ACS OMEGA

Article

Cite This: ACS Omega 2018, 3, 18753−18761

http://pubs.acs.org/journal/acsodf

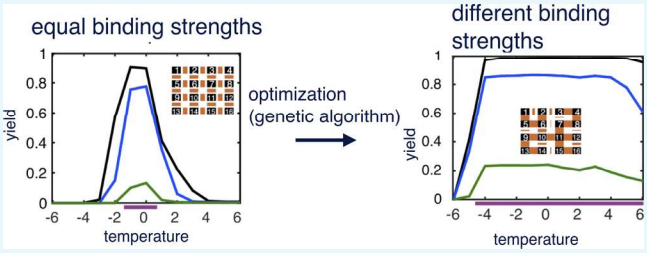
Optimizing Component−Component Interaction Energies in the Self-Assembly of Finite, Multicomponent Structures

John Zenk,[†] Matthew Billups,[†] and Rebecca Schulman*[,†,‡]

[†]Chemical and Biomolecular Engineering and [‡]Computer Science, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, United States

Supporting Information

ABSTRACT: Components of multicomponent biological self-assembly processes generally have interfaces that are optimized to enable the process to proceed quickly and reliably. Analogous principles for designing such interfaces for engineered multicomponent self-assembly processes, such as those involving nucleic acid components, are still being developed. Inspired by biological systems, here, we use stochastic kinetic computer simulations to understand how to tune the strength of interfaces to achieve maximal yields in an addressable self-assembly process. We find that high yields of a desired product can be achieved across a broad range of isothermal assembly conditions by appropriately choosing the interaction energies between components and that heterogeneous (i.e., nonuniform) component−component interaction energies can improve self-assembly outcomes, especially because uniform interfacial energies have unusually low assembly rates. In particular, the structures that assemble with the highest yields under isothermal conditions often include a strong−strong−strong−weak interface motif in closed ring substructures. We apply this process for optimizing interfacial energies to a previously characterized set of components for a self-assembly process using measured kinetic and thermodynamic parameters and show that the initial design of the assembling complex could be improved (i.e., made to assemble with higher yield) by 20−60% using different interface designs. This work suggests that this type of iterative, computational optimization can improve the design cycle for an engineered complex by suggesting next-generation complex designs and preventing the need to experimentally test many different designs.



INTRODUCTION

Nature elegantly assembles RNA and protein complexes, such as the nuclear pore complex,[1,2] kinetochore,[3] and ribosome,[4] from individual components with high yields and with control over where and when components assemble.[4−7] Millions of years of selection pressures have largely determined the biophysical properties of complexes, including characteristics such as quaternary structures,[8,9] thermostability,[10] interaction energies between components,[4,6,7] and assembly pathways.[11] Complex formation highlights nature's strong control over both the folding and assembly landscape of biomolecules.

A goal of biomolecular engineers is the design of self-assembly processes that rapidly produce desired structures with high yields from a set of components that can be easily synthesized. DNA is a promising material for such processes because the reaction rates, specificity of interactions, and thermostability can be tuned by altering the lengths and sequences of the interacting DNA segments.[12,13] In the case of DNA nanostructures, it is increasingly possible to design a range of such components so that design space may be systematically explored.[14] Given a fixed set of components, the effect of varying the assembly protocol (i.e., the reaction conditions) has been explored extensively.[15−19] Common protocols include isothermal assembly, such as rapid folding of DNA origami,[20] altering the salinity of the assembly mixture,[21] and annealing (i.e., cooling the solution temperature from above to below the structure's melting temperature). Annealing is a common protocol because many DNA self-assembly processes have a narrow temperature window where rapid, reversible assembly occurs and this temperature window is often encountered during annealing.[20,22] Annealing is thus used to assemble 2D and 3D DNA "brick" structures consisting of hundreds of short ssDNA components[15,16] and to fold DNA origami nanostructures.[23] However, annealing is slow and often does not produce high yields despite the required waiting time. Further, annealing may not be practical when assembly is required to occur in situ. Currently, highly optimized thermal protocols for assembling DNA nanostructures are often nontrivial (such as a thermal zig-zag) and many still produce low yields and/or nonideal assembly outcomes such as aggregates and intermediates.[14,17,18,24]

An alternative to designing a thermal protocol to optimize assembly yield during annealing is to design the interaction strengths between components.[14,19] A common strategy for

Received: September 6, 2018
Accepted: December 12, 2018
Published: December 28, 2018

