Combinatorial protein–protein interactions on a polymerizing scaffold

Andrés Ortiz-Muñoz[a,1] , Héctor F. Medina-Abarca[b,1] , and Walter Fontana[b,2,3]

[a]Division of Biology and Bioengineering, California Institute of Technology, Pasadena, CA 91125; and [b]Department of Systems Biology, Harvard Medical School, Boston, MA 02115

Scaffold proteins organize cellular processes by bringing signaling molecules into interaction, sometimes by forming large signalosomes. Several of these scaffolds are known to polymerize. Their assemblies should therefore not be understood as stoichiometric aggregates, but as combinatorial ensembles. We analyze the combinatorial interaction of ligands loaded on polymeric scaffolds, in both a continuum and discrete setting, and compare it with multivalent scaffolds with fixed number of binding sites. The quantity of interest is the abundance of ligand interaction possibilities—the catalytic potential Q—in a configurational mixture. Upon increasing scaffold abundance, scaffolding systems are known to first increase opportunities for ligand interaction and then to shut them down as ligands become isolated on distinct scaffolds. The polymerizing system stands out in that the dependency of Q on protomer concentration switches from being dominated by a first order to a second order term within a range determined by the polymerization affinity. This behavior boosts Q beyond that of any multivalent scaffold system. In addition, the subsequent drop-off is considerably mitigated in that Q decreases with half the power in protomer concentration than for any multivalent scaffold. We explain this behavior in terms of how the concentration profile of the polymer-length distribution adjusts to changes in protomer concentration and affinity. The discrete case turns out to be similar, but the behavior can be exaggerated at small protomer numbers because of a maximal polymer size, analogous to finite-size effects in bond percolation on a lattice.

polymerizing scaffold | combinatorial assembly | pleiomorphic ensemble

Protein–protein interactions underlying cellular signaling systems are mediated by a variety of structural elements, such as docking regions, modular recognition domains, and scaffold or adapter proteins (1, 2). These devices facilitate the evolution and control of connectivity within and among pathways. In particular, the scaffolding function of a protein can be conditional upon activation and serve to recruit further scaffolds, thus creating opportunities for network plasticity in real time. Scaffolds are involved in the formation of signalosomes, which are transient protein complexes that process and propagate signals. A case in point is the so-called "destruction complex" that tags β-catenin for degradation in the canonical Wnt pathway. β-Catenin is modified by CK1α and GSK3β without binding any of these kinases directly but interacting with them through an Axin scaffold (3, 4). In addition, the DIX domain in Axin allows for oriented Axin polymers (5, 6), while APC (another scaffold) can bind multiple copies of Axin (7), yielding Axin–APC aggregates to which kinases and their substrates bind. By virtue of their polymeric nature, such scaffold assemblies have no defined stoichiometry and may only exist as a heterogeneous combinatorial ensemble (8, 9)—also called "pleiomorphic ensemble" (10)—rather than a single well-defined complex. Deletion of Axin's DIX domain abolishes degradation of β-catenin (5), and mutations in APC that drive familial adenomatous polyposis map to truncations reducing the number of SAMP repeats at which APC binds Axin (11). These observations suggest a possible link between the size distribution of scaffolding aggregates and disease.

Interest in intracellular phase separation phenomena has increased since the discovery of P body dissolution/condensation in *Caenorhabditis elegans* (12). Much attention has been given to the physics underlying sol–gel transitions and polymerization (13–15). Here, we focus on the combinatorial aspects of ligand interactions on a (noncovalently) polymerizing scaffold, in particular, since scaffold-mediated interactions are subject to the prozone or "hook" effect (16–18): low scaffold concentrations promote interactions between ligands, but high concentrations oppose them by isolating ligands on different scaffold molecules. The main objective of our analysis is to gain insight into how this effect plays out in the context of a polymerizing-scaffold system and whether it affords opportunities for regulation. We proceed by defining and analyzing a simple model at a level of abstraction that only encapsulates combinatorial features without explicitly taking into account spatial constraints arising from polymer conformation.

