

APPLIED SCIENCES AND ENGINEERING

The sequence of events during folding of a DNA origami

Fabian Schneider, Natalie Möritz, Hendrik Dietz*

We provide a comprehensive reference dataset of the kinetics of a multilayer DNA origami folding. To this end, we measured the folding kinetics of every staple strand and its two terminal segments during constant-temperature assembly of a multilayer DNA origami object. Our data illuminate the processes occurring during folding of the DNA origami in fine detail, starting with the first nucleating double-helical domains and ending with the fully folded DNA origami object. We found a complex sequence of folding events that cannot be explained with simplistic local design analysis. Our real-time data, although derived from one specific DNA origami object, through its sheer massive detail, could provide the crucial input needed to construct and test a quantitatively predictive, general model of DNA origami assembly.

INTRODUCTION

DNA origami enables the bottom-up self-assembly of discrete objects (1–3) with subnanometer precise features (4), dimensions ranging from the nanometer to the micrometer scale (2, 4–14), and molecular masses up to the gigadalton scale (15, 16). In DNA origami, hundreds of “staple” DNA oligonucleotides fold a long “scaffold” DNA single strand into complex, heavily intertwined objects stabilized by thousands of base pairs. DNA origami forms in one-pot reactions along defined pathways (17) in a cooperative folding transition that can occur isothermally and that appears to require nucleation (18). Currently, it is not clear how to rationally choose sequences or strand routing strategies over others and how these choices affect yield and could be used to steer assembly toward desired target conditions. The conditions at which a particular design could be produced optimally currently cannot be predicted. Thus, DNA nanoengineers rely on experience and use inefficient iterative procedures to design and refine DNA origami. A first principle-based, design-dependent model of DNA origami folding that can make quantitatively correct predictions would offer the possibility to rationalize design and production procedures, therefore moving the field beyond tedious trial and error-based approaches. To cope with the compositional and structural complexity in DNA origami, a predictive model of folding will likely need to be coarse-grained, but it is currently not clear what level of detail would be required to correctly describe DNA origami folding. The stochastic Markov chain operating on the level of double-helical domains, as described by Turberfield and co-workers (17, 19), may provide a viable starting point here.

The goal of this work is to provide an experimental reference dataset of the kinetics of a multilayer DNA origami folding to enable theorists to make appropriate approximations and to calibrate rate constants to quantitatively describe the manifold reactions occurring during folding of a DNA origami. Rather than studying many different DNA origami designs in less detail, we decided to investigate the behavior of a single DNA origami test object with the greatest experimental detail that we could realize at this point. To this end, we prepared thousands of distinct self-assembly reactions

to reveal the folding kinetics of every staple strand during constant-temperature assembly of the test object. As a result, we obtained a massive dataset that, through its sheer detail, should achieve general importance, in the sense that a predictive model that can correctly capture the many aspects displayed in our reference dataset will likely have generally predictive powers.

By tradition in the field, researchers predominantly use temperature-ramp annealing to self-assemble DNA origami objects. Informed by this tradition, previous folding studies monitored aspects of DNA origami folding during annealing (20) or investigated outcomes after annealing (17). However, chemical reaction rates become accessible only by monitoring reactions as a function of time at constant temperature since reaction rates themselves depend on temperature. Fortunately, the temperature-ramp annealing is not necessary; DNA origami objects also self-assemble at constant temperature (18) (so far, we have found no exception from this rule), which allows the undertaking of quantitative studies of folding to extract kinetic rate constants.

RESULTS

Experiment design

To carry out our study, we developed two complementary assays: one that is sensitive to the proximity of the 5' and 3' termini of the same strand (“intrastrand terminal proximity folding assay”) and one that probes the proximity of two selected termini of two different strands (“interstrand terminal proximity folding assay”). We applied these assays in several thousand distinct self-assembly reactions to reveal the folding kinetics of every staple strand during constant-temperature assembly of a multilayer DNA origami test object. We chose a multilayer DNA origami object with a brick-like shape to perform our folding studies (Fig. 1A and fig S1). Variants of this object have been investigated previously in folding studies (18, 21) and defect assays (22) and used as test sample to advance DNA origami processing methods (23, 24). In this object, 42 helices are packed in a honeycomb-type lattice. The object comprises 5880 base pairs formed between 140 staple strands and 1 scaffold strand. The design encodes 747 double-helical DNA domains delimited either by cross-overs or staple strand backbone nicks. On average, each staple sequence has ~5.3 segments forming double-helical DNA domains with the scaffold. Furthermore, the design defines 1132 staple crossovers

Copyright © 2019
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

Physik Department and Institute for Advanced Study, Technische Universität München, Garching near Munich, Germany.