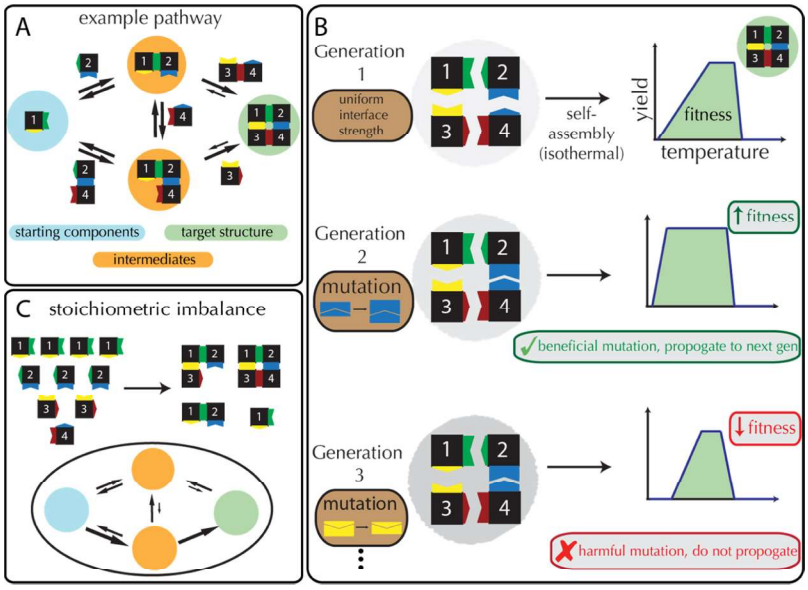


Figure 1. Simulated multicomponent self-assembly and component optimization shown with a 2 × 2 grid system. (A) Examples of possible reactions in a 2 × 2 grid system, where starting components interact with each other starting components or assembly intermediates and assembly intermediates can also interact with one another. Yield at a particular point is measured as the fraction of the total material that is incorporated into target structures. The color and shape of an interface on a square component indicate bond type and complementarity, respectively, while the illustrated size of an interface indicate bond strength. Components and assemblies cannot be rotated or bind to themselves and are numbered by type. Arrow sizes qualitatively indicate reaction propensities. (B) Directed evolution algorithm used to find interfaces that lead to high-yield assembly across a range of assembly conditions. The algorithm begins with generation 1, where all interfaces have equal strength. We simulate isothermal self-assembly at many temperatures for a given period of time and plot assembly yield as a function of temperature. The fitness of a particular set of components is defined as the area under this curve (see Model and Methods). Then, we mutate the strength of one interface in the complex and measure the fitness of the new structure. If the fitness increases from the previous generation, as shown from generation 1−2, the mutation is retained, otherwise the interface returns to its previous strength. This process continues for a set number of generations or different arrangements of interface strengths (typically 500). (C) Component stoichiometry shapes the process of assembly (depicted in oval below the arrow) and thereby yields. When all components are initially present at the same concentration, all starting materials can be incorporated into the target structure. However, in practice, there are slight differences in such concentrations, which can lead to imbalances. The depicted reaction outcomes are hypothetical.

designing the interfaces between the components of DNA nanostructures involves choosing a set base pair (bp) length for all component interfaces and choosing sequences such that each interface has roughly the same energy of interaction.[14−16,18,19,23,24] However, this naïve approach has often resulted in low yields even with optimized assembly protocols.[18,19] More recently, studies of the self-assembly of DNA bricks are demonstrating the benefits of assembling components that have a wide distribution of interaction energies[15,16,22] because uniform interaction energies result in detrimental assembly phenomenon such as kinetic trapping.[22] Similarly, a recent study of DNA-origami folding has shown that by using longer, stronger "staple" strands to influence early formation of long-range interactions, the assembly pathways of the origami structure can be rationally designed, resulting in more robust assembly.[25,26] These increasing improvements in assembly outcomes through the use of heterogeneous interaction strengths between components in a finite-sized structure highlight the potential importance of such heterogeneous interactions. Biological complexes also make use of a wide range of interaction energies between the components of a given complex.[8,27]

These insights have led to interest in developing means for assembling specific final structures that exploit heterogeneity of binding interfaces to achieve high yield quickly and robustly. General insights for design of such structures include the idea that bond strengths should vary[22] and that coordination number influences the yield of assembly;[22] one idea explored recently is that the use of an excess of components on the boundary of a structure might improve yields.[28,29] Recently, methods for improving the predicted yield of a specific structure by changing the interfaces have also been proposed, both within a general statistical mechanical formalism[30] and within a simulation focused on oriented interactions of patchy particles.[31] In this paper, we explore how we might use a simple protocol to design the energies of interaction between components of addressable, finite-size structures to maximize yields over a wide range of assembly protocols using a computational approach. While analytical approaches can suggest general principles for designing interfaces, in practice interfaces must be realized as molecules, which places complex constraints on their nature. Our approach is compatible with these constraints. Here, we demonstrate this approach and use it to suggest a new generation of designs for a reported self-assembly system with measured kinetic and thermodynamic parameters. In particular, we show how this type of design can be used to produce structures which assemble robustly with close to perfect yields across a very wide range of assembly conditions.

The goal in our approach to design is to develop a set of components that can rapidly assemble into the target structures with high yield across a broad range of isothermal conditions for assembly. We then use assembly simulations[32] to optimize yields by making an iterative series of small-scale

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761

improvements akin to those made in a directed evolution algorithm. We also investigate the generality of this approach by asking under what conditions a system of self-assembling components whose interfaces are optimized in this fashion will assemble well.

Our studies suggest that, independent of the assembly protocol, multicomponent structures with multidimensional topology are most efficiently formed when interfaces in a target structure have a wide distribution of energies, suggesting that hierarchical assembly is a good strategy for complex design; to test our protocol's use in practice, we tested the optimization protocol using an experimental, DNA origami-based complex and corresponding interface designs with measured kinetic and thermodynamic parameters.[14,19] This process suggested that specific interface designs which simulations predict should significantly enhance the assembly yield. This work thus suggests that system-level design of the interfaces for self-assembly may be an important method toward improving assembly yields and it may also be possible using real-world data to improve the yield of a set of existing structures whose interactions cannot be tuned arbitrarily.

MODEL AND METHODS

To computationally study biomolecular self-assembly processes, we adapted a coarse-grained kinetic model from ref 32. This model describes the assembly kinetics of biomolecular complexes, using finite-sized rectangular lattices in 2D and a cubic lattice in 3D as prototypical structures. It assumes short-ranged, specific, pairwise interactions between assembling structures, additive interface strengths, and no interface cross talk, nonspecific interactions, or component rotation (see Figure 1A). Assembly begins with an initial, fixed number of each component. Components are depleted as the reaction progresses. All possible intermediate assemblies (i.e., structures that have a connected subset of components in a complex) can form and assembly reactions are allowed between two monomers, a monomer and an assembly, and two assemblies. We use regular rectangular lattices in 2D and a regular cubic lattice in 3D as prototypical structures, which we expect to be good representative structures to a general class of such self-assembly problems.

The model implements a Gillespie algorithm[33] to describe reaction kinetics and assumes a constant, temperature-independent macroscopic forward reaction rate constant, k_{on}, similar to experimentally measured values for reactions involving biomolecules and biomolecular assemblies[34−37] and reverse reaction rate constants which depend on the standard Gibbs free energy, defined as $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, of the reaction. We set the standard enthalpy, $\Delta H^\circ = -51.2$ kcal/mol, and standard entropy, $\Delta S^\circ = -0.150$ kcal/mol/K, from experimentally measured values of two 5 base-pair DNA−DNA hybridization reactions through "sticky ends" (SEs) in a DNA nanostructure assembly process,[38] which have values comparable to other biomolecular interactions.[39−41] The reverse reaction rate constant is mathematically defined as

$$k_{off,b} = k_{on} \exp\left(\frac{b\Delta G^\circ}{RT}\right) \quad (1)$$

where R is the universal gas constant, ΔG° is the standard Gibbs free energy (i.e., binding strength) through a single interface, T is the reaction temperature, and the interface strength factor, b, is a parameter that determines the strength of interaction between two components. For a temperature of 37 °C, these parameters give $\Delta G^\circ/RT = -7.68$ kcal/mol. We limited $0 \le b \le 2$, which is a conservative estimate of the interaction strengths that are achievable with straightforward manipulation of the interface, for example, by altering the number or type of contacts, which for DNA components would involve altering the length or sequence of the interacting nucleotides in an interface.[14] The energy of interaction between components or intermediates that involved the contact of multiple was assumed to be the sum of the energies of the contacting interfaces. This energy was then used to determine the off rate in these cases.

