Possibilities and limits of DNA-enabled programmable 2D self-assembly

Nicholas Tjahjono[1], Evgeni S. Penev[1], and Boris I. Yakobson[1,2]*

[1]*Department of Materials Science and NanoEngineering, Rice University, Houston, Texas 77005, United States*
[2]*Department of Chemistry, Rice University, Houston, Texas 77005, United States*

Abstract
Programmable self-assembly provides a promising avenue to improve upon traditional synthesis and create multi-component materials with emergent properties and arbitrary nanoscale complexity. However, its most successful realizations utilizing DNA often use complicated arduous procedures that result in low yields. Here, we employ coarse-grained molecular dynamics to uncover the ranges of temperatures and misbinding strengths needed for successful one-pot self-assembly of generic, two-dimensional (2D), and distinguishable blocks. Analysis of the energies associated with a single-stranded DNA interacting with all other sequences within a mixture revealed that the success of DNA-based assembly is primarily determined by the strongest misbinding a given sequence can encounter with a sequence highly similar to its reverse complement. This enabled us to design optimized sequence ensembles with acceptably weak and consequently rare misbinding. An estimate is provided for the maximum size of, and complexity of sequences needed to synthesize self-assembled structures with high accuracy and yield, with potential relevance for DNA-functionalized low-dimensional materials for electronics and energy storage.

Keywords: *programmable self-assembly, misbinding, error-free DNA-mediated assembly, DNA sequence design, coarse-grained molecular dynamics, DNA-functionalized 2D materials, aperiodic nanostructures*

Advances in materials synthesis have enabled increasingly complex nanostructures. Such complexity is often necessary to achieve emergent material properties unavailable in simpler arrangements. Examples range from tunable bandgaps and intrinsic carrier mobilities due to local edge geometries in graphene nanoribbons[1–3], to high strength and toughness due to complex hierarchical structures in biological composites.[4–7] However, current synthesis methods utilizing top-down and bottom-up approaches have limited control to produce complicated aperiodic structures with localized nanoscale features.[8]

Programmable self-assembly, in which each component is "informed" of its specific position in the target structure due to its chemical and/or shape specificity, provides a promising avenue to engineer structures with aperiodic complexity and molecular resolution. Some of most successful realizations utilize DNA: each building block of the target structure contains a unique short DNA segment(s) that can specifically bind to nucleotide strands from other blocks. The high chemical specificity of Watson–Crick (WC) complementary base-pairing facilitates the accurate assembly of these blocks, and has led to the development of a wide range of truly programmed and self-assembled nanostructures ranging from DNA bricks and origami, to DNA-prescribed nanoparticle arrays with tailored optical and plasmonic responses for single-molecule manipulation and detection in biophysical studies and diagnostics.[9–23] Recent advances in DNA-functionalization of graphene and nanotubes also suggest that this method may be expedient in on-surface synthesis of low-dimensional architectures with complex nanoscale features for flexible electronics, biosensors, and optical computing.[24–36]

To ensure that only desired blocks interact, DNA-based structures self-assembly may use *hierarchical* methods:[9,10,37,38] a couple of blocks are combined at each step to form intermediates, which then act as larger building blocks for the next assembly stage. While such stage-by-stage assembly has been shown, in experiment and theory, to be successful in suppressing some spurious interactions between blocks,[9,10,37,38] the yield of the target structure is low when assembling multiple blocks.[9,10] The yield scales as p^{N-1} (p is the yield per step, N is the number of distinguishable components[9,10]), which can reach as low as 3% for $N = 64$ despite a high $p = 0.95$.[10] The experimental procedure also is quite cumbersome, as combining k subunits per step requires $\log_k N$ stages of assembly[10] that must each be precisely executed. Even if the design and experimental process can be parallelized and automated, the low yield and

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

complexity will still hamper such hierarchical protocol. The decreased yield may be linked to the stronger non-native interactions or "*misbinding*", increasing with the interface between any two blocks, whose size gets larger with each successive stage.[37] This makes the hierarchical approach susceptible to kinetic traps if non-native attractions cannot be suppressed,[37] and ultimately limits the achievable complexity and size of target structures.

An alternative approach that may mitigate the aforementioned problems is to assemble all distinguishable components at the same time in a *one-pot* reaction. This non-hierarchical route has not only been able to assemble DNA blocks into a variety of complex shapes,[9,11,12,15–17,20,23] but also resulted in comparable or higher yields than hierarchical schemes.[9,11,15,20,23] One-pot assembly may thus be superior to stage-by-stage assembly,[37] but requires the use of larger and more complex DNA tiles and origami.[11,20] The design of the tiles, particularly the DNA segments at the binding interfaces, may be simplified if the mechanisms of one-pot assembly are understood. However, while much theoretical and computational work has been done in rationalizing the success of DNA brick and tile structures,[37,39,40] the models tend to be highly generic and do not capture the more subtle details of DNA binding. In particular, they do not consider the associated DNA hybridization energy scales and misbinding interactions arising from the interactions between non-complementary single-stranded DNA segments (or sticky ends), crucial for understanding non-hierarchical programmed assembly.

Here, we use a model approach, based on coarse-grained (CG) molecular dynamics simulations and theory, to elucidate the conditions and DNA sequence complexity necessary to achieve error-free, one-pot self-assembly of 2D distinguishable blocks into arbitrary patterns. Our model suggests the existence of a domain in the temperature/misbinding diagram where complete error-free self-assembly is possible. We then analyze the interaction energies between sequences in DNA-brick experiments[12,39] and uncover a multi-modal energy distribution. Utilizing this energy spectrum into our CG model allows us to propose a rational design scheme capable of generating collections of sequences with low misbinding. Moreover, such a framework can provide a quantitative estimate for the *length*, *number,* and *complexity* of DNA sequences needed to synthesize, and *maximum size* of, structures that can be self-assembled with high accuracy and yield. This optimization of DNA-guided 2D assembly can provide facile routes to achieve precise alignments of graphene and other nanosheets during liquid-phase processing, thereby potentially improving the electrical conductivity achievable in energy storage applications,[41–43] among other benefits.

Results and discussion

Generalized model. Our approach is based on a minimalist coarse-grained model in which programmable assembly of (structurally identical) planar square-shaped units is realized through chemical specificity (or "color"[44]). Such building blocks are rigid bodies consisting of nine particles (with a fixed nearest neighbor distance of 1 in dimensionless LJ units), where at least one particle on their edges is "colored" (Figure 1a). Each colored particle, which is assigned a unique index of i, in a block can only interact with colored particles on other blocks via Lennard-Jones (LJ) potentials. Maximum affinity/attraction (of LJ strength ε) occurs solely between particles of matching colors, which allows addressability, i.e., specifying the position of a block in the target assembly. Undesired coupling between mismatched colors, or "misbinding", is penalized by its weaker strength $\varepsilon' < \varepsilon$. Under the defined interaction scheme, if two colored particles, i and j, are within a distance r below the LJ cutoff distance of $r_c = 2.3$, their energy is

$$V_{i,j}(r) = \delta(c_i,c_j)\, U_{\mathrm{LJ}}(r,\varepsilon) + [1 - \delta(c_i,c_j)]\, U_{\mathrm{LJ}}(r,\varepsilon'), \quad (1)$$

where c_i corresponds to the color of particle i, $\delta(n,m)$ is the Kronecker delta, and $U_{\mathrm{LJ}}(r,\varepsilon)$ is the standard 12-6 LJ potential with $\sigma = 1$ and minimum energy at $-\varepsilon$. The summation over the energies of all unique pairs of colored particles gives the potential energy of the system, as $E = \Sigma_{i<j} V_{i,j}(r)$. This is equivalent to the summation over all pairwise energies between particles that form interfaces, due to the designed exclusion of interactions between colored particles on the same block (i.e., $V_{i,j} = 0$ if i and j are on the same block). The lowest energy is thus achieved only for a *complete* and *correct* structure, $E_{\min} = -n\varepsilon$, where n is the number of interfaces in the target structure, and all interfaces are formed from particles of matching colors, $c_i = c_j$, separated at their optimal LJ distance of $r_0 = 2^{1/6}$.