The Polymerizing-Scaffold System

Let S (the scaffold) be an agent with four distinct binding sites {a,b,x,y}. At site y, agent S can reversibly bind site x of another S with affinity σ, forming (oriented) chains. For the time being, we exclude the formation of rings. Sites a and b can reversibly bind an agent of type A (the enzyme) and of type B (the substrate) with affinities α and β, respectively. All binding interactions are independent. When the system is closed, the total concentrations

of A, B, and S are given by t_A, t_B, and t_S. This setup allows for a variety of configurations, such as shown in Fig. 1*A*. We posit that each enzyme A can act on each substrate B bound to the same complex. We refer to the number pq of potential interactions enabled by a configuration with sum formula $A_p S_n B_q$ as that configuration's "catalytic potential" Q. By extension, we will speak of the catalytic potential Q of a mixture of configurations as the sum of their catalytic potentials weighted by their concentrations.

If we assume that the assembly system equilibrates rapidly, the rate of product formation is given by $Q k_{cat}$, with k_{cat} the catalytic rate constant and Q the equilibrium abundance of potential interactions between A and B agents. Rapid equilibration is a less realistic assumption than a quasi-steady state but should nonetheless convey the essential behavior of the system. We first provide a continuum description of equilibrium Q in terms of concentrations (which do not impose a maximum polymer length) and then a discrete statistical physics treatment for the average equilibrium Q (where t_S is a natural number and imposes a maximum polymer length).

In the present context, molecular species Y_i assemble from $T = 3$ distinct building blocks ("atoms") X_j through reversible binding interactions. The Y_i have a graphical (as opposed to geometric) structure (Fig. 1). We denote the number of atoms X_j in species Y_i with $\mu_{i,j}$. The equilibrium concentration y_i of any species Y_i can be obtained by recursion over the assembly reactions as $y_i = \varepsilon_i \prod_{j=1}^{T} x_j^{\mu_{i,j}}$, where $\varepsilon_i = 1/\omega_i \prod_{r \in \mathcal{P}} K_r$ is the exponential of the free-energy content of Y_i. ω_i denotes the number of symmetries of Y_i, which, in our case, is always 1 because the polymers are oriented. The product runs over a sequence of reactions r that form an assembly path $\mathcal{P}$ of Y_i. In equilibrium, it is irrelevant which $\mathcal{P}$ one chooses. $K_r \in \{\alpha, \beta, \sigma\}$ is the equilibrium constant of the rth reaction, and the x_j are the equilibrium concentrations of free atoms of type j. Aside from the symmetry correction, ε_i is the exponential of a sum of binding energies, for example, $\varepsilon_i = \alpha^p \beta^q \sigma^r$ for a Y_i that contains p bonds between A and S, q bonds between B and S, and r bonds between S protomers.

Consider first the polymerization subsystem. From what we just laid out, the equilibrium concentration of a polymer of length l is $\sigma^{l-1} s^l$, where s is the equilibrium concentration of monomers of S. Summing over all polymer concentrations yields the total abundance of entities in the system,

$$W(s) = \sum_{l=1}^{\infty} \sigma^{l-1} s^l = s/(1 - \sigma s). \qquad [1]$$

$W(s)$ yields the conservation relation, $t_S = s\, dW(s)/ds$, from which we obtain s as:

$$s = \frac{1}{4\sigma} \left(\sqrt{4 + 1/(\sigma t_S)} - \sqrt{1/(\sigma t_S)} \right)^2. \qquad [2]$$

Using Eq. 2 in $\sigma^{l-1} s^l$ yields the dependence of the polymer-size distribution on parameters t_S and σ. $W(s)$ has a critical point at $s_{cr} = 1/\sigma$, at which the concentrations of all length classes become identical. It is clear from Eq. 2 that s can never attain that critical value for finite σ and t_S.

