Molecular computation at equilibrium via programmable entropy

Boya Wang[1,3], Cameron Chalk[1,4], David Doty[2], and David Soloveichik[1,*]

[1] *Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, USA*
[2] *Computer Science, University of California, Davis, CA 95616, USA*
[3] *Current address: Bioengineering, California Institute of Technology, Pasadena, CA 91125, USA*
[4] *Current address: Computation and Neural Systems, California Institute of Technology, Pasadena, CA 91125, USA*

Abstract

Synthetic molecular information processing is typically designed through programming kinetic pathways, so that molecules bind, unbind, or incur conformational changes in some desired order. In contrast, thermodynamic programming focuses solely on the desired end-state rather than the path, often allowing simpler reasoning and requiring fewer parameters. Thermodynamic programming also naturally avoids energetically-favored, yet undesired, "error" states that often frustrate kinetic approaches. Here we demonstrate a thermodynamics-first paradigm based on the Thermodynamic Binding Networks (TBN) model, where the minimum free-energy configuration maximizes the number of separate complexes. We construct signal propagation circuits including fan-in and fan-out, seeded-assembly systems that perform Boolean logic computation, and systems for synthesis of concatemers of size quadratic in that of the substrates (by computing their least common multiple). Our work may enable new ways to engineer complex molecular behaviors and help inform the understanding of the computational power of kinetics versus thermodynamics for molecular systems.

1 Introduction

Engineered molecular information processing has shown great promise with applications in nanoscience, synthetic biology, and therapeutics. As a biocompatible engineering material, DNA has been widely used for computation, diagnostics, nanorobotics, self-assembling structures, and super-resolution imaging tools[1–3]. Most of the molecular information processing systems are designed through programming a desired kinetic pathway: Molecules bind, unbind, or incur conformational changes to either release a kinetically-trapped output molecule, or assemble to the desired state through a specific binding order or partial order. Such kinetic programming matches the standard imperative programming practice of computer science. Although effective, such information processing often suffers from undesired side reactions that also occur, albeit at a slower rate. Mechanisms designed to reduce error usually rely on engineering error-correcting kinetic pathways, while errors are still thermodynamically preferred[4–7]. Engineering molecular systems that function correctly

at thermodynamic equilibrium could, in contrast, have robust guarantees of correctness, since the desired state is also the energetically most favorable. Moreover and more broadly, programming algorithmic behavior encoded in thermodynamics could make programming certain tasks easier by alleviating the need to reason about exponentially-many possible kinetic pathways, or could enable entirely new types of computational behaviours.

The first-principles understanding of the thermodynamics of Watson-Crick-Franklin base pairing has enabled simulation tools such as Mfold[8], NUPACK[9], oxDNA[10], allowing the engineering of diverse thermodynamic functionalities. For example, molecular probes with carefully designed free-energies were adapted for molecular diagnostics such as differentiating single nucleotide polymorphisms[11]. Controlling the equilibrium of a DNA-based simple molecular mechanism allowed programming the phase transition of nanoparticles[12]. By engineering the activation energy of intermediate species, a reversible logic circuit was designed to respond to the change of inputs and recompute the output[13]. Designing the equilibrium states of molecular systems at different temperatures allowed resetting molecular configurations through thermal cycling[14]. Carefully tuned weak interactions between pairs of strands have been shown to be capable of surprisingly complex digital and analog behavior[15]. However, we currently lack a sufficiently general abstraction at the thermodynamic level for systematic programming of systems with wide-ranging functionalities. Most designs are ad hoc for specific tasks, and understanding the thermodynamic behaviors relies heavily on the power of computational simulation, often at the computationally intensive level of individual nucleotides. To establish the design principles for systematically engineering molecular computation at thermodynamic equilibrium, it is necessary to develop a theoretical framework that is at the right level of abstraction such that the model is not far away from the underlying physics, but also simple enough for systematic investigation and rigorous mathematical proofs.

Recently the *thermodynamic binding networks* (TBNs)[16,17] model has been proposed to aid in engineering functional molecules at thermodynamic equilibrium. In typical chemical systems, the thermodynamic equilibrium is determined by a large variety of chemical, structural, geometric, and combinatorial factors. The TBN model simplifies analysis of energy landscapes by coarsely modeling molecules in a well-mixed solution under a range of experimental regimes where certain model abstractions hold. First, bonds between complementary domains are strongly favored (and non-complementary domains are assumed to be *orthogonal*, i.e., to have zero binding affinity). Second, ties between multiple configurations with the same number of (domain-level) bonds are broken in favor of configurations with more separate complexes[1]. (Intuitively, more separate complexes results in larger entropy due to additional microstates describing the position of each separate complex.) Configurations which are maximally bonded and have the largest number of separate complexes among the maximally bonded configurations capture the thermodynamically preferred equilibrium in the model. This formalization of thermodynamic stability is valid in the regime in which bonds are too strong to dissociate spontaneously (e.g., DNA domains with 15 or more base pairs at room temperature), and also sufficiently dilute that the free energy of association penalty is significant. This simple energy abstraction makes it feasible to understand and analyze the robustness of molecular systems[18–20]. In addition, complex molecular be-

[1] In this work, the term 'complex' encompasses both multi-stranded structures and the degenerate case of a single-stranded DNA molecule.

haviors have been theoretically constructed with thermodynamic guarantees of correctness, such as arbitrary Boolean circuit computation[16,21], and engineering the intermediate configurations of the kinetic pathway in catalysis[22].

The goal of this paper is to experimentally establish the TBN model as a design paradigm for engineering complex molecular systems. We start with the simplest non-trivial TBN system, demonstrating the feasibility of engineering the molecular equilibrium such that the maximally bonded configuration with the most separate complexes is strongly preferred. Then we engineer a signal propagation cascade that propagates information both "forward" and "backward", including both fan-in and fan-out. Thermodynamic control also immediately accords new functionality to the same components: for example, fan-out in reverse generates AND logic. Next we focus on algorithmic self-assembly, which, in contrast to non-algorithmic self-assembly such as DNA origami[23], is typically programmed using imperative reasoning: Molecules are required to add in a particular order, or at least corresponding to the mathematical notion of a partial order[24]. In contrast we experimentally test a seeded algorithmic self-assembly system computing Boolean logic at thermodynamic equilibrium. This self-assembly design provides another way to compute any Boolean logic circuits as well as perform algorithmic self-assembly with provable guarantees of correctness[21].

As our last construction, we develop a novel method to engineer concatemers (DNA strands consisting of repeated domains) with precisely defined size. We show that a "least common multiple" computation embedded in the thermodynamics of our system can be used to ligate a specific number of short concatemers together, allowing the total size of the resulting concatemer to be quadratic in the sizes of the starting concatemers. By programming short and cheap strands to assemble into long concatemers of specified size, our technique could be used, for example, as the template for the formation of periodic nano-structures.

In implementing the various TBN constructions we develop a methodology for faithfully realizing the TBN abstraction with DNA strands. In particular, we arranged domains on strands in an order which ensures geometric feasibility of resulting complexes (unpseudoknotted). Sequence design algorithms were used to ensure domain orthogonality and strong binding. Finally, when needed, we introduced single-base deletions to compensate for geometric energy penalties outside the scope of the TBN abstraction. By exploring the constraints the TBN model encountered for physical implementation, and finding methods to overcome these constraints, we provide a way to bridge the gap between idealized theory and realistic experimental implementation.

