

Assembly Factors Resolve Speed and Encoding Bottlenecks in Multifarious Self-Assembly

Félix Benoist[✉] and Pablo Sartori*

Gulbenkian Institute of Molecular Medicine, Oeiras, Portugal



(Received 7 January 2026; accepted 15 July 2026; published 27 August 2026)

Cytoplasm is a heterogeneous mixture containing many types of proteins that self-assemble into a wide variety of complexes. The accuracy and speed of cytoplasmic self-assembly is astonishing because it involves the correct identification of components shared among different structures, despite pervasive thermal fluctuations. Typical toy models of self-assembly are based on the specificity of binding energies among components. However, kinetics plays a key role in biological self-assembly, often catalyzed by a plethora of assembly factors. Building on this observation, we extend a previous heteropolymer growth model to describe the retrieval of two-dimensional structures via quasi-two-dimensional growth. We find that the self-assembly of structures in this model is subject to strong speed and encoding bottlenecks. Moreover, we show that these bottlenecks can be suppressed by increasing the connectivity of a small fraction of components. This mechanism of kinetically controlling a small number of critical binding events provides a simple explanation for the timely assembly of large heteromeric protein complexes and suggests a unifying principle for the role of assembly factors.

DOI: [10.1103/5w6h-1931](https://doi.org/10.1103/5w6h-1931)

I. INTRODUCTION

Many proteins assemble into heteromeric complexes, i.e., complexes composed of multiple protein species, to carry out essential cellular functions. During assembly, protein complexes must accurately identify their specific components from a crowded mixture of thousands of species. Assembly must be not only precise but also rapid, so that cellular functions proceed in a timely manner. Two remarkable examples are spliceosomal complexes and ribosomal subunits, each composed of dozens of proteins [1,2], with ribosomal subunits assembling in under 2 min [3]. Advances in high-throughput cryogenic electron microscopy have steadily expanded our knowledge of the molecular and structural details of such complexes. Yet, the generic physical principles that enable high assembly speed and low error rates to coexist in heterogeneous cellular environments remain elusive.

What are the key biological ingredients of cellular self-assembly? We identify two such ingredients, assembly factors and assembly geometry, which we describe below. Assembly factors are proteins that transiently assist the growth of complexes [2,4,5], akin to chaperones in protein folding [see the schematic in Fig. 1(a)]. Assembly factors bind at specific locations and times during the growth of protein complexes [see examples in Figs. 1(b) and 1(c)]. Moreover, their number is often comparable to that of the proteins in the complex itself [see Fig. 1(d)]. Their specific molecular functions are diverse: Some block premature binding, others catalyze specific

events [4,8,9,17] and yet others act as proofreading enzymes [2,4,18]. Overarching this diversity is the general observation that depletion of assembly factors slows down self-assembly and leads to the accumulation of incomplete intermediates [17–20]. Another common feature is that assembly factors are dissipative, hydrolyzing ATP or GTP to perform their function [4,9,10,17]. Thus, assembly factors are a key ingredient in protein complex assembly, ensuring both speed and accuracy by strongly driving specific binding events.

A second fundamental ingredient is the diversity of protein shapes and connectivities within heteromeric complexes. Even at a coarse-grained level, some proteins (or RNA components) are large and elongated, while others are small and globular [21,22]. Consequently, their connectivity varies widely: Some proteins bind to many other proteins in the final complex, while others bind to few [see Fig. 1(e) for examples across complexes]. This structural diversity has important kinetic consequences. In particular, the kinetic cooperativity of neighboring proteins is known to largely accelerate self-assembly [23,24]. We thus expect that highly connected proteins can be kinetically discriminated and quickly assemble without errors. Overall, unlike large homomeric complexes such as actin filaments [25,26], microtubules [27], or viral capsids [28], which are highly symmetric and require few assembly factors, large heteromeric complexes feature far more intricate geometries and involve many catalyzing assembly factors.

In this paper, we develop a toy model of heteromeric multifarious self-assembly to investigate the role of the two ingredients above. Our goal is not to undertake the detailed molecular dynamics of heteromeric self-assembly [29,30], but rather to design simplified models that illuminate fundamental principles of nonequilibrium self-assembly and geometry, and their implications in biological and synthetic systems [31–34]. Building on previous work in self-assembly [31,35–37] and nonequilibrium information processing [38–42], we explore

*Contact author: pablo.sartori@gimm.pt

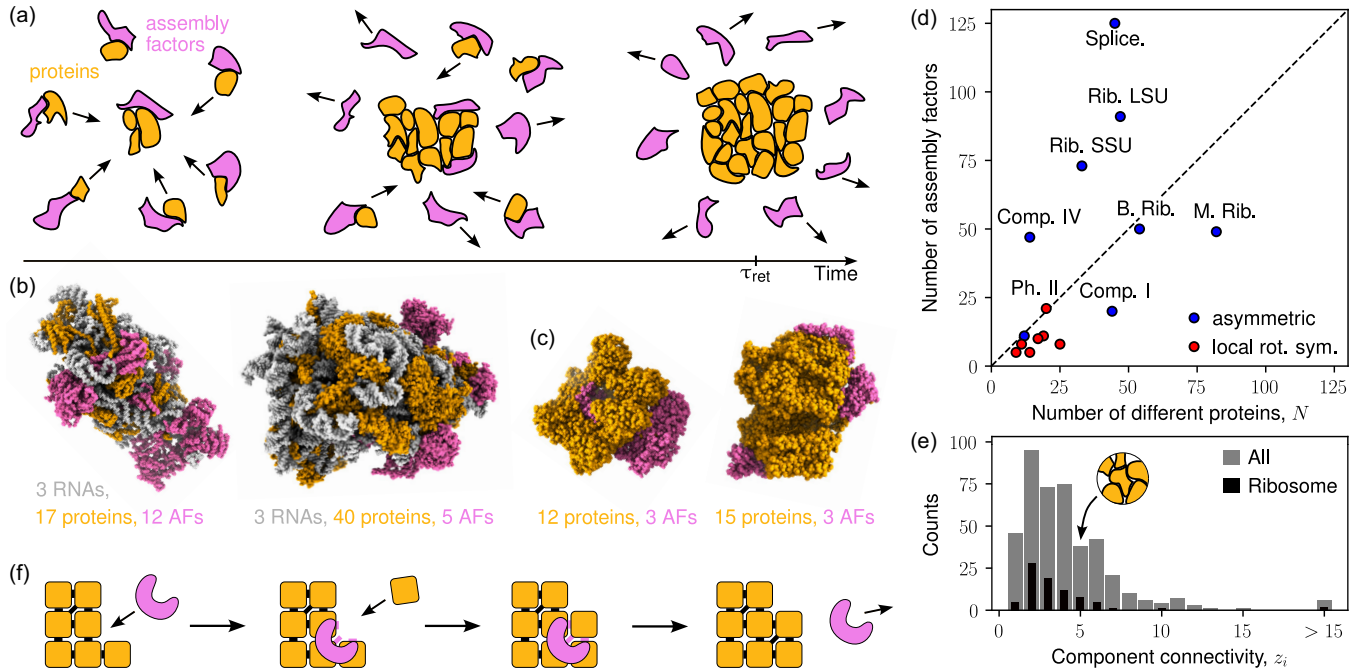


FIG. 1. Formation of large heteromeric complexes rely on many assembly factors. (a) Schematic representation of the self-assembly of a protein complex. The self-assembly of protein subunits is driven by assembly factors, which have the ultimate role of favoring speed and accuracy of complex formation. The complete protein assembly contains no assembly factor. (b) Structural snapshots of assembly intermediates of the yeast ribosome large subunit. As we can see, assembly factors are transiently bound to the RNA-protein assembly (Protein Data Bank IDs 6EM3 and 6LSR [6]). (c) Intermediates of the human proteasome (8QZ9) and of the bacterial photosystem II (7NPQ). (d) In purely asymmetric complexes, numbers of assembly factors roughly increase with the numbers of protein subunits. By contrast, complexes with a local rotational symmetry require fewer assembly factors. Data are detailed in Sec. S1 of Ref. [7] and Refs. [2,4,5,8–16] therein. (e) The connectivity of components in an assembly is highly variable. Highly connected components correspond to RNAs. A detailed plot is given in Sec. S1 of Ref. [7]. (f) Our kinetic encoding model consists in the acceleration of correct binding events via kinetic cooperativity and catalyzing assembly factors.

how kinetic discrimination and variable connectivity enable fast and accurate assembly in the presence of confounding interactions.