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

First, we map the (T,ε')-domain where the complete one-pot self-assembly of a single chain of N distinguishable square blocks is possible. The quality of the resultant structure is judged based on its *completeness*, i.e., the length of the assembled structure divided by N, and its *correctness*, i.e., the fraction of correct interfaces within the cluster. These are used here to define an *addressability score* α = (completeness × correctness), evaluated at the end of a simulation. Three categories are introduced to describe the sampling outcome based on the addressability scores: (i) *addressable*, indicating a successfully assembled target structure (α = 1), (ii) *partial*, indicating that some portions of the target structure were successfully assembled ($0.8 \leqslant \alpha \leqslant 0.99$), and (iii) *failed*, indicating that the blocks were unable to bind and/or the developed structure contains multiple incorrect interfaces (α < 0.8).

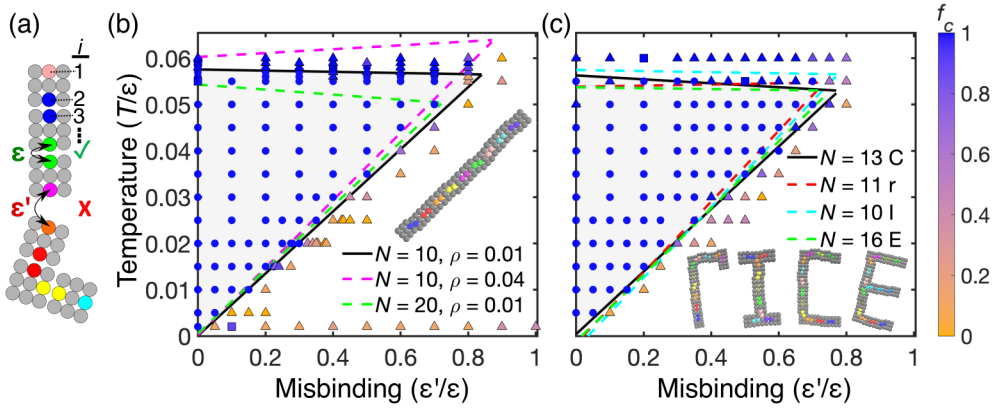

Figure 1. CG-MD simulations of self-assembled programmable 2D structures. **(a)** Schematic of the generalized CG model. **(b-c)** Addressability of various structures, made of N distinguishable blocks, and block area densities, ρ, as a function of temperature and misbinding (normalized by ε = 1, const). **(b)** Chains of N = 10 or 20 blocks and ρ = 0.01 or 0.04 **(c)** Letters "r", "I","C", and E" of N = 11, 10, 13, and 16 blocks at ρ = 0.0069, 0.01, 0.0081, and 0.01, respectively. Data points are sampled (T,ε')-combinations from the MD simulation for a chain of N = 10 blocks at ρ = 0.01 in (b) and the "C" of N = 13 blocks at ρ = 0.0081 in (c). The shading inside the solid black boundaries in (b) and (c) mark the region of complete addressability (α = 1) for the same two systems. Symbols represent addressability level: ○ – addressable, □ – partial, ▵ – failed. Colors indicate the average fraction of correct interfaces within each cluster at the end of the simulation, f_c. Correctly assembled structures are shown in the inset.

For illustration, α(T,ε') is sampled for a chain of N = 10. This reveals a limited (T,ε')-domain, bounded by $T \simeq$ constant (const) ≡ T_c, and T/ε' ≃ const, where individual components can be successfully assembled into a completely addressable (α = 1) structure (Figure 1b). At $T > T_c$ the assembled structures are correct, but incomplete, unable to reach the desired target size of N = 10. This suggests that native bonds are unstable for $T > T_c$, akin to melting: the energy gain (-ε) in forming correct bonds is less than the entropic cost associated with such assembly, $< T\Delta s$, implying $T_c = -\varepsilon/\Delta s$. The weak sensitivity of T_c to changes in ε', along with the appearance of only correct bonds at the final simulation state above this limit (Figure 1b), confirms the instability of ε bonds above T_c. Thus T_c serves as the "ceiling temperature", above which assembly is impossible. This is analogous to the balance between entropy and enthalpy (ΔH_p) governing the T_c for addition polymerization through Dainton's equation, $T_c \sim -|\Delta H_p|/\ln(\rho)$ or $T_c/|\Delta H_p| \sim 1/\ln(V)$, where ρ and V are the monomer concentration and system volume.[45] Notably, this equation suggests that T_c should increase with increasing concentration, which is also seen in our simulations utilizing different block area concentrations, $\rho = N A_p/A$ (A_p and A are the areas of one block and of simulation box). For example, the T_c for a chain of N = 10 is higher at ρ = 0.04 compared to ρ = 0.01 (Figure 1b). However, our particles are distinguishable in contrast to the indistinguishable monomers in addition polymerization, which entails that each particle incorporated into the growing structure is associated with an additional entropy cost of $\Delta s_{dist} \sim -\ln(N!)/N$.[37,46] Thus, a lower T_c is expected for structures with larger N, as observed in our simulations: T_c for a chain of N = 20 is lower than that of a chain of N = 10 at the same ρ = 0.01 (Figure 1b).

The low-T limit, on the other hand, characterized by T^*/ε' ≃ const, can be attributed to the formation of kinetic traps (Figure 1b). While random collisions can lead to misbinding between two blocks, these

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

mistakes can be corrected by thermal fluctuations, provided that ε' is sufficiently small[47]. However, when T is too low or ε'/ε too close to one, for undesired bonds to be corrected, partial reversibility of bonds is hindered and results in kinetically-trapped structures.[47,48] This phenomenon is clearly demonstrated in the MD simulations: incorrect bonds formed during the assembly can be "healed" when $T \gtrsim T^*$ and/or $\varepsilon'/\varepsilon \lesssim \varepsilon'^*/\varepsilon$, whereas assembly outside these conditions results in kinetic-traps, structures with multiple incorrect bonds (Figure 1b and S2). This suggests that the low-T limit arises because the entropic gain, $T\Delta s$, for a chain of N blocks fragmenting into two shorter pieces is balanced by the increase in bond energy, $\Delta u = \varepsilon'$. The change in free energy is thus $\Delta f = \Delta u^* - T^*\Delta s = 0$, thereby $T^*/\varepsilon'^* \sim 1/\Delta s$. As this entropy change should be relatively insensitive to small changes in concentration, size, and target structure shape, we observed a nearly constant T^*/ε'^* for all our simulated structures of slightly varying N, ρ, and shapes (Figure 1b-c, Table S1).

While our simulations are performed with only a single copy of a structure, practical self-assembly involves synthesizing multiple copies at once.[9–11,15] However $\alpha(T,\varepsilon')$ should be similar regardless of the number of copies, as all interactions between blocks are identical between the two except for the possibility of self-interaction in the latter. We confirmed such behavior through simulations of nine copies of the N = 10 chain, which displays a similar $\alpha(T,\varepsilon') = 1$ domain as that of a single copy of the chain (Figure S3a, Table S1). Moreover, experiments often create "closed" structures,[9–11,15] as opposed to the "open" structures (such as chains, and shapes without closed loops) we have thus far examined. The assembly outcome of such closed structures is harder to predict, with success highly dependent on the sub-structures that develop during the synthesis pathway. For example, along the pathway to forming a 3X3 square tile typical of DNA tiles, smaller correct structures can form, but are unable to join together due to steric hindrance (Figure S4). Nevertheless, we observe that closed structures such as the 3×3 tile also exhibit a broad triangular domain of $\alpha(T,\varepsilon') = 1$. Thus the behavior from our CG-MD results should be broadly applicable to practical self-assembly experiments.