*Corresponding author. Email: dietz@tum.de

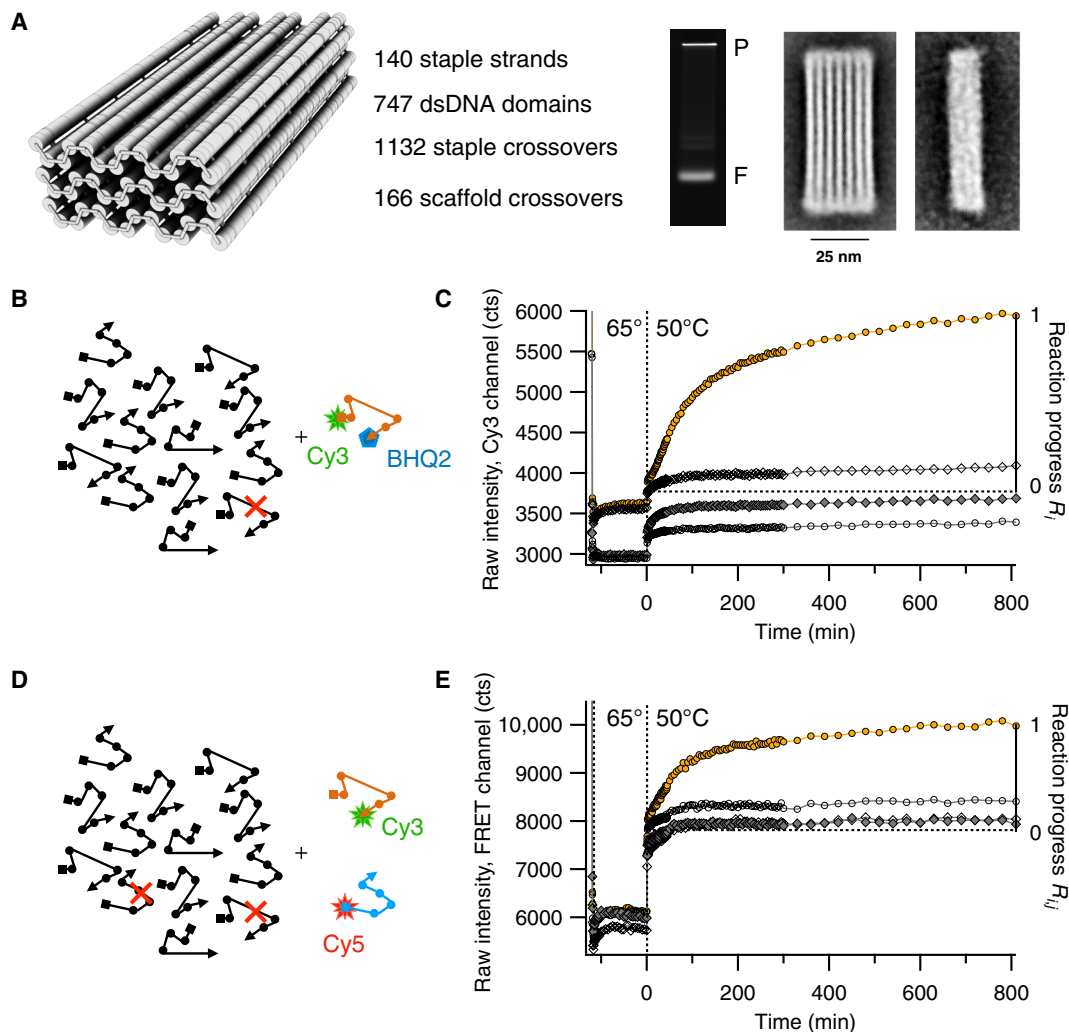


Fig. 1. Illustration of the DNA origami test object and the fluorometric folding assays with sample data. (A) Left: CanDo model (30,31) of the investigated DNA origami test object, a 42-helix bundle, designed in honeycomb lattice consisting of 140 staple strands. Middle: Image of an agarose gel lane on which folding products of the 42-helix bundle were electrophoresed. F, folded species; P, gel pocket. Right: Class averages of negative-staining TEM images. dsDNA, double-stranded DNA. (B) Illustration of the intrastrand terminal proximity folding assay. Segmented arrows indicate individual staple strands; segments illustrate sequence stretches that will form continuous double-helical domains with the scaffold. Cy3, cyanine 3; BHQ2, Black Hole Quencher 2. (C) Raw fluorescent data of sample folding reactions of the intrastrand assay. (D) Illustration of the interstrand terminal proximity folding assay. Cy5, cyanine 5. (E) Raw fluorescent data of sample folding reactions of the interstrand assay. FRET, fluorescence resonance energy transfer. (C and E) Orange circles, complete folding reaction mixtures; gray symbols, folding reaction mixtures missing essential components (empty circles, scaffold missing; empty diamonds, full set of staple strands missing; gray diamonds, full set of staple strands and scaffold missing). The fluorescently labeled staple strands are always included. cts, counts.

and 166 scaffold crossovers. The object self-assembles with satisfactory yield and few by-products when subjected to constant temperature incubation at 50°C in the presence of 20 mM MgCl₂, as seen by agarose gel electrophoresis and direct imaging with negative-staining transmission electron microscopy (TEM) (Fig. 1A).

In the intrastrand terminal proximity folding assay, the scaffold strand and all staple strands as defined in the DNA origami design, except for one staple strand with index i , were all mixed in one pot. We then supplemented the mixture with a variant of the thus far missing strand i carrying a cyanine 3 (Cy3) fluorescent dye modification on the 5' terminus and a Black Hole Quencher (BHQ2) on the 3' terminus. With such double modification at its termini, the strand becomes a molecular beacon (25) that is sensitive to conformational changes of the strand during folding at constant tempera-

ture (Table 1). Since our test object contains 140 unique staple strands, we prepared 140 distinct folding reaction mixtures, where, in each reaction, a different staple strand was doubly chemically modified. Three independent replicates of each reaction were prepared and analyzed.

In the interstrand terminal proximity folding assay, the scaffold strand and all staple strands defined in the design, except for two strands with index i and j , are mixed in one pot. We then supplement the mixture with modified variants of the two missing strands i and j , where one modified strand carries a terminal acceptor dye [cyanine 5 (Cy5)] and the other carries a terminal donor dye (Cy3) modification. In this setup, the two dyes can report via fluorescence resonance energy transfer (FRET) (26, 27) on the evolution of the distance of the two labeled termini of the strands i and j during

Table 1. Temperature settings for the folding reactions. Each folding reaction consisted of three temperature plateaus. The first step serves instrument control; the second step serves strand annealing, while the folding takes place during the 50°C incubation.

Temperature (°C)	Time (min)
25	2.25
65	120
50	840

folding. The strong and short-ranged distance dependence of FRET can ensure that signal is produced only in states where the two termini are actually colocalized on the same scaffold molecule. A similar setup has been applied previously to thermally map a small subset of strands in a DNA origami object during annealing (20). For our experiments, we obtained five variants of the entire set of 140 staple strands: unlabeled; labeled with Cy5 on the 5' terminus or on the 3' terminus, respectively; and labeled with Cy3 on the 5' terminus or on the 3' terminus, respectively. We decided to investigate 3172 unique combinations of donor and acceptor sites. One selection criterion was whether the terminal dye-to-dye distance $d_{i,j}$ in the fully folded state would be shorter than 5 nm on the basis of the designed geometry of the DNA origami brick. In addition, we ensured that each strand terminus appears in separate reaction mixtures at least once labeled with a donor dye and once labeled with an acceptor dye (see fig. S2 for statistics on reaction pairings and strand coverage). We identify label pairings in the reaction mixtures via two indices i and j with the following conventions: Indices run from 0 to 279 where 0 to 139 denote 5' label positions whereas 140 to 279 denote 3' label positions. The first versus the second index refers to the donor versus acceptor carrier, respectively. Four independent replicates of each reaction mixture were prepared to give a total of 12,688 separately prepared folding reaction mixtures. To facilitate pipetting of the large number of reaction mixtures (fig. S3), each including 141 distinct DNA strands, the reaction mixtures actually contained unlabeled and labeled variants of the same staple strand in 1:1 stoichiometry, with the total strand concentration matching those of all the other strands in the mixture.

The intrastrand and the interstrand terminal proximity folding reaction mixtures were all analyzed in four 96-well quantitative polymerase chain reaction (qPCR) thermal cycling devices. We subjected the reaction mixtures to a temperature jump protocol with two phases. First, a denaturation phase at 65°C breaks base-paired structures that could have formed during reaction mixture preparation. Self-assembly of the DNA origami objects then proceeds in the folding phase at a constant temperature of 50°C. During incubation, the fluorescence intensity is recorded continuously, with typical signals obtained from the intrastrand and the interstrand terminal proximity folding assay shown exemplarily in Fig. 1 (C and E, respectively). The signals obtained from the full folding reaction mixtures showed pronounced increases in fluorescence over time, which eventually leveled out, whereas control reaction mixtures lacking key components such as the scaffold strand did not show such signal increases. We subtracted background and normalized the raw intensity data curves obtained from each reaction such that the intensity signal upon entering the folding phase is set to 0 ("folding start"), whereas it is set to 1 ("folding completion") after ending the reaction monitoring. The thus treated

curves represent the reaction progress R_i and $R_{i,j}$ in the intrastrand and interstrand terminal proximity folding reaction mixtures, respectively.

The sequence of events and two-phase reaction kinetics

We averaged the data curves obtained from the replicate reaction mixtures. As a result, we obtained 140 reaction progress curves from the intrastrand proximity folding assay and 2781 curves containing at least two successful replicates in the case of the interstrand terminal proximity folding assay. For the intrastrand assay, we find that the time to reach folding completion in the reaction mixtures varied from ~20 to ~200 min. The shape of the individual data curves obtained from the intrastrand assay and also from the interstrand assay may be described satisfactorily in most cases by a superposition of two exponential functions (Fig. 2B and fig. S3). The individual data curves may be sorted according to the mean folding time (i.e., mean time of reaching signal saturation; fig. S4), thus revealing a defined sequence of events (Fig. 2, A and B). For the interstrand assay, the time to reach folding completion per reaction mixture covers similar time spans, as seen in the intrastrand assay, and the data may also be sorted according to a mean folding time (Fig. 2C). Therefore, similar to the intrastrand assay, the interstrand assay also reveals a defined sequence of events during DNA origami folding. The shape of the individual data curves obtained from the interstrand assay may also be described satisfactorily by a superposition of two exponential functions (see Fig. 2, E and F, and fig. S5).