2 The simplest TBN

The TBN model abstracts programming the thermodynamics of a chemical system in terms of the formal *enthalpy* and *entropy* of configurations, where enthalpy is equated with the number of bonds, and entropy with the number of separate complexes.[2] Bonds are strictly favored—if one configuration has more bonds than another then the first is always preferred no matter how many more separate complexes the second has. A configuration that is

[2]TBN formal quantities of enthalpy and entropy do not unambiguously map onto physical enthalpy and entropy, but are intentional engineering simplifications. For instance, the free energy of base-pair formation has significant entropic terms, which would be lumped in TBN "enthalpy". Further, note that no temperature dependence is considered.

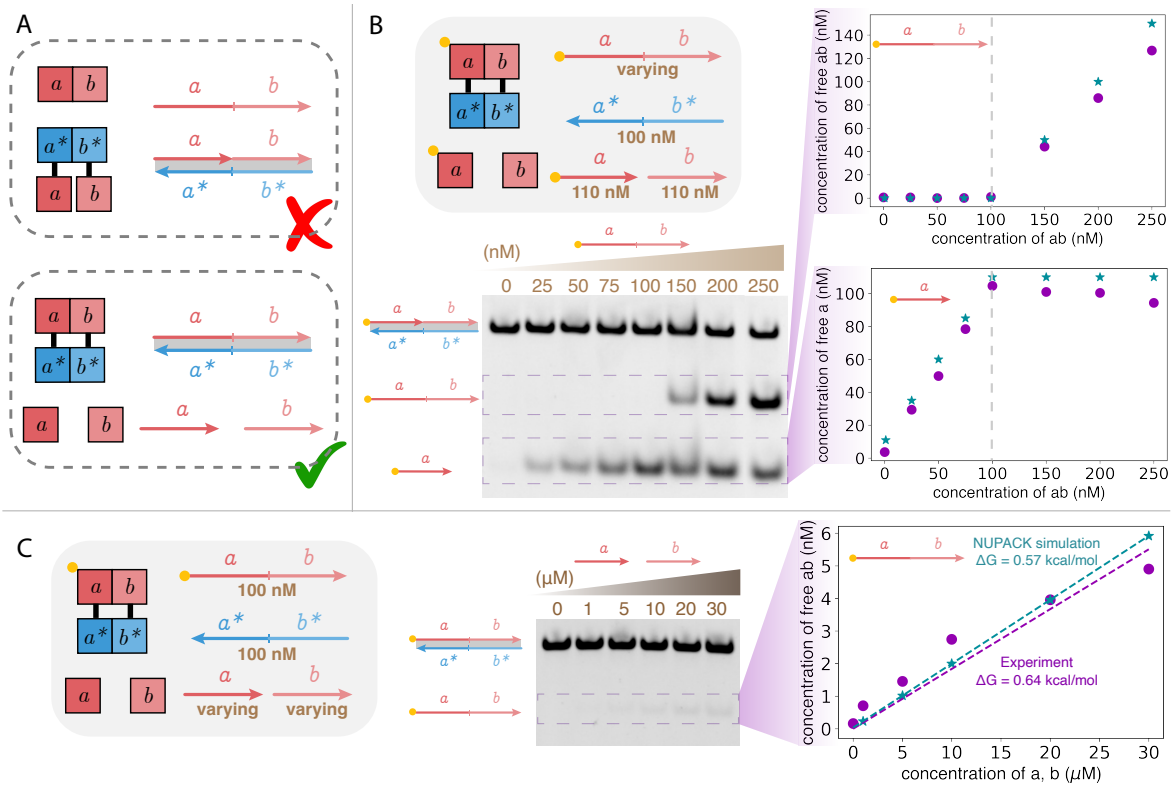

Figure 1: Experimental verification of the simplest non-trivial TBN. (A) Among the two saturated (maximum bond) configurations, the bottom configuration is stable (maximum bonds and also maximum number of separate complexes). Each box corresponds to a *domain* (consecutive DNA nucleotides) that can bind with its complementary domain (labeled by a star). (B-C) Unstained 12% native PAGE gel varying (B) strand ab or (C) strands a and b after annealing (95 ℃ to 25 ℃ within 1.5 h). The yellow dot represents the fluorophore. The concentration of each strand in the annealing reaction is labeled in the figure. The circled data are experimental measurements of band intensity in the gel after normalization (full gels with calibration controls are in Figure S1). The starred data are NUPACK simulation results (maximum complex size set to 3). In (B), a 10% excess of strands a and b was added to ensure that all starred domains are bound (strand a^*b^*) in case of pipetting or concentration-measuring errors. Likewise, to account for stoichiometric imprecision in (C), complex $ab : a^*b^*$ was purified prior to mixing with a and b in order to ensure no excess of ab.

maximally bonded is called *saturated*. The configuration with the largest number of separate complexes among saturated configurations is called *stable*.

To use the TBN model as an effective programming paradigm, systems need to be engineered to satisfy the model's abstraction. For DNA systems, TBN bonds correspond to domain-level Watson-Crick-Franklin base-paring and standard sequence design is performed to ensure orthogonal domain binding. Domain lengths are chosen long enough such that saturated configurations are strongly preferred. The order of domains has to be chosen to ensure feasible geometry of desired complexes (discussed later).

In this section, we study the arguably simplest non-trivial TBN containing four indivisible molecules a, b, ab and a^*b^*. (Abstract TBN molecules[3] are drawn using the "box" notation in the figures.) Bonds can form between starred and unstarred boxes with the same label (complementarity). The two configurations shown in Figure 1A have the maximum possible bonds and thus are saturated. The bottom configuration has 3 separate complexes compared with 2 for the top configuration. Since there are no saturated configurations with more separate complexes, the bottom configuration is stable according to the TBN model and is expected to be thermodynamically preferred.

We call the free-energy difference of one separate complex "*one unit of entropy*". For DNA strands, the standard free energy at 37 ℃ of associating two molecules is estimated as 1.96 kcal/mol[25]. At typical concentrations (250 nM, 37 ℃) calculations show that one unit of entropy should be approximately equal to the enthalpy of a 8 base-pair bond[19]. The penalty of association can be further increased through reducing reaction concentration. Thus it is expected that experimental systems and protocols can be designed to implement the TBN model with high fidelity.

To experimentally confirm that a single unit of entropy is sufficient to strongly differentiate the two configurations in Figure 1A at concentrations typical for DNA nanotechnology, we mapped the TBN to a set of DNA molecules as described in Section S1.1 and Section S4. As a proxy for thermodynamic equilibrium, we annealed the strands at concentrations shown in Figure 1(B)-(C) as described in the caption, and performed native polyacrylamide gel electrophoresis (PAGE). For greater sensitivity, we labeled strands a and ab by fluorophores; gel experiments without fluorophores were consistent with fluorophore-labeled gels (see Figure S1A for a comparison). Strands a and ab were labeled by identical fluorophores to ensure that any fluorophore-dependent energy contribution would be equal. As shown in Figure S1(B), we varied the amount of ab, confirming that up 100 nM (the concentration of strand a^*b^*), there is almost no measurable free ab, while free a linearly increases up to saturation. These experimental results are close to the NUPACK[9] simulation of the same system. These data confirm the strong thermodynamic preference for the bottom configuration in Figure 1A.