Application to DNA-based interaction specificity. To account for the interaction energy distribution characteristic of DNA-based self-assembly, we first analyze the energies associated with a specific single-stranded DNA (ssDNA) sequence binding to all other ssDNA sequences in typical one-pot DNA self-assembly (Figure 2). The DNA *hybridization free energy*, i.e., the free energy ΔG associated with two ssDNA sequences from solution forming a DNA duplex, was determined using the nearest-neighbor model from ref.[49], allowing for internal and terminal mismatches, and dangling ends (see Supporting Information, Section S3 for more details).

As a representative set, we use unique and non-self-complementary nucleotide (nt) sequences from ref.[39]. Each brick is a 32 nt ssDNA molecule containing 4 potential binding interfaces, each 8 nt long. To analyze ΔG, we assume that the four 8-nt sequences are separated from each other, each being free to interact with any other of the 3992 sequences (including itself). However, being mainly concerned with the highest misbinding a particular oligonucleotide sequence could encounter within the one-pot mixture, we do not analyze the energies between all possible pairs. Instead, we use the DNA alignment program MAFFT (v7)[50] to determine the top 100 sequences that can bind to the target sequence in alignments that are most similar to the correct native alignment/base-pairing (i.e. similar nucleobase identity) between the target sequence and its reverse complement (Figure S5). We assert that this alignment achieves the lowest ΔG any one sequence can have with the target sequence, thus resulting in the strongest misbinding (see Figure S5 and Supporting Information, Section S3 for more details).

An example of the ΔG distribution associated with one sequence binding to the other sequences is shown in Figure 2a. It reveals that most non-native interactions, of strengths $\{\varepsilon''\}$, yield noticeably higher free energy than the native with correct WC base-pairing, ε. However, one sequence (or a few) tends to be very similar to the reverse complement, differing by only a few nucleotides. The energy -ε' of this "closest non-native interaction" is lower than for any other "random" interaction -ε" (where -ε" tends to be positive due to lack of complementarity between the sequences), and is closest to that of, but distinct from the native interaction, -ε. This smallest, but always present nonzero "energy gap" between -ε' and -ε is essential, ensuring that errors-mutations can only occur infrequently, and is deeply related to the quantum nature of

https://doi.org/10.26434/chemrxiv-2024-q1pfb ORCID: https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

molecular interactions, as noted by E. Schrödinger in his seminal book.[51] The multimodal energy distribution is universal not only among all sequences within this chosen ensemble (Figure 2b), but also in other randomly-generated sequences of various nucleotide lengths (Figure 2c, S8). Importantly, such analysis yields a quantitative measure of both the largest misbinding ε'/ε, and random misbinding ε"/ε, a sequence can encounter (Figure 2c). Both kinds of misbinding decrease with increasing temperature (Figure S7) and increasing sequence length (Figure 2c). The former is due to a greater entropic cost of duplex formation at higher T ($-T\Delta S$, $\Delta S < 0$), and the latter is expected as the probability of nucleotide mismatches between two sequences increases with longer sequences.

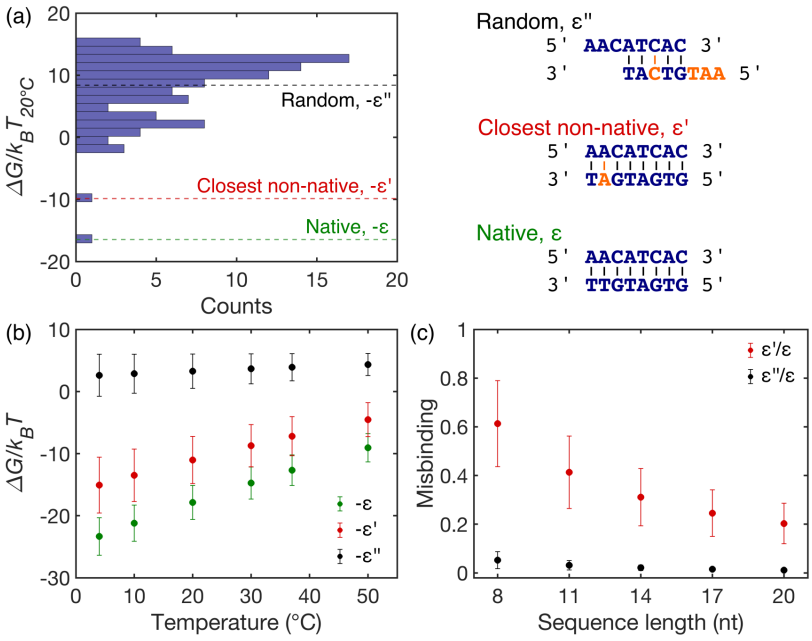

Figure 2. DNA hybridization energy distribution of sequences in typical one-pot self-assembly. **(a)** Histogram of the free energy ΔG (at T = 20°C) of a specific 8 nt ssDNA sequence (5'-AACATCAC-3') bound to other sequences. Distinct energy ranges correspond to different types of sequences, as shown on the right. **(b)** Average -ε, -ε', and -ε" at various T, obtained by performing the analysis in (a) for all 3333 (out of 3992) unique and non self-complementary 8-nt sequences from ref.[39] **(c)** Mean "largest misbinding" (ε'/ε) and mean "random misbinding" (ε"/ε) as a function of sequence length at 20°C, achieved by appending random nucleotides to the original 8 nt sequences. Error bars indicate standard deviation.

The three energy parameters (ε, ε', and ε") are then incorporated in our CG models to better capture the "multimodal" ΔG distribution associated with DNA-based assembly. The revised interaction potential between colored particles i and j reads

$$V_{i,j}(r) = \{\delta(c_i,c_j)\, U_{\mathrm{LJ}}(r,\varepsilon) + [1 - \delta(c_i,c_j)]\, U_{\mathrm{LJ}}(r,\varepsilon")\}\, [1 - \delta(j,m^*(i))] + \delta(j,m^*(i))\, U_{\mathrm{LJ}}(r,\varepsilon'), \qquad (2)$$

which is an extension of Equation 1 to include an additional ε'-term: a one-to-one mapping was created between each colored particle i and a randomly assigned, but non-matching colored, particle $m^*(i)$. Consequently, a colored particle i can form one of three interactions (Figure 3a): 1) correct native interaction between particles of matching colors, $c_i = c_j$, conducive to forming the designed target structure with a strength of ε, 2) non-native interaction between most edges of different colors, with a strength of ε", and 3) non-native interaction with only one differently colored particle, between i and $j = m^*(i)$, with a strength of ε' (mimicking the tendency of a sequence to have only a single or few ε' interactions with other sticky end sequences in the assembly of DNA-tiles).

The extended CG model is used to sample $\alpha(T,\varepsilon',\varepsilon")$ for a chain of N = 10 at ρ = 0.01 (Figure 3b-d). Complete addressability is now achieved within a polyhedral domain in the $(T,\varepsilon',\varepsilon")$ space (Figure 3b), formed primarily from three planes. The high-T plane, $T_c \simeq 0.054$, is associated with the ceiling temperature described previously. There are now two low-T limit planes, associated with measuring addressability at low or high ε'/ε. These correspond to the $T^*/\varepsilon"^*$ and T^*/ε'^* planes respectively, as the competition between ε' and ε" determines the extent of addressability. For example, at ε"/ε = 0.05, we observed that the critical T^* is largely dominated by the T^*/ε'^* plane: since ε' < ε" (|ε'| > |ε"|), all incorrect bonds tend to be ε'- rather than ε"-bonds (Figure 3c). Consequently, T^*/ε'^* is due to a kinetic trap from incorrect ε' bonds. As ε"/ε is increased to 0.4, ε"-bonds begin to dominate the incorrect structures whenever ε"/ε > ε'/ε, thereby shrinking the α = 1 domain (Figure 3d). Practically, however, ε'/ε tends to be much larger than ε"/ε and ε"/ε is near 0,

https://doi.org/10.26434/chemrxiv-2024-q1pfb ORCID: https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

according to our DNA hybridization energy analysis, irrespective of T and sequence length (Figure 2c, S8c). Thus, the T^*/ε''^* plane can be ignored, and addressability for DNA-enabled self-assembly will be determined by T_c and T^*/ε'^*, akin to Figure 3c.