The Chemostatted Case

In a chemostatted system, s can be fixed at any desired value, including the critical point $1/\sigma$. At this point, ever more protomers are drawn from the S reservoir into the system to compensate for their incorporation into polymers. We next consider the system with ligands A and B held fixed at concentrations a and b. Let $A_p S_n B_q$ be the sum formula of a scaffold polymer of length n with p A agents and q B agents. There are $\binom{n}{p}\binom{n}{q}$ such configurations, each with the same catalytic potential $Q = pq$. Summing up the equilibrium abundances of all configurations yields

$$W(s, a, b) = a + b + \frac{s(1 + \alpha a)(1 + \beta b)}{1 - \sigma s(1 + \alpha a)(1 + \beta b)}. \qquad [3]$$

Eq. 3 corresponds to the $W(s)$ of ligand-free polymerization, Eq. 1, by a coarse-graining that erases the ligand-binding state of scaffolds, i.e., by dropping terms not containing s and substituting s for $s(1 + \alpha a)(1 + \beta b)$. Eq. 3 indicates that, at constant chemical potential for A, B, and S, the presence of ligands lowers the critical point of polymerization to $s_{cr} = 1/(\sigma(1 + \alpha a)(1 + \beta b))$ because, in addition to polymerization, free S is also removed through binding with A and B.

Q_{poly}, the Q of the system, is obtained by summing up the Q of each configuration weighted by its equilibrium concentration (*SI Appendix*, section 1). Using W, we compute Q_{poly} as

$$Q_{poly} = ab \frac{\partial^2}{\partial a \partial b} W = \alpha a \beta b\, s\, \frac{1 + \sigma s(1 + \alpha a)(1 + \beta b)}{(1 - \sigma s(1 + \alpha a)(1 + \beta b))^3}. \qquad [4]$$

Note that Q_{poly} inherits the critical point of W. The behavior of the chemostatted continuum model is summarized in Fig. 2. Q_{poly} (red) diverges as the polymerization system approaches the critical point. Fig. 2*A*, *Inset* shows the scaffold-length distribution at the black dot on the Q_{poly} profile. The red dotted curve reports the length distribution in the presence
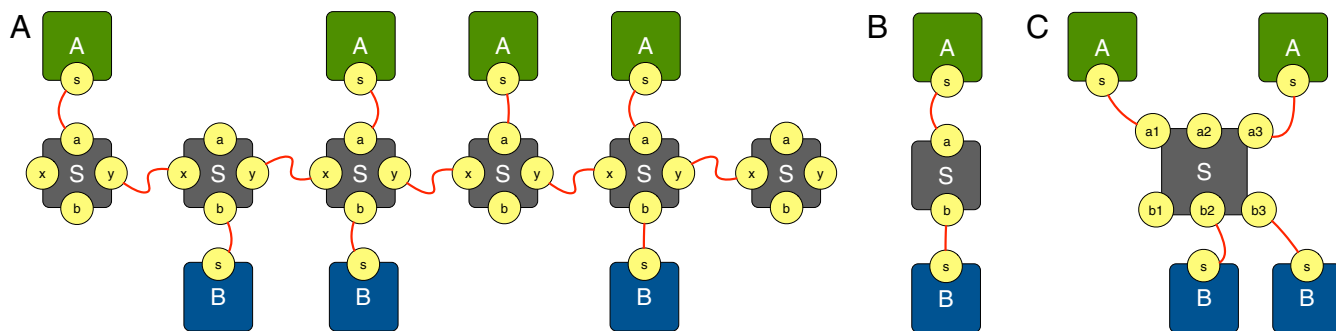
Fig. 1. Scaffold types. (*A*) Protomers S, each binding an enzyme A and a substrate B with affinities α and β, respectively, polymerize with affinity σ to yield a distribution of complexes. (*B*) A monovalent scaffold does not polymerize and has only one binding site for A and B each. (*C*) A n-valent (or multivalent) scaffold is a nonpolymerizing scaffold with n binding sites for A and B each. Here, $n = 3$. The catalytic potential Q of a configuration is the number of possible interactions between A and B agents bound to the same complex: 12 in *A*, 1 in *B*, and 4 in *C*.