The paper is structured as follows. In Sec. II, we introduce a minimal lattice model for far-from-equilibrium heteromeric self-assembly. In Sec. III, we study the assembly of a single square target with fixed connectivity using Monte Carlo simulations and identify speed bottlenecks. In Sec. IV, we extend the model by encoding an additional structure with the same components and show that the speed bottlenecks also act as encoding bottlenecks. In Sec. V, we derive conditions for rapid assembly of multiple structures and propose alleviating bottlenecks by increasing the connectivity of a small subset of components. In Sec. VI, we numerically validate these analytical predictions. In Sec. VII, we show that increased connectivity yields combinatorial scaling of the encoding capacity. In Sec. VIII, we summarize, discuss our findings, and offer prospects for future work.

II. MODEL SETUP

Within our framework, a set of square tiles with varying connectivity can reversibly bind at specific locations on the boundary of a growing structure, aided by catalyzing assembly factors [see Fig. 1(f)]. We abstract the action of assembly factors and the kinetic cooperativity discussed in

the Introduction by accelerating binding events that lead toward the formation of a target two-dimensional (2D) structure representing a protein complex. The corresponding unbinding events are likewise accelerated so that the target structure is encoded purely kinetically. To capture the promiscuous nature of proteins [43,44], which may lead to erroneous bonds between monomers not in contact in the final structure, we allow multiple target structures to be encoded simultaneously from the same components. This corresponds to the multifarious regime [35,36,45–47]. In this setting, monomer promiscuity increases uniformly with the number of encoded structures. Our model thus constitutes an abstraction of protein complex assembly, which we now describe in greater mathematical detail.

We consider a set of $N \gg 1$ component species labeled $i = 1, \dots, N$ [see Fig. 2(a)]. Each component corresponds to a tile with a fixed orientation on a square lattice, which can bind with its neighbors. All tiles have the usual nearest-neighbor binding sites (north, n; south, s; east, e; and west, w), and some tiles may have additional next-nearest-neighbor binding sites (north-east, ne; north-west, nw; south-east, se; and south-west, sw). The components are held at fixed chemical potentials $\mu_i = \mu$ and may self-assemble into $S \geq 1$ heteromeric square target structures of side length $\ell = \sqrt{N}$. Each target corresponds to a reshuffling of all N species, so S determines the degree of promiscuity. Crucially, reuse

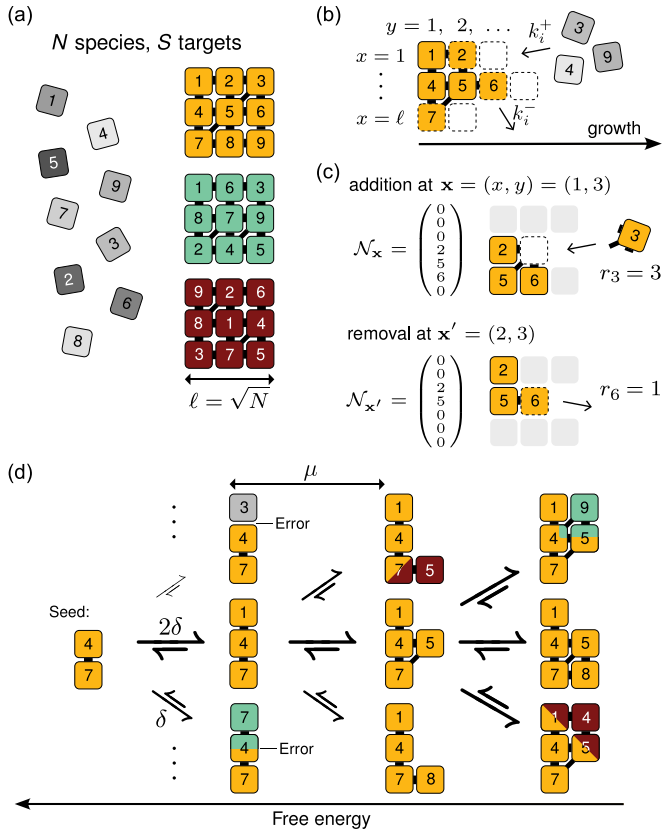


FIG. 2. Kinetic encoding of multiple target structures. (a) A mixture of $N = 9$ component species can assemble into $S = 3$ structures of size $N = \ell \times \ell$. Components bind via the black rectangles. Vertical periodic boundaries imply that, e.g., in the yellow target, components 1, 2, and 3 also bind to 7, 8, and 9. (b) Growth occurs by adding/removing monomers at the boundary locations $\mathbf{x} = (x, y)$ (dashed contour). (c) Addition (k_i^+) and removal (k_i^-) rates depend on the number of correct partners r_i in the neighborhood $\mathcal{N}_{\mathbf{x}}$ (0 denotes empty sites). Here component 3 binds to the three neighbors as in the yellow target, while component 6 binds to only one of its two neighbors. Note that bonds serve to accelerate binding, rather than to stabilize it. (d) Example self-assembly pathway of the yellow structure from a small seed. Each monomer binding decreases free energy via μ . The acceleration of correct binding/unbinding events via δ [Eq. (1)] is sketched as larger arrows. Binding of component 3 at the first assembly step is a thermal error, due to thermal fluctuations, whereas binding of component 7 is a combinatorial error, due to component reuse across targets.

of components across different structures compromises the reliability of self-assembly, a phenomenon previously described as multifarious [35,36,45–47].

To study retrieval of a structure, we model growth from a nucleation seed as a boundary-driven process, in which components bind/unbind at lattice locations $\mathbf{x} = (x, y)$ of a single boundary of the assembly [48,49] [see Fig. 2(b) as well as Sec. S3 in Ref. [7]]. The perpendicular lattice direction has periodic boundary conditions connecting the lattice locations $x = \ell$ to $x = 1$. This setup corresponds to a quasi-2D assembly process. The rates $k_i^\pm$ for adding and removing component i at $\mathbf{x}$ depend on the identity of the components in the neighborhood $\mathcal{N}_{\mathbf{x}}$, which we write as a seven-dimensional vector

$\mathcal{N}_{\mathbf{x}} = (i_{ne}, i_n, \dots, i_{se})$ [see Fig. 2(c) for examples]. (The eastern neighbor is not taken into account since components with an eastern neighbor are not part of the boundary.) To enable fast and reliable self-assembly of multiple targets, we encode structures in the binding kinetics, instead of the energetics. Following earlier work on polymerization [37,39,40], we define

$$\begin{aligned} k_i^+(\mathcal{N}_{\mathbf{x}}) &= \exp[r_i(\mathcal{N}_{\mathbf{x}})\delta], \\ k_i^-(\mathcal{N}_{\mathbf{x}}) &= \exp[r_i(\mathcal{N}_{\mathbf{x}})\delta - \mu]. \end{aligned} \quad (1)$$

In this expression, δ is a kinetic discrimination parameter that quantifies differences in kinetic barriers, and $r_i(\mathcal{N}_{\mathbf{x}})$ counts the number of neighbors in $\mathcal{N}_{\mathbf{x}}$ that also neighbor i at the corresponding relative locations in any target structure (not necessarily the one being assembled) (see Ref. [7] for the explicit expression). Thus, binding of i at location $\mathbf{x}$ creates $r_i(\mathcal{N}_{\mathbf{x}})$ bonds. Note that, since $k_i^+/k_i^- = \exp(\mu)$, the free-energy change due to growth contains no structural information, unlike in Refs. [35,36,50–53]. Instead, Eq. (1) accelerates addition/removal events that are consistent with target neighborhoods, so the structures are kinetically encoded [see the schematics in Fig. 2(d)]. In Sec. S2 of Ref. [7], we show how Eq. (1) can be interpreted within a simple framework combining kinetic cooperativity and assembly factor catalysis.