To characterize the scaling of the assembly process, we use dimensionless variables. We define dimensionless time, τ

$$\tau = k_{on}[X]_0 t \quad (2)$$

which depends upon the initial component concentration, $[X]_0$, the macroscopic forward reaction rate constant, and the dimensional time t in seconds. Further, we define a dimensionless reaction temperature, η, where high values of η correspond to low temperatures and vice versa

$$\eta \equiv \log_{10}\left(\frac{k_{on}}{k_{off,1}}[X]_0\right) \quad (3)$$

We simulate both isothermal assembly processes and assembly via annealing. Isothermal assembly takes place at integral values of η in the range of −6 to 6. During annealing, the solution cools linearly from $\eta = -6$ to $\eta = 6$ in a stepwise fashion such that the solution is subject to 100 different values of η, each for durations of $\tau_{anneal}/100$, where τ_{anneal} is the total dimensionless time spent annealing. The dimensionless timescales and temperatures we use in these simulations are realistic, corresponding to a range of temperatures from about $T = 8$ °C ($\eta = -6$) to $T = 58$ °C ($\eta = 6$) for and timescales from 30 min ($\tau = 10$) to 2 days ($\tau = 1000$) for 10 nM of components, for example. We define *yield* as the fraction of the total starting material incorporated into the target structure, which, for example, can easily be determined by gel assay.[15,16]

Directed Evolution of Interfaces. To explore how to design a set of interfaces between components for an addressable assembly process to achieve optimal yields, we implemented an iterative optimization algorithm termed the directed evolution of interfaces. The goal of this algorithm is to take a simple starting structure and iteratively mutate the interface strengths such that the optimized structure is one that assembles rapidly and produces high yields across a broad range of isothermal conditions. The algorithm starts by determining the "fitness" of a simple structure where all interface strength factors are equal ($b = 1$). Fitness is defined here as the integral of the *yield* versus η curve where the yield at each η is the yield of an isothermal assembly process lasting $\tau = 1000$ time units. The algorithm then "mutates" the strength of one interface in the structure by changing the value of b to a new value drawn uniformly from all allowed values of b. All interfaces have the same probability of being chosen to mutate. The mutation is propagated to the next generation if the fitness of the mutated components is higher than the current fitness; otherwise, the mutations are ignored (see Figure 1B). The algorithm repeats this mutation → selection → propagation process for 500 generations. All components have initial concentration $[X]_0$ (i.e., balanced stoichiometry).

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761

Directed Evolution of Interfaces of a DNA Origami-Based Complex with Experimentally Measured Kinetic and Thermodynamic Parameters. To develop a method for designing high-yield assembly processes using a library of molecular components, we modified the optimization algorithm above to choose mutations so that new interfaces would be chosen from the predetermined library of potential molecular interfaces. The specific algorithm we developed optimizes the tetramer (2 × 2) DNA origami-based complex using a set of 24 potential interfaces whose forward and reverse rate constants were previously characterized[14] (see Tables S1 and S2 for measured rate constants). To more accurately model the assembly of these structures, the computational model for assembly of these components uses a bond additivity, α = 0.58, that is consistent with existing measurements[19] rather than the simple addition of bond energies in the cases where components can interact with more than one interface simultaneously (see Supporting Information, Note 2 for additivity calculation). To reflect specific experimental conditions, we use dimensional temperature, time, and concentration values for these simulations. The algorithm isothermally assembles the components for 72 h at the four different temperatures where self-assembly was characterized in ref 14 (30, 35, 40, and 45 °C). A structure's fitness is defined as the area under the yield versus temperature curve (see Supporting Information, Note 2 for more details). This definition was chosen with the goal of finding interface strengths that would permit very high yield assembly across the broadest range of temperatures. The algorithm iterated through 500 designs. All components had initial concentrations of 5 nM.

Stoichiometric Imbalance. Variations in synthesis, concentration measurement, or mixing often mean that the concentrations of components can vary significantly from one assembly process to another; such variation can influence assembly outcome[15,16,23] and in practice a set of components should not be unusually sensitive to such variation. To determine how assembly yields are affected by random noise in component concentrations, we simulate the self-assembly of structures and measure the assembly yield for the cases where the initial concentrations of components are unequal (see Figure 1C). The initial concentration of each component j was chosen as $\gamma_j[X]_0$, where $[X]_0$ is the desired concentration and γ_j is a Gaussian random variable with mean μ = 1 and standard deviation σ constrained to $\gamma_j > 0$. We consider values of σ between 0 and 0.25. Stochastic stoichiometric simulations are performed in replicates of 10.

RESULTS AND DISCUSSION

We began our study of the optimization of component interaction energies for addressable self-assembly processes by simulating the directed evolution of interfaces (see Model and Methods) for regular grid structures with various dimensionalities and numbers of components. We studied a 1 × 5 line structure in 1D (Figure S1), 2 × 2, 3 × 3, and 4 × 4 square grid structures in 2D, and a 2 × 2 × 2 cubic structure in 3D (Figure 2).