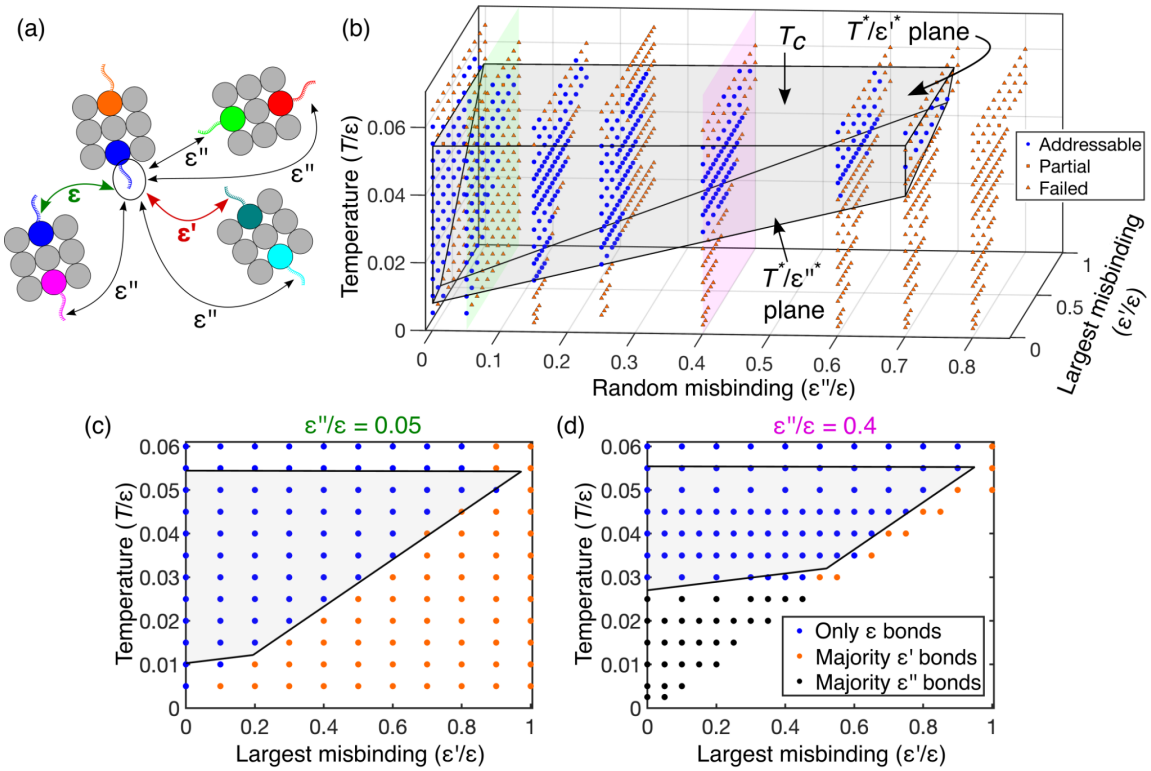

Figure 3. CG-MD simulation of a chain with 10 distinguishable square blocks, with an energy distribution informed by DNA-mediated one-pot self-assembly experiments. **(a)** Schematic of the revised CG model. **(b)** 3D plot of $\alpha(T,\varepsilon',\varepsilon'')$ at constant $\varepsilon \equiv 1$, and $|\varepsilon''|, |\varepsilon'| \leqslant |\varepsilon|$. Data points are sampled $(T,\varepsilon',\varepsilon'')$ combinations from the MD simulations of the same structure at $\rho = 0.01$. Symbols represent addressability classifications: ○ – addressable, □ – partial, ▵ – failed. The green and magenta planes on the right plot correspond to planes of constant $\varepsilon''/\varepsilon = 0.05$ and $\varepsilon''/\varepsilon = 0.4$, respectively. **(c-d)** Types of bonds formed at constant $\varepsilon''/\varepsilon = 0.05$ (c) or $\varepsilon''/\varepsilon = 0.4$ (d). Blue dots indicate that only correct (ε) bonds are present at the end of the simulation, orange dots indicate that there are more ε'- than ε''-strength bonds in the final structure, and black dots indicate the opposite. Solid black lines are defined by the three planes indicated in (b).

The DNA-informed CG models enabled us to start from considering all possible 4^l sequences of a specific length l, and establish rules that will narrow down this complete ensemble to create smaller ensembles prime for programmable self-assembly (Figure 4, S9). Every sequence within these optimized ensembles has its ε'/ε less than a selected cutoff, $(\varepsilon'/\varepsilon)_c = 0.6$, with this value indicated by the T^*/ε'^* boundary of the $\alpha = 1$ region in our addressability plots (Figure 1b-c, 3c). We also chose annealing temperatures of T/ε between 0.040 and 0.053 (Figure 1b-c, 3c) as a higher annealing T slightly below T_c would help suppress the nucleation of incorrect partial structures that could frustrate and prolong successful assembly.[52] Assuming the average energy for a WC nearest neighbor pair[49] at the appropriate temperatures, this imposes an annealing temperature at 20°C for 9 nt and 8 nt sequences (corresponding to $T/\varepsilon \sim 0.041$ and $T/\varepsilon \sim 0.047$, respectively), 10°C for 7 nt sequences ($T/\varepsilon \sim 0.047$), and 4°C for 6 nt sequences ($T/\varepsilon \sim 0.052$), all of which fit within the desired T/ε range.

Some rules in the sequence design procedure are based on the identity of each DNA sequence. Palindromic sequences must be eliminated, as they could allow a distinct block to bind to another block in an incorrect reversed orientation. Self-complementary sequences must also be excluded, as a distinct block decorated with a self-complementary sequence could erroneously bind to another block of the same type when multiple copies of blocks are present. Other rules are based on the differences between one sequence

https://doi.org/10.26434/chemrxiv-2024-q1pfb ORCID: https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. License: CC BY 4.0

and all the other sequences in the ensemble. For example, if a sequence is highly similar to another sequence but differs only by a couple of nucleotides, there may be a high chance of misbinding between the former sequence and the latter sequence's reverse complement. These problematic sequences can be eliminated by using the nearest neighbor model[49] to estimate the critical number of nucleotide mismatches, below which $\varepsilon'/\varepsilon > (\varepsilon'/\varepsilon)_c$ (specifically the critical number of single internal (m_s*), double internal (m_d*), and terminal (m_t*) mismatches, see Figure 5a and Supporting Information, Section S5 for the derivation of Equations S2–S4). We also eliminated sequences which could form dangling ends shorter than or equal to a critical value, $d_l \leq d_l$*, as it would lead to duplexes within the ensemble with $\varepsilon'/\varepsilon > (\varepsilon'/\varepsilon)_c$ (see Equation S5, and Supporting Information, Section S5).