In the case where components participate in multiple targets ($S > 1$), their connectivity may differ from one structure to another. We thus define z_i as the connectivity of component i within the seeded structure. Components with higher connectivity (z_i large) can achieve larger $r_i(\mathcal{N}_{\mathbf{x}})$, enabling faster binding and stronger discrimination than a weakly connected component (z_i small). For instance, in Fig. 2, component $i = 5$ has $z_5 = 6$ in the yellow target due to two diagonal bonds. As a consequence, its addition rate at the correct location $\mathbf{x} = (2, 2)$ in the growing yellow structure of Fig. 2(d) is $k_5^+(\mathcal{N}_{\mathbf{x}}) = \exp(2\delta)$, compared to $k_5^+(\mathcal{N}_{\mathbf{x}'}) = \exp(\delta)$ at a location $\mathbf{x}' = (3, 2)$ consistent with the red structure.

We track assembly size by counting the number of monomers on the lattice in units of N , and we define two measures of accuracy of tiles, a global and a local one. Global accuracy is measured as the fraction of correctly placed tiles relative to the seeded target within the square $\mathbf{x} = (1, 1)$ to (ℓ, ℓ) [Fig. 2(b)]. A tile contributes to the global accuracy of the seeded structure regardless of its local neighborhood. By contrast, local accuracy is defined through bonds: A tile is locally accurate if $r_i(\mathcal{N}_{\mathbf{x}}) \geq 1$. This measure is used to identify chimeric structures. We moreover characterize assembly dynamics via two characteristic timescales. The retrieval time τ_{ret} corresponds to the time it takes to retrieve a target structure, i.e., the time for the global accuracy to reach unity. Likewise, we define the lifetime τ_{life} as the time it takes to add one extra monomer to the completed structure.

In the following, we analyze retrieval of kinetically encoded structures under the rates of Eq. (1) for mixtures of components with varying connectivity.

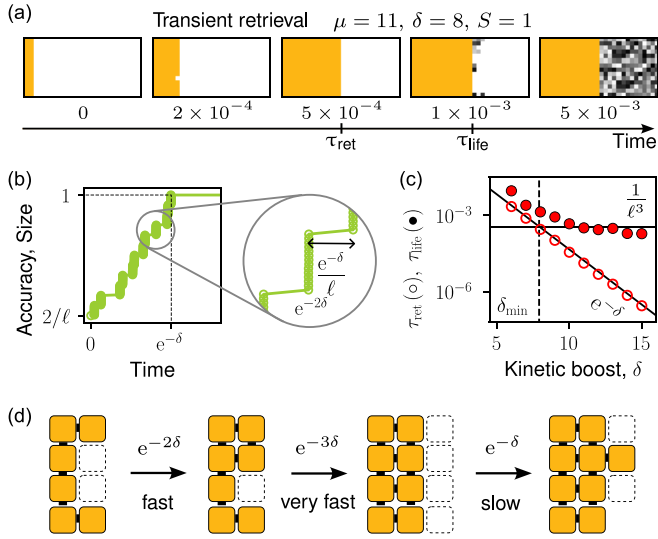


FIG. 3. The nearest-neighbor case yields slow retrieval. (a) Assembly snapshots showing retrieval of a single square target of side length $\ell = 14$ tiles from a two-layer nucleation seed. Assembly is allowed to grow beyond the target size ℓ up to 2ℓ . White denotes empty locations, yellow shows correctly placed tiles, and shades of gray denote incorrect tiles of different species (actually corresponding to chimeras of the target). Other growth regimes are described in Sec. S4 of Ref. [7]. (b) The global retrieval accuracy, i.e., the fraction of correctly placed tiles (line), follows the normalized assembly size (circles) over time and exhibits speed bottlenecks. (c) The target retrieval time τ_{ret} from Eq. (5) separates from the target lifetime τ_{life} for $\delta > \delta_{\text{min}}$ [Eq. (4)]. (d) In the irreversible high-discrimination regime, each addition event maximizes the number of bonds. The mean waiting time is dominated by the inverse of the addition rate [Eq. (1)], so additions with few bonds slow down the assembly.

III. SPEED BOTTLENECKS

As a starting point, we consider a mixture of N components, whose kinetics encode a single target structure $S = 1$ with bulk connectivity $z = 4$. (While bulk components have $z_i = 4$ encoded bonds, boundary components only have $z_i = 3$.) This corresponds to a simple model of heteromeric 2D self-assembly [31,34–36,45,51–53]. Based on previous work on kinetic encoding [37,40,42], successful self-assembly is expected when the kinetics are both highly irreversible, large μ , and highly discriminatory, large δ . Figure 3(a) shows snapshots of such a growing assembly from a small nucleation seed. In this regime, the target structure (yellow) is retrieved within a time τ_{ret} , but because it is not a free-energy minimum, i.e., not thermodynamically stable, incorrect tiles are eventually incorporated after a lifetime τ_{life} .

A closer analysis of the retrieval dynamics reveals that τ_{ret} is not governed by a typical timescale but instead by a few rare, slow events [see Fig. 3(b)], unlike in the case of heteropolymer growth [37]. Specifically, these bottleneck events have a characteristic time $\exp(-\delta)$, so that the retrieval time scales as $\tau_{\text{ret}} \approx \exp(-\delta)$ [Fig. 3(c)]. By contrast, the target lifetime τ_{life} is asymptotically independent of δ , ensuring kinetic stability of the structure. Yet, a much larger timescale separation is prevented due to the few events that act as speed bottlenecks.

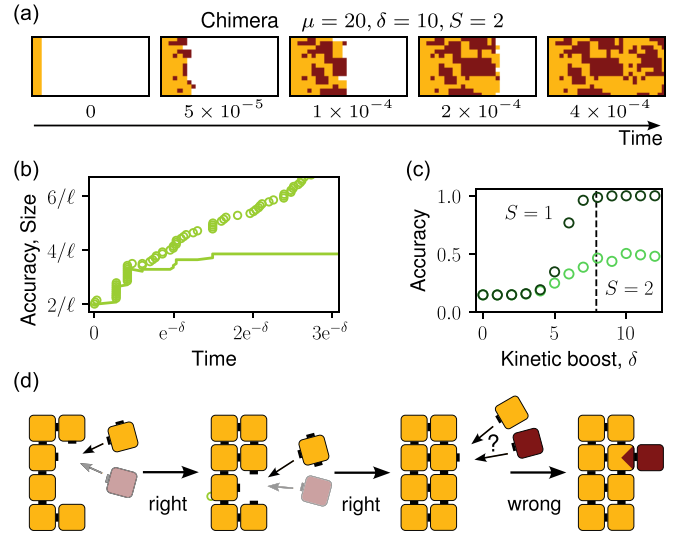


FIG. 4. Speed bottlenecks are also encoding bottlenecks. (a) Assembly snapshots for $S = 2$ arbitrary targets (yellow and red) show the formation of a chimeric structure (detailed description in Sec. S4 of Ref. [7]). Here tiles are colored according to the target with which they share the most correct neighbors, irrespective of global accuracy. (b) Errors make the global retrieval accuracy (line) depart from the normalized assembly size (circles). (c) For $S = 2$ targets, the maximal accuracy at large δ is around 50%. (d) Since each component has two different partners, kinetic discrimination is impossible at addition events with only one bond.

The origin of the bottlenecks can be understood intuitively as follows (a formal derivation is carried out in Sec. V). In the highly irreversible regime, the assembly dynamics is governed by the addition rates k_i^+ . For any boundary configuration, the most probable next additions are those with the largest k_i^+ . Assuming error-free growth, addition events fall into three classes: $O(N)$ addition events correspond to lateral layer growth, with $k_i^+ = \exp(2\delta)$; $O(\sqrt{N})$ events close a layer, with $k_i^+ = \exp(3\delta)$; and $O(\sqrt{N})$ events nucleate a new layer, with $k_i^+ = \exp(\delta)$. As illustrated in Fig. 3(d), the slowest of these, i.e., nucleation of new layers, sets the overall scaling of the retrieval time. Thus, although most growth events are fast, lasting a mean time $\exp(-2\delta)$, the rare bottleneck events dominate the assembly time, leading to $\tau_{\text{ret}} \approx \exp(-\delta)$.