Ortiz-Muñoz et al.
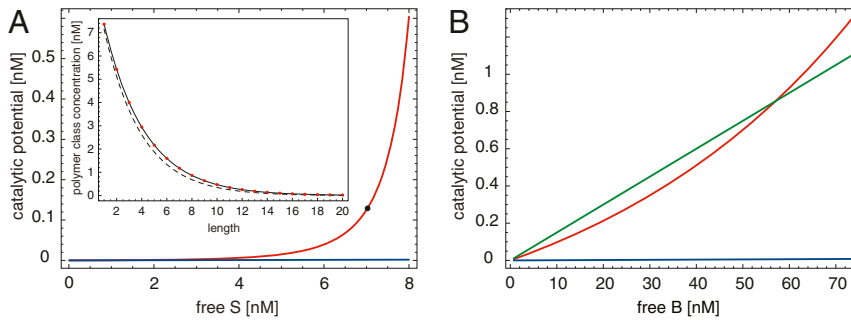
Fig. 2. Catalysis in a chemostatted polymerizing-scaffold system. (*A*) The red graph shows the catalytic potential Q as a function of chemostatted s according to Eq. **4** for $\alpha = \beta = 10^6$ M^{-1}, $\sigma = 10^8$ M^{-1}, and $a = b = 15 \cdot 10^{-9}$ M (about 9,000 molecules in a volume of 10^{-12} L). The blue curve is the special case of $\sigma = 0$, which is the monovalent scaffold system, $Q = \alpha a \beta b s$. *Inset* shows the scaffold-length distribution at $s = 7.15$ nM, corresponding to Q at the black filled circle. The critical point in this example is $s_{cr} \sim 9.7$ nM. (*B*) The catalytic potential at $s = 7.15$ nM as a function of clamped b (the substrate); other parameters are as in *A*. Red: polymerizing-scaffold system; blue: monovalent scaffold; green: chemostatted Michaelis–Menten in which A binds directly to B with affinity α.

of ligands, $[\{A_* S_k B_*\}] = \sigma^{-1}(\sigma s(1 + \alpha a)(1 + \beta b))^k$, whereas the black dotted curve reports the length distribution in the absence of ligands, $s_k \equiv [S_k] = \sigma^{k-1} s^k$. The presence of A and B shifts the distribution to longer chains. The blue curve in Fig. 2*A* shows the catalytic potential of the monovalent scaffold, $\sigma = 0$. It increases linearly with s but at an insignificant slope compared with the polymerizing case, which responds by raising the size distribution, thus drawing in more S from the reservoir to maintain a fixed s; this, in turn, draws more A and B into the system. In Fig. 2*B*, s is fixed and b, the substrate concentration, is increased. The green straight line is the Michaelis–Menten case, which consists in the direct formation of an AB complex and whose $Q = \alpha \, a \, b$ is linear in b. The red line is the polymerizing-scaffold system whose s_{cr} can be attained by just increasing b

(Eq. **4**). All else being equal, there is a b at which more substrate can be processed than through direct interaction with an enzyme. The slope of the monovalent scaffold (blue) is not noticeable on this scale.