Uniform Interaction Strength between Components Leads to High Yields of 1D Structures But Not 2D or 3D Structures. The interfaces that evolved between the components self-assembling 1D structures were all strong, consistent with previous observations that such interfaces maximize the yield of these structures under a variety of assembly conditions.[27,32] In the 1 × 5 line structure, the optimized structure had uniform interface strength factors of ~2 (the highest possible), which increased the fitness of the structure by more than 130% (see Figures S1 and S2). However, such uniformly strong interfaces do not lead to optimal yields of higher dimensional structures. The optimized 2 × 2 line structure had three strong interfaces and one weak interface ($b_1 = b_2 = 2$, $b_3 = 1.99$, $b_4 = 0.78$). Structures with

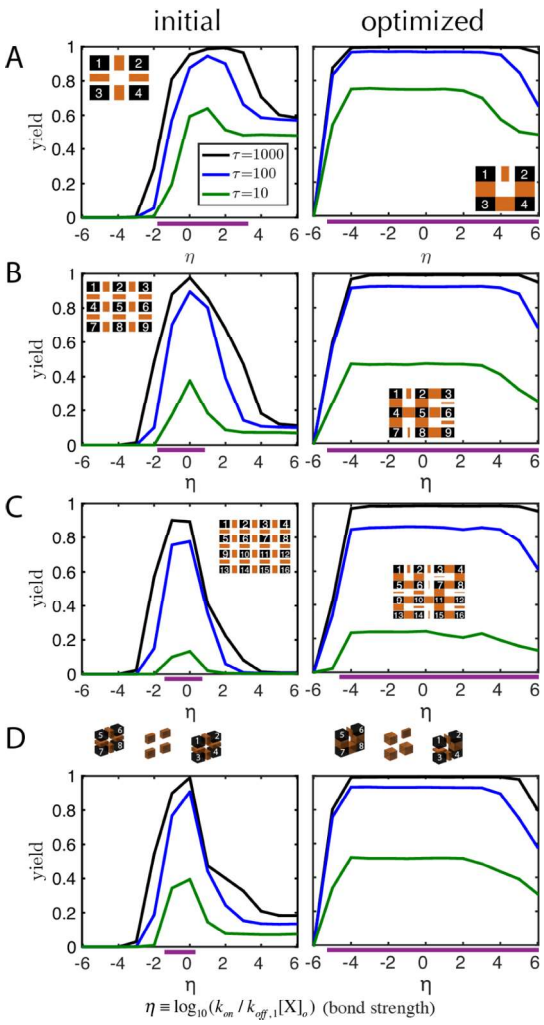


Figure 2. A mixture of strong and weak interfaces between components can enhance the fitness (a measure of yields under different conditions) of 2D and 3D structures. Yield is plotted as a function of assembly temperature η for a structure with interfaces all having the same strength of interaction (left) and the corresponding structure with optimized interfaces between components (right) for (A) 2 × 2 grid, (B) 3 × 3 grid, (C) 4 × 4 grid, and (D) 2 × 2 × 2 cube target structures. The inset diagrams show the structures of each complex: black squares are the components and the thicknesses of the brown bands indicate the strengths of interaction between them. Different component types each have different numbers. The 2 × 2 × 2 diagram is split into three separate cartoon diagrams to more easily visualize the interface strengths, where the left and right diagrams depict the strengths of interfaces between the four components of the top and bottom rows, respectively, of the target structure and the middle diagram depicts the strengths of interfaces between the two rows. The size of the brown block is proportional to the interface strength. The purple lines underneath plots indicate the high-yield assembly regime, defined as where the target structure *yield* >80% after τ = 1000.

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761

these interfaces were 80% fitter than the original structures, assembling with virtually 100% yield at almost all temperatures considered. Similarly, mixtures of strong, medium, and weak interface strengths were observed in the optimized 3 × 3, 4 × 4, and 2 × 2 × 2 assemblies; these structures increased in fitness by 120, 220, and 180%, respectively, in comparison to the initial structures (see Figures 2B–D and S2). Thus, higher dimensional target structures self-assemble more reliably after the strengths of interaction between the components are optimized. While structures with uniform interaction strengths assemble well only in a narrow range of temperatures and this temperature range decreases with the number of components (Figure 2, purple bars), structures containing components with optimized component structures assemble well across almost all temperatures and the range of temperatures where efficient assembly occurs does not shrink for large structures.

Optimized Structures Assemble Faster or Just as Fast as Initial Structures. Our next goal was to determine whether optimizing a complex using the selection criteria in the directed evolution of interfaces simulations, that is, yield across a broad range of isothermal conditions at a particular long assembly time, also increased the speed of structure formation. To do this, we compared the average rate of formation, defined mathematically as $\mathrm{d}(yield)/\mathrm{d}\tau$, across the range of isothermal conditions for the optimized assemblies to those of the initial assemblies at the three reaction times tested and found that optimized assemblies assemble faster (in some cases up to $\sim 10^4$ faster) or just as fast as initial assemblies, at all isothermal conditions for all structures tested (see Figure S3). This suggests that by optimizing structures using the selection criteria, structures gain the added bonus of increasing the speed of assembly for free.

Strong–Strong–Strong–Weak Interface Strengths in a Four-Component Ring Is a Common Motif within Efficient Assemblers. Turning a 2D assembly problem into a pseudo-1D problem has been shown to be a good design strategy for multicomponent structures. In protein rings, which have evolved to contain one significantly weaker interface than the rest, assembly occurs through the pathway where components with strong interfaces first form the chain and finally the component with the weak interface completes the ring.[27] Our findings here not only corroborate these results in a simple 2 × 2 structure (see Figure 2B) but also in more complex 2D and 3D structures. We found that the optimal interface design commonly includes three strong interfaces and one significantly weaker interface in closed-ring substructures in a complex. For example, the optimized 3 × 3 square grid structure has four of four possible (4/4) closed four-component rings which share this motif (see Figure 2B), the optimized 4 × 4 square grid has 5/9 (see Figure 2C), and the optimized 2 × 2 × 2 cube has 5/7 (see Figure 2D). In the optimized 4 × 4 structure, for example, the four, four-component rings (quadrants) at the corners share this motif but there are minimal interfaces between the quadrants, suggesting how assembly might occur hierarchically through the formation of the four quadrants that then assemble into a larger structure.