IV. ENCODING BOTTLENECKS

So far, we have shown that for large μ and δ a single target structure can be reliably assembled, albeit with speed bottlenecks. We now turn to the case in which the same mixture encodes two distinct assemblies $S = 2$. (Bulk components thus have a set of $z_i = 4$ partners in each structure.) Figure 4(a) shows that, in this setting, a seed corresponding to a given target (yellow) does not grow reliably. Instead, the resulting structure is a chimera: Fragments of the alternative target (red) appear interwoven with those of the seeded target (yellow). Consequently, although the assembly grows over time, retrieval accuracy remains low, around 50% [see Fig. 4(b)]. Importantly, this low fidelity cannot be remedied by increasing the discrimination energy δ [see Fig. 4(c)]. The

reason is that the errors are not of thermal origin, but of combinatorial origin, as already noted for equilibrium assemblies [35,36].

The emergence of chimeras for $S = 2$ is directly linked to the speed bottlenecks identified above. To see this, recall that in the fully irreversible regime the added components are those with maximal rate k_i^+ . For components of the seeded target (yellow), these rates scale as $\exp(\delta)$, $\exp(2\delta)$, and $\exp(3\delta)$ for nucleating, growing, and closing a layer, respectively. By contrast, erroneous additions from the confounding target (red) almost never benefit from multiple coincident neighbors, and thus their rate scales as $\exp(\delta)$. As a result, at the onset of each new assembly layer, the correct and incorrect components compete on equal footing, leading to a roughly 50% chance of error [see Fig. 4(d)]. These rare but slow nucleation events thus have two implications: They limit the overall growth speed and, at the same time, render kinetic discrimination ineffective, producing encoding bottlenecks.

In the next section, we characterize the general conditions required for both fast and accurate retrieval and propose a simple change in our model that resolves these dual bottlenecks.

V. CONDITIONS FOR FAST RETRIEVAL

So far, we have seen that fast retrieval of target structures is possible, but limited by encoding and speed bottlenecks (see Figs. 3 and 4). In this section, we determine the conditions on the chemical potential μ , the discrimination barrier δ , the number of encoded structures S , and the assembly size N that allow for fast retrieval of multiple structures. To this end, we unravel the origin of both speed and encoding bottlenecks so that they can be resolved.

A. Speed and encoding bottlenecks

We begin by defining speed and encoding bottlenecks more precisely. Retrieval of kinetically encoded targets is most effective in the strongly irreversible and highly discriminatory limit $\mu \rightarrow \infty$ and $\delta \rightarrow \infty$. In this scenario, the assembly of a structure of size N from a seed of size N_{seed} proceeds through exactly $N_{\text{tot}} = N - N_{\text{seed}}$ successive addition events, without removals [Eq. (1)].

We label these N_{tot} subsequent addition events in order of occurrence by a “time” index $t = 1, 2, \dots, N_{\text{tot}}$, distinct from the physical time τ . Each event is characterized by a pair $(i^t, \mathbf{x}^t)$ indicating the component i^t added at location $\mathbf{x}^t$. Given that, at time $t - 1$, the assembly is characterized by neighborhoods $\mathcal{N}_{\mathbf{x}}^{t-1}$ at each location $\mathbf{x}$, the event at time t is among those that maximize the addition rate:

$$(i^t, \mathbf{x}^t) \in \arg \max_{(i, \mathbf{x})} [k_i^+ (\mathcal{N}_{\mathbf{x}}^{t-1})]. \quad (2)$$

Typically, multiple components and locations satisfy Eq. (2), so there are multiple kinetically equivalent events at each time. We further denote the rate associated to the addition event $(i^t, \mathbf{x}^t)$ by k^t , and the corresponding number of bonds is defined as $n^t = \ln(k^t)/\delta$ [see Eq. (1)].

Critical events correspond to “correct” addition events that are much slower than the rest, potentially as slow as an “in-

correct” addition. We therefore compute the critical addition rate as the mini-max

$$k_c = \min_t k^t = \min_t \max_{(i, \mathbf{x})} [k_i^+ (\mathcal{N}_{\mathbf{x}}^{t-1})] \quad (3)$$

and accordingly define $n_c = \ln(k_c)/\delta$. A speed bottleneck occurs when k_c corresponds to few events, so these few slow additions determine the overall retrieval time. An encoding bottleneck occurs when the critical rate k_c is equal to the maximal rate of incorrect additions so that incorrect and correct monomers are indistinguishable.

As an example, consider the case in which structures have a bulk connectivity $z = 4$, already discussed before. As the assembly grows from a seed, the number of bonds n^t established by each added component changes over time. As sketched in Fig. 3(d), layer nucleation events create $n^t = 1$ bond, layer growth events create $n^t = 2$ bonds, and layer completion events create $n^t = 3$ bonds. Because there are many fewer layer nucleation events than the total number of events, and because these events are the slowest, they act as speed bottlenecks, with a critical addition rate of $k_c = \exp(\delta)$. Furthermore, if $S = 2$ structures are encoded, incorrect additions also occur at rate $\exp(\delta)$, and therefore these nucleation events also constitute encoding bottlenecks [see Fig. 4(d)].

B. Conditions for accuracy at high speed

To investigate the conditions for accurate high-speed assembly, we focus on a mixture that encodes a single structure, and so $S = 1$. In this scenario, only one correct monomer can be added at each boundary location, and so $\mu > 0$ is necessary to ensure growth. To determine the minimum discrimination strength necessary for reliable assembly, $\delta_{\min}$, we focus on the $N_c \leq N_{\text{tot}}$ critical events with $k^t = k_c$ that may act as speed bottlenecks. For large but finite δ , each critical event may correspond to either a correct or an incorrect addition. Correct additions respect the condition in Eq. (2) and have an addition rate $k_c = \exp(n_c \delta)$, while incorrect additions have a rate $\exp(0) = 1$. Denoting the multiplicity of correct and incorrect additions at critical events by Ω_C and Ω_I , we show in Sec. S6 of Ref. [7] that the minimum discrimination barrier necessary for accurate high-speed retrieval is

$$\delta_{\min} = \frac{1}{n_c} \ln \frac{N_c \Omega_I}{\Omega_C}. \quad (4)$$

For the simple scenario with connectivity $z = 4$ and a square target with side length $\ell = \sqrt{N}$ (Fig. 3), we have that addition events that nucleate a new layer create only $n_c = 1$ bond. At each of these $N_c = \ell$ critical events (minus the initial seed length), there is one correct monomer and $N - 1$ incorrect monomers at each of the ℓ locations. This corresponds to a multiplicity of correct additions $\Omega_C = \ell$ and a multiplicity of incorrect additions $\Omega_I \approx \ell(N - 1) \approx \ell^3$. Equation (4) then yields $\delta_{\min} \approx 3 \ln \ell$.

Conditions $\mu > 0$ and $\delta > \delta_{\min}$ result in accurate retrieval of a target structure, as confirmed in Figs. 3(a) and 3(b). We can further estimate the time it takes to retrieve a structure from the sum of the mean waiting times of all N_c critical events, since these are exponentially slower than the remaining ones. For each critical event, we approximate the mean

waiting time as $1/\Omega_C k_c$. We therefore have

$$\tau_{\text{ret}} \approx \frac{N_c}{\Omega_C} \exp(-n_c \delta), \quad (5)$$

which reduces to $\tau_{\text{ret}} \approx \exp(-\delta)$ for the case $z = 4$ shown in Fig. 3(c). By contrast, the lifetime of the target structure τ_{life} , defined as the time it takes to add a few incorrect monomers, reads $\tau_{\text{life}} \approx 1/\Omega_I$, because monomers at the target boundary do not have encoded bonds with further monomers. At large δ , these two timescales separate, ensuring kinetic stability of the target. Notice, however, that in the thermodynamic limit $N \rightarrow \infty$, the minimum discrimination energy diverges logarithmically and the lifetime vanishes. Therefore, kinetic encoding only applies to structures of finite size.

C. Conditions for combinatorial encoding

In this section, we discuss the conditions for encoding several targets from the same components. When the components of the target being assembled are shared among $S \geq 2$ targets, some incorrect additions will also be accelerated via δ . These combinatorial errors arise whenever two components have the same neighbors in different targets. For instance, in Fig. 2(d) at time $t = 2$, both components 5 and 8 can bind to component 7 with rate $\exp(\delta)$, because they are partners of 7 in the red and yellow structures, respectively. As a result, components 5 and 8 become effectively indistinguishable to the growing assembly. This implies the existence of an encoding capacity S_{max} , above which errors cannot be avoided by increasing δ .