The Continuum Case in Equilibrium

We turn to the system with fixed resources t_S, t_A, and t_B, expressed as real-valued concentrations. Eq. **4** for Q_{poly} is now evaluated at the equilibrium concentrations s, a, and b of the free atoms. These are obtained by solving the system of conservation equations, $t_S = s \, \partial W/\partial s$, $t_A = a \, \partial W/\partial a$, and $t_B = b \, \partial W/\partial b$ (solutions in *SI Appendix*, section 1). The orange curve in Fig. 3*A* depicts the saturation curve of the catalytic potential Q_{direct} of the Michaelis–Menten mechanism for a fixed concentration t_A
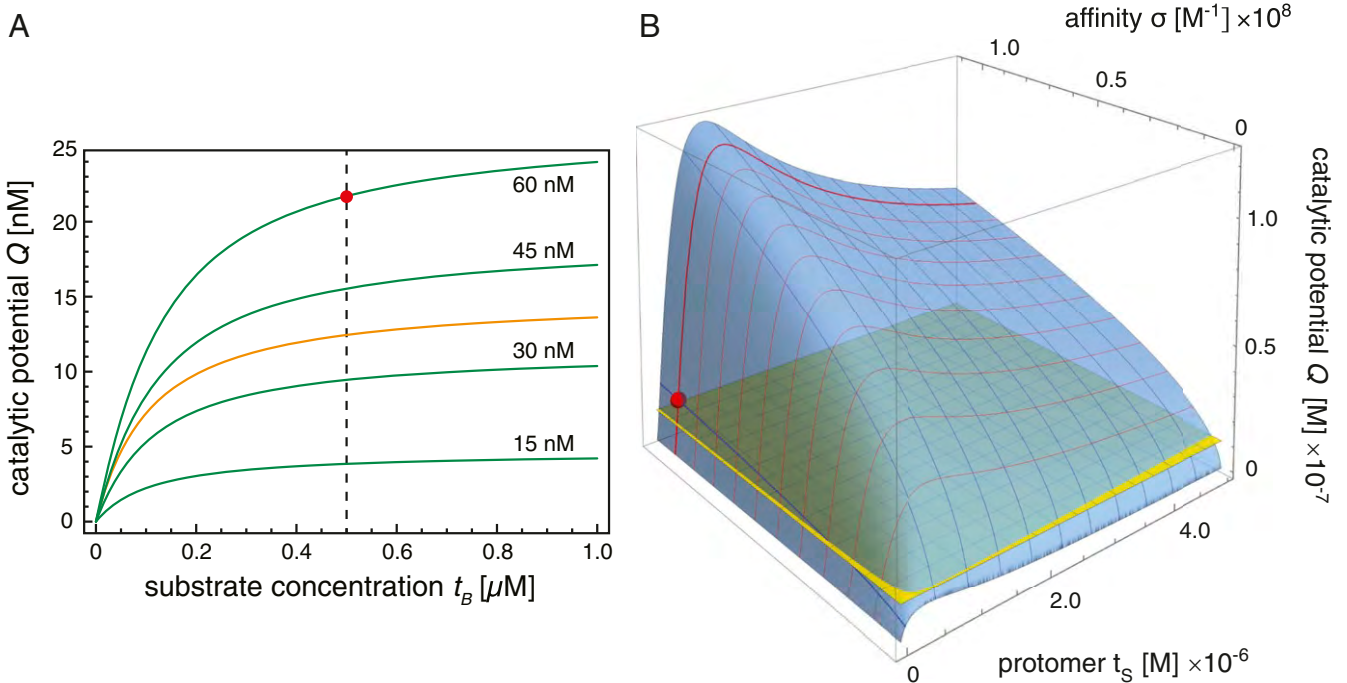
Fig. 3. Catalysis in an equilibrated polymerizing-scaffold system. (*A*) The orange curve shows the saturation of catalytic potential Q of the direct Michaelis–Menten type enzyme–substrate interaction as a function of total substrate t_B for $\beta = 10^7$ M^{-1} and $t_A = 15 \cdot 10^{-9}$ M. The green curves depict the saturation curves for Q of the polyscaffold with affinities $\alpha = \beta = 10^7$ M^{-1} and $\sigma = 10^8$ M^{-1} at various protomer abundances t_S. (*B*) The catalytic-potential surface for the polyscaffold as a function of t_S and σ; other parameters are as in *A*. The red ball corresponds to the conditions marked by the red dot in *A* ($t_B = 5 \cdot 10^{-7}$ M). The flat yellow surface is the Q for the direct enzyme–substrate interaction (i.e., the intersection of the vertical dotted line in *A* with the orange curve).

of enzyme as a function of substrate t_B. The green curves are saturation profiles of the polymerizing-scaffold system at varying protomer abundances t_S under the same condition. As in the chemostatted case, beyond some value of t_S, the catalytic potential of the polymerizing system exceeds that from direct interaction.

Q_{poly} can be modulated not only by the protomer concentration t_S but also the protomer affinity σ (Fig. 3B). Increasing t_S improves Q_{poly} dramatically at all affinities up to a maximum after which enzyme and substrate become progressively separated due to the prozone effect. At all protomer concentrations, in particular, around the maximizing one, Q_{poly} always increases with increasing affinity σ.

Comparison with Multivalent Scaffold Systems. With regard to Q, a polymer chain of length n is equivalent to a multivalent scaffold agent $S_{(n)}$ with n binding sites for A and B each. It is therefore illuminating to compare the polymerizing system with multivalent scaffolds and their mixtures.