Different Directed Evolution of Interface Simulations Converge to Similar Optimized Structures. A common result in using genetic algorithms is that structures discovered through the optimization represent "local optima" such that different, fitter variants may be discovered by rerunning the process. We simulated the directed evolution of the 3 × 3 square grid three times from the same initial conditions (i.e., uniform interaction strength) to see whether some iterations would produce fitter variants than others. In all cases, the fitness values improved by approximately 120%. The specific interaction strengths are widely heterogeneous for all structures. However, while they differ in their specifics, observation of the interfaces suggest that all structures form via hierarchical assembly pathways (see Figure S4) and most fit structures have a strong–strong–strong–weak interface strength motif. This suggests that a simple design strategy of including strong–strong–strong–weak loops for interface strengths in higher dimensional structures or more generally designing structures that assemble via hierarchical pathways is an important general design strategy for components in addressable structure strengths.

Structures with High Fitness for Isothermal Assembly Also Assemble Well When Annealed. Many self-assembly processes rely on nonisothermal assembly conditions to achieve robust assembly[15,16,22,23] and there is a growing body of computational studies, largely guided by principles observed in practice, that annealing is the best method for promoting error-free, robust assembly of finite structures because conditions that promote robust assembly are essentially guaranteed to be found,[22,32,42] with longer annealing protocols typically resulting in higher yields.[16] With this knowledge, our goal was to determine if optimized structures or those with high fitness could also assemble with high yields when annealed. To test this, we simulated the annealing of all 500 structures (generations) in the directed evolution of interface simulation for the 2 × 2, 3 × 3, and 2 × 2 × 2 systems (see Model and Methods for the annealing protocol) and found that assemblies with high fitness values also have high yields after intermediate and long annealing times (see Figure 3). After long annealing times ($\tau_{\mathrm{anneal}} = 1000$), yields approached 1 for all structures with fitness values approximately >0.5, >0.75, and >1 for the 2 × 2, 3 × 3, and 2 × 2 × 2 assemblies, respectively. This suggests that annealing is an excellent method to assemble structures, as even less fit assemblies can produce high yields. However, assemblies with very low fitness values ($\lesssim$0.5), resulting from mutations that were not retained during the assembly process, do not assemble with high yield even at long assembly times likely due to the fact that one or more of their bonds are too weak to form, even at low temperatures. For assemblies with intermediate fitness values subject to shorter anneals, yields vary widely and increased fitness does not necessarily mean a higher yield. The least efficient assembler for those structures with fitness $\lesssim$1 at intermediate times is the structure with uniform interaction strengths.

Structure Formation Rates during Annealing and Isothermal Assembly Can Vary Widely. We next wanted to determine which protocol, annealing or isothermal assembly, produced structures most rapidly for the initial and optimized structures. We investigated the rates of structure formation for the two assembly methods for the initial and optimized structures in the 2 × 2, 3 × 3, and 2 × 2 × 2 assemblies and found that when isothermally assembled under moderate conditions ($-2 \lesssim \eta \lesssim 2$) for short to intermediate reaction times, isothermal assembly was faster than annealing. In contrast, annealing was faster than isothermal assembly at extreme temperatures $\eta \lesssim -4$ and $\eta \gtrsim 4$ (see Figure S5).

Increasing Stoichiometric Imbalance Decreases the Fitness of All Structures in a Uniform Fashion. To

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753–18761

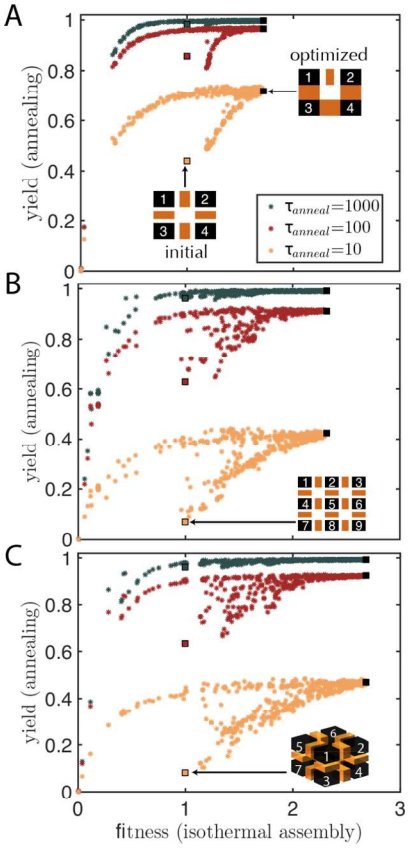


Figure 3. Structures with high fitness assemble via annealing with high yields. Yields of the target complex after annealing as a function of fitness (for isothermal assembly conditions) after various annealing assembly times, τ_{anneal}. (A) 2 × 2, (B) 3 × 3, and (C) 2 × 2 × 2 systems are depicted. Interface strengths are equal initially, depicted in the inset diagrams and indicated by square markers outlined with black. All other generations are shown as dots except the most fit generation, which is shown as with filled-in black squares. Structures with fitness below 1 that are shown are mutations that were not retained during the optimization process. The cartoon diagram for the optimized 2 × 2 depicted structure depicted in (A). Annealing protocol starts at $\eta = -6$ and steps toward $\eta = 6$ over a series of 100 steps (100 values of linearly increasing η), spending $\tau_{anneal}/100$ at each value of η. Yields are recorded at the end of the anneal ($\eta = 6$). Components are numbered to indicate uniqueness.

investigate the interplay between fitness and stoichiometric imbalance, a real-world experimental factor that could affect assembly outcomes, we selected a subset of 50 assemblies at random from the 500 generations of the 3 × 3 square grid assemblies, each with different interfacial interaction energies and thus presumably different values of fitness, and tested their fitness when the assembly began with a stoichiometric imbalance of components. Specifically, we varied the number of starting components (at random, see Model and Methods) such that the initial concentration of each component was set by multiplying $[X]_0$ by a sample from a normal random distribution with mean 1 and a standard deviation σ (see Figure 1C). The fitness was then computed using the area under the yield curve where the yield was calculated using the initial component concentration $[X]_0$. A stoichiometric imbalance of starting components would be expected to decrease the fitness of the assemblies, but it was not clear whether this decreases would be determined mostly by the fact that there were insufficient quantities of some components or whether the yields of structures with heterogeneous, optimized bond strengths would be more or less affected than structures with equal energies of interaction between all matching components. We found that there is a strong linear correlation ($r^2 > 0.99$) between the fitness of isothermally assembled structures with stoichiometric balance and the fitness of isothermally assembled structures with stoichiometric imbalance ($\sigma > 0\%$), suggesting that the penalty for stoichiometric imbalance for most structures is essentially the same. Increasing the value of σ (the extent of imbalance) decreased the slope (all <1) of the linear trend line (see Figure 4). Still,