To elucidate the origin of these errors, we focus again on the critical events with $k^t = k_c$ that can act as encoding bottlenecks. In the simple case $z = 4$, each critical addition involves only $n_c = 1$ bond, which is insufficient to discriminate between components. Indeed, most components have different neighbors in different targets [Fig. 4(d)]. Consequently, already for $S = 2$ targets chosen at random, the assembled structure accumulates numerous errors [see Fig. 4(a)]. Thus, for $n_c = 1$, the encoding capacity is limited to $S_{\text{max}} = 1$. By contrast, for $n_c \geq 2$, we find

$$S_{\text{max}} \sim N^{1-2/n_c} \quad (6)$$

(see Sec. S6 in Ref. [7]). The scaling ranges from $O(1)$ for $n_c = 2$ to $O(N)$ in the fully connected case with $n_c = N \gg 1$. The corresponding increase in encoding capacity results from a drastic reduction in errors. This is most visible in the low-promiscuity regime, where S exceeds 1 but remains well below S_{max} . In that case, we estimate the minimum error fraction at large δ as

$$e_{\text{min}} \sim N_c/N^{n_c-1} \quad (7)$$

(see Sec. S6 in Ref. [7]), which decreases rapidly with increasing n_c .

Next we examine how reusability of components affects the characteristic timescales of self-assembly. While the retrieval time is still expected to follow Eq. (5), the target lifetime is shortened. This is because components at a target boundary generally belong to the bulk of other targets, and therefore completion of one target accelerates subsequent incorrect additions, yielding $\tau_{\text{life}} \lesssim \exp(-\delta)$. More specifically,

as the number of structures S grows, they will overlap more, which will decrease τ_{life} and thus impede timescale separation.

In summary, the conditions $\mu > 0$, $\delta > \delta_{\text{min}}$, and $S < S_{\text{max}}$ are required to ensure forward growth devoid of thermal end combinatorial errors. Ultimately, the speed, encoding capacity, and error fraction of Eqs. (5)–(7) depend acutely on the number of bonds at critical events, n_c . In the next two sections, we demonstrate how controlling n_c allows us to resolve speed and encoding bottlenecks.

VI. RESOLVING BOTTLENECKS

To increase robustness in our self-assembly model, we focus on the critical addition events. These typically manifest at assembly boundaries and at the onset of new layers. By contrast, additions that extend a layer tend to create more bonds and thus proceed faster [Eq. (1)]. In this manner, growth tends to proceed layer by layer, naturally favoring layered growth. In the following, we analyze the dependence of the critical number of bonds per addition n_c on the monomer connectivity z .

In the case of 1D self-assembly with nearest-neighbor interactions, each added monomer forms $n_c = 1$ bond, which hampers monomer discrimination and limits assembly speed [37]. Achieving $n_c = 2$ requires that each monomer interacts with its next-nearest neighbors, increasing the bulk connectivity from $z = 2$ to 4. By contrast, for quasi-2D assemblies which are the focus of this paper, n_c can be increased without significantly increasing the connectivity of the whole assembly. Indeed, because the critical events correspond to the creation of new layers, one additional diagonal bond per layer in each target suffices to reach $n_c = 2$ [see Fig. 5(a)]. Therefore, among the $N = \ell \times \ell$ assembly components, only 2ℓ need an increased connectivity $z_i = 5$ instead of $z_i = 4$, which only marginally increases the average bulk connectivity of structures to $z = 4 + 2/\ell$ (hereafter written as 4^+). From Eqs. (5)–(7), we expect drastic improvements in retrieval time τ_{ret} , encoding capacity S_{max} , and minimum error e_{min} , even though the change in connectivity of structures is small.

This minor connectivity modification raises the rate of critical additions to $k_c = \exp(2\delta)$ and increases their number from $N_c = \ell$ to $\ell(\ell - 1)$ so that there are no more speed bottlenecks (see Sec. S5 in Ref. [7]). Moreover, the multiplicity of correct additions at critical events collapses from $\Omega_C = \ell$ to 1 or 2, while the multiplicity of incorrect additions Ω_I remains unchanged. This enables fast and reliable retrieval of the seeded target despite components being reused in $S = 2$ targets [see Fig. 5(b)]. Separation between retrieval time and lifetime occurs above the reduced discrimination threshold $\delta_{\text{min}} = \frac{5}{2} \ln \ell$ [Fig. 5(c) and Eq. (4)]. While the lifetime shortens to $\tau_{\text{life}} \sim \exp(-\delta)$, retrieval becomes exponentially faster due to the elimination of bottlenecks, consistent with the prediction $\tau_{\text{ret}} \approx N \exp(-2\delta)$ [Eq. (5)]. Furthermore, Eq. (6) correctly predicts a nonscaling encoding capacity $S_{\text{max}} = O(1)$ for $n_c = 2$ [see Fig. 5(d) herein and Fig. S10 in Ref. [7] regarding error fraction].

To conclude, increases in local connectivity at assembly bottlenecks funnels growth toward the seeded target. This

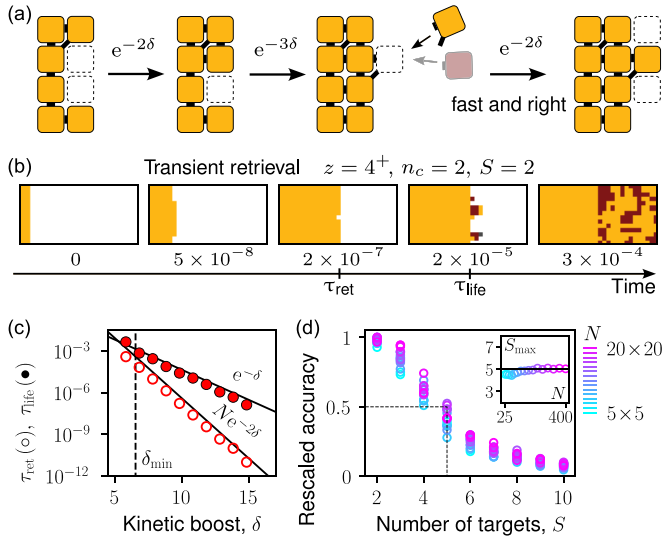


FIG. 5. Local increases in connectivity resolve assembly bottlenecks. (a) Schematic showing that adding one diagonal bond per layer enables fast and accurate retrieval of $S = 2$ targets by ensuring $n_c = 2$. (b) Assembly snapshots show faithful retrieval for the same settings as in Fig. 4(a), except that $z = 4^+$. The coloring is the same as in Fig. 4(a). (c) Retrieval time and lifetime separate at large δ , as predicted. (d) Accuracy, rescaled between 0 and 1, decreases sigmoidally with S . The inset shows that the sigmoid midpoint $S_{\max}$ depends only weakly on N [Eq. (6)].

results in drastic improvements in retrieval speed and accuracy, conferring robustness against moderate component promiscuity.

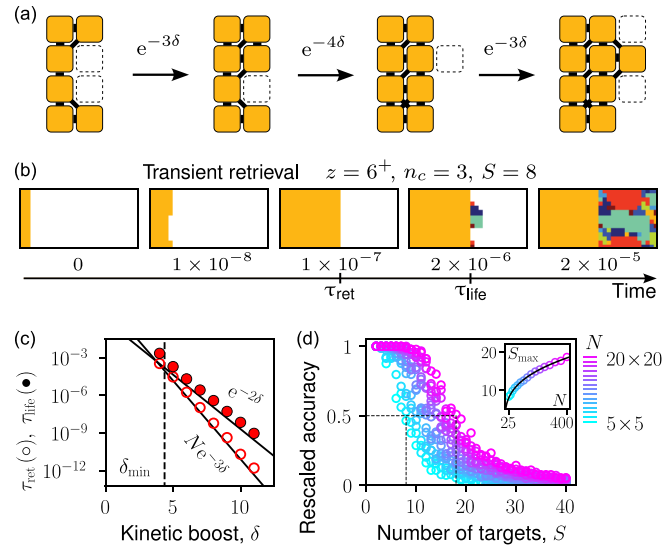


FIG. 6. Larger increases in connectivity enable combinatorial encoding of target structures. (a) The connectivity is raised to $z = 6^+$, ensuring a minimum number of bonds per addition $n_c = 3$. (b) Assembly snapshots show faithful retrieval for $S = 8$ targets. Here $\mu = 20$ and $\delta = 7$. (c) Retrieval time and lifetime separate once again at large δ , ensuring kinetic stability. (d) For $n_c = 3$, $S_{\max}$ scales as $N^{1/3}$, in agreement with Eq. (6). (e) To interpolate between the cases $z = 4^+$ and 6^+ , we gradually increase the number of diagonal bonds in target structures as shown. (Tile contours and nondiagonal bonds are omitted for clarity.) (f) Dependence of timescales on connectivity $z > 4$ for $S = 2$, $\mu = 20$, $\delta = 10$, and target side lengths $\ell = 7, 14, 20$. The retrieval time τ_{ret} follows Eq. (5) (lines). Cases with connectivities $z = 4^+$ and 6^+ are indicated by $+$ markers. (g) The encoding capacity increases with target size depending on connectivity. The fits (lines) use the scaling exponent $\alpha = 1 - 4/z$ from Eq. (8).