The equilibrium concentration of configurations $A_p S_{(n)} B_q$ for an n-valent scaffold can be calculated by exploiting the independence of binding interactions (*SI Appendix*, section 2). The calculation yields as a general result that the catalytic potential for an arbitrary scaffolding system consists of two factors:

$$Q = \underbrace{p(t_{sit}, t_A, \alpha)p(t_{sit}, t_B, \beta)}_{I}\underbrace{Q_{max}(\vec{t}_S)}_{II}. \qquad [5]$$

The dimensionless function $p(t_{sit}, t_X, \gamma)$ denotes the equilibrium fraction of X-binding sites, with total concentration t_{sit}, that are occupied by ligands of type X, with total concentration t_X, interacting with affinity γ:

$$p(t_{sit}, t_X, \gamma) = \frac{\gamma t_X - \gamma t_{sit} - 1 + \sqrt{4\gamma t_X + (\gamma t_X - \gamma t_{sit} - 1)^2}}{\gamma t_X - \gamma t_{sit} + 1 + \sqrt{4\gamma t_X + (\gamma t_X - \gamma t_{sit} - 1)^2}}.$$

This expression is the well-known dimerization equilibrium, computed at the level of sites rather than scaffolds and taken relative to t_{sit} (*SI Appendix*, section 2).

Factor I depends on the total concentration of ligand-binding sites (for each type) but not on how these sites are partitioned across the agents providing them. For example, a multivalent scaffold $S_{(n)}$, present at concentration $t_{S_{(n)}}$, provides $t_{sit} = n t_{S_{(n)}}$ binding sites and the probability that a site of any particular agent is occupied is the same as the probability that a site in a pool of $n t_{S_{(n)}}$ sites is occupied. For a heterogeneous mixture of multivalent scaffold agents, we have $t_{sit} = \sum_{i=1}^{n} i\, t_{S_{(i)}}$; for a polymerizing system in which each protomer S exposes one binding site, we have $t_{sit} = t_S$.

Factor II is the maximal Q attainable in a scaffolding system. This factor depends on how sites are partitioned across scaffold agents with concentrations $\vec{t}_S = (t_{S_{(1)}}, \ldots, t_{S_{(n)}})$ but does not depend on ligand-binding equilibria. For example, a system of multivalent agents at concentrations $\vec{t}_S$ has $Q_{max} = \sum_{i=1}^{n} i^2 t_{S_{(i)}}$. The polymerizing-scaffold system is analogous, but $n = \infty$ and the $t_{S_{(i)}}$ are determined endogenously by aggregation: $t_{S_{(i)}} = s_i = \sigma^{i-1}s^i$. This yields simple expressions for the catalytic potential of a polymerizing scaffold, Q_{poly}, and multivalent scaffold, Q_{multi}:

$$Q_{poly} = p(t_S, t_A, \alpha)p(t_S, t_B, \beta)\frac{s(1+\sigma s)}{(1-\sigma s)^3} \qquad [6]$$

$$Q_{multi} = p(n\, t_{S_{(n)}}, t_A, \alpha)p(n\, t_{S_{(n)}}, t_B, \beta)n^2 t_{S_{(n)}},$$

with s in Eq. 6 given by Eq. 2. Eq. 6 is equivalent to Eq. 4. While Eq. 4 requires solving a system of mass conservation equations to obtain a, b, and s, Q_{poly}, as given by Eq. 6, does not refer to a and b but only to s, as determined by the ligand-free polymerization subsystem. The Q that shapes the Michaelis–Menten rate law under the assumption of rapid equilibration of enzyme–substrate binding has the same structure as Eq. 5: $Q_{direct} = p(t_A, t_B, \alpha)t_A$, where t_A and t_B are the total enzyme and substrate concentration, respectively. The presence of a second concurrent-binding equilibrium in Eq. 5 characterizes the prozone effect.

Adding sites, all else being equal, necessarily decreases the fraction p of sites bound. Specifically, factor I tends to zero like $1/t_{sit}^2$ for large t_{sit}. In contrast, Q_{max} increases monotonically, since adding sites necessarily increases the maximal number of interaction opportunities between A and B. For a multivalent scaffold, Q_{max} diverges linearly with t_{sit}. For the polymerizing system, Q_{max} diverges like $t_{sit}^{3/2}$ (*SI Appendix*, section 5).