To more quantitatively measure the thermodynamic driving force of one unit of entropy, in Figure 1C we increased the amount of strands a and b to the point where free ab could be detected. Freeing 5% of ab required 3000 times higher concentration of strands a and b than ab. The standard free energy of association derived from these data (0.64 kcal/mol) is close to the effective value fitting NUPACK simulation (0.57 kcal/mol) (methods in Section S2.3). Note that this deviates from the theoretical standard free energy of association at 25 ℃ (1.89 kcal/mol) computed from the literature[25] value of 1.96 kcal/mol at 37 ℃ (Section S2.2).

[3] Prior TBN literature referred to indivisible molecules as monomers[16,17].

The discrepancy is likely caused by weak spurious interactions of strands and the free energy contribution of nick in the three-stranded complex $a : b : a^* b^*$. Note that the discrepancy only results in 11% difference in predicted concentration at 250 nM [4].

3 Signal propagation circuits

Intermolecular signalling circuits are common in biology, and their recapitulation with first-principles design has been a powerful test-case for dynamic DNA nanotechnology. As a prototypical example, strand displacement cascades modulate the accessibility of the region used for the initiation of displacement (toehold sequestering), and can achieve intermolecular signal propagation and computation on digital and analog signals[1]. Toehold sequestering is a form of kinetic control resulting in metastable complexes. Indeed, the desired output is typically at odds with the thermodynamic equilibrium of the system resulting in so called "leak" (see refs.[19,20] for exceptions). In contrast, systems designed through thermodynamics cannot rely on metastable complexes and the toehold sequestering mechanism. Nonetheless, we can engineer complex signal propagation including combinatorial logic using the TBN abstraction.

Based a previously proposed TBN AND gate[16,17], we experimentally implemented a circuit reversibly converting between the pair of signals A ($a_1 a_2$) and B ($b_1 b_2$), and signal C ($c_1 c_2$). As shown in Figure 2A, the system has 3 stable configurations that are expected to be favored at thermodynamic equilibrium. Since among these configurations C can be free only if A and B are sequestered, this system implements AND gate behavior ($A + B \rightarrow C$) in the forward direction, and fanout ($C \rightarrow A + B$) in the backward. Note that there is no direct interaction between signal strands and the information processing is conceptually similar to allosteric action: input binding results in change of configuration which indirectly influences whether the output molecule is bound. Importantly, there is no forced sequence complementarity or identity between the signal strands, allowing for composition into longer cascades taking the output as the input to a downstream circuit[16].

To experimentally test the signal propagation circuit, we annealed all the strands with and without the inputs, and used native PAGE analysis to determine the amount of free output (fluorophore labeled). For the circuit implementing the reaction $A + B \rightarrow C$ (Figure 2B), the amount of free output C was 8 times more when both inputs A and B were present than when only one or none of the inputs was present. When we used the same circuit to implement the reaction $C \rightarrow A + B$ (Figure 2C), the amount of output was at least 3 times more when the input C was present. (In separate experiments either output A or output B had a fluorophore label.) The background signal could be due to imperfect stoichiometry between starred and unstarred strands caused by pipetting errors. In addition to gel experiments, we obtained similar results in fluorophore-quencher experiments in a plate reader (Figure S3B). NUPACK simulations confirm the overall pattern of the data, albeit with almost no background signal and somewhat higher correct free output concentration.

The middle stable configuration of Figure 2A has all signals sequestered; thus, depending on the concentrations of the constituent strands we expect less output signal(s) to be released than input signal(s) added. While we used identical concentrations of all strands,

[4] Free energy at concentration c M is calculated by $\Delta G^\circ + RT \ln(1/c)$.

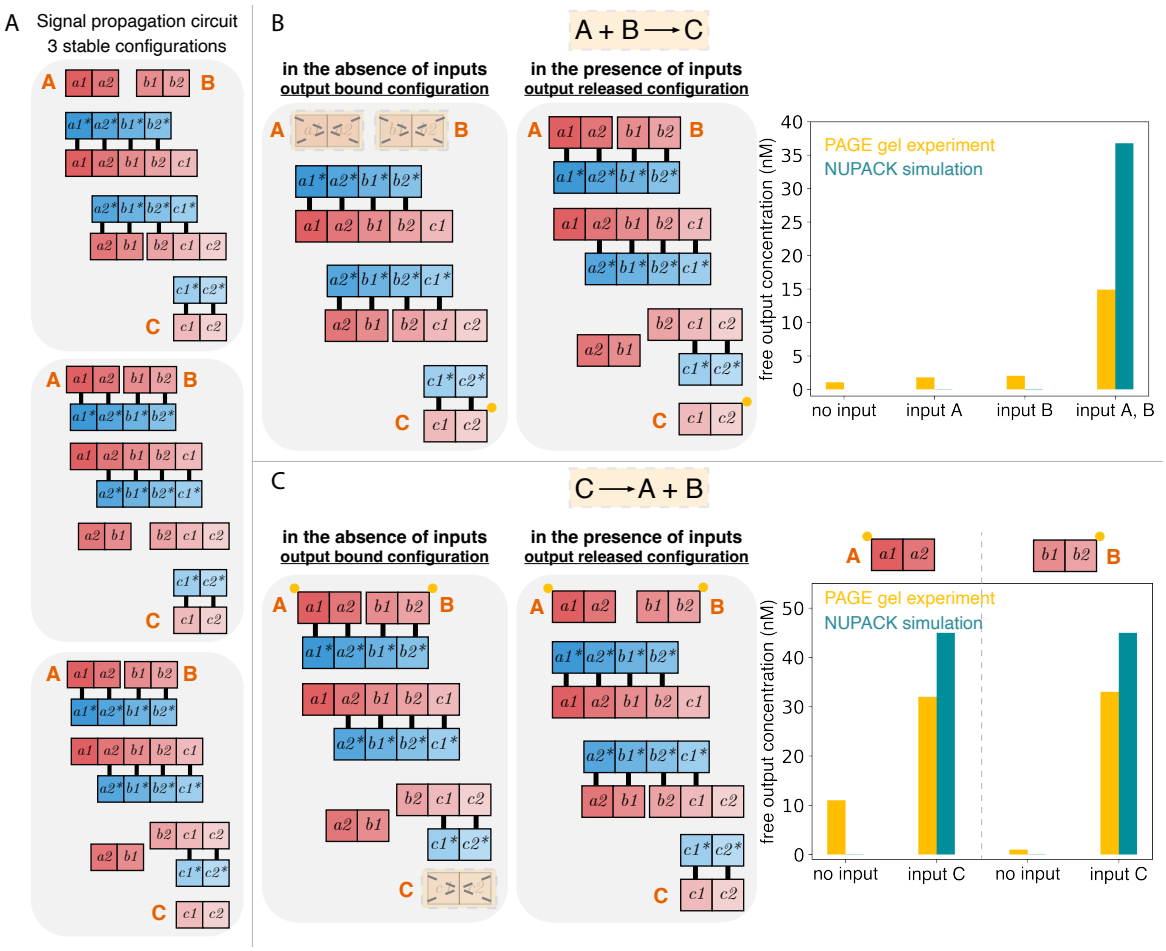

Figure 2: Experimental results for the signal propagation circuits. (A) The circuit contains 3 stable configurations. (B) Controlling the configuration of the circuits through adding the inputs A and B. The output C is labeled by a fluorophore (yellow dot). (C) Controlling the configuration of the circuits through adding the input C. In the experiment either output A or output B has a fluorophore label (yellow dot). All the species are 100 nM and slowly annealed from 90 ℃ to 25 ℃ with a rate of 0.1 ℃ per min. The yellow bar plots are experimental measurement of band intensity in native PAGE gels after normalization (full gels are in Section S3). The NUPACK simulation (25 ℃) used maximum complex size cutoff of 3.

in principle the equilibrium concentration of free output can be increased by appropriately increasing the concentrations of non-input strands (Section S3.4). More sophisticated chemical information processing will require larger chemical circuits, and we explore potential scaling issues in the Discussion.