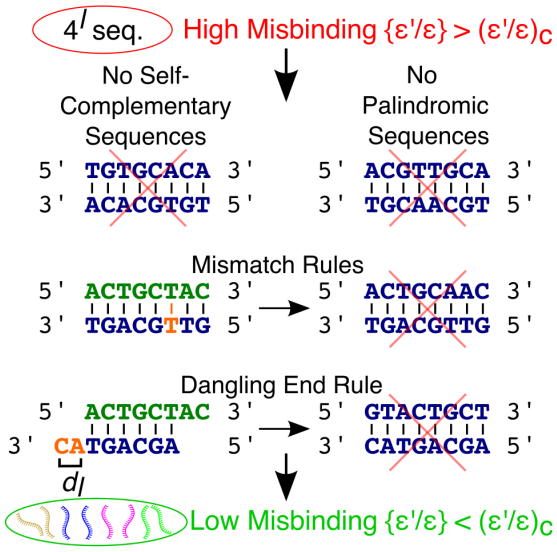

Figure 4. Schematic of DNA sequence design rules to generate sequence ensembles with acceptably low misbinding. Rules filter the complete statistical ensemble of sequences of a given length, l, to extract a set of sequences that could ensure programmable self-assembly, where each sequence has its strongest misbinding, ε'/ε, below the desired cutoff, $(\varepsilon'/\varepsilon)_c$. Sequence highlighted in green illustrates that a target sequence in the ensemble is randomly chosen to be kept, and all other sequences that bind to it too strongly, $\varepsilon'/\varepsilon > (\varepsilon'/\varepsilon)_c$, are removed (their complementary sequences are also removed).

Application of these rules to the complete ensemble of 4^l sequences not only informs the design of this optimal ensemble of different sequences with minimal misbinding, but also provides an estimate for the largest number of unique sequences that can be combined during one-pot self-assembly (Table 1, see the full procedure in Supporting Information, Section S5). Interestingly, application of the terminal mismatch rule results in the same number of sequences for all the hundred sequence ensembles generated for each l, yet each of them contain a unique set of sequences. This may be because application of the rule to a specific sequence corresponds to eliminating the other 15 (out of 16) different ways nucleotides can be arranged at the terminal positions. We also noticed that due to the discrete nature of the rules in Equations S2–5, there is a non-monotonic reduction of sequences across different lengths. For example, the average number of permitted sequences after application of the double internal mismatch rule for 6 nt sequences is higher than that for 7 nt (46 vs. 39, respectively) as m_d* jumps from 1 to 2, respectively. This results in the final number of sequences within a collection with $\varepsilon'/\varepsilon \leq (\varepsilon'/\varepsilon)_c$ for an ensemble of 7 nt sequences to be comparable with that of 6 nt sequences.

To verify that our optimal set does have low misbinding, we took the largest optimized ensemble of 8 nt sequences we found, containing 48 sequences, and analyzed ΔG of all the possible interactions (Figure 5b-c). Figure 5b reveals a nearly ideal energy distribution for self-assembly, with the native interaction ε being generally well-separated in energy from both non-specific interactions ε' and ε''. Importantly, all sequences within this ensemble satisfy $\varepsilon'/\varepsilon < (\varepsilon'/\varepsilon)_c = 0.6$ as desired (Figure 5c). However, since the sequence design scheme ignores the potentially high misbinding that could result after the formation of the dangling end, a small fraction of sequences can exhibit $\varepsilon'/\varepsilon > (\varepsilon'/\varepsilon)_c$ (Figure S10). This would still allow for reliable self-assembly if these problematic sequences were removed. Nevertheless, our design scheme is able to create an ensemble of sequences with much smaller misbinding values (mean of $\varepsilon'/\varepsilon = 0.21$ in Figure 5c) than ensembles that are randomly generated (mean of $\varepsilon'/\varepsilon = 0.61$ in Figure 2c), an approach commonly used in experiments.[10–12,53] Thus our rational design procedure is expected to greatly improve the chance of successful programmable self-assembly of various structures using DNA (collection of sequences used in Figures 5c and S10 is provided in Supporting Information, Section S6).

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Table 1: Process of creating ensembles of sequences with low misbinding and corresponding limits on the size of desired target structures.

Side of block sequence length ***l*** nt	6	7	8
Annealing temperature, T/ε	4°C (0.052)	10°C (0.047)	20°C (0.047)
Number of sequences in Complete Statistical Ensemble, 4^l	4096	16384	65536
Number of non- palindromic or self-complementary sequences	3968	16128	65024
Average number of sequences after terminal mismatch rule (± Standard Deviation, SD)	240 (±0)	1024 (±0)	4032 (±0)
Average number of sequences after single internal mismatch rule (minimum-maximum)	71 (58–84)	94 (82–108)	737 (696–770)
Average number of sequences after double internal mismatch rule (minimum-maximum)	46 (32–56)	39 (30–52)	511 (478–550)
Average Number of sequences after dangling ends rule/ Average Number of sequences in an ensemble with $\varepsilon'/\varepsilon \leq (\varepsilon'/\varepsilon)_c$ (minimum-maximum)	3 (0–8)	6 (0–12)	33 (16–48)
Largest chain of size ***N*** possible	5	7	25
Largest square tile of size ***N*** × ***N*** possible[a]	2 × 2	2 × 2	4 × 4

[a]Since the development of square tiles is prone to steric hindrance (Figure S4), these optimal sequence ensembles may not be able to guarantee assembly of multiple copies of square tiles. However, such limits could still improve the yield, as one factor that frustrates successful self-assembly, misbinding, is mitigated.

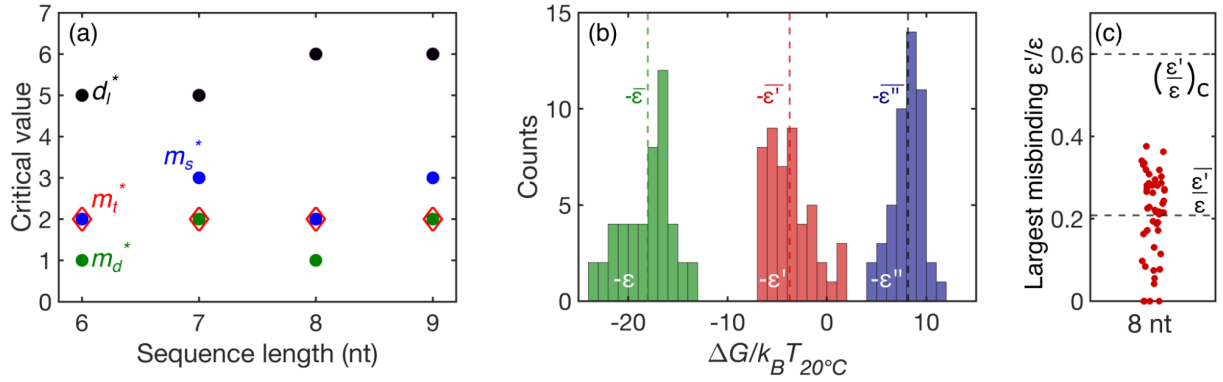

Figure 5. Designed ensembles of sequences with low misbinding. **(a)** Plot of sequence length (number of nucleotides) vs. critical number of single internal (m_s^*, blue dots), double internal (m_d^*, green dots), terminal mismatches (m_t^*, red diamonds), and critical length of dangling ends (d_l^*, black dots), above which programmable self-assembly is possible as $\varepsilon'/\varepsilon < (\varepsilon'/\varepsilon)_c$. The points are derived from Equations S2–S5, respectively, with a cutoff of $(\varepsilon'/\varepsilon)_c = 0.6$ and T/ε between 0.04 and 0.053. **(b)** DNA-hybridization energy histogram of -ε, -ε', and -ε'' each sequence within a designed ensemble of 48X8 nt sequences would encounter. Dashed lines indicate the overall average -ε, -ε', and -ε''. **(c)** Plot of largest misbinding, ε'/ε, associated with each sequence in the ensemble. Points are horizontally offset for clarity. Dashed lines indicate the mean and cutoff largest misbinding.