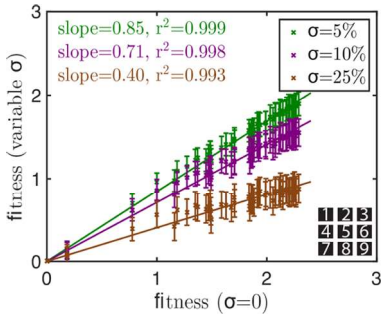


Figure 4. Fitness decreases by a constant, multiplicative factor when there is variation in component stoichiometry (as characterized by a standard deviation of concentration σ). The plot relates the fitness of the same structures with and without random variation in the stoichiometry of components for the 3 × 3 square grid system. The inset diagram depicts the target structure. Components are numbered to indicate uniqueness. Lines are linear fits to the data, with fit statistics inset in the plot. Ten simulations were performed for each assembly condition. Error bars indicate one standard deviation of the reported quantity.

structures with high fitness values (>1) maintain their superior fitness when subject to stoichiometric imbalance, even when the imbalance is as high as σ = 25%. This suggests that by selecting for structures with greater fitness, structures also gain some robustness to stoichiometric imbalance.

Further investigation of stoichiometric imbalance of in the formation of different initial structures suggested that systems with high stoichiometric imbalances should be assembled under conditions that slightly favor reverse reactions (i.e., nucleation limited). For example, increasing σ from 0 to 25% decreased the mean yield by about 10−60% at neutral and forward-driven assembly conditions $\eta \geq 0$ but remained approximately constant in nucleation-limited conditions for all initial structures tested (see Supporting Information Figures S6−S15). To understand whether the yields were being affected by a limiting single component or a general bias in the assembly pathways, we plotted the ratio of the number of target structures formed to the number of components for the component type which started the simulation with the fewest quantity and found that the ratio as a function of η changed little with increasing stoichiometric imbalance and were almost identical to such curves for systems with stoichiometric balance. This suggests that the component with the lowest starting concentration indeed limits the assembly process. It is worth noting that here we randomly altered the number of starting components for each component type; however, ref 43 suggests a way to increase yields by intentionally introducing stoichiometric variability between different types of compo-

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761

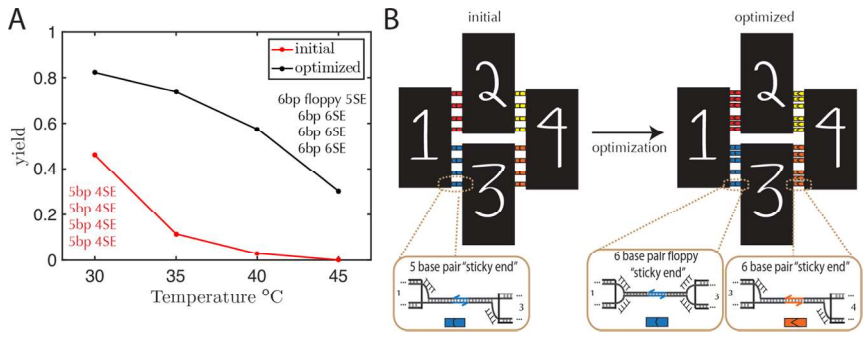


Figure 5. Optimizing the interfaces within a self-assembly system where four different DNA-origami components assemble into a tetramer complex. Using experimentally measured values of the additivity between interfaces, as reported in ref 19, and kinetic parameters such as the forward (k_{on}) and reverse (k_{off}) reaction rate constants at several different temperatures for a wide variety of interface designs, as reported in ref 14, we computationally determined a set of interfaces for this system to optimize assembly yields (see Supporting Information Note 1). The types of interfaces available in this design space vary in the number of linkers (i.e., dsDNA connections that form) in an interface, the length of the "SE" (ssDNA region through which two components can specifically bind to one another) of the linker, and whether the linker has a flexible region (or is "floppy"). The naïve design approach assumes that all interfaces have the 5 bp 4SE design, which means that there are four linkers, each with 5 bp SEs. (A) Yield of the tetramer complex at a range of temperatures after 72 h of isothermal assembly for the naïve design and the optimized designed after 500 generations. The inset text indicates the designs of the four interfaces. (B) Cartoon schematic of the tetramer complex. Black components with white letters represent the four DNA origami components and hairpins on the surface for component differentiation. The left diagram depicts the naïve structure, with all interfaces having the same, 5 bp 4SE design, and the right diagram shows the optimized structure, with three interfaces having the 6 bp 6SE design and one interface having the 6 bp floppy 5SE design. Linkers are colored according to the interface. The zoomed-in view shows the linking architecture of the linkers in an interface and cartoon schematic, with the SE colored. The 3′ end of DNA is indicated with arrows. The initial component concentration, $[X]_0$ = 5 nM, is the same as in the described experiments. These simulations assumed no component stoichiometric variability.

nents in a target structure for a process self-assembling finite-sized structures.

Optimizing a DNA Origami-Based Biomolecular Complex Using Measured Kinetic and Thermodynamic Parameters. One of the benefits of interface design using the methods we have developed in this paper is that the methods can be applied not only to optimize interface structures within a continuous range but also to select among a library of existing interface choices. Our final goal in this work was to determine how well such a process would work by taking a known, reported set of interfaces with measured energies of interaction and additivity and to optimize the selection of interfaces for the components to increase yields using our directed evolution of interfaces algorithm. We used the structure reported in ref 19, which is a tetrameric complex made from four different DNA origami components designed to have four approximately equal interface strengths (i.e., same number of dsDNA linkers per interface, each with the same length of "SE", or single-stranded region used to specifically bind a linker with the complementary sequence). On the basis of the reported free energies of reaction,[19] this structure has a bond additivity between interfaces of $\alpha = 0.58$ (see Supporting Information Section 3.1), which means that increasing the number of bonds slightly increases the overall strength of interaction between components. We select from possible interface strengths using the interface designs for such origami components reported in ref 14, with measured forward (k_{on}) and reverse (k_{off}) reaction rate constants at different temperatures (see Supporting Information Section 3.2). The simulation parameters here are informed by the experimental details described in refs 14 amd 19 and reported in dimensional quantities (see Model and Methods, Supporting Information Section 3.3 for computational details).