4 Seeded-assembling circuits

Both in evolved biology and engineered nanotechnology, molecular self-assembly is used to create complex structures from simple parts in a bottom-up manner[3]. For example, in DNA nanotechnology, DNA origami achieves the desired shape after annealing to thermodynamic equilibrium[26]. While the computational abilities of DNA origami are limited , algorithmic tile-assembly is effectively guided by algorithms encoded in molecular bonds. Unlike DNA origami, algorithmic assembly is not thermodynamically favored over error pathways, and to prevent spurious binding between molecules precise control over binding kinetics—such as adjusting binding strength and temperature—is essential[27].

As an example of *thermodynamic* algorithmic self assembly, we experimentally realize self-assembling logic circuits proposed in prior work[21]. All the inputs are encoded by the choice of domains on a single (seed) strand, and the thermodynamic equilibrium determines the choice of other strands that bind and perform computation. The circuits (Figure 3A) are designed to have exactly 1 stable configuration, and this configuration has the desired output.

We started with a 2-input 1-output AND gate and NAND gate (Figure 3B). The input strands have domains w^*, a_i, and b_j, with the two input values represented by $i, j \in \{0, 1\}$. The output strands have domains y_k^* and w, with the output value represented by $k \in \{0, 1\}$. The logic value of the truth table of the gate is hard coded into the gate molecules $a_i^* b_j^* y_k$. Strands $a_i b_j$ and y_i, termed caps, ensure that the only large complex that forms is with the seed (input). Although for full analysis we refer the reader to prior work, without caps the unused gate strands will [can?] bind together into complexes growing with the size of the circuit. Domain w is used to ensure the correctness of computation. [Without the domain w, all configurations of partially assembled circuits have the same entropy as the complete circuit. With the domain w, the complete circuit gains one unit of entropy.] The output of the logic gate is represented by which output molecule is in the same complex as the input. The correct computation always has one more separate complexes than that of the incorrect computation (example of input 11 in Figure 3B).

To measure the circuit output, we labeled the input strand by a quencher and the 0 and 1 outputs by two distinct fluorophores. Which of the two fluorescence signals is quenched indicates the result of the computation. To avoid potentially geometrically infeasible structures (Figure 3C), we followed a systematic procedure in determining the domain order on strands [SI ref]. In each experiment, we annealed one chosen input with gates, caps and outputs and then measured the fluorescence signals for both fluorophores. We tested all possible inputs for both the AND and NAND gates. In each case the fluorescence signal for the correct output is almost fully quenched, and the fluorescence signal for the other output is greater than half of the maximum signal. (Maximum signal corresponds to the fluorescence of the output strands $y_0^* w$ or $y_1^* w$ alone.)

We next scaled up self-assembling circuits, in the process further exploring TBN model

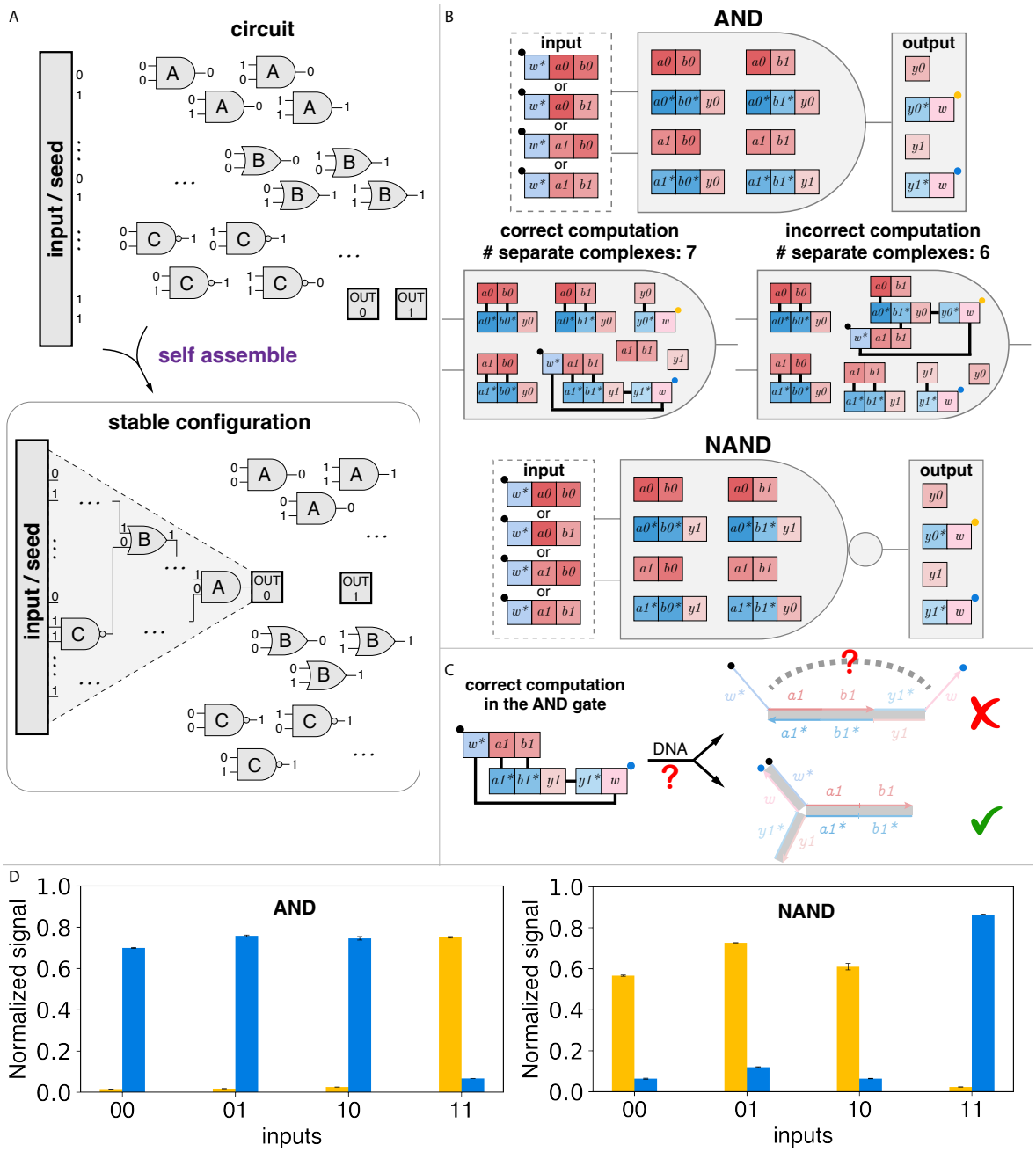

Figure 3: Experimental results for the seeded-assembling circuits. (A) Seeded-assembling circuits designed through the TBN model in principle have the thermodynamic guarantee that the stable configuration has the correct computation circuit. (B) Construction for the AND and NAND gates. The logic information for each gate is hard coded into the gate molecules. The input has a quencher attached, and each output strand has one of two different fluorophores. Which fluorophore is quenched is a proxy for which output is assembled with the input. As an example, the configuration of the correct computation for the input 11 has the output 1 in the same complex as the input, and incorrect computation has output 0 in the same complex as the input. The configuration with the correct computation has 1 more separate complex than any configuration with incorrect computation. (C) Converting the abstract TBN notation to DNA molecules may require organizing domain order in molecules to make it geometrically feasible (methods in Section S4). (D) Fluorescence signals of the DNA implementation for a NAND and a AND gates with all inputs at thermodynamic equilibrium after annealing. Data were normalized according to the method in Section S5.3. It is not clear why why positive signals only reach to 0.8. All the species are 100 nM and slowly annealed from 90 ℃ to 25 ℃ with a rate of 0.1 ℃ per min.