Hence, within the resolution of the data, each folding reaction displays approximately two-phase kinetics (reaction progress = $1 - \{A_1 \cdot \exp[-(t - t_0)/\tau_1] + A_2 \cdot \exp[-(t - t_0)/\tau_2]\}$). For the intrastrand assay, we find that the time constants τ for the two processes extracted from fitting the individual data curves both increase linearly with the staple index when the data are sorted according to the mean folding time (Fig. 2G). The fast process occurs with time constants ranging from ~2 to ~50 min, whereas the slow process time constant covers the range from ~100 to ~350 min. By contrast, the time constants extracted from fitting the interstrand assay data display quite different behavior: The time constants cluster on three (potentially four) distinct levels in a plot in which the reaction mixtures are sorted according to the mean folding time (Fig. 2H). The three levels are ~0.5, ~30, and ~300 min.

Interpreting the interstrand terminal proximity data

In the interstrand assay, signal arises when the labeled termini of two separate strands come into close proximity. Since all strands are represented in multiple terminal labeling variations and in different combinations with other labeled strands, it is possible to directly extract information about the relative order of events. For example, the data curves obtained from a set of reactions featuring a particular staple strand labeled with a donor but in which different staple strands were labeled with an acceptor can all overlap closely (Fig. 2E). In such cases, we conclude that the strand with the donor dye incorporated with the same rate as or a slower rate than the staples carrying the acceptor dyes. We can contrast this situation with cases in which the data curves obtained from a set of mixtures with a particular donor-labeled staple, but varying acceptor staple strands, do not overlap at all (Fig. 2F). Here, we can conclude that the donor dye incorporated at a faster rate than the acceptor dye-carrying strands did; therefore, the relative sequence at which the different acceptor strand termini came into proximity of the donor strand

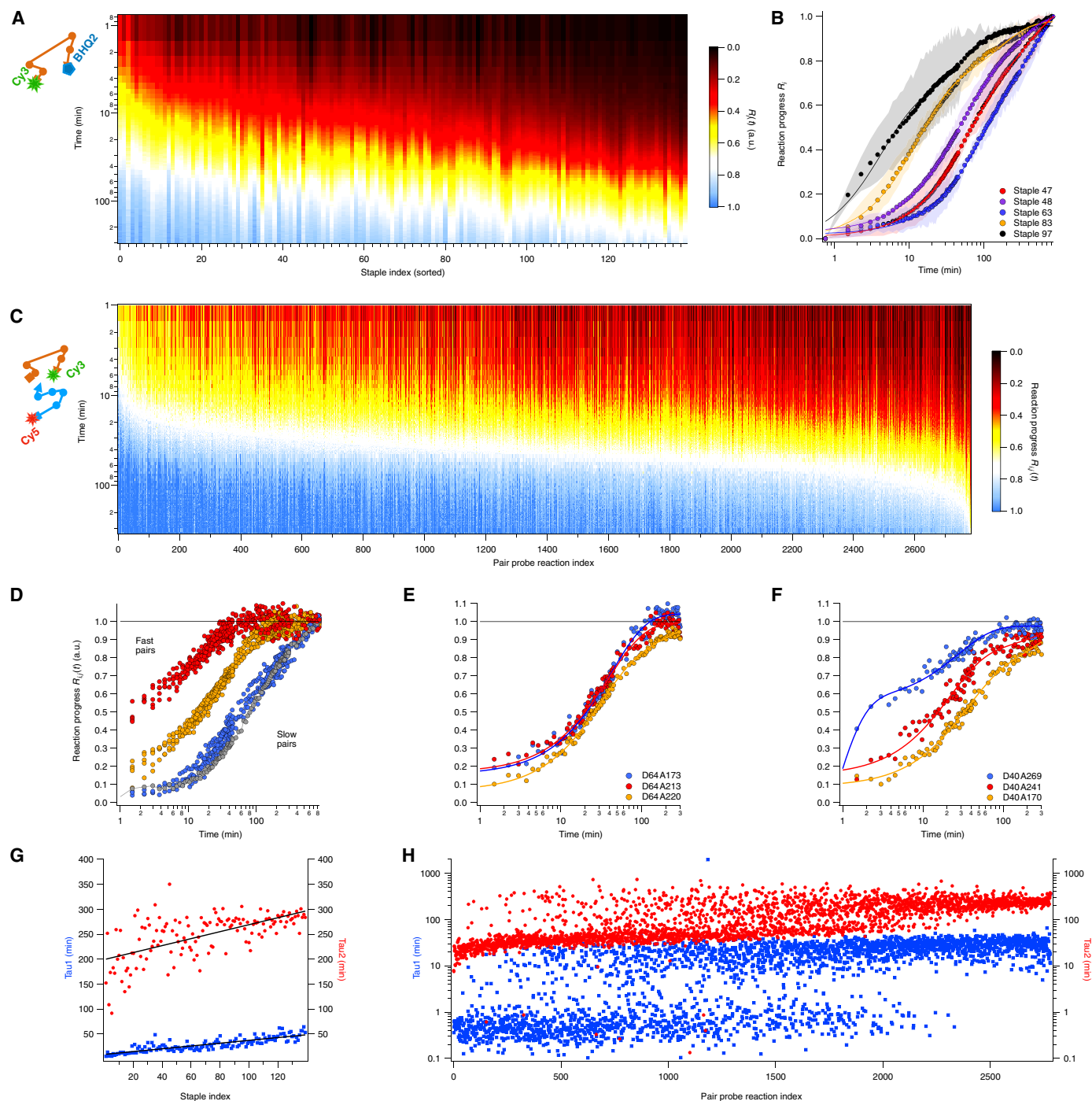


Fig. 2. Time-dependent reaction progress and fit parameters. (A) Measured reaction progress data obtained with the intrastrand folding assay for the 42-helix bundle test objects, sequence permutation-0. Each column is the average of three replicates of fluorescence time traces obtained for a particularly labeled folding reaction mixture. Traces are sorted according to the mean folding time. a.u., arbitrary units. (B) Sample curves obtained from the intrastrand assay to illustrate the variety of incorporation behavior. Shaded areas indicate the SD of the averaged data. (C) Measured reaction progress data obtained with the interstrand folding assay. Data are sorted according to the time needed to reach 80% of the maximum intensity. (D) Sample traces obtained with the interstrand assay to illustrate the variety of incorporation behavior. Gray circles, a single trace to illustrate noise level. (E) A set of sample pair probe measurements with the same staple strand labeled with a Cy3 and varying Cy5-labeled staple strands in which the donor-carrying strand limits the kinetics of signal evolution. (F) A set of sample pair probe measurements in which the kinetics of the acceptor-carrying strand can be discriminated and sorted (relative to the donor carrier). (G) Time constants τ_{1} (blue) and τ_{2} (red) obtained by fitting double exponentials to the intrastrand assay data. (H) Time constants τ_{1} (blue) and τ_{2} (red) values obtained by fitting double-exponential fits to the interstrand assay data.

terminus can be directly read off from the respective set of data curves (Fig. 2F).