Fig. 4A provides a wide-range comparison of Q_{poly} (red) with Q_{multi} for various valencies (blue) at the same site concentration $t_{sit} = t_S$. On a log–log scale, scaffolds of arbitrary valency n exhibit a Q_{multi} whose slope as a function of t_{sit} is 1, with offset proportional to n, until close to the peak. For the polymerizing scaffold, the first-order term of the series expansion of Q_{poly} is independent of the affinity σ (*SI Appendix*, section 5), whereas the second-order term is linear in σ. Hence, for small t_{sit}, the polymerizing system behaves like a monovalent scaffold, and any multivalent scaffold offers a better catalytic potential. However, as t_S increases, the equilibrium shifts markedly toward polymerization, resulting in a slope of 2, which is steeper than that of any multivalent scaffold. The steepening of Q_{poly} is a consequence of longer chains siphoning off ligands from shorter ones (*SI Appendix*, section 4). All n-valent scaffolds reach their maximal Q_{multi} at the same abundance of sites $t_{sit} = n\, t_{S_{(n)}} = t_S$ and before Q_{poly}. The superlinear growth in Q_{max} of the polymerizing system softens the decline of Q_{poly} to an order $t_S^{-1/2}$ for large t_S. In contrast, the decline of Q_{multi} is of order t_{sit}^{-1}. In sum, the polymerizing-scaffold system catches up with any multivalent scaffold, reaches peak-Q later, and declines much slower.

