal conditions which allow observation of the blocked state for varying design motifs.

For instance, one can imagine that a significant portion of internal, long-distance, folds enter the blocked state during a thermal anneal, only to subsequently undergo strand displacement. This also may explain why isothermal anneals result in more rapid origami formation[36], as they may not have to undergo as many strand displacement reactions. Generally, as an origami folds, any blocked location will eventually be brought into close enough proximity to engage in strand displacement. The cooperative reduction in entropic penalty that occurs as the structure folds makes the $^3/_2\,R$ gain in entropy as a blocking strand is released into solution a sufficiently large driving force for strand displacement to occur and assembly to progress.

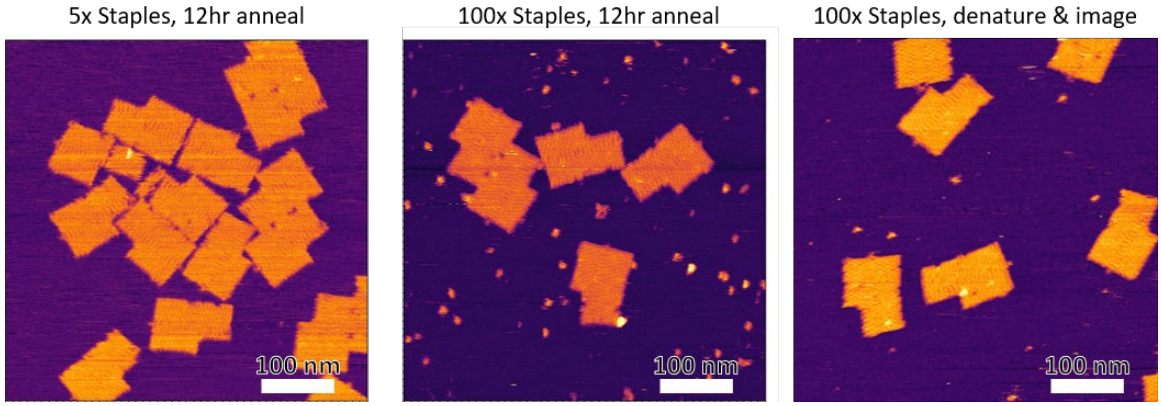

Fig. 6. Example AFM images of 12 hr anneals of notched rectangle origami at 5x and 100x staples for twelve hour anneals, and for a denaturing step at 95 °C, where cooling occurred in transit to AFM.

The above picture suggest that, if the blocked state is a kinetic 'speed-bump' in whole origami assembly that is removed by strand displacement, as driven by cooperativity, it would make sense that the origami would nucleate in areas having short fold distances, giving rise to a growth front that propagates by strand displacement along axes of increasing cooperative contributions.

In 3D origami, compared to 2D origami, one might expect the blocked state to be more prevalent since the distribution of fold distances tends, on average, to be more heavily weighted towards longer ones. These would exacerbate the effect of the blocked state, suggesting that it could play a more significant role in the misfolding of 3D structures. However, the effect that staple motif can have in mitigating the misfolding of 3D structures would indicate that there is more than a single factor at play[10,11].

The notion of the blocked state provides an interesting perspective from which to view DNA tiles. Specifically, the failure mode of infinitely expanding tiles as a function of starting stoichiometry can be thought of as a different manifestation of the blocked state: as the tile system grows to full hybridization, strands of its constituent species are consumed in stoichiometric ratios. Work by Green et al. provides an interesting characterization of

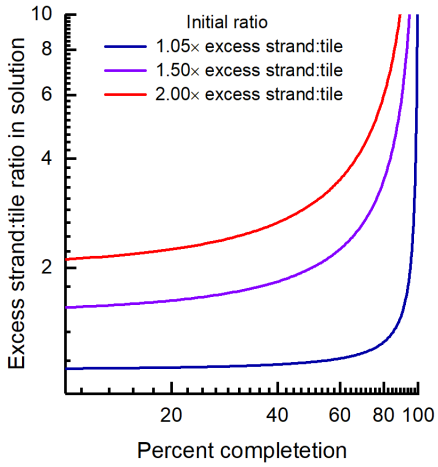

Fig. 7. Analytical model of relative excess of the overabundant oligomer as a function of reaction completion.

this behavior[37]. If a single oligomer is slightly higher in the initial concentrations relative to the other strands, the relative proportion of that strand increases geometrically as the reaction progresses[6] as illustrated in Fig. 7,with equations as described in the SI. For this purpose, we consider the excess strand:tile ratio as analogous to the staple:scaffold ratio. This leads to an increasing fraction of blocked sites around the perimeter of the tile as it increases in size, where the inter-strand cooperative base stacking from neighbors is also at a minimum, ultimately terminating growth.