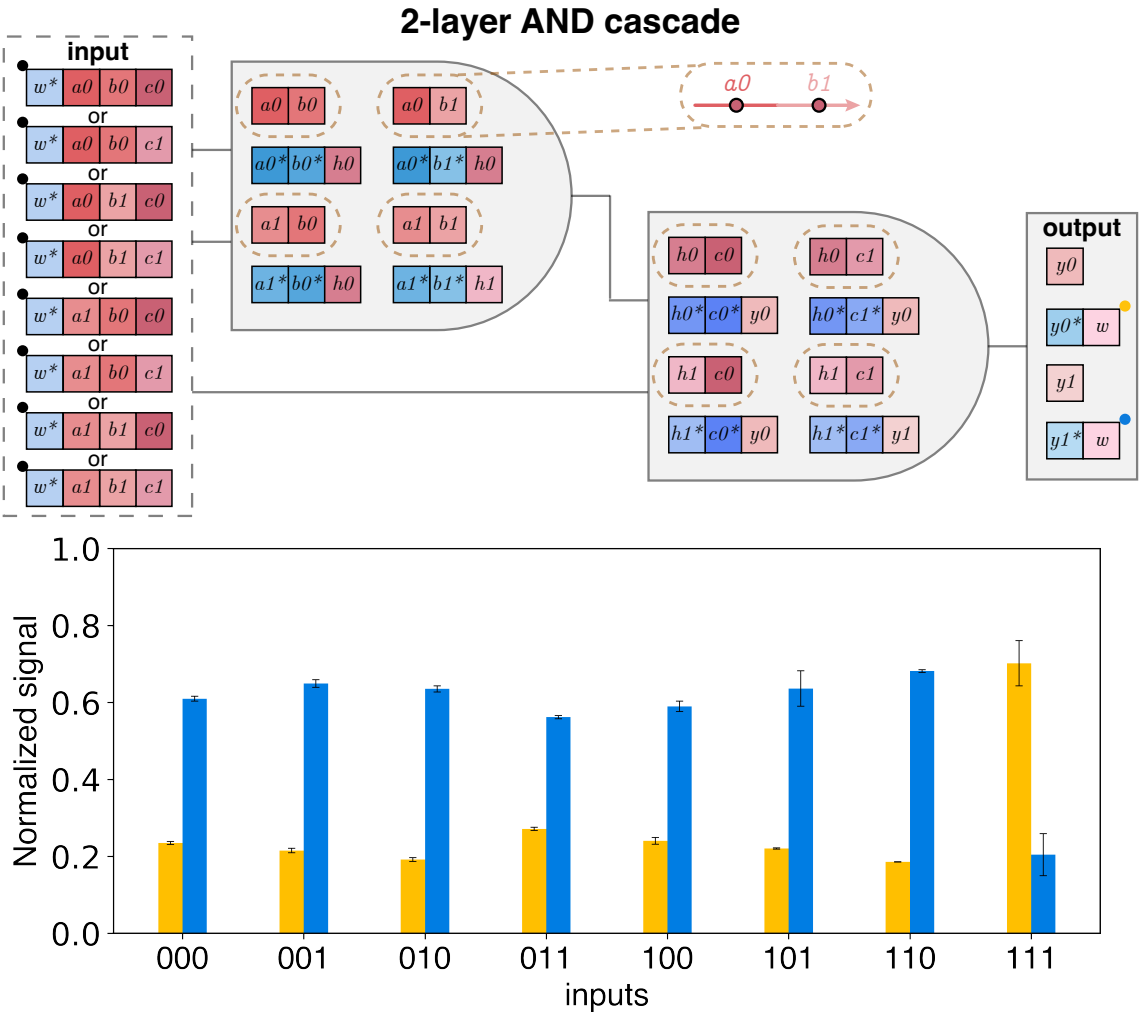

Figure 4: Construction and fluorescence signals for a seeded-assembling circuit of a 2-layer AND cascade with all inputs. The cap strands of the gates (highlighted by dashed lines) had single-base deletions; the deletions for $a_0 b_1$ are in the call-out (red circles with black outline). Data were normalized according to the method in Section S5.3. All the species are 100 nM and slowly annealed from 90 ℃ to 25 ℃ with a rate of 0.1 ℃ per min.

non-idealities. One of the free energy contributions ignored by the TBN model is the penalty due to formation of multi-arm junctions. The previous system with one AND gate only had one 3-arm junction and we ignored it in our design methodology. Can the TBN model still be used as an effective programming paradigm when more such junctions are unavoidable and their total free energy penalty could be larger than one unit of entropy?

The self-assembling circuit construction with two cascaded AND gates (Figure 4) has two 3-arm junctions in the desired (assembled) configuration (Figure S13A). As a result, the free energy of the assembled structure is roughly 8-12 kcal/mol higher. To compensate, we introduced deletions in the "cap" molecules to increase the free energy of the configuration without assembly (Figure S13B). Deletions significantly improved the circuit performance compared to systems without deletions. Figure 4 shows the fluorescence output of the two AND gate construction for all eight inputs. For every input, the fluorescence signal for the correct output is about 20% of fully quenched, and the fluorescence signal for the incorrect output is greater than 60% of the maximum signal.

5 Size-controlled concatemers

Synthetic nucleic acids have been used to construct finely controlled molecular structures[3], and concatemers have been used as templates for engineering periodic structures[29]. Concatemers have been synthesized through enzymatic amplification[30], or ligation of periodically hybridized DNA oligos, but it is hard to finely control the product size[31]. The apparent difficulty in self-assembling concatemers of programmed size can be articulated as follows: since the units are repeating, how does the system know when to stop growing?

Here we use the TBN model to construct concatemers of programmed size where the size is controlled by simple computation at thermodynamic equilibrium. The TBN construction contains two strands, composed of m and n repeated domains respectively. The domains between the two strands are complementary (x and x^*). When the ratio of the strands is $n : m$, the number of domains x^* in the largest complex in the stable configuration is the least common multiple of m and n: $\mathrm{LCM}(m, n)$ (proof in Section S6). Note that the largest concatemer can be constructed from the smallest starting molecules when m and n differ by 1. In this way we can construct a concatemer of size quadratic in the sizes of starting substrates.