An initial design for a structure used all interfaces having the same linking architecture, specifically four linkers with 5 bp SEs (or in shorthand "5 bp 4SE"). The optimization process suggested a structure where three interfaces have 6 bp 6SE and one interface has a 6 bp 5SE with a "floppy" portion in the linking architecture (Figure 5). With this optimization, simulations indicate that yield could be improved by 30–60% across the range of temperatures tested. Furthermore, for an analogous 3 × 3 structure, similar to the type of structure reported in ref 18, yields could be improved by up to 50% (see Figure S16). Interestingly, both the tetramer and the 3 × 3 design include multiple types of interface architectures in the optimized design, suggesting that even for measured values of additivity and thermodynamic and kinetic parameters, heterogeneity in interfaces is a good design strategy.

CONCLUSIONS

In this paper we propose and test a simple computational method for guiding the design cycle of engineered complexes that could potentially eliminate the need to design and experimentally test many different structural combinations. An iterative optimization approach like the one described here for finite-sized structures is feasible for almost any structure in practice where something about the energy of interaction of different interfaces is known or could be estimated. This method can be used for structures with a wide variety of starting conditions (e.g., numbers and types of components, interface designs, and dimensionality) and can be subject to different numbers of iterations (generations) to improve a structure in a more practical situation. For example, one could optimize isothermal assembly at a single temperature or ask whether this optimization would lead to efficient assembly via annealing. Further, instead of considering final yields, one might seek to use such an approach to optimize nucleation processes or the pathway of assembly to follow specific rules. More generally, one might use such an approach to design not only assembly processes but also the design of processes that have specific dynamics as a result of a large number of dynamic parameters.

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761

We showed that iterative optimization of local contacts in a structure can enable structures to isothermally assemble faster, achieve higher yields, and do so across a wider range of temperatures, effectively eliminating the need to guess and check or finely tune assembly protocols. With such isothermal optimization, structures also assemble well under an annealing protocol and can do so just as rapidly at longer assembly times, where even many less fit structures can achieve rapid, high-yield assembly.

Designing target structures with uniform interaction strengths is a poor design strategy, unless designing a 1D structure, because the resulting landscape is vast and relatively flat: many potential intermediate structures can readily be formed at even slightly forward driven conditions. In comparison, well-designed (heterogeneous) interaction strengths enable rapid, high-yield assembly, suggesting hierarchal assembly pathways.

Our coarse-grained model makes simplifying assumptions for the self-assembly process which might result in an incomplete theory in practice. Here, we do not consider other nonspecific or off-target interactions which could affect the assembly outcomes in ways this model does not predict. Furthermore, modeling interface additivity using a single value might not be the best method, where real-world systems often exhibit cooperative[44,45] or noncooperative binding[19] and depend on the number and type of ligands present.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.8b02303.

Additional simulations; and additional description of simulation methods (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rschulman@jhu.edu (R.S.).

ORCID

Rebecca Schulman: 0000-0003-4555-3162

Author Contributions

J.Z. and R.S. conceived of the study. J.Z. and M.B. performed the research and performed the analysis and J.Z. and R.S. wrote the manuscript.

Funding

The authors would also like to acknowledge funding from NSF CCF grant 1161941 and the DOE BES award DE-SC0010426 that helped support the authors and this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank Dr. Deepak Agrawal, Angelo Cangialosi, Joshua Fern, Sam Schaffter, and Dominic Scalise for their helpful discussions in this work as well as Dr. Jeff Gray and Dr. Marc Ostermeier for their valuable input and feedback.

■ REFERENCES

(1) Suntharalingam, M.; Wente, S. R. Peering through the Pore. *Dev. Cell* **2003**, *4*, 775−789.

(2) Alber, F.; Dokudovskaya, S.; Veenhoff, L. M.; Zhang, W.; Kipper, J.; Devos, D.; Suprapto, A.; Karni-Schmidt, O.; Williams, R.; Chait, B. T.; Sali, A.; Rout, M. P. The molecular architecture of the nuclear pore complex. *Nature* **2007**, *450*, 695−701.

(3) Cheeseman, I. M.; Desai, A. Molecular architecture of the kinetochore-microtubule interface. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 33−46.

(4) Talkington, M. W. T.; Siuzdak, G.; Williamson, J. R. An assembly landscape for the 30S ribosomal subunit. *Nature* **2005**, *438*, 628−632.

(5) Mulder, A. M.; Yoshioka, C.; Beck, A. H.; Bunner, A. E.; Milligan, R. A.; Potter, C. S.; Carragher, B.; Williamson, J. R. Visualizing Ribosome Biogenesis: Parallel Assembly Pathways for the 30SSubunit. *Science* **2010**, *330*, 673−677.

(6) Synowsky, S. A.; van den Heuvel, R. H. H.; Mohammed, S.; Pim Pijnappel, W. W. M.; Heck, A. J. R. Probing genuine strong interactions and post-translational modifications in the heterogeneous yeast exosome protein complex. *Mol. Cell. Proteomics* **2006**, *5*, 1581−1592.

(7) Sharon, M.; Taverner, T.; Ambroggio, X. I.; Deshaies, R. J.; Robinson, C. V. Structural organization of the 19S proteasome lid: Insights from MS of intact complexes. *PLoS Biol.* **2006**, *4*, e267.

(8) Levy, E. D.; Erba, E. B.; Robinson, C. V.; Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **2008**, *453*, 1262−1265.

(9) Marsh, J. A.; Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **2015**, *84*, 551−575.

(10) DePristo, M. A.; Weinreich, D. M.; Hartl, D. L. Missense meanderings in sequence space: A biophysical view of protein evolution. *Nat. Rev. Genet.* **2005**, *6*, 678−687.