To retrieve the absolute single-strand terminal incorporation dynamics from the measured pair probe dynamics, we fitted a global incorporation probability model to the data (Fig. 3, A to C, and fig. S4). To this end, we assume that the measured reaction progress $R_{ij}(t)$ is proportional to the product of single-strand terminal incorporation probabilities $P_i(t)$ and $P_j(t)$ at all times, which is equivalent to saying that, at any time point, the incorporation probabilities were independent of each other. Since DNA origami folding globally shows signs of cooperativity, this description represents an oversimplification. However, one appealing aspect of our description is that it is highly overdetermined by our data: At each time point, we fit in total 280 P_i to the ~ 2800 experimentally determined $R_{ij}(t)$ values. The thus computed single terminal incorporation probability curves, recovered separately for the two termini of each strand (Fig. 3B versus Fig. 3C), still exhibit clear differences in the time needed to reach folding completion. Hence, in addition, these data can be sorted according to the mean folding time and reflect a defined sequence of events. The shape of the resulting single-strand terminal incorporation traces can also be described by double exponential function (fig. S6). We also inverted the model and computed the pair probe reaction progress R_{ij} from the fitted single terminal incorporation probabilities P_i (fig. S5). The thus backward-generated R_{ij} appears as a smoothed version of the measured R_{ij} .

5' and 3' terminal strand segments display independent dynamics

One important observation is that the 5' strand terminal data display a completely different sequence of events than the 3' terminal data, which becomes apparent by sorting the 5' terminal incorporation probability curves according to the mean folding time and then displaying the data curves obtained from the corresponding 3' strand termini with the same index sorting (Fig. 3B versus Fig. 3C). Whereas the 5' dataset shows a clear gradient from early to late incorporating termini, the 3' dataset has no apparent order. There is very little correlation between the mean folding times obtained for the same staple strands when the dynamics are probed via a 5' label versus a 3' label (Fig. 3D). Therefore, the data indicate that strand termini incorporate independently of each other during DNA origami folding. Apart from the lack of correlation, we note that the 3' terminal dataset reflects slightly accelerated incorporation dynamics compared to the 5' dataset (Fig. 3E). We will discuss a possible explanation for this finding further below.

Segment truncation experiments support sequence of events

To test the validity of the incorporation sequence determined for individual strand termini as extracted by fitting the measured interstrand terminal proximity folding data, we performed complementary time-resolved folding experiments in which subsets of staple strands were replaced with variants whose 5' terminal segments were truncated. Specifically, we selected the 20 fastest and slowest incorporating terminal segments, as determined in Fig. 3B, and synthesized variants of the corresponding staple strands that lacked these segments. The length of the truncated segments varied from 7 to 14 bases with an average of 8 bases. We studied the folding behavior at 50°C in a time-resolved fashion using a reaction quench followed by agarose gel electrophoretic shift assays, as described

previously (Fig. 3F and fig. S7) (18). With the default staple set, the unfolded state disappears immediately upon incubation at 50°C and converts into an intermediate, compacted species (Fig. 3F, top). The fully folded species emerges shortly after at around 30 min of incubation. When we deleted the 20 fastest incorporating terminal segments, however, the intermediate no longer occurs. The unfolded state prevails for hours, and the fully folded state emerges only with substantial hour-long delay (Fig. 3F, middle). By contrast, when we deleted the 20 slowest incorporating terminal segments (Fig. 3F, bottom), the initial folding trajectory is again very similar to the behavior seen for the default staple set: The unfolded species vanishes immediately, and a compacted intermediate appears. In contrast to the default staple set, however, when deleting the “late” segments, the intermediate now persists for much longer times and the appearance of the fully folded state is delayed. Hence, deleting “early” segments eliminated a distinct species that also occurs early during folding, whereas deleting late segments caused delays in late phases of folding. Therefore, the truncation experiments, which we also performed for 3' terminal truncations (fig. S7), support the notion that the sequence of terminal incorporation events that we inferred from the interstrand folding data by using a model correctly captures the actual sequence of events.

The interstrand folding data report on phenomena that alter the distance between two separate strand termini. In the complementary intrastrand folding assay data, the signal is sensitive to the conformational changes that an entire staple strand undergoes during folding. We found above that the 5' and 3' terminal segments show independent incorporation dynamics. Hence, we may expect that our dataset obtained from the intrastrand terminal proximity folding assay will display dynamics on similar time scales but not necessarily with strong correlation to either of the two terminal segment incorporation datasets, which is what we observe (Fig. 4A). Overall, the average mean folding time of entire strands as seen in the intrastrand assay data is ~ 40 min, whereas it is ~ 35 min for the 5' terminal segments and ~ 30 min for the 3' terminal segments, as retrieved from the interstrand assay data.

Incorporation times weakly depend on sequence

We sought to identify design features that correlate with the experimentally determined sequence of events. For example, as speculated previously (18), staple strands or staple segments with high free energy of hybridization (equal to high thermal stability) could potentially incorporate earlier than other strands. However, we see little correlation between the free energy of hybridization of entire staple strands with their mean folding time (Fig. 4B), and there is also little correlation between the free energy of hybridization of the terminal staple segments with their respective mean folding time (Fig. 4C). Hence, it is not the staple (segment) sequence alone that independently determines when a particular staple strand or staple segment will incorporate. To confirm this finding, we performed another series of folding experiments in which we truncated segments in subsets of staple strands and investigated the consequences on the folding behavior (Fig. 4E and fig. S7). Deleting terminal staple segments with either the largest or the smallest free energy of hybridization resulted in very similar effects on the folding pathway: Formation of the early compacted intermediate state and conversion to the fully folded state were much delayed compared to the default staple set. We also simply eliminated all the long terminal staple segments (11 segments each with 21 bases), again with very similar results