VII. COMBINATORIAL SELF-ASSEMBLY

In the preceding section, we showed that encoding bottlenecks can be overcome by adding an extra bond to $2\ell \ll N$ components. However, the maximal number of stored structures did not scale with N . Equation (6) predicts that the encoding capacity $S_{\max}$ scales with N for $n_c \geq 3$. In this section, we explore how to achieve this combinatorial scaling in the number of target structures.

To guarantee that $n_c = 3$, we consider assemblies with a bulk connectivity $z = 6$, whereby all components form diagonal bonds with their north-west and south-east neighbors in each target. We then repeat the strategy in the preceding section and add an additional bond for one component per layer (in the south-west direction) so that these components and their partners have $z_i = 7$. This modification raises the average bulk connectivity to $z = 6 + 2/\ell$ (hereafter denoted by 6^+) and ensures that $n_c = 3$ [see Fig. 6(a)].

The higher connectivity leads to faithful retrieval for $S \geq 2$ targets [Fig. 6(b)] above the reduced discrimination threshold $\delta_{\min} \approx \frac{5}{3} \ln \ell$ [Eq. (4)]. As shown in Fig. 6(c), for S near $S_{\max}$, the large overlap between targets shortens lifetimes to $\tau_{\text{life}} \approx \exp(-2\delta)$, while $\tau_{\text{ret}} \approx N \exp(-3\delta)$ [Eq. (5)], ensuring kinetic stability of targets. Likewise, the encoding capacity $S_{\max}$ follows the predicted scaling $N^{1/3}$ from Eq. (6) [Fig. 6(d)]. Alternatively, for just $S = 2$ targets at $z = 6^+$, the error fraction decreases with size as $e_{\min} \sim 1/N$ [Eq. (7)] (see Fig. S10 in Ref. [7]). This validates our predictions and points to robust target designs.

To further assess the effect of connectivity, we interpolate between $z = 4^+$ and 6^+ by gradually adding diagonal bonds in the square lattice [see Fig. 6(e)]. Despite this gradual

increase in connectivity, the critical number of bonds n_c changes discretely, e.g., from 2 to 3 as z increases from 6 to 6^+ (see Sec. S5 in Ref. [7]). We thus expect corresponding jumps in the downward trend of retrieval times τ_{ret} [Eq. (5)] and in the upward trend of encoding capacity S_{max} . Indeed, τ_{ret} drops sharply with connectivity [see Fig. 6(f) for $S = 2$ and Fig. S7 in Ref. [7] for S near S_{max}]. However, the maximum number of targets S_{max} , governed by the scaling exponent $\alpha = 1 - 2/n_c$, grows smoothly with z [see Fig. 6(g)]. Interpolating n_c in the expression for α as $z/2$ (see Sec. S6 in Ref. [7]) yields the combinatorial scaling

$$S_{\text{max}} \sim N^{1-4/z} \quad \text{for } z \geq 4, \quad (8)$$

which captures well the observed increase in encoding capacity with increasing connectivity. This expression shows that for $z = 4$, the encoding capacity does not scale with system size, as previously discussed, unlike in previous models in which structures are encoded through binding energies [35,36]. In contrast, the scalings for $z = 6$ and 8 are $O(N^{1/3})$ and $O(N^{1/2})$, respectively.

In summary, kinetic encoding enables fast retrieval far from equilibrium. Within a simple quasi-2D geometry, gradually raising the connectivity of structures progressively accelerates retrieval and buffers it against thermal and combinatorial errors.

VIII. DISCUSSION

Summary. In this article we have shown that a large heteromeric structure can rapidly self-assemble from a small seed through kinetic discrimination of its components. This extends our previous work on polymerization [37] to the case of quasi-2D growth. Nonetheless, in simple assembly models with fixed connectivity, few specific additions weaken discrimination and decrease assembly speed. These speed and encoding bottlenecks constitute a key difference from 1D growth, where assembly steps have comparable timescales and error rates. We found that adding a single bond per assembly layer alleviates these bottlenecks, yielding faster and more robust assembly. In particular, these localized changes in connectivity greatly reduce errors due to component promiscuity. More broadly, gradual increases in connectivity lead to progressive improvements in assembly speed, accuracy, and encoding capacity.

Relationship to previous theoretical work. This study builds on recent advances showing that mixtures with many components can encode higher-order structures, both solid [35,36] and liquid [54,55], as minima of a high-dimensional free-energy landscape, akin to pattern retrieval in Hopfield neural networks [56–58]. Beyond classical free-energy minimization, achieving rapid far-from-equilibrium growth requires careful design of assembly pathways [59,60].

Within the framework of nonequilibrium self-assembly, selectively accelerating specific binding events emerges as a powerful mechanism to kinetically trap assemblies along pre-encoded target pathways [37]. Such kinetic discrimination has proven effective in diverse contexts, including copolymerization [39,41,61,62], minimal DNA self-assembly [63], error correction [34,40,64], random walks [65], and associative memory [42]. Across these systems, optimal retrieval

occurs far from equilibrium, thereby overcoming the intuitive speed-accuracy tradeoff [39,40,66].

Previous self-assembly studies have highlighted the importance of monomer geometry and connectivity [31–33]. Yet, most 2D lattice models, relying on a single monomer type, restrict bulk connectivity to discrete values: Triangles bind to three neighbors [31], squares to four [31,36,45,48,49,52,53], and hexagons to six [31,67]. Our approach goes beyond these geometric constraints: By tuning the connectivity of selected components, we enable a continuous and versatile route toward optimal assembly performance.

Overall, the present work provides a new instance of kinetic encoding, in which kinetic discrimination and tailored connectivity cooperate to robustly guide growth along predefined target pathways.

Implications for biological self-assembly. As shown in the data analysis of Fig. 1, both kinetic discrimination and variable connectivity are key features of the assembly of large heteromeric protein complexes. Interestingly, several of our results parallel biological assembly principles.

First, the speed bottlenecks that naturally arise in our toy model are similarly observed during *in vitro* ribosome assembly, where they manifest via the accumulation of assembly intermediates [17–20]. Assembly speed and robustness are particularly sensitive to intervention at these bottlenecks. In our toy model, we increased the connectivity of components that participate to these bottlenecks to enhance their kinetic discrimination, leveraging cooperative binding. An alternative strategy is to catalyze these bottlenecks by way of assembly factors (see Sec. S2 in Ref. [7]). This may provide a simple explanation for the increase in the number of assembly factors with assembly size [Fig. 1(d)]. Second, we find that assemblies with low-promiscuity components achieve very low error rates at high connectivity. This feature is especially relevant for ribosomes, whose components are seldom shared with other complexes [68], but whose flawless assembly is vital for accurate protein translation [69]. Third, kinetic encoding, whether arising from cooperative binding events or the action of assembly factors, tends to reduce the number of possible assembly pathways (see Sec. S7 in Ref. [7]). This mirrors the ordered assembly observed in many complexes [5,8,12,20,70], where growth proceeds through well-defined intermediates.

In conclusion, our toy model offers a unifying framework that helps simultaneously explain the high assembly speed, the small number of pathways, and the need for many assembly factors in large heteromeric protein complexes.