(11) Marsh, J. A.; Hernández, H.; Hall, Z.; Ahnert, S. E.; Perica, T.; Robinson, C. V.; Teichmann, S. A. Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. *Cell* **2013**, *153*, 461−470.

(12) Seelig, G.; Soloveichik, D.; Zhang, D. Y.; Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **2006**, *314*, 1585−1588.

(13) Zhang, D. Y.; Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **2009**, *131*, 17303−17314.

(14) Zenk, J.; Tuntivate, C.; Schulman, R. Kinetics and Thermodynamics of Watson-Crick Base Pairing Driven DNA Origami Dimerization. *J. Am. Chem. Soc.* **2016**, *138*, 3346−3354.

(15) Wei, B.; Dai, M.; Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **2012**, *485*, 623−626.

(16) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012**, *338*, 1177−1183.

(17) Endo, M.; Sugita, T.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. Programmed-Assembly System Using DNA Jigsaw Pieces. *Chem.—Eur. J.* **2010**, *16*, 5362−5368.

(18) Rajendran, A.; Endo, M.; Katsuda, Y.; Hidaka, K.; Sugiyama, H. Programmed Two-Dimensional Self-Assembly of Multiple DNA Origami Jigsaw Pieces. *ACS Nano* **2011**, *5*, 665−671.

(19) Fern, J.; Lu, J.; Schulman, R. The Energy Landscape for the Self-Assembly of a Two-Dimensional DNA Origami Complex. *ACS Nano* **2016**, *10*, 1836−1844.

(20) Sobczak, J.-P. J.; Martin, T. G.; Gerling, T.; Dietz, H. Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature. *Science* **2012**, *338*, 1458−1461.

(21) Gerling, T.; Wagenbauer, K. F.; Neuner, A. M.; Dietz, H. Dynamic DNA devices and assemblies formed by shape-complementary, non-base pairing 3D components. *Science* **2015**, *347*, 1446−1452.

(22) Jacobs, W. M.; Reinhardt, A.; Frenkel, D. Rational design of self-assembly pathways for complex multicomponent structures. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 6313−6318.

(23) Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **2006**, *440*, 297−302.

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761

(24) Liu, W.; Zhong, H.; Wang, R.; Seeman, N. C. Crystalline Two-Dimensional DNA-Origami Arrays. *Angew. Chem., Int. Ed.* **2011**, *50*, 264−267.
(25) Dunn, K. E.; Dannenberg, F.; Ouldridge, T. E.; Kwiatkowska, M.; Turberfield, A. J.; Bath, J. Guiding the folding pathway of DNA origami. *Nature* **2015**, *525*, 82−86.
(26) Marras, A. E.; Zhou, L.; Kolliopoulos, V.; Su, H.-J.; Castro, C. E. Directing folding pathways for multi-component DNA origami nanostructures with complex topology. *New J. Phys.* **2016**, *18*, 055005.
(27) Deeds, E. J.; Bachman, J. A.; Fontana, W. Optimizing ring assembly reveals the strength of weak interactions. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2348−2353.
(28) Wayment-Steele, H. K.; Frenkel, D.; Reinhardt, A. Investigating the role of boundary bricks in DNA brick self-assembly. *Soft Matter* **2017**, *13*, 1670−1680.
(29) Sajfutdinow, M.; Jacobs, W. M.; Reinhardt, A.; Schneider, C.; Smith, D. M. Direct observation and rational design of nucleation behavior in addressable self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, E5877−E5886.
(30) Miskin, M. Z.; Khaira, G.; de Pablo, J. J.; Jaeger, H. M. Turning statistical physics models into materials design engines. *Proc. Natl. Acad. Sci.* **2016**, *113*, 34−39.
(31) Madge, J.; Miller, M. A. Optimising minimal building blocks for addressable self-assembly. *Soft Matter* **2017**, *13*, 7780−7792.
(32) Zenk, J.; Schulman, R. An Assembly Funnel Makes Biomolecular Complex Assembly Efficient. *PLoS One* **2014**, *9*, e111233.
(33) Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **1977**, *81*, 2340−2361.
(34) Wetmur, J. G. DNA Probes: Applications of the Principles of Nucleic Acid Hybridization. *Crit. Rev. Biochem. Mol. Biol.* **1991**, *26*, 227−259.
(35) Camacho, C. J.; Kimura, S. R.; DeLisi, C.; Vajda, S. Kinetics of desolvation-mediated protein-protein binding. *Biophys. J.* **2000**, *78*, 1094−1105.
(36) Evans, C. G.; Hariadi, R. F.; Winfree, E. Direct Atomic Force Microscopy Observation of DNA Tile Crystal Growth at the Single-Molecule Level. *J. Am. Chem. Soc.* **2012**, *134*, 10485−10492.
(37) Recht, M. I.; Williamson, J. R. Central domain assembly: thermodynamics and kinetics of S6 and S18 binding to an S15-RNA complex 1 1Edited by D. Draper. *J. Mol. Biol.* **2001**, *313*, 35−48.
(38) Schulman, R.; Winfree, E. Synthesis of crystals with a programmable kinetic barrier to nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15236−15241.
(39) Horton, N.; Lewis, M. Calculation of the free energy of association for protein complexes. *Protein Sci* **1992**, *1*, 169−181.
(40) Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Simple, Intuitive Calculations of Free Energy of Binding for Protein−Ligand Complexes. 1. Models without Explicit Constrained Water. *J. Med. Chem.* **2002**, *45*, 2469−2483.
(41) SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460−1465.
(42) Reinhardt, A.; Frenkel, D. Numerical evidence for nucleated self-assembly of DNA brick structures. *Phys. Rev. Lett.* **2014**, *112*, 238103.
(43) Murugan, A.; Zou, J.; Brenner, M. P. Undesired usage and the robust self-assembly of heterogeneous structures. *Nat. Commun.* **2015**, *6*, 6203.
(44) Perutz, M. F. Mechanisms of Cooperativity and Allosteric Regulation in Proteins. *Q. Rev. Biophys.* **1989**, *22*, 139.
(45) Williamson, J. R. Cooperativity in macromolecular assembly. *Nat. Chem. Biol.* **2008**, *4*, 458−465.

DOI: 10.1021/acsomega.8b02303
ACS Omega 2018, 3, 18753−18761