While the limitations of infinitely growing tiles are well known[6], contrasting them with origami can provide new insights for tile design. Origami systems have three sources of cooperativity which favor folding – progressive reductions in loop-entropy penalty, unimolecular equilibrium constants for the folded state, and inter-staple base stacking for neighboring DNA strands on the same helix. Tile systems only have the third source of cooperativity, making the analogous blocked state much more problematic. If other sources of cooperativity could be engineered to drive tile systems away from blocked states, e.g., metastable hairpins on each oligomer that increase local concentration near the growth front enough to enhance strand displacement, it is possible that yield of tiling systems could be significantly improved.

Conclusions

In this work, we have shown strong evidence – the turnover and decay of yield *versus* staple concentration – of a predicted but previously unobserved equilibrium blocked state in single-fold systems. As this turnover occurs at staple:scaffold ratios between 1 and 10, it is a probable intermediate state in the typical assembly of DNA origami structures. This state is likely a major contributor to thermodynamic limits to yield of nanoswitch fabrication[13]. We find that the blocked state is sensitive to loop distance and staple:scaffold ratio but not significantly to staple motif. While the blocked state can dramatically reduce the yield of single folds, it does not reduce origami yield to the same degree, presumably due to cooperative effects and strand displacement, although it may well contribute to yield failures in 3D origami structures. In both 2D and 3D origami, we hypothesize that the role of cooperativity is to facilitate strand displacement and removal of the blocked state.

The degree to which the blocked state will act as a barrier, either thermodynamic or kinetic, will depend heavily on the distribution of fold *j*-factors, initial concentrations, and on the annealing regime.

The blocked state, acting as a kinetic trap in the folding process, can, in principle be monitored in real time by varying annealing rates. Additionally, the multichannel reporting mechanisms we suggest for tracking other states in the folding diagram will provide more insight into both cooperative effects and the unexpected yield/T_m trends we observe here. These approaches, coupled with the fact that the influence of the blocked state is expected to be highly dependent on structure design (i.e., loop distance distribution), makes measurement of the blocked state a useful tool with which to probe cooperative effects and gain a foothold in understanding what constitutes “good” DNA origami design.

Finally, our experimental confirmation of the blocked state provides critical context for previous work. While DNA origami is, inherently, sequence specific, it is probable that the strand displacement reactions correcting the blocked state and other misfolds, can be slow and are only favorable along a ‘nucleation front’, where beneficial cooperative effects outweigh the entropic penalties associated with folding. This nucleation front interpretation is consistent with previous AFM imaging work on DNA origami formation[4,8,15,26] and indicates more nuanced behavior in origami assembly beyond simple melt

temperature reduction as a function of conformational entropy penalties. Understanding the competition between the blocked state, other misfolding mechanisms, and cooperative effects, as mediated by strand displacement, presents a meaningful opportunity to derive DNA origami design rules.

[1] P. W. K. Rothemund, *Nature* **2006**, *440*, 297–302.

[2] F. Dannenberg, K. E. Dunn, J. Bath, A. J. Turberfield, T. E. Ouldridge, M. Kwiatkowska, A. J. Turberfield, T. E. Ouldridge, *J. Chem. Phys.* **2015**, *143*, 1–19.

[3] M. T. Strauss, F. Schueder, D. Haas, P. C. Nickels, R. Jungmann, *Nat. Commun.* **2018**, *9*, 1–7.

[4] J.-M. Arbona, J.-P. Aimé, J. Elezgaray, *J. Chem. Phys.* **2013**, *138*, 015105.

[5] J. Zenk, R. Schulman, *PLoS One* **2014**, *9*, DOI 10.1371/journal.pone.0111233.

[6] N. C. Seeman, *Structural DNA Nanotechnology*, Cambridge University Press, Cambridge, **2016**.

[7] B. Wei, M. Dai, P. Yin, *Nature* **2012**, *485*, 623–6.

[8] J. M. Majikes, J. A. Nash, T. H. LaBean, *Nanoscale* **2017**, *9*, 1637–1644.

[9] K. E. Dunn, F. Dannenberg, T. E. Ouldridge, M. Kwiatkowska, A. J. Turberfield, J. Bath, *Nature* **2015**, DOI 10.1038/nature14860.