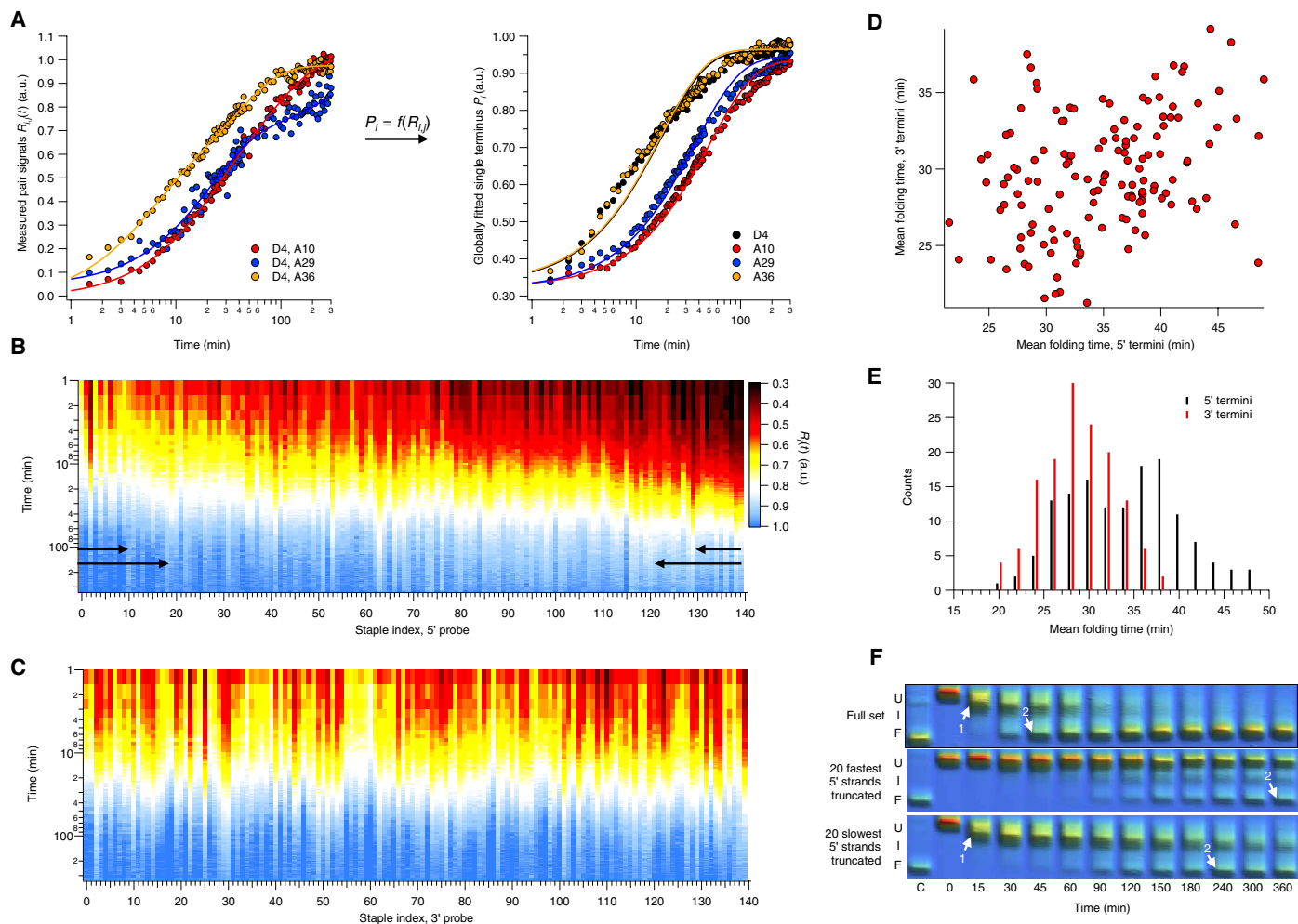


Fig. 3. Single segment binding probabilities. (A) Sample time traces illustrating the inference of single-strand terminal binding probabilities (right) from the pair probe interstrand assay reaction progress data (left). (B and C) Inferred single-strand 5' and 3' terminal binding probabilities, respectively. Data are sorted according to the mean folding time of the 5' termini. Arrows indicate the 10 (20) fastest (slowest) staple strand termini. (D) The mean folding time of the 3' terminal staple strand segments versus 5' termini mean folding time. (E) Histogram of the distribution of mean folding time for the 5' termini (black) and the 3' termini (red). (F) Time-resolved gel electrophoretic folding analysis of the default folding reaction (top), with the 20 fastest (middle) and 20 slowest (bottom) 5' staple strand termini truncated. Three-dimensional (3D) representation was chosen to increase visibility of weak bands. Arrows indicate the occurrence of a compacted intermediate (1) and folded species (2).

(Fig. 4E, bottom). Since the effects of deleting high free-energy or low free-energy segments can hardly be distinguished, we conclude that the folding pathway does not strongly depend on the strand segment binding strength.

To further investigate the relevance of the particular staple (segment) sequences for the sequence of events along the folding pathway, we repeated the intrastrand terminal proximity folding assay with a circular scaffold sequence permutation of the DNA origami test brick design (figs. S3 and S8), leading to two independent datasets: one for the original design (“permutation-0”) and one for the new sequence variant (“permutation-1”). In the permutation-1 object, all the staple strands had completely different sequences compared to the permutation-0 object, but the internal topology of connectivity remained identical compared to permutation-0. The mean folding times per staple strand that we determined for the two DNA origami brick variants correlated more (Fig. 4D) than, e.g., the mean folding times between the terminal segments of the object variant (Fig. 3D). We also performed “knockout” experiments in

which we omitted early or late incorporating strands, as determined by the interstrand assay for the original permutation-0 object, but the corresponding strand identities were omitted in folding reaction mixtures for both permutation-0 and permutation-1 objects (fig. S9). The omissions had a similar impact on the folding behavior (Fig. 4F); that is, when early strands were omitted, the intermediate, as discussed above (cf. Fig. 3F), was no longer formed and folding was delayed for both sequence permutations. By contrast, when the late strands were omitted, the intermediate did form, also for both sequence permutations (Fig. 4F). These data show that both the local sequences and internal topology of connectivity are two important factors for assembly behavior.

Mapping the sequence of events to three-dimensional shape and chain connectivity

We projected our experimentally determined sequence of events onto a three-dimensional (3D) layer-by-layer model of the DNA origami test brick to determine whether any correlations may be identified