Figure 5A shows an example system with $m = 3$ and $n = 2$. To generate the concatemers, we annealed the two strands, followed by ligation and gel separation. To account for possible pipetting imprecision, we tested a range of concentration ratios by fixing the concentration of x^*x^* and varying the concentration of xxx. Native PAGE gel confirmed that the products formed after annealing are sensitive to the initial concentrations (Figure 5B). After normalization, the phase change of the concentrations of the products as changing the initial xxx concentration is consistent with NUPACK prediction, and the peak of the desired complex is close to 200 nM—the ideal concentration ratio is $2 : 3$. Note that we highlight two bands on the native gel to correspond to the desired product. Indeed, there are two different secondary structures with the same size satisfying the TBN model, one linear and one branched, which we expect are responsible for the two bands (Figure S15A). The denatured PAGE gel analysis shows the absence of products larger than the expected size. The smaller side product is hypothesized to be $x^*x^*x^*x^*$, caused by ligation of branched

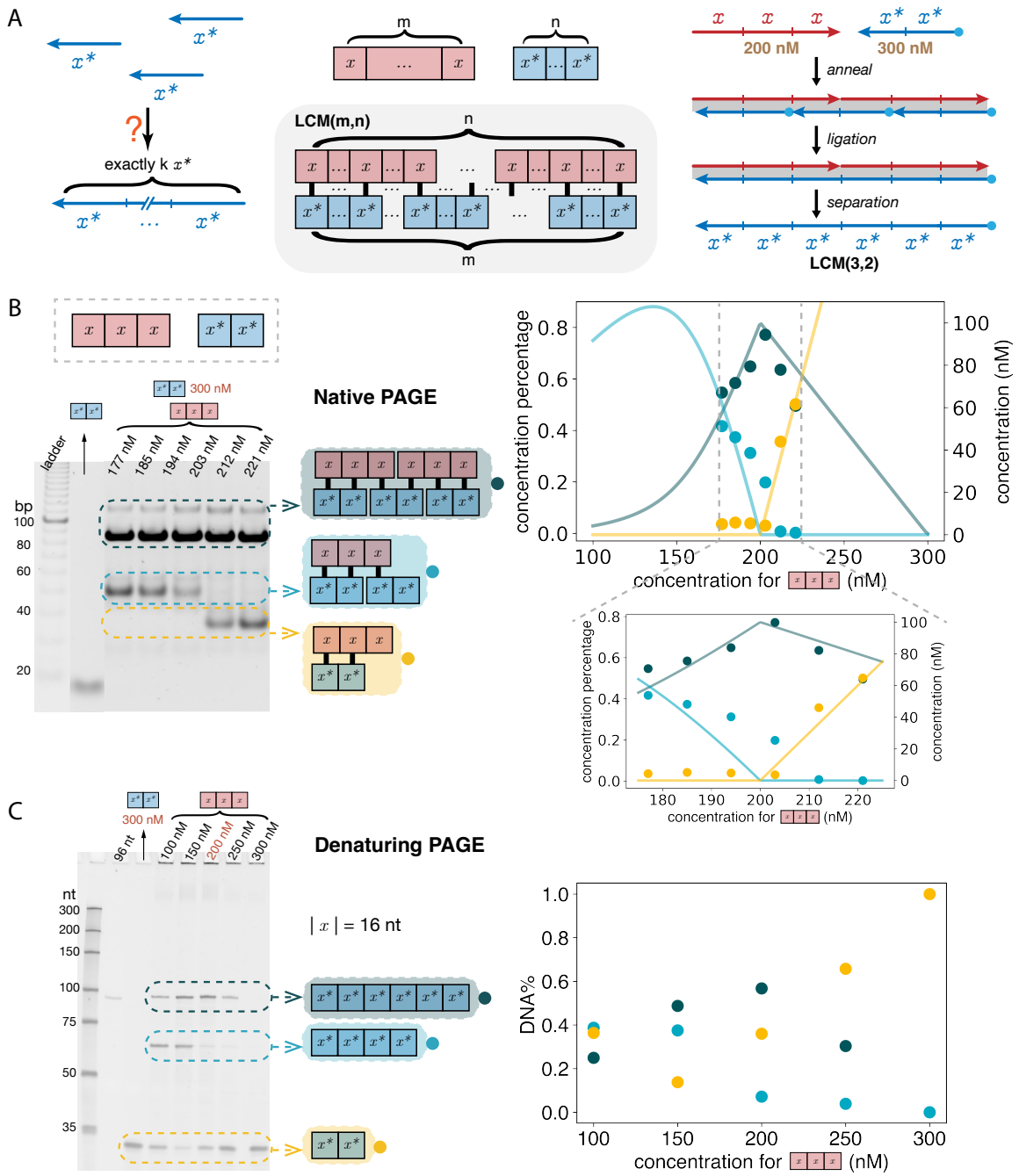

Figure 5: Size-controllable concatemers. (A) Unseeded self-assembly under the TBN model can be used to produce size-controllable concatemers. A TBN system with a pair of molecules with m and n repeated complementary domains has the stable configuration with the maximum possible number $\mathrm{LCM}(m, n)$ of domains for each molecule. As an example, when the ratio of xxx and x^*x^* is $2 : 3$, the stable configuration only contains the complex with $\mathrm{LCM}(3, 2) = 6$ bound domain pairs. The 6-domain concatemer can be obtained after annealing, ligation and gel separation. The blue dot represents a phosphate label on 5'-end for ligation purpose. (B) The native 12% PAGE analysis of the annealed product with fixed x^*x^* concentration and a range of different xxx concentrations. A range concentration of xxx were annealed with a fixed amount of x^*x^* (300 nM). NUPACK simulation (line) is compared with the normalized native PAGE results (scatter plot). The color of the plots match the products showed in the gel. The NUPACK simulation has the maximum number of strands in a complex set to be 5. (C) The denaturing PAGE analysis of the ligated product after annealing a fixed amount of x^*x^* with different concentrations of xxx. The x domain is 16 nt. Note that the x domain is composed of the ATC alphabet, which does not get stained by Sybr Gold[28], so it does not get imaged in the denaturing gel.

structure product (Figure S15).

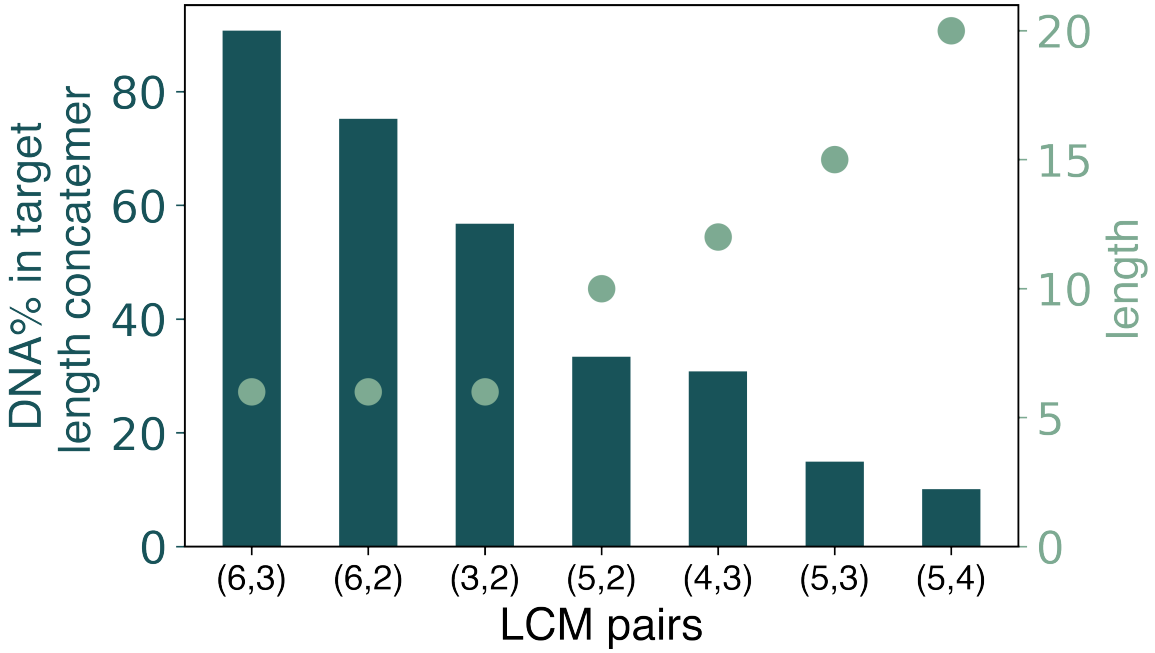

Figure 6: Generating size-controllable concatemers with different sizes of the substrate strands. Dots indicate the target size (LCM of the sizes of the substrates). The bars show the percent of DNA in the target size concatemer computed from band intensities in the denaturing gels (gels in Section S6.2).