[10] Y. Ke, G. Bellot, N. V. Voigt, E. Fradkov, W. M. Shih, *Chem. Sci.* **2012**, *3*, 2587.

[11] T. G. Martin, H. Dietz, *Nat. Commun.* **2012**, *3*, 1103–1106.

[12] B. E. K. Snodin, F. Romano, L. Rovigatti, T. E. Ouldridge, A. A. Louis, J. P. K. Doye, *ACS Nano* **2016**, *10*, 1724–1737.

[13] K. Halvorsen, M. E. Kizer, X. Wang, A. R. Chandrasekaran, M. Basanta-Sanchez, *Anal. Chem.* **2017**, acs.analchem.7b00791.

[14] J. Lee Tin Wah, C. David, S. Rudiuk, D. Baigl, A. A. A. Estevez-Torres, J. L. T. Wah, C. David, S. Rudiuk, D. Baigl, A. A. A. Estevez-Torres, *ACS Nano* **2016**, *10*, 1978–1987.

[15] J. Song, J.-M. Arbona, Z. Zhang, L. Liu, E. Xie, J. Elezgaray, J.-P. Aime, K. V. Gothelf, F. Besenbacher, M. Dong, *J. Am. Chem. Soc.* **2012**, *134*, 9844–7.

[16] J. SantaLucia, D. Hicks, *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.

[17] P. J. Mikulecky, A. L. Feig, *Biopolymers* **2006**, *81*, 38–58.

[18] J. M. Arbona, J. Elezgaray, J. P. Aimé, *EPL (Europhysics Lett.* **2012**, *100*, 28006.

[19] J. P. K. Doye, T. E. Ouldridge, A. A. Louis, F. Romano, P. Šulc, C. Matek, B. E. K. Snodin, L. Rovigatti, J. S. Schreck, R. M. Harrison, W. P. J. Smith, *Phys. Chem. Chem. Phys.* **2013**, *15*, 20395–20414.

[20] F. Schneider, N. Möritz, H. Dietz, *Sci. Adv.* **2019**, *5*, 1–11.

[21] P. Yakovchuk, E. Protozanova, M. D. Frank-Kamenetskii, *Nucleic Acids Res.* **2006**, *34*, 564–574.

[22] J. M. Majikes, P. N. Patrone, D. Schiffels, M. Zwolak, J. Kearsley, S. P. Forry, J. A. Liddle, *Nucleic Acids Res.* **2020**, 1–13.

[23] S. C. Riemer, V. A. Bloomfield, *Biopolymers* **1978**, *17*, 785–794.

[24] D. Mathur, I. L. Medintz, *Anal. Chem.* **2017**, acs.analchem.6b04033.

[25] X. Wei, J. Nangreave, S. Jiang, H. Yan, Y. Liu, *J. Am. Chem. Soc.* **2013**, *135*, 6165–6176.

[26] J. M. Majikes, J. A. Nash, T. H. LaBean, *New J. Phys.* **2016**, *18*, 115001.

[27] R. Kosinski, A. Mukhortava, W. Pfeifer, A. Candelli, P. Rauch, B. Saccà, *Nat. Commun.* **2019**, *10*, 1–12.

[28] J. Tellinghuisen, *Anal. Biochem.* **2003**, *321*, 79–88.

[29] P. Patrone, A. Kearsley, J. Majikes, J. A. Liddle, *Anal. Biochem.* **2020**, DOI doi.org/10.1016/j.ab.2.

[30] J. L. Viovy, *Rev. Mod. Phys.* **2000**, *72*, 813–872.

[31] J. S. Simske, S. Scherer, *Nucleic Acids Res.* **1989**, *17*, 4359–4365.

[32] J. D. McGhee, P. H. von Hippel, *Biochemistry* **1975**, *14*, 1281–1296.

[33] S. D. Levene, S. M. Giovan, A. Hanke, M. J. Shoura, **2013**, 513–518.

[34] A. O. Nwokeoji, P. M. Kilby, D. E. Portwood, M. J. Dickman, *Anal. Chem.* **2017**, *89*, 13567–13574.

[35] M. K. Johansson, H. Fidder, D. Dick, R. M. Cook, *J. Am. Chem. Soc.* **2002**, *124*, 6950–6956.

[36] J.-P. J. Sobczak, T. G. Martin, T. Gerling, H. Dietz, *Science* **2012**, *338*, 1458–61.

[37] C. M. Green, K. Schutt, N. Morris, R. M. Zadegan, W. L. Hughes, W. Kuang, E. Graugnard, *Nanoscale* **2017**, *9*, 10205–10211.