Moreover, we can estimate the limits of the size of target structures that can be successfully assembled (Table 1). For reference, a chain containing N blocks requires $2(N\text{-}1)$ unique sequences (N-1 sequences and N-1 reverse complements) and a square tile of size $N \times N$ blocks requires $4N(N\text{-}1)$ unique sequences. Interestingly, this sequence design scheme and analysis suggest that having longer sequences does not immediately entail their larger number, viz. blocks, that can be used for self-assembly. For

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

example, after application of our established rules, we find that the number of sequences in the 6 nt ensembles, which can contain anywhere from 0 to 8 unique sequences (0 means that $\varepsilon'/\varepsilon < (\varepsilon'/\varepsilon)_c = 0.6$ was not possible), is comparable to that of the 7 nt ensembles, which can contain 0 to 12 sequences. Consequently, the sequences in some 7 nt ensembles may not be able to create chains as large as those in some 6 nt ensembles. However, longer sequences can still generally increase the number of optimal sequences within the curated ensemble and accordingly the maximum size of the target structure: we found that an ensemble of 8 nt can contain a maximum of 48 optimal sequences after sequence optimization, which could create a 4 × 4 tile that is larger than the 2 × 2 tile possible using a 6 or 7nt sequence. We note that the maximum structure sizes denoted here are highly conservative since they were based on utilizing the most stringent thermodynamic parameters that would allow for reliable and complete addressability with perfect yield, i.e., $\alpha = 1$. Larger structures could be synthesized by application of the rules with larger $(\varepsilon'/\varepsilon)_c$, but at the expense of yield.

Conclusions

CG MD simulations combined with analytical models elucidate the range of temperatures and misbinding energies that allow for the successful one-pot self-assembly of distinguishable blocks into intrinsically pre-programmed structures. Accounting for the multimodal energy distribution (characteristic of DNA-based assembly) in the CG model shows that the success of programmable assembly is determined not by the random background misbinding, $\varepsilon''/\varepsilon$, but rather the strongest (closest to native) misbinding a sequence encounters, ε'/ε. Since ε'/ε is associated with sequences that are highly similar to the reverse complement of a chosen sequence, we were able to establish rules for generating optimal collections of sequences with low misbinding, and to estimate the maximum possible sizes of DNA-based structures that can be self-assembled with high fidelity and yield.

Importantly, our design scheme can not only increase the efficiency of one-pot self-assembly, but in hierarchical assembly as well. Instead of combining, for example, two blocks per step during step-by-step assembly,[9,10] our findings suggest that one can combine more blocks into each step, provided that the development of smaller target substructures are not prone to steric hindrance problems. This would dramatically improve the low yields of current hierarchically assembled structures, and perhaps decrease the complexity involved in such assembly.

The present theoretical exploration provides initial guidance for experimental realization of programmability, via ssDNA "functionalization", for achieving self-assembled carbon-based architectures for nanoelectronics.[54] Indeed, early attempts indicate that DNA functionalization of nanocarbons, such as graphene[24] or CNTs[25] may be possible although a synthetically challenging task. Quantum transport calculations also suggest that such components, even in presence of hydrogen-bonded junctions, may adequately function as electron transport media between GNRs.[55] DNA-functionalization may also improve upon existing liquid-phase processing methods in creating highly compact and aligned nanosheets for electronics.[41–43] This molecular precision of nanosheet assembly can be accomplished both in 2D and potentially 3D via the incorporation of multiple different DNA sequences on one edge[56] and the use of flexible linkers that enable additional shape (instead of solely chemical) complementarity.[57] Our results also have an interesting parallel to naturally assembled structures. Years of evolution have determined that three nucleotides are optimal and sufficient to specify *codons* that can generate an overwhelming diversity of proteins and biological structures. Similarly, we suggest that the energy scales associated with DNA specify an optimal size and heterogeneity of sequences that can reliably self-assemble into a diverse gamut of complex nanostructures.

Methods

Coarse-Grained Molecular Dynamics Simulations. NVT simulations are done at constant *T*, constrained in 2D, and under periodic boundary conditions. Blocks are treated as rigid bodies in LAMMPS.[58] We use a standard 12-6 LJ potential between colored beads, with $\sigma = 1$ and a cutoff of 2.3σ. The latter is used to improve the alignment of blocks as they are incorporated into the growing structure. For each simulation, all correct interactions were kept constant at a strength of $\varepsilon = 1$, while all incorrect interactions were set at

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

a constant strength of ε', ε'' < ε = 1. Simulations were performed at different (T, ε') and block area concentrations ρ (through changing the simulation box sizes) to test if the desired target structure can be successfully assembled under those conditions. To ensure that the target structure has enough time to develop and is stable, simulations run until fluctuations in the potential energy $\delta U < \varepsilon$ for a minimum number of timesteps t_c, with integration timestep $\delta t = 0.022$ (Figure S1). As the more complex structures require longer times to fully develop, the t_c were increased as necessary: $t_c \sim 20–50 \times 10^6$ timesteps for one copy of the chains and simulated letters, but can reach as high as 10^8 timesteps to assemble 9 copies of the $N = 10$ chain.

Calculation of DNA hybridization energies and misbinding. The DNA hybridization free energy, ΔG is calculated using the nearest neighbor model established by SantaLucia Jr and colleagues,[49,59] which provides experimental thermodynamic parameters for $\Delta H_{37°C}$ and $\Delta S_{37°C}$ of correct WC base-pairing, single, double and terminal mismatches, and dangling-ends. These quantities can be used to calculate ΔG at the various temperatures used in our study, $\Delta G(T) = \Delta H_{37°} - T\Delta S_{37°C}$, as it's widely-accepted that ΔH and ΔS are largely independent of temperature for DNA[49] (see Supporting Information, Section S3, for more details). We note that ΔG for two sequences with numerous mismatched base pairs, even in its most energetically favorable alignment can lead to $\Delta G > 0$, meaning -ε'' or -ε' can be greater than 0. In this case, the corresponding misbinding, ε'/ε and ε''/ε are assumed to be 0 since thermodynamically unfavorable binding would not result in attractive misbinding.

Design and application of rules to create optimized sequence collections. Equations for the selection rules (such as the terminal and various internal mismatch rules), are formulated based on the of the weakest complementary $|\Delta G|$ and the strongest possible mismatch/dangling ends $|\Delta G|$, which would consequently lead to the highest misbinding (see Supporting Information, Section S5 and Table S2 for the detailed procedure, parameters, and derivation of Equations S2–S5 corresponding to the selection rules). All the selection rules described are applied sequentially to every sequence. For example, the single internal mismatch rule is applied in relation to the first sequence in the remaining ensemble (after application of the previous rules). After eliminating all sequences in the collection that have $m_s \leq m_s^*$ the same rule is applied to the second sequence and its problematic sequences are removed. Once the rule is applied to the last remaining sequence in the ensemble, the next rule is then used on the first sequence, and this sequential process is repeated until all the rules are applied. Due to the sequential procedure, the sequence order in the latter greatly influences the size and sequence types in the final list of optimal sequences. The "earlier" sequences in the list will be used as a reference to discard others that appear later due to the high ε'/ε that can occur. Thus, after obtaining a set of non-palindromic and non- self-complementary sequences from the complete statistical ensembles, this list was randomly shuffled 100 times and then subjected to the remaining rules to obtain 100 different ensembles per sequence length tested (Table 1).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/…

> Additional examples and details on the coarse-grained molecular dynamics simulations of programmable structures; details on the calculation of DNA hybridization energies and alignments; full procedure to generate, and examples of, optimized sequence ensembles (PDF)

Notes

The authors declare no competing financial interest.

Acknowledgements

This work was supported by the Office of Naval Research Grants N00014-19-1-2191, N00014-22-1-2753, and by the NASA Space Technology Graduate Research Opportunity, Grant 80NSSC21K1299, for N.T. Computer resources were provided through allocation DMR100029 from the ACCESS program, supported by NSF grants #2138259, #2138286, #2138307, #2137603, and #2138296.