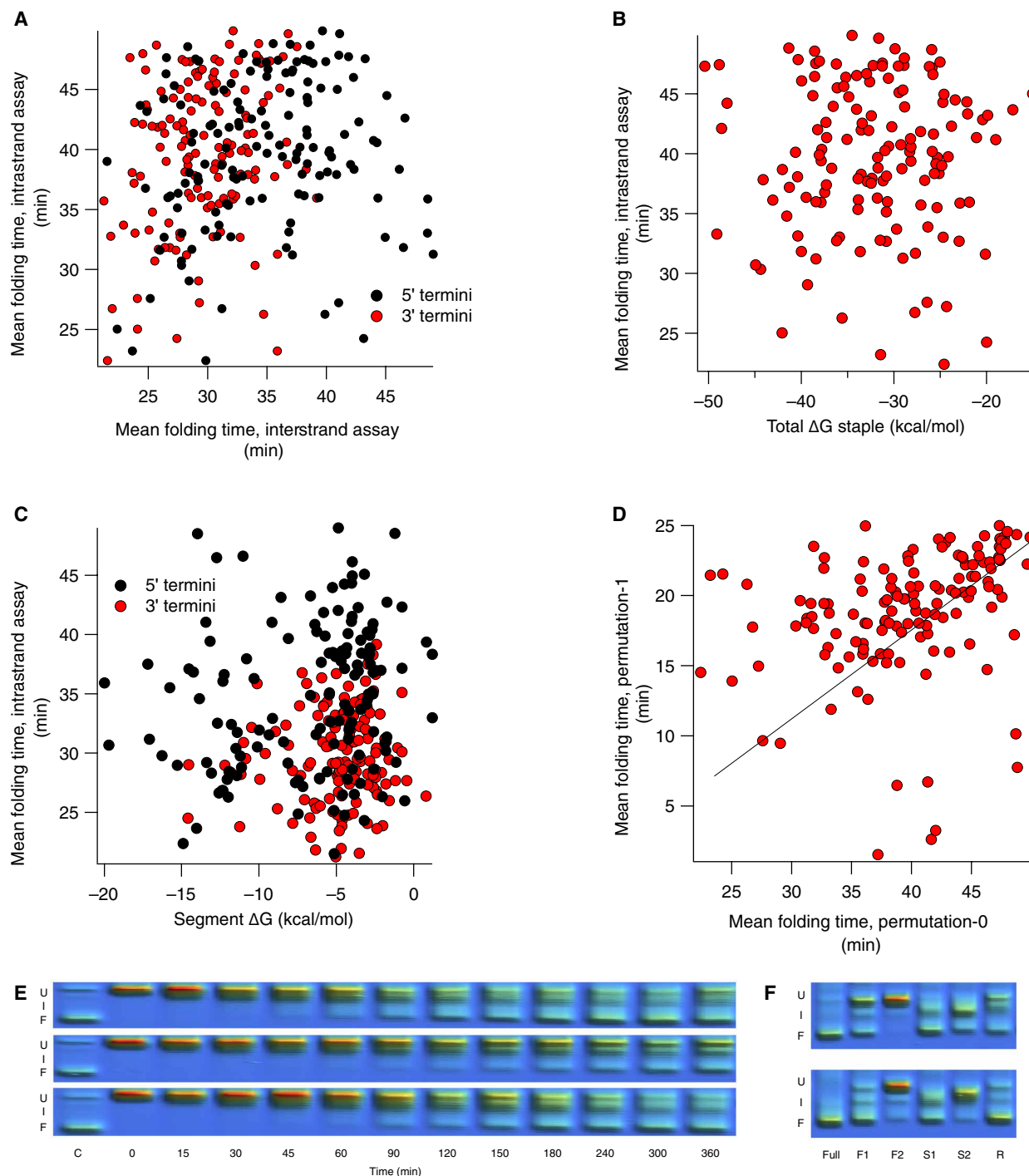


Fig. 4. Analysis of the mean folding times. (A) The intrastrand assay mean folding time versus the mean folding time resolved from the interstrand assay including 5' and 3' termini incorporation. (B) The intrastrand assay mean folding time versus the total binding energy of the staple strands. (C) The mean folding time of 5' and 3' terminal staple strands resolved from the interstrand assay versus the binding energy of the terminal staple strand segments. (D) The mean folding time of the intrastrand assay mean folding times obtained for circular sequence permutation-1 versus permutation-0 objects. The solid line is merely a guide to the eye. (E) Time-resolved gel electrophoresis of folding reactions where the 20 5' terminal staple strand segments with the highest (top) and lowest (middle) binding energy have been truncated. Lower gel shows a folding reaction missing the 11 longest (21 bases) staple strand segments. (F) Gel electrophoresis of permutation-0 (top) and permutation-1 (bottom) after 6 hours of incubation at 50°C. Folding reactions have been prepared without the 10 (20) fastest (F1 and F2) and slowest (S1 and S2) staple strands, respectively, and with 10 randomly omitted staple strands.

between strand segment location and 3D structure (Fig. 5A). We cannot discern global spatial patterns such as strands from the brick's interior binding systematically earlier or later compared to strands from the brick's exterior. Although, at first impulse, it seems natural to search for particular early or late coming "hotspots" or other correlations between the object's shape and the sequence of events, the lack of such correlations may not surprise too much: The fully compacted 3D structure only emerges after all strands have successfully incorporated into the structure. The "exterior" or the "interior" of the object does not really exist during much of the folding.

By contrast, a pattern becomes apparent when we project our mean folding time data onto a graph in which the scaffold strand is shown as a ring and a subset of strand crossovers is drawn as arcs that bridge particular scaffold base indices (Fig. 5B). Only the crossovers between the terminal staple strand segment and the next-to-terminal segment on the same staple are drawn. The terminal segments from which the arcs emanate are marked as discs. The arcs and discs were colored according to the experimentally measured folding time. Subsets of arcs and discs cluster in curvature and color. In these clusters, there are terminal segments adjacent in

the scaffold's primary sequence that have similar incorporation times. The crossovers emanating from these segments point across the ring to distant scaffold regions featuring groups of other terminal strand segments that incorporate at similar times like the segments from which the crossovers emanated. This indicates that certain distant regions on the scaffold that, once connected by one or more crossovers, fold cooperatively or at least in close succession. This finding makes sense by considering the entropic cost of forming scaffold loops: Once one big loop is formed that bridges to distant scaffold sites, other staple strand crossovers in the vicinity may form more easily in succession. We note that when the ring graphs are drawn separately for the 5' terminal segments and the 3' terminal segments, it becomes apparent that crossovers emanating from 5' terminal segments close systematically longer loops than those emanating from the 3' terminal segments (fig. S10). The larger 5' looping penalties may provide an explanation for the experimental observation that the mean folding times for the 5' termini were systematically longer than those of the 3' termini (as discussed in Fig. 3D). The difference in loop length for the 5' and 3' termini arises coincidentally from the particular staple and scaffold routing in the design.

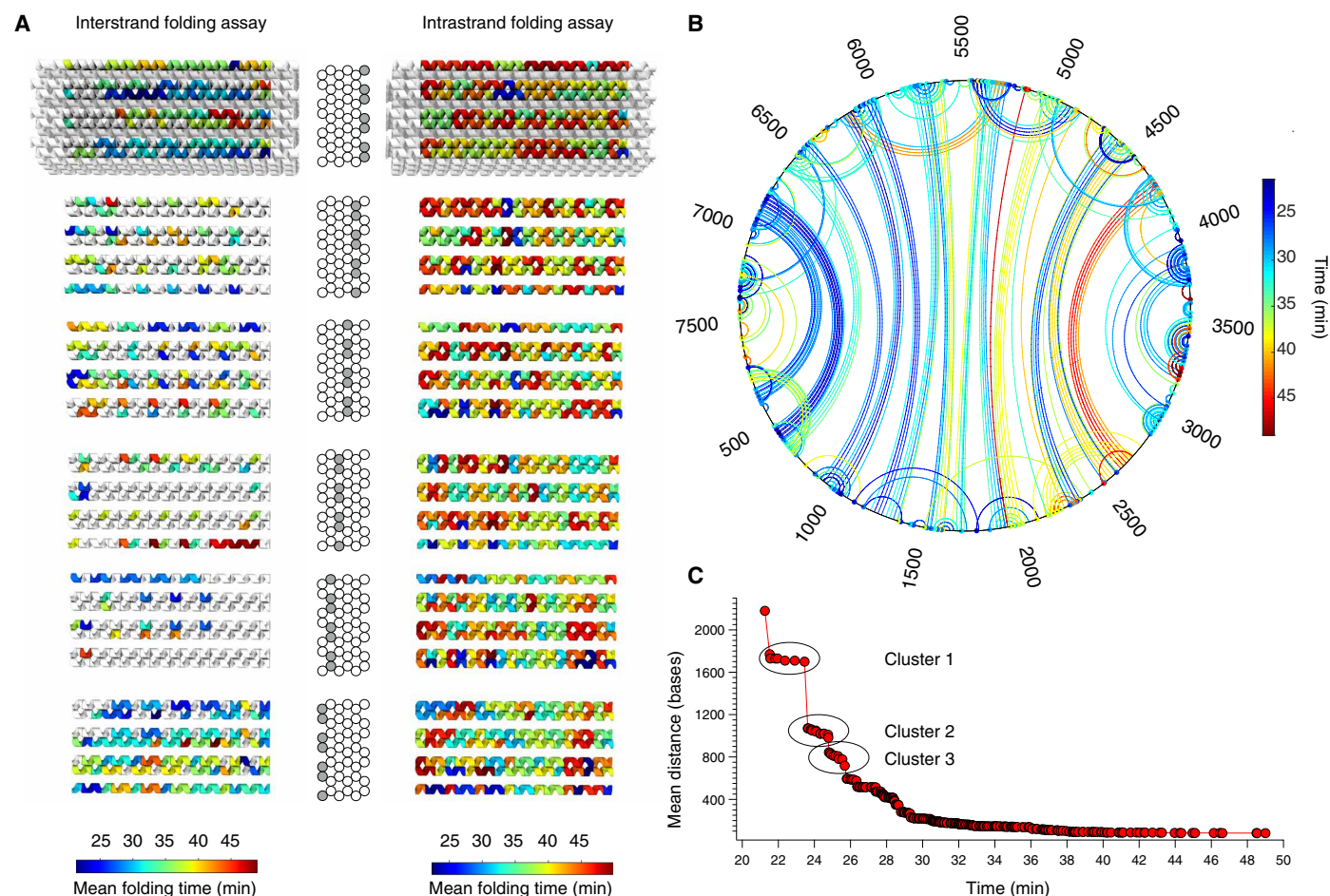


Fig. 5. DNA origami strand routing and folding pathway. (A) Layer-by-layer 3D models of the 42-helix bundle, colored according to the experimentally measured mean folding times of the staple strand segments. (B) Arcs denote staple strand crossover bridging the terminal staple strand segments to the next-to-terminal segment. Discs indicate the terminal position. Numbers give scaffold strand base indices. Color code indicates the mean folding time of the terminal staple strand segment resolved from the interstrand assay. (C) Estimation of the evolution of the average interbase distance on the scaffold over time. Each circle indicates a new compaction event produced by binding of a new staple segment at the mean folding time obtained from the interstrand assay.