Next we generated different sizes of concatemers starting with substrates with different m and n values (Figure 6, Section S6.2). In every case, there was no apparent concatemer of length longer than the LCM. However, especially for longer target lengths, a substantial fraction of the total DNA was in shorter side-products. Further, when multiple pairs have the same LCM values (pairs $(6,3)$, $(6,2)$, and $(3,2)$) starting with smaller substrates seemed to result in less DNA of full-length target size. Overall this TBN construction shows that a surprisingly simple computation performed by thermodynamic molecular self assembly can be effectively used to program concatemer assembly.

6 Discussion

Much of experimental work in molecular programming focuses on kinetic approaches aiming to prescribe a desired sequence of molecular interactions (notable exceptions include DNA origami[26] and non-complementary strand commutation networks[15]). However, kinetic approaches might unnecessarily shoehorn an ill-fitting programming paradigm motivated by the analogy to electronic computation where a program specifies a sequence of instructions. Unlike electronics, chemical systems naturally explore large configuration spaces with intrinsically random collisions of molecules. Here we have used an approach that embraces thermodynamic programming and leverages it as a fundamental design principle in the form of the TBN model. We designed and implemented a range of systems showing the flexibility of the approach: signal propagation circuits demonstrating conjunction and fan-out, seeded

algorithmic self-assembly, and a system for assembling size-controlled concatemers based on the computation of least common multiple.

The demonstrated TBN constructions have thermodynamic guarantees relying on one unit of entropic difference (one separate complex) between desired and undesired behaviors. For example, in the signal propagation circuit, the output can be freed in the absence of the input at the cost of one entropy unit as shown in XXX. Similarly, the seeded-assembly circuit can assemble incorrectly also at the cost of one fewer separate complexes as shown in Figure 3(b). For larger systems—especially if the number of undesired configurations scales with the size of the system faster than the number of desired configurations—this one unit of entropy may not be large enough, resulting in a higher concentration of the undesired complexes. As an example, cascading signal propagation units increases the amount of leak. Since one unit of entropy difference has a stronger effect at lower concentrations, one way to ensure desired behavior is via dilution. Nonetheless, since we must maintain reasonable experimental conditions, going beyond one unit of entropy is desired. Luckily, many previously described TBN designs, including for logic circuits and signal amplifiers, have a programmable units of entropy difference between the desired and undesired configurations[16,22,32]. It remains open, however, how other TBNs such as the self-assembly system can be made more robust in a similar way.

The TBN model captures thermodynamic favorability through characterizing the stable configurations. When every species is at a single copy, the probability of being in one desired stable configuration is, all else being equal, inversely proportional to the total number of stable configurations. However, in bulk experiment where species have multiple copies, extra entropic contribution (entropy of mixing) makes it harder to reason about the exact equilibrium concentration of each species without simulation. When the scale of the system increases, it could be computationally expensive for the current simulation tool to predict the behaviors. It remains an open question to better determine the equilibrium concentration of bulk TBN systems.

The TBN model applies to systems without geometric constraints, but using the model as an engineering paradigm requires considering geometric feasibility. We ordered domains on strands to ensure that desired structures are unpseudoknotted and introduced point deletions to compensate for the free energy penalty of multi-arm junctions. Alternatively, adding a couple of nucleotides at the base of the 3-arm junction could be an alternative way to stabilize it. More generally, adding “spacer” domains (e.g., poly-T regions) may be another approach to overcoming geometric constraints, and it may be feasible to allow pseudoknotted structures.

The original TBN model is specifically targeted at the regime where the enthalpic energy of forming a bond is strongly favored over the entropic energy of separating two molecules[16,17], although more recent work allows for tuning the weight of enthalpic and entropic contributions in the model[22]. Recent experimental work has explored the computational behaviors achievable through designing weak bonds that bind different partners with various strengths[15], yet the design principle of utilizing weak bonds to engineer systems with thermodynamic guarantees has not been thoroughly explored. In addition, thermodynamic equilibrium is a function of temperature. Incorporating consideration of temperature into the model could provide insights about engineering resettable DNA circuits through thermocycling. Programming the weight of both enthalpy and entropy and exploring the effect of

temperature on preferred configurations would provide a better understanding of the thermodynamics in bioengineering and can lead to a more complete picture of programming thermodynamic systems.

A potentially important advantage of thermodynamics-first systems is their reusability. If the input is changed (e.g., if the original input strands are removed and new input strands are added), the thermodynamic equilibrium will naturally conform to the new input. Consider the contrast to systems whose correctness relies the interactions of meta-stable complexes: for example, DNA strand displacement systems are typically "use-once" and the initial DNA complexes have to be externally re-made in order for computation to proceed on a different set of inputs. Figure S12 shows input resetting in the NAND-gate self-assembly system. To switch the input, we deactivated the original input strand with its complement, and subsequently added the new input. (The significance of kinetic trace is described below.)

Throughout this work we have used an annealing protocol to drive the system to thermodynamic equilibrium in reasonable time.[5] While thermal annealing is applicable for some applications such as paper-device sensing [cite] or molecular self-assembly [cite], other applications require isothermal operation. For example, smart drugs are intended to operate within the body detecting and triggering on particular combinations of molecular cues [cite]. Accordingly most of the strand displacement literature considers such a biologically compatible isothermal regime. We hypothesize that by partially destabilizing the domain binding we can achieve isothermal operation of TBNs. The kinetic trace in Figure S12 shows that the correct output can be produced on the order of minutes without annealing. Follow-up work is necessary to test the limits of input switching as well as to optimize the isothermal kinetics in a way that's broadly applicable to TBNs.

Biology demonstrates that sophisticated natural molecular information processing is achievable, suggesting no inherent barriers to similar advancements in engineered molecular programming. Although a thermodynamics-first approach is established in many cases such as protein folding, it is not clear whether biological information processing can be more broadly understood as an equilibrium phenomenon. Nonetheless, de novo molecular engineering, while inspired by nature, does not have to follow identical design principles. Counting the number of separate complexes provides a particularly clean abstraction aligned with bottom up engineering practices.

Funding:

B.W., C.C. and D.S. were supported by NSF awards 2200290, 1901025, DARPA grant W911NF-18-2-0032, and the Alfred P. Sloan Foundation. D.D. was supported by NSF awards 2329909, 2211793, 1900931, 1844976. D.S. and D.D. were also supported by DoE EXPRESS award SC0024467.