Model extensions. We now review some possible extensions of our toy model. The core assumption behind our model is cooperative monomer binding: Just as a component with two bonds has twice the binding energy under energetic discrimination, its kinetic barrier shrinks twice under kinetic discrimination. Extensions could explore more general dependences of binding acceleration on the number of bonds. In Sec. S4 of Ref. [7], we in fact discuss the case of saturation in binding acceleration. The model further assumes uniform chemical potential and discrimination strength across component species. We find that small variations of these quantities do not qualitatively alter the main results (see Sec. S8 in Ref. [7]).

Additional extensions include fully two-dimensional growth [35,36] and less constrained off-lattice geometries [32]. Other important directions include the incorporation of allosteric interactions [71] beyond the next-nearest interactions considered here, as well as energetic discrimination mechanisms [35,36,51] alongside the kinetic discrimination studied in this work. More realistic connectivity patterns [Fig. 1(e)], as well as component diffusion [72] and depletion effects [73], also constitute relevant extensions (see the extended discussion in Sec. S9 of Ref. [7] and Refs. [74–77] therein). More broadly, an important direction for future work is to determine how the present results extend to more realistic three-dimensional models of large heteromeric complex self-assembly [29,30].

The catalysis of correct monomer binding events has brought about a new high-speed self-assembly regime. Beyond catalysis, other roles of assembly factors, such as proofreading [34,40,78], could be integrated into future

models. More broadly, insights from biological assembly principles may guide the design of synthetic systems. For instance, integrating assembly-factor-like mechanisms into DNA tile assemblies [79–81] could substantially boost both yield and speed.

ACKNOWLEDGMENTS

This work was partly supported by La Caixa Foundation Grant No. LCF/BQ/PI21/11830032 and core funding from the Gulbenkian Foundation.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

- [1] H. Khattner, A. G. Myasnikov, S. K. Natchiar, and B. P. Klaholz, Structure of the human 80S ribosome, *Nature (London)* **520**, 640 (2015).
- [2] M. C. Wahl, C. L. Will, and R. Lührmann, The spliceosome: Design principles of a dynamic RNP machine, *Cell* **136**, 701 (2009).
- [3] S. S. Chen, E. Sperling, J. M. Silverman, J. H. Davis, and J. R. Williamson, Measuring the dynamics of *E. coli* ribosome biogenesis using pulse-labeling and quantitative mass spectrometry, *Mol. BioSyst.* **8**, 3325 (2012).
- [4] K. Dörner, C. Ruggeri, I. Zemp, and U. Kutay, Ribosome biogenesis factors—from names to functions, *EMBO J.* **42**, EMBJ2022112699 (2023).
- [5] I. Vercellino and L. A. Sazanov, The assembly, regulation and function of the mitochondrial respiratory chain, *Nat. Rev. Mol. Cell Biol.* **23**, 141 (2022).
- [6] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The Protein Data Bank, *Nucl. Acids Res.* **28**, 235 (2000).
- [7] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/5w6h-1931> for additional details.
- [8] A. Rousseau and A. Bertolotti, Regulation of proteasome assembly and activity in health and disease, *Nat. Rev. Mol. Cell Biol.* **19**, 697 (2018).
- [9] F. Altegoer and G. Bange, Undiscovered regions on the molecular landscape of flagellar assembly, *Curr. Opin. Microbiol.* **28**, 98 (2015).
- [10] T. Wild and P. Cramer, Biogenesis of multisubunit RNA polymerases, *Trends Biochem. Sci.* **37**, 99 (2012).
- [11] J. Woolford and S. J. Baserga, Ribosome biogenesis in the yeast *Saccharomyces cerevisiae*, *Genetics* **195**, 643 (2013).
- [12] E. Lavdovskaia *et al.*, A roadmap for ribosome assembly in human mitochondria, *Nat. Struct. Mol. Biol.* **31**, 1898 (2024).
- [13] S. Heinz, P. Liauw, J. Nickelsen, and M. Nowaczyk, Analysis of photosystem II biogenesis in cyanobacteria, *Biochim. Biophys. Acta Bioenerg.* **1857**, 274 (2016).
- [14] H. Yang, J. Liu, X. Wen, and C. Lu, Molecular mechanism of photosystem I assembly in oxygenic organisms, *Biochim. Biophys. Acta Bioenerg.* **1847**, 838 (2015).
- [15] T. Rühle and D. Leister, Assembly of F_1F_0 -ATP synthases, *Biochim. Biophys. Acta Bioenerg.* **1847**, 849 (2015).
- [16] J. P. Armitage and R. M. Berry, Assembly and dynamics of the bacterial flagellum, *Annu. Rev. Microbiol.* **74**, 181 (2020).
- [17] J. H. Davis and J. R. Williamson, Structure and dynamics of bacterial ribosome biogenesis, *Philos. Trans. R. Soc. B* **372**, 20160181 (2017).
- [18] A. Seffouh, R. Nikolay, and J. Ortega, Critical steps in the assembly process of the bacterial 50S ribosomal subunit, *Nucl. Acids Res.* **52**, 4111 (2024).
- [19] A. M. Mulder, C. Yoshioka, A. H. Beck, A. E. Bunner, R. A. Milligan, C. S. Potter, B. Carragher, and J. R. Williamson, Visualizing ribosome biogenesis: Parallel assembly pathways for the 30S subunit, *Science* **330**, 673 (2010).
- [20] K. Sheng, N. Li, J. N. Rabuck-Gibbons, X. Dong, D. Lyumkis, and J. R. Williamson, Assembly landscape for the bacterial large ribosomal subunit, *Nat. Commun.* **14**, 5220 (2023).
- [21] H. P. Erickson, Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy, *Biol. Proced. Online* **11**, 32 (2009).
- [22] S. Reuveni, M. Ehrenberg, and J. Paulsson, Ribosomes are optimized for autocatalytic production, *Nature (London)* **547**, 293 (2017).
- [23] A. E. Bunner, A. H. Beck, and J. R. Williamson, Kinetic cooperativity in *E. coli* 30S ribosomal subunit reconstitution reveals additional complexity in the assembly landscape, *Proc. Natl. Acad. Sci. USA* **107**, 5417 (2010).
- [24] J. R. Williamson, Cooperativity in macromolecular assembly, *Nat. Chem. Biol.* **4**, 458 (2008).
- [25] E. S. Chhabra and H. N. Higgs, The many faces of actin: Matching assembly factors with cellular structures, *Nat. Cell Biol.* **9**, 1110 (2007).
- [26] J. von der Ecken, M. Müller, W. Lehman, D. J. Manstein, P. A. Penczek, and S. Raunser, Structure of the F-actin–tropomyosin complex, *Nature (London)* **519**, 114 (2015).
- [27] A. Fellous, J. Francon, A.-M. Lennon, and J. Nunez, Microtubule assembly in vitro, *Eur. J. Biochem.* **78**, 167 (1977).
- [28] M. G. Mateu, Assembly, stability and dynamics of virus capsids, *Arch. Biochem. Biophys.* **531**, 65 (2013).