To visualize the progression of compaction associated with crossover formation, we computed the mean shortest path length between any two bases on the scaffold as a function of time (Fig 5C and fig. S10). The compaction proceeds initially in several distinct steps, which may be attributed to binding of multiple segments that contribute crossovers that connect the same scaffold loop region as seen in the ring graph. The mean distance decreases rapidly from initially >2000 bases to ~50 to 100 bases. Since these are the distances that would need to be closed when forming new crossovers, we see that the entropic penalties associated with forming loops in the DNA origami structure vanish rapidly.

DISCUSSION

Together, we uncovered with molecular resolution the sequence of events along the folding pathway of our DNA origami test object. Each full strand and its two terminal segments, as defined in the design, can now be annotated with kinetic parameters (fig. S1). Mapping the sequence of events onto a ring graph presentation revealed nucleating staple segment clusters connected by long-distance crossovers and their location on the scaffold primary sequence. Both the intrastrand and interstrand folding progress data could be described with two-phase exponential kinetics for most of the strands. From the interstrand folding assay, we learned that the two termini of the same strand display independent binding kinetics. This observation indicates that folding occurs via individual double-stranded DNA (dsDNA) domain formation, rather than entire staple strand binding. We found little to no correlation between sequence properties of the domains and their binding kinetics, indicating that cooperative, nonlocal effects determine which domain would bind at which time. By applying the intrastrand folding assay to two circular sequence permutations of the same DNA origami test design, we found that the two respective sequences of binding events, i.e., the folding pathways, correlated strongly. This observation supports the view that the folding pathway does not depend so much on the local sequence but on cooperative effects induced by the global topology of connectivity. Auxiliary strand segment truncation experiments showed that the folding proceeds via a compacted intermediate formed by a set of nucleating double-helical DNA domains. If these cannot be formed because the respective strand segments are missing, then folding may still occur, but it proceeds via another, slower pathway. Previously, we stipulated that the nuclei for folding could consist of the dsDNA domains with the highest independent thermal stability as defined by design (18), but analysis of our current folding data with respect to sequence properties shows that this view is too simplistic.

An inconvenient consequence of the observed folding behavior is that the sequence of events cannot be easily tied to (and, thus, also not controlled by) simple, local design rules. A description that considers the topology of connectivity together with the domain sequences, their edge interactions, and possibly also other factors such as off-target interactions will be necessary. Our real-time dataset with its thousands of curves (tables S1 to S5) provides a comprehensive set of constraints to challenge microscopic models of DNA origami assembly. Our experiments also suggest DNA origami as a playground for studying principles of biomolecular self-assembly phenomena, in loose analogy to the idea of using colloids as big atoms (28). Cooperativity and defined folding pathways with compacted intermediates play important roles not only in DNA origami

but also, for example, in protein folding, despite the fact that proteins typically fold from single chains, whereas DNA origami folding is arguably more complex with its hundreds of intertwined chains. Resolving the sequence of side-chain contact formation along a protein folding pathway, which would be equivalent to resolving the sequence of strand segment binding as described here for DNA origami, can currently be performed only in silico (29). Whereas the primary sequence and the global topology of connectivity can be uncoupled in DNA origami, they are tightly coupled for proteins. Hence, even small interventions such as mutations may affect the protein's global fold. By contrast, we could modify the strength of local interactions by sequence permutations without affecting the global topology, and we could perform surgical operations such as shutting off particular local interactions through strand truncations and study the effects thus produced on the folding pathway.

MATERIALS AND METHODS

Molecular self-assembly with scaffolded DNA origami

The reaction mixtures for the folding of our test structures contained 20 nM scaffold and each staple strand at 100 nM concentration. The scaffold and staple strands were mixed with reaction buffer containing 5 mM Tris, 1 mM EDTA, 5 mM NaCl (pH 8), and 20 mM MgCl₂.

All oligonucleotides were synthesized by Eurofins Genomics (Ebersberg, Germany). All fluorometric assays were observed using a real-time PCR (RT-PCR) system (Agilent Mx3000P), while the folding reactions for gel electrophoresis were incubated in Tetrad thermal cycler (MJ Research, now Bio-Rad). The folding reactions were subjected to the thermal profile described in Table 1.

Gel electrophoresis and image processing

Gel electrophoresis of DNA origami structures was performed using 2% agarose gels including 11 mM MgCl₂ and ethidium bromide. Structures folded for cryogenic time-resolved experiments were thawed at 30°C for 15 min in a Tetrad thermocycler before gel loading. A voltage of 75 V was applied to the ice water-cooled gels for 3 hours. Afterward, gels were scanned using a Typhoon FLA 9500 laser scanner (GE Healthcare), and the 16-bit tagged image file format files were processed using Photoshop CS6 and analyzed using Igor Pro 7.

TEM and image processing

Folded products were incubated for 45 s on CU400 TEM grids (Science Services, Munich, Germany) and stained using 2% uranyl formate solution containing 25 mM sodium hydroxide. Images were acquired on a Philips CM100 at 100 kV using an AMT 4-megapixel charge-coupled device camera at 28,500× magnification. Particles were manually picked and averaged using Xmipp mlf_2Dalign routine.

Ensemble FRET measurements using RT-PCR devices

All fluorometric assays were measured in four identical Mx3005P RT-PCR systems (Agilent). Each folding reaction (50 µl) was pipetted into 96-well plates, and each column was sealed individually using 8-cap strips. Cy3, Cy5, and FRET emission intensities were acquired for the interstrand assay, while only Cy3 was recorded for the intra-strand assay. Data were processed and analyzed using MATLAB 2017 and Igor Pro 7.

Extraction of the binding probabilities from the interstrand assay

The set of 2781 individual traces $P_{ij}(t)$ was used to calculate the binding probability of each staple strand terminus i using the nonlinear least-squares solver implemented in MATLAB for each data point individually, assuming that $R_{ij}(t) = P_i(t) * P_j(t)$.