[5] It should be noted that certain systems do not readily converge to thermodynamic equilibrium even with slow annealing such as, e.g. HCR[33,34], although we do not believe that our systems have this property. In particular, during cooling of HCR, hairpins form first, trapping the system in a meta-stable state, and there are no hairpins designed in our system.

Authors contributions:

B.W. performed experiments and analyzed data. C.C. performed some experiments on the seeded-assembling circuit. D.D. proposed the size-controllable concatemer design. All authors discussed the results. B.W. wrote the initial draft of the manuscript. D.S. provided guidance throughout the project.

Competing interests:

The authors declare no competing or financial interests.

References

[1] Friedrich C. Simmel, Bernard Yurke, and Hari R. Singh. Principles and Applications of Nucleic Acid Strand Displacement Reactions. *Chemical Reviews*, 119(10):6326–6369, May 2019.

[2] Dominic Scalise and Rebecca Schulman. Controlling Matter at the Molecular Scale with DNA Circuits. *Annual Review of Biomedical Engineering*, 21(1):469–493, June 2019.

[3] Amelie Heuer-Jungemann and Tim Liedl. From DNA Tiles to Functional DNA Materials. *Trends in Chemistry*, 1(9):799–814, December 2019.

[4] Xiaoping Olson, Shohei Kotani, Jennifer E. Padilla, Natalya Hallstrom, Sara Goltry, Jeunghoon Lee, Bernard Yurke, William L. Hughes, and Elton Graugnard. Availability: A metric for nucleic acid strand displacement systems. *ACS Synthetic Biology*, 6(1):84–93, 2017.

[5] Lulu Qian and Erik Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196–1201, June 2011.

[6] Paul W. K Rothemund, Nick Papadakis, and Erik Winfree. Algorithmic Self-Assembly of DNA Sierpinski Triangles. *PLoS Biology*, 2(12):e424, December 2004.

[7] Rebecca Schulman, Christina Wright, and Erik Winfree. Increasing Redundancy Exponentially Reduces Error Rates during Algorithmic Self-Assembly. *ACS Nano*, 9(6):5760–5771, June 2015.

[8] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, July 2003.

[9] Mark E. Fornace, Jining Huang, Cody T. Newman, Nicholas J. Porubsky, Marshall B. Pierce, and Niles A. Pierce. NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems. Preprint, Chemistry, November 2022.

[10] Erik Poppleton, Roger Romero, Aatmik Mallya, Lorenzo Rovigatti, and Petr Šulc. OxDNA.org: A public webserver for coarse-grained simulations of DNA and RNA nanostructures. *Nucleic Acids Research*, 49(W1):W491–W498, July 2021.

[11] Sherry Xi Chen, David Yu Zhang, and Georg Seelig. Conditionally fluorescent molecular probes for detecting single base changes in double-stranded DNA. *Nature Chemistry*, 5(9):782–789, September 2013.

[12] W. Benjamin Rogers and Vinothan N. Manoharan. Programming colloidal phase transitions with DNA strand displacement. *Science*, 347(6222):639–642, February 2015.

[13] Anthony J. Genot, Jonathan Bath, and Andrew J. Turberfield. Reversible Logic Circuits Made of DNA. *Journal of the American Chemical Society*, 133(50):20080–20083, December 2011.

[14] Jaeseung Hahn and William M Shih. Thermal cycling of DNA devices via associative strand displacement. *Nucleic Acids Research*, 47(20):10968–10975, November 2019.

[15] Maxim P. Nikitin. Non-complementary strand commutation as a fundamental alternative for information processing by DNA and gene regulation. *Nature Chemistry*, 15(1):70–82, January 2023.

[16] David Doty, Trent A. Rogers, David Soloveichik, Chris Thachuk, and Damien Woods. Thermodynamic Binding Networks. In Robert Brijder and Lulu Qian, editors, *DNA Computing and Molecular Programming*, volume 10467, pages 249–266. Springer International Publishing, Cham, 2017.

[17] Keenan Breik, C. Thachuk, Marijn Heule, and D. Soloveichik. Computing properties of stable configurations of thermodynamic binding networks. *Theoretical Computer Science*, 785:17–29, September 2019.

[18] Chris Thachuk, Erik Winfree, and David Soloveichik. Leakless DNA strand displacement systems. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9211, pages 133–153. Springer Verlag, 2015.

[19] Boya Wang, Chris Thachuk, Andrew D. Ellington, Erik Winfree, and David Soloveichik. Effective design principles for leakless strand displacement systems. *Proceedings of the National Academy of Sciences of the United States of America*, 115(52):E12182–E12191, December 2018.

[20] Boya Wang, Chris Thachuk, and David Soloveichik. Speed and correctness guarantees for programmable enthalpy-neutral DNA reactions. *ACS Synthetic Biology*, 12(4):993–1006, April 2023.

[21] Cameron Chalk, Jacob Hendricks, Matthew J. Patitz, and Michael Sharp. Thermodynamically favorable computation via tile self-assembly. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10867 LNCS, pages 16–31. Springer Verlag, 2018.

[22] Keenan Breik, Cameron Chalk, David Doty, David Haley, and David Soloveichik. Programming Substrate-Independent Kinetic Barriers With Thermodynamic Binding Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(1):283–295, January 2021.

[23] Andre V. Pinheiro, Dongran Han, William M. Shih, and Hao Yan. Challenges and opportunities for structural DNA nanotechnology. *Nature Nanotechnology*, 6(12):763–772, December 2011.

[24] Paul W. K. Rothemund and Erik Winfree. The program-size complexity of self-assembled squares (extended abstract). In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 459–468, Portland Oregon USA, May 2000. ACM.

[25] John SantaLucia and Donald Hicks. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33:415–440, 2004.

[26] Paul W.K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, March 2006.

[27] Constantine G. Evans and Erik Winfree. Physical principles for DNA tile self-assembly. *Chemical Society Reviews*, 46(12):3808–3829, 2017.

[28] Xutiange Han, Erchi Wang, Yixiao Cui, Yikai Lin, Hui Chen, Ran An, Xingguo Liang, and Makoto Komiyama. The staining efficiency of cyanine dyes for single-stranded DNA is enormously dependent on nucleotide composition. *ELECTROPHORESIS*, 40(12-13):1708–1714, July 2019.

[29] Weian Zhao, M. Monsur Ali, Michael A. Brook, and Yingfu Li. Rolling Circle Amplification: Applications in Nanotechnology and Biodetection with Functional Nucleic Acids. *Angewandte Chemie International Edition*, 47(34):6330–6337, August 2008.

[30] M. Monsur Ali, Feng Li, Zhiqing Zhang, Kaixiang Zhang, Dong-Ku Kang, James A. Ankrum, X. Chris Le, and Weian Zhao. Rolling circle amplification: A versatile tool for chemical biology, materials science and medicine. *Chemical Society Reviews*, 43(10):3324, 2014.

[31] Lu Sun and Björn Åkerman. Characterization of self-assembled DNA concatemers from synthetic oligonucleotides. *Computational and Structural Biotechnology Journal*, 11(18):66–72, August 2014.

[32] Joshua Petrack, David Soloveichik, and David Doty. Thermodynamically driven signal amplification. In *29th International Conference on DNA Computing and Molecular Programming (DNA 29)(2023)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2023.

[33] Robert M Dirks and Niles A Pierce. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences*, 101(43):15275–15278, 2004.

[34] Justin S Bois. *Analysis of interacting nucleic acids in dilute solutions*. California Institute of Technology, 2006.