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

References

(1) Cai, J.; Ruffieux, P.; Jaafar, R.; Bieri, M.; Braun, T.; Blankenburg, S.; Muoth, M.; Seitsonen, A. P.; Saleh, M.; Feng, X.; Müllen, K.; Fasel, R. Atomically Precise Bottom-up Fabrication of Graphene Nanoribbons. *Nature* **2010**, *466* (7305), 470–473.

(2) Moreno, C.; Vilas-Varela, M.; Kretz, B.; Garcia-Lekue, A.; Costache, M. V.; Paradinas, M.; Panighel, M.; Ceballos, G.; Valenzuela, S. O.; Peña, D.; Mugarza, A. Bottom up Synthesis of Multifunctional Nanoporous Graphene. *Science* **2018**, *360* (6385), 199–203.

(3) Mehdi Pour, M.; Lashkov, A.; Radocea, A.; Liu, X.; Sun, T.; Lipatov, A.; Korlacki, R. A.; Shekhirev, M.; Aluru, N. R.; Lyding, J. W.; Sysoev, V.; Sinitskii, A. Laterally Extended Atomically Precise Graphene Nanoribbons with Improved Electrical Conductivity for Efficient Gas Sensing. *Nat. Commun.* **2017**, *8* (1), 1–9.

(4) Wegst, U. G. K.; Bai, H.; Saiz, E.; Tomsia, A. P.; Ritchie, R. O. Bioinspired Structural Materials. *Nat. Mater.* **2015**, *14* (1), 23–36.

(5) Eder, M.; Amini, S.; Fratzl, P. Biological Composites—Complex Structures for Functional Diversity. *Science* **2018**, *362* (6414), 543–547.

(6) Yang, W.; Quan, H.; Meyers, M. A.; Ritchie, R. O. Arapaima Fish Scale: One of the Toughest Flexible Biological Materials. *Matter* **2019**, *1* (6), 1557–1566.

(7) Nepal, D.; Kang, S.; Adstedt, K. M.; Kanhaiya, K.; Bockstaller, M. R.; Brinson, L. C.; Buehler, M. J.; Coveney, P. V.; Dayal, K.; El-Awady, J. A.; Henderson, L. C.; Kaplan, D. L.; Keten, S.; Kotov, N. A.; Schatz, G. C.; Vignolini, S.; Vollrath, F.; Wang, Y.; Yakobson, B. I.; Tsukruk, V. V.; Heinz, H. Hierarchically Structured Bioinspired Nanocomposites. *Nat. Mater.* **2023**, *22* (1), 18–35.

(8) Cademartiri, L.; Bishop, K. J. M. Programmable Self-Assembly. *Nat. Mater.* **2015**, *14* (1), 2–9.

(9) Park, S. H.; Pistol, C.; Ahn, S. J.; Reif, J. H.; Lebeck, A. R.; Dwyer, C.; LaBean, T. H. Finite-Size, Fully Addressable DNA Tile Lattices Formed by Hierarchical Assembly Procedures. *Angew. Chem. - Int. Ed.* **2006**, *45* (5), 735–739.

(10) Tikhomirov, G.; Petersen, P.; Qian, L. Fractal Assembly of Micrometre-Scale DNA Origami Arrays with Arbitrary Patterns. *Nature* **2017**, *552* (7683), 67–71.

(11) Wei, B.; Dai, M.; Yin, P. Complex Shapes Self-Assembled from Single-Stranded DNA Tiles. *Nature* **2012**, *485* (7400), 623–626.

(12) Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012**, *338* (6111), 1177–1183.

(13) Bhatia, D.; Wunder, C.; Johannes, L. Self-Assembled, Programmable DNA Nanodevices for Biological and Biomedical Applications. *ChemBioChem* **2021**, *22* (5), 763–778.

(14) Rogers, W. B.; Shih, W. M.; Manoharan, V. N. Using DNA to Program the Self-Assembly of Colloidal Nanoparticles and Microparticles. *Nat. Rev. Mater.* **2016**, *1*.

(15) Wang, W.; Lin, T.; Zhang, S.; Bai, T.; Mi, Y.; Wei, B. Self-Assembly of Fully Addressable DNA Nanostructures from Double Crossover Tiles. *Nucleic Acids Res.* **2016**, *44* (16), 7989–7996.

(16) Lewis, D. J.; Zornberg, L. Z.; Carter, D. J. D.; Macfarlane, R. J. Single-Crystal Winterbottom Constructions of Nanoparticle Superlattices. *Nat. Mater.* **2020**, *19* (7), 719–724.

(17) Tian, Y.; Lhermitte, J. R.; Bai, L.; Vo, T.; Xin, H. L.; Li, H.; Li, R.; Fukuto, M.; Yager, K. G.; Kahn, J. S.; Xiong, Y.; Minevich, B.; Kumar, S. K.; Gang, O. Ordered Three-Dimensional Nanomaterials Using DNA-Prescribed and Valence-Controlled Material Voxels. *Nat. Mater.* **2020**, *19* (7), 789–796.

(18) Kuzyk, A.; Schreiber, R.; Fan, Z.; Pardatscher, G.; Roller, E.-M.; Högele, A.; Simmel, F. C.; Govorov, A. O.; Liedl, T. DNA-Based Self-Assembly of Chiral Plasmonic Nanostructures with Tailored Optical Response. *Nature* **2012**, *483* (7389), 311–314.

(19) Acuna, G. P.; Möller, F. M.; Holzmeister, P.; Beater, S.; Lalkens, B.; Tinnefeld, P. Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas. *Science* **2012**, *338* (6106), 506–510.

(20) Wintersinger, C. M.; Minev, D.; Ershova, A.; Sasaki, H. M.; Gowri, G.; Berengut, J. F.; Corea-Dilbert, F. E.; Yin, P.; Shih, W. M. Multi-Micron Crisscross Structures Grown from DNA-Origami