Calculation of the mean distance on the scaffold

MATLAB was used to calculate the mean distance $R_g(t)$ between the bases of the scaffold using the formula supplied in fig. S16. The 7560 × 7560 distance matrix was updated with every binding event of a staple strand terminal segment, assuming that these events short-circuit the base-to-base distance on the scaffold by forming a crossover to the neighboring staple segment.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/5/5/eaaw1412/DC1>

- Fig. S1. Design diagram of the brick-like test object variant 0, prepared using cadnano.
 Fig. S2. Schematic representation of the interstrand proximity assay with measured pairs.
 Fig. S3. Data obtained from the intrastrand proximity assay and double-exponential fits.
 Fig. S4. Calculation of mean folding time and binding probabilities.
 Fig. S5. Data obtained from the interstrand proximity assay and double-exponential fits.
 Fig. S6. Binding probabilities derived from the interstrand proximity assay and double-exponential fits.
 Fig. S7. Laser-scanned images of 2% agarose gels placed in a water bath.
 Fig. S8. Design diagram of the brick-like test object variant 1, prepared using cadnano.
 Fig. S9. Laser-scanned image of 2% agarose gel placed in a water bath.
 Fig. S10. Ring graph representation of the scaffold strand and distance calculation.
 Table S1. Averaged raw data of the intrastrand proximity assay.
 Table S2. Averaged raw data of the interstrand proximity assay.
 Table S3. Reaction progress data of the intrastrand proximity assay.
 Table S4. Reaction progress data of the interstrand proximity assay.
 Table S5. Staple strand sequences of the used DNA origami test objects.
 References (32, 33)

REFERENCES AND NOTES

- S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W. M. Shih, Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
- H. Dietz, S. M. Douglas, W. M. Shih, Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725–730 (2009).
- E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. P. Oliveira, J. S. Pedersen, V. Birkedal, F. Besenbacher, K. V. Gothelf, J. Kjems, Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76 (2009).
- X.-c. Bai, T. G. Martin, S. H. W. Scheres, H. Dietz, Cryo-EM structure of a 3D DNA-origami object. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20012–20017 (2012).
- F. Hong, F. Zhang, Y. Liu, H. Yan, DNA origami: Scaffolds for creating higher order structures. *Chem. Rev.* **117**, 12584–12640 (2017).
- D. Han, S. Pal, J. Nangreave, Z. Deng, Y. Liu, H. Yan, DNA origami with complex curvatures in three-dimensional space. *Science* **332**, 342–346 (2011).
- R. Veneziano, S. Ratanalert, K. Zhang, F. Zhang, H. Yan, W. Chiu, M. Bathe, Designer nanoscale DNA assemblies programmed from the top down. *Science* **352**, 1534 (2016).
- E. Benson, A. Mohammed, J. Gardell, S. Masich, E. Czeizler, P. Orponen, B. Högberg, DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523**, 441–444 (2015).
- D. Han, S. Pal, Y. Yang, S. Jiang, J. Nangreave, Y. Liu, H. Yan, DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412–1415 (2013).
- T. Gerling, K. F. Wagenbauer, A. M. Neuner, H. Dietz, Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components. *Science* **347**, 1446–1452 (2015).
- J. J. Funke, H. Dietz, Placing molecules with Bohr radius resolution using DNA origami. *Nat. Nanotechnol.* **11**, 47–52 (2016).

- F. Zhang, S. Jiang, S. Wu, Y. Li, C. Mao, Y. Liu, H. Yan, Complex wireframe DNA origami nanostructures with multi-arm junction vertices. *Nat. Nanotechnol.* **10**, 779–784 (2015).
- B. Wei, M. Dai, P. Yin, Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623–626 (2012).
- Y. Ke, L. L. Ong, W. M. Shih, P. Yin, Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
- K. F. Wagenbauer, C. Sigl, H. Dietz, Gigadalton-scale shape-programmable DNA assemblies. *Nature* **552**, 78–83 (2017).
- L. L. Ong, N. Hanikel, O. K. Yaghi, C. Grun, M. T. Strauss, P. Bron, J. Lai-Kee-Him, F. Schueder, B. Wang, P. Wang, J. Y. Kishi, C. Myhrvold, A. Zhu, R. Jungmann, G. Bellot, Y. Ke, P. Yin, Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature* **552**, 72–77 (2017).
- K. E. Dunn, F. Dannenberg, T. E. Ouldridge, M. Kwiatkowska, A. J. Turberfield, J. Bath, Guiding the folding pathway of DNA origami. *Nature* **525**, 82–86 (2015).
- J.-P. Sobczak, T. G. Martin, T. Gerling, H. Dietz, Rapid folding of DNA into nanoscale shapes at constant temperature. *Science* **338**, 1458–1461 (2012).
- F. Dannenberg, K. E. Dunn, J. Bath, M. Kwiatkowska, A. J. Turberfield, T. E. Ouldridge, Modelling DNA origami self-assembly at the domain level. *J. Chem. Phys.* **143**, 165102 (2015).
- X. Wei, J. Nangreave, S. Jiang, H. Yan, Y. Liu, Mapping the thermal behavior of DNA origami nanostructures. *J. Am. Chem. Soc.* **135**, 6165–6176 (2013).
- T. G. Martin, H. Dietz, Magnesium-free self-assembly of multi-layer DNA objects. *Nat. Commun.* **3**, 1103 (2012).
- K. F. Wagenbauer, C. H. Wachauf, H. Dietz, Quantifying quality in DNA self-assembly. *Nat. Commun.* **5**, 3691 (2014).
- E. Stahl, T. G. Martin, F. Praetorius, H. Dietz, Facile and scalable preparation of pure and dense DNA origami solutions. *Angew. Chem.* **53**, 12735–12740 (2014).
- F. Praetorius, B. Kick, K. L. Behler, M. N. Honemann, D. Weuster-Botz, H. Dietz, Biotechnological mass production of DNA origami. *Nature* **552**, 84–87 (2017).
- S. Tyagi, F. R. Kramer, Molecular beacons: Probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14**, 303–308 (1996).
- T. Förster, Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437**, 55–75 (1948).
- L. Stryer, R. P. Haugland, Energy transfer: A spectroscopic ruler. *Proc. Natl. Acad. Sci. U.S.A.* **58**, 719–726 (1967).
- W. Poon, Physics. Colloids as big atoms. *Science* **304**, 830–831 (2004).
- K. Lindorff-Larsen, S. Piana, R. O. Dror, D. E. Shaw, How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
- C. E. Castro, F. Kilchherr, D.-N. Kim, E. L. Shiao, T. Wauer, P. Wortmann, M. Bathe, H. Dietz, A primer to scaffolded DNA origami. *Nat. Methods* **8**, 221–229 (2011).
- D.-N. Kim, F. Kilchherr, H. Dietz, M. Bathe, Quantitative prediction of 3D solution shape and flexibility of nucleic acid nanostructures. *Nucleic Acids Res.* **40**, 2862–2868 (2012).
- S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, W. M. Shih, Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001–5006 (2009).
- M. Fixman, Radius of gyration of polymer chains. *J. Chem. Phys.* **36**, 306–310 (1962).

Acknowledgments: We thank J.-P. Sobczak for support with data analysis and C. Wachauf, J. J. Funke, and Jonathan Doye for discussions. We are grateful for scaffold production by F. Praetorius and F. Engelhardt. **Funding:** This project was supported by European Research Council starting grant no. 256270, consolidator grant no. 724261, and the Deutsche Forschungsgemeinschaft through grants provided by the Gottfried Wilhelm Leibniz Program and the SFB863 TP A9. **Author contributions:** H.D. designed the research. F.S. and N.M. performed the research and analyzed data. All authors discussed data. H.D. and F.S. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 21 November 2018

Accepted 20 March 2019

Published 3 May 2019

10.1126/sciadv.aaw1412

Citation: F. Schneider, N. Möritz, H. Dietz, The sequence of events during folding of a DNA origami. *Sci. Adv.* **5**, eaaw1412 (2019).

The sequence of events during folding of a DNA origami

Fabian Schneider, Natalie Möritz and Hendrik Dietz

Sci Adv 5 (5), eaaw1412.
DOI: 10.1126/sciadv.aaw1412

ARTICLE TOOLS

<http://advances.sciencemag.org/content/5/5/eaaw1412>

SUPPLEMENTARY MATERIALS

<http://advances.sciencemag.org/content/suppl/2019/04/29/5.5.eaaw1412.DC1>

REFERENCES

This article cites 33 articles, 11 of which you can access for free
<http://advances.sciencemag.org/content/5/5/eaaw1412#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).