- [29] I. Kurisaki and S. Tanaka, Computational prediction of heteromeric protein complex disassembly order using hybrid Monte Carlo/molecular dynamics simulation, *Phys. Chem. Chem. Phys.* **24**, 10575 (2022).
- [30] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, Molecular dynamics simulations of large macromolecular complexes, *Curr. Opin. Struct. Biol.* **31**, 64 (2015).
- [31] F. M. Gartner and E. Frey, Design principles for fast and efficient self-assembly processes, *Phys. Rev. X* **14**, 021004 (2024).
- [32] M. Lenz and T. Witten, Geometrical frustration yields fibre formation in self-assembly, *Nat. Phys.* **13**, 1100 (2017).
- [33] V. N. Manoharan, Colloidal matter: Packing, geometry, and entropy, *Science* **349**, 1253751 (2015).
- [34] R. Ravasio, K. Husain, C. G. Evans, R. Phillips, M. Ribezzi, J. W. Szostak, and A. Murugan, A minimal scenario for the origin of non-equilibrium order, [arXiv:2405.10911](https://arxiv.org/abs/2405.10911).
- [35] A. Murugan, Z. Zeravcic, M. P. Brenner, and S. Leibler, Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures, *Proc. Natl. Acad. Sci. USA* **112**, 54 (2015).
- [36] P. Sartori and S. Leibler, Lessons from equilibrium statistical physics regarding the assembly of protein complexes, *Proc. Natl. Acad. Sci. USA* **117**, 114 (2020).
- [37] F. Benoist and P. Sartori, High-speed combinatorial polymerization through kinetic-trap encoding, *Phys. Rev. Lett.* **134**, 038402 (2025).
- [38] K. A. Dill, K. M. Fiebig, and H. S. Chan, Cooperativity in protein-folding kinetics, *Proc. Natl. Acad. Sci. USA* **90**, 1942 (1993).
- [39] C. H. Bennett, Dissipation-error tradeoff in proofreading, *Biosystems* **11**, 85 (1979).
- [40] P. Sartori and S. Pigolotti, Kinetic versus energetic discrimination in biological copying, *Phys. Rev. Lett.* **110**, 188101 (2013).
- [41] Y.-C. Tsai and K. A. Johnson, A new paradigm for DNA polymerase specificity, *Biochemistry* **45**, 9675 (2006).
- [42] F. Benoist, L. Peliti, and P. Sartori, Content-addressable memory with a content-free energy function, *Phys. Rev. Lett.* **135**, 237102 (2025).
- [43] A.-C. Gavin *et al.*, Proteome survey reveals modularity of the yeast cell machinery, *Nature (London)* **440**, 631 (2006).
- [44] S. Kühner *et al.*, Proteome organization in a genome-reduced bacterium, *Science* **326**, 1235 (2009).
- [45] S. Osat and R. Golestanian, Non-reciprocal multifarious self-organization, *Nat. Nanotechnol.* **18**, 79 (2023).
- [46] M. C. Hübl, T. E. Videbæk, D. Hayakawa, W. B. Rogers, and C. P. Goodrich, A polyhedral structure controls programmable self-assembly, *Nat. Phys.* **22**, 294 (2026).
- [47] T. Diep *et al.*, Universal 3D Nanoscale Assembly with DNA Polycubes, [ChemRxiv](https://chemrxiv.org/abs/2025010001) (2025).
- [48] S. Whitlam, L. O. Hedges, and J. D. Schmit, Self-assembly at a nonequilibrium critical point, *Phys. Rev. Lett.* **112**, 155504 (2014).
- [49] M. Nguyen and S. Vaikuntanathan, Design principles for nonequilibrium self-assembly, *Proc. Natl. Acad. Sci. USA* **113**, 14231 (2016).
- [50] F. Chen and W. M. Jacobs, Dynamical phase transition in the growth of programmable polymorphic materials, *Phys. Rev. Mater.* **9**, 053403 (2025).
- [51] G. Bisker and J. L. England, Nonequilibrium associative retrieval of multiple stored self-assembly targets, *Proc. Natl. Acad. Sci. USA* **115**, E10531 (2018).
- [52] A. Bupathy, D. Frenkel, and S. Sastry, Temperature protocols to guide selective self-assembly of competing structures, *Proc. Natl. Acad. Sci. USA* **119**, e2119315119 (2022).
- [53] M. Holmes-Cerfon and M. Wyart, Hierarchical self-assembly for high-yield addressable complexity at fixed conditions, *Proc. Natl. Acad. Sci. USA* **122**, e2500245122 (2025).
- [54] R. B. Teixeira, G. Carugno, I. Neri, and P. Sartori, Liquid Hopfield model: Retrieval and localization in multicomponent liquid mixtures, *Proc. Natl. Acad. Sci. USA* **121**, e2320504121 (2024).
- [55] W. M. Jacobs, Self-assembly of biomolecular condensates with shared components, *Phys. Rev. Lett.* **126**, 258101 (2021).
- [56] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [57] D. Krotov and J. J. Hopfield, Dense associative memory for pattern recognition, in *Proceedings of the Annual Conference on Advances in Neural Information Processing Systems 29, Barcelona, 2016*, edited by D. D. Lee, M. Sugiyama, U. von Luxburg, I. Guyon, and R. Garnett (Curran Associates, Red Hook, 2016).
- [58] H. Ramsauer *et al.*, Hopfield networks is all you need, *Proceedings of the Ninth International Conference on Learning Representations (ICLR, Appleton, 2021)*, paper SLSWGHA21.
- [59] S. Whitlam and R. L. Jack, The statistical mechanics of dynamic pathways to self-assembly, *Annu. Rev. Phys. Chem.* **66**, 143 (2015).
- [60] A. McMullen, M. Muñoz Basagoiti, Z. Zeravcic, and J. Bruijic, Self-assembly of emulsion droplets through programmable folding, *Nature (London)* **610**, 502 (2022).
- [61] K. Banerjee, A. B. Kolomeisky, and O. A. Igoshin, Elucidating interplay of speed and accuracy in biological error correction, *Proc. Natl. Acad. Sci. USA* **114**, 5183 (2017).
- [62] S. Pigolotti and P. Sartori, Protocols for copying and proofreading in template-assisted polymerization, *J. Stat. Phys.* **162**, 1167 (2016).
- [63] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce, Programming biomolecular self-assembly pathways, *Nature (London)* **451**, 318 (2008).
- [64] J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity, *Proc. Natl. Acad. Sci. USA* **71**, 4135 (1974).
- [65] B. Lefebvre and C. Maes, Frenetic steering in a nonequilibrium graph, *J. Stat. Phys.* **190**, 90 (2023).
- [66] D. Andrieux and P. Gaspard, Nonequilibrium generation of information in copolymerization processes, *Proc. Natl. Acad. Sci. USA* **105**, 9516 (2008).
- [67] L. Koehler, P. Ronceray, and M. Lenz, How do particles with complex interactions self-assemble? *Phys. Rev. X* **14**, 041061 (2024).
- [68] J. R. Warner and K. B. McIntosh, How common are extraribosomal functions of ribosomal proteins? *Mol. Cell* **34**, 3 (2009).
- [69] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. (Garland Science, New York, 2015).
- [70] M. Bai, X. Zhao, K. Sahara, Y. Ohte, Y. Hirano, T. Kaneko, H. Yashiroda, and S. Murata, In-depth analysis of the lid

- subunits assembly mechanism in mammals, [Biomolecules](#) **9**, 213 (2019).
- [71] J. Metson, Designing complex behaviors using transition-based allosteric self-assembly, [Phys. Rev. Res.](#) **7**, 033044 (2025).
- [72] T. S. Harmon and F. Jülicher, Molecular assembly lines in active droplets, [Phys. Rev. Lett.](#) **128**, 108102 (2022).
- [73] A. Murugan, J. Zou, and M. P. Brenner, Undesired usage and the robust self-assembly of heterogeneous structures, [Nat. Commun.](#) **6**, 6203 (2015).
- [74] I. H. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems* (Wiley, New York, 1975), Chap. 9, pp. 505–845.
- [75] M. Beck, A. Schmidt, J. Malmstroem, M. Claassen, A. Ori, A. Szymborska, F. Herzog, O. Rinner, J. Ellenberg, and R. Aebersold, The quantitative proteome of a human cell line, [Mol. Syst. Biol.](#) **7**, MSB201182 (2011).
- [76] D. L. J. Lafontaine, J. A. Riback, R. Bascetin, and C. P. Brangwynne, The nucleolus as a multiphase liquid condensate, [Nat. Rev. Mol. Cell Biol.](#) **22**, 165 (2021).
- [77] J. D. Lewis and D. Tollervy, Like attracts like: Getting RNA processing together in the nucleus, [Science](#) **288**, 1385 (2000).
- [78] Q.-Z. Zhu, C. X. Du, E. M. King, and M. P. Brenner, Proofreading mechanism for colloidal self-assembly, [Phys. Rev. Res.](#) **6**, L042057 (2024).
- [79] B. Wei, M. Dai, and P. Yin, Complex shapes self-assembled from single-stranded DNA tiles, [Nature \(London\)](#) **485**, 623 (2012).
- [80] C. G. Evans, J. O’Brien, E. Winfree, and A. Murugan, Pattern recognition in the nucleation kinetics of non-equilibrium self-assembly, [Nature \(London\)](#) **625**, 500 (2024).
- [81] R. D. Barish, R. Schulman, P. W. K. Rothmund, and E. Winfree, An information-bearing seed for nucleating algorithmic self-assembly, [Proc. Natl. Acad. Sci. USA](#) **106**, 6054 (2009).