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

Slats. *Nat. Nanotechnol.* **2023**, *18* (3), 281–289.
(21) Derr, N. D.; Goodman, B. S.; Jungmann, R.; Leschziner, A. E.; Shih, W. M.; Reck-Peterson, S. L. Tug-of-War in Motor Protein Ensembles Revealed with a Programmable DNA Origami Scaffold. *Science* **2012**, *338* (6107), 662–665.
(22) Ross, M. B.; Ku, J. C.; Vaccarezza, V. M.; Schatz, G. C.; Mirkin, C. A. Nanoscale Form Dictates Mesoscale Function in Plasmonic DNA–Nanoparticle Superlattices. *Nat. Nanotechnol.* **2015**, *10* (5), 453–458.
(23) Ong, L. L.; Hanikel, N.; Yaghi, O. K.; Grun, C.; Strauss, M. T.; Bron, P.; Lai-Kee-Him, J.; Schueder, F.; Wang, B.; Wang, P.; Kishi, J. Y.; Myhrvold, C.; Zhu, A.; Jungmann, R.; Bellot, G.; Ke, Y.; Yin, P. Programmable Self-Assembly of Three-Dimensional Nanostructures from 10,000 Unique Components. *Nature* **2017**, *552* (7683), 72–77.
(24) Bonanni, A.; Ambrosi, A.; Pumera, M. Nucleic Acid Functionalized Graphene for Biosensing. *Chem. - Eur. J.* **2012**, *18* (6), 1668–1673.
(25) Pei, H.; Sha, R.; Wang, X.; Zheng, M.; Fan, C.; Canary, J. W.; Seeman, N. C. Organizing End-Site-Specific SWCNTs in Specific Loci Using DNA. *J. Am. Chem. Soc.* **2019**, *141* (30), 11923–11928.
(26) Zou, X.; Yakobson, B. I. An Open Canvas - 2D Materials with Defects, Disorder, and Functionality. *Acc. Chem. Res.* **2015**, *48* (1), 73–80.
(27) Zhi, C.; Bando, Y.; Wang, W.; Tang, C.; Kuwahara, H.; Golberg, D. DNA-Mediated Assembly of Boron Nitride Nanotubes. *Chem. - Asian J.* **2007**, *2* (12), 1581–1585.
(28) Hwang, Y. J.; Yu, H.; Lee, G.; Shackery, I.; Seong, J.; Jung, Y.; Sung, S.-H.; Choi, J.; Jun, S. C. Multiplexed DNA-Functionalized Graphene Sensor with Artificial Intelligence-Based Discrimination Performance for Analyzing Chemical Vapor Compositions. *Microsyst. Nanoeng.* **2023**, *9* (1), 28.
(29) Yang, S.; Zhang, F.; Wang, Z.; Liang, Q. A Graphene Oxide-Based Label-Free Electrochemical Aptasensor for the Detection of Alpha-Fetoprotein. *Biosens. Bioelectron.* **2018**, *112*, 186–192.
(30) Chen, L.; Li, G.; Yang, A.; Wu, J.; Yan, F.; Ju, H. A DNA-Functionalized Graphene Field-Effect Transistor for Quantitation of Vascular Endothelial Growth Factor. *Sens. Actuators B Chem.* **2022**, *351*, 130964.
(31) Shin, B.; Kim, W.-K.; Yoon, S.; Lee, J. Duplex DNA-Functionalized Graphene Oxide: A Versatile Platform for miRNA Sensing. *Sens. Actuators B Chem.* **2020**, *305*, 127471.
(32) Liu, S.; Fu, Y.; Xiong, C.; Liu, Z.; Zheng, L.; Yan, F. Detection of Bisphenol A Using DNA-Functionalized Graphene Field Effect Transistors Integrated in Microfluidic Systems. *ACS Appl. Mater. Interfaces* **2018**, *10* (28), 23522–23528.
(33) Sinitskii, A.; Dong, G.; Seeman, N. C.; Canary, J.; Crommie, M. F.; Lyding, J.; Aluru, N. *DNA-Enabled Hierarchical Assembly of Graphene Electronics*. Office of Naval Research Nanoelectronics Program, Virtual Meeting, July 6–9, 2020.
(34) LaBoda, C.; Duschl, H.; Dwyer, C. L. DNA-Enabled Integrated Molecular Systems for Computation and Sensing. *Acc. Chem. Res.* **2014**, *47* (6), 1816–1824.
(35) Liu, Z.; Liu, B.; Ding, J.; Liu, J. Fluorescent Sensors Using DNA-Functionalized Graphene Oxide. *Anal. Bioanal. Chem.* **2014**, *406* (27), 6885–6902.
(36) Hu, K.-M.; Guo, W.; Deng, X.-L.; Li, X.-Y.; Tu, E.-Q.; Xin, Y.-H.; Xue, Z.-Y.; Jiang, X.-S.; Wang, G.; Meng, G.; Di, Z.-F.; Lin, L.; Zhang, W.-M. Deterministically Self-Assembled 2D Materials and Electronics. *Matter* **2023**, *6* (5), 1654–1668.
(37) Whitelam, S. Hierarchical Assembly May Be a Way to Make Large Information-Rich Structures. *Soft Matter* **2015**, *11* (42), 8225–8235.
(38) Haxton, T. K.; Whitelam, S. Do Hierarchical Structures Assemble Best via Hierarchical Pathways? *Soft Matter* **2013**, *9* (29), 6851–6861.
(39) Reinhardt, A.; Frenkel, D. Numerical Evidence for Nucleated Self-Assembly of DNA Brick Structures. *Phys. Rev. Lett.* **2014**, *112* (23), 1–5.
(40) Hedges, L. O.; Mannige, R. V.; Whitelam, S. Growth of Equilibrium Structures Built from a Large Number of Distinct Component Types. *Soft Matter* **2014**, *10* (34), 6404–6416.

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0

(41) Umezaki, U.; Smith McWillams, A. D.; Tang, Z.; He, Z. M. S.; Siqueira, I. R.; Corr, S. J.; Ryu, H.; Kolomeisky, A. B.; Pasquali, M.; Martí, A. A. Brownian Diffusion of Hexagonal Boron Nitride Nanosheets and Graphene in Two Dimensions. *ACS Nano* **2024**, *18* (3), 2446–2454.
(42) Zhong, J.; Sun, W.; Wei, Q.; Qian, X.; Cheng, H.-M.; Ren, W. Efficient and Scalable Synthesis of Highly Aligned and Compact Two-Dimensional Nanosheet Films with Record Performances. *Nat. Commun.* **2018**, *9* (1), 3484.
(43) Xin, G.; Zhu, W.; Deng, Y.; Cheng, J.; Zhang, L. T.; Chung, A. J.; De, S.; Lian, J. Microfluidics-Enabled Orientation and Microstructure Control of Macroscopic Graphene Fibres. *Nat. Nanotechnol.* **2019**, *14* (2), 168–175.
(44) Huntley, M. H.; Murugan, A.; Brenner, M. P. Information Capacity of Specific Interactions. *Proc. Natl. Acad. Sci.* **2016**, *113* (21), 5841–5846.
(45) Dainton, F. S.; Ivin, K. J. Reversibility of the Propagation Reaction in Polymerization Processes and Its Manifestation in the Phenomenon of a 'Ceiling Temperature'. *Nature* **1948**, *162* (4122), 705–707.
(46) Frenkel. Order Through Entropy. *Nat. Mater.* **2015**, *14* (1), 9–12.
(47) Whitelam, S.; Jack, R. L. The Statistical Mechanics of Dynamic Pathways to Self-Assembly. *Annu. Rev. Phys. Chem.* **2015**, *66*, 143–163.
(48) Whitelam, S. Control of Pathways and Yields of Protein Crystallization through the Interplay of Nonspecific and Specific Attractions. *Phys. Rev. Lett.* **2010**, *105* (8), 1–4.
(49) SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.
(50) Katoh, K.; Rozewicki, J.; Yamada, K. D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief. Bioinform.* **2018**, *20* (4), 1160–1166.
(51) Schrödinger, E. *What Is Life? The Physical Aspect of the Living Cell ; with, Mind and Matter ; & Autobiographical Sketches*; Cambridge University Press: Cambridge ; New York, 1992.
(52) Jacobs, W. M.; Frenkel, D. Self-Assembly of Structures with Addressable Complexity. *J. Am. Chem. Soc.* **2016**, *138* (8), 2457–2467.
(53) Tikhomirov, G.; Petersen, P.; Qian, L. Programmable Disorder in Random DNA Tilings. *Nat. Nanotechnol.* **2017**, *12* (3), 251–259.
(54) Programmable Graphene Molecular Architecture. Multidisciplinary Research Program of the University Research Initiative, ONR Announcement #N00014-18-S-F006; Topic 2, Arlington, VA, 2018.
(55) Huang, Y.; Altalhi, T.; Yakobson, B. I.; Penev, E. S. Nucleobase-Bonded Graphene Nanoribbon Junctions: Electron Transport from First Principles. *ACS Nano* **2022**, *16* (10), 16736–16743.
(56) Tanaka, H.; Dotera, T.; Hyde, S. T. Programmable Self-Assembly of Nanoplates into Bicontinuous Nanostructures. *ACS Nano* **2023**, *17* (16), 15371–15378.
(57) Zhou, W.; Li, Y.; Je, K.; Vo, T.; Lin, H.; Partridge, B. E.; Huang, Z.; Glotzer, S. C.; Mirkin, C. A. Space-Tiled Colloidal Crystals from DNA-Forced Shape-Complementary Polyhedra Pairing. *Science* **2024**, *383* (6680), 312–319.
(58) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; In 'T Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.
(59) Allawi, H. T.; Santalucia, J. Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry* **1997**, *36* (34), 10581–10594.

https://doi.org/10.26434/chemrxiv-2024-q1pfb **ORCID:** https://orcid.org/0000-0003-0743-3581 Content not peer-reviewed by ChemRxiv. **License:** CC BY 4.0