The mitigation of the prozone effect begs for a mechanistic explanation, in particular, since a prozone could occur not only within each length class but also between classes. To assess the within-class prozone, we think of a length class k as if it were an isolated k-valent scaffold population at concentration $t_{S_{(k)}} = s_k = \sigma^{k-1}s^k$ with $Q_{multi} = p(k\, s_k, t_A, \alpha)p(k\, s_k, t_B, \beta)k^2 s_k$. Assuming equal affinity α for both ligands A and B, Q_{multi} peaks at $\hat{t}_{S_{(k)}} = k^{-1}(\alpha^{-1} + (t_A + t_B)/2)$. However, when established through a polymerization system, $t_{S_{(k)}} = s_k \leq 1/\sigma$ for any k and any t_S (*SI Appendix*, section 2 and Fig. S1A). This means that for k up to $\sigma/\alpha + \sigma\,(t_A + t_B)/2$, the concentration s_k of polymers of length k can never exceed the concentration required for the prozone peak $\hat{t}_{S_{(k)}}$. For the parameters used in the red curve of Fig. 4B, this value of k is about 35. To put this in perspective, in Fig. 4B at the yellow marker and at peak-Q_{poly}, 98 and 68%, respectively, of all sites are organized in length classes below 10. Thus, the most populated lengths avoid the within-class prozone entirely (for example, $k = 3$ in Fig. 4C, green solid line). However, the actual behavior of the kth length class occurs in the context of all other classes, i.e., at site concentration t_S, not just $k\, s_k$. In that frame, the class does exhibit a prozone (Fig. 4C, red solid line). Hence, the overall prozone of the polymerizing-scaffold system is mainly due to the distribution and ensuing isolation of ligands across length classes, not within. This "heterogeneity prozone"
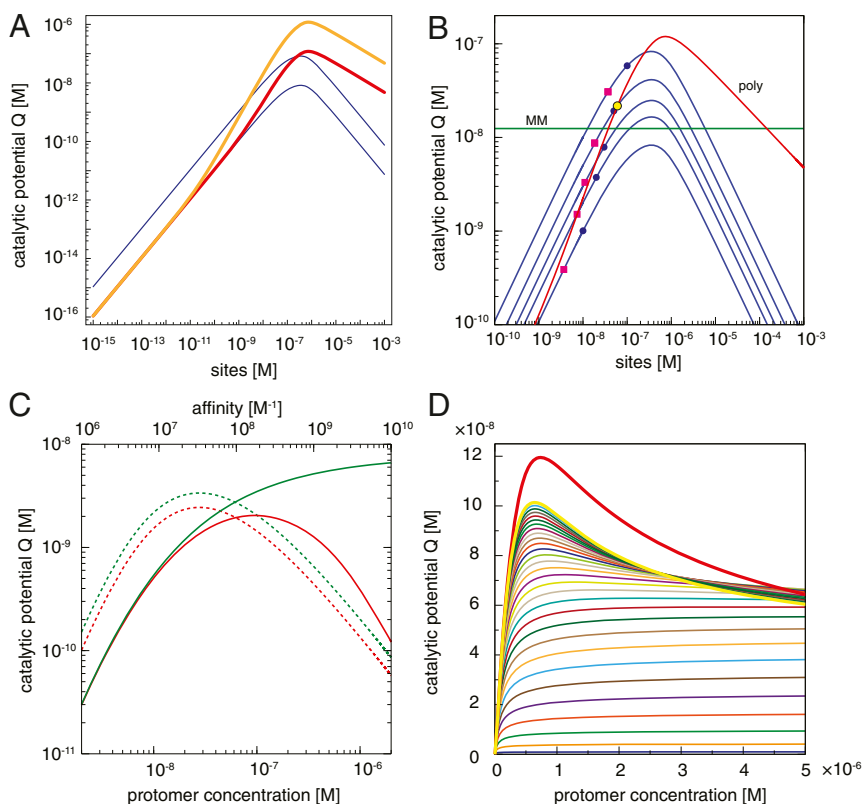
Fig. 4. Multivalent scaffolds and polymerizing scaffold. (*A*) Large-scale view of the catalytic potential Q as a function of site concentration t_S. The blue curves depict Q_{multi} for a monovalent (lower) and 10-valent scaffold. The location of the Q_{multi} peak is independent of the valency n when expressed as a function of $t_{sit} = t_S$ (*SI Appendix*, section 5, Eq. **38**). The red and orange curves depict Q_{poly} for two affinities, $\sigma = 10^8$ M^{-1} and $\sigma = 10^{10}$ M^{-1}, respectively. Other parameters: $\alpha = \beta = 10^7$ M^{-1}, $t_A = 1.5 \cdot 10^{-8}$ M, and $t_B = 5 \cdot 10^{-7}$ M. On a log–log scale, the up-slope of Q_{poly} is 1 initially—as for multivalent scaffolds—and increases to 2 prior to reaching the prozone peak. The down-slope is $-1/2$, whereas it is -1 for multivalent scaffolds (*SI Appendix*, section 5). (*B*) Close-up of the peak region in *A* for the red curve; multivalent scaffolds were added for $n = 2, 3, 5$. The slight asymmetry in the Q profiles of multivalent scaffolds stems from the differences in ligand concentrations of our running example (*SI Appendix*, section 11). The yellow dot on the Q_{poly} curve corresponds to the red dot in Fig. 3. A pink square on a blue curve of valency n marks Q_{multi} when the scaffold concentration $t_{S(n)}$ is the same as s_n at the t_S at which the length class n dominates the polymerizing system (*SI Appendix*, section 3 and Fig. S2B). The blue dots mark Q_{multi} at scaffold concentrations $t_{S(n)} = 1/\sigma$, the asymptotic (and maximal) value of s_n for all n in the limit of infinite t_S. These markers show that the prozone peak is never reached within the most populated length classes. "MM" labels the Michaelis–Menten case of Fig. 3 for comparison. (*C*) The solid lines in the graph exemplify the absence of a prozone within the isolated length class $n = 3$ and the occurrence of a prozone in the context of all other classes. Green solid: Q_{multi} for $n = 3$ using $t_{S(3)} = s_3$ and $t_{sit} = 3 t_{S(3)}$. Red solid: Q_{multi} for $n = 3$ using $t_{S(3)} = s_3$ but $t_{sit} = t_S$. The dotted lines illustrate the situation for $n = 3$ as a function of affinity σ (upper abscissa). In this dimension, the bending of the curves is not due to a prozone effect, since the number of sites does not increase. (*D*) Cumulative sums from $i = 1$ to $n = 30$ of Q_{multi} with $t_{S(i)} = s_i$ and $t_{sit} = \sum_{i=1}^{n} i \, t_{S(i)}$.

becomes noticeable only when including all length classes up to relatively high k because the majority of sites are concentrated at low k, where they are even jointly insufficient to cause a prozone (Fig. 4*D*).

At constant t_S and in the limit $\sigma \to \infty$, s_k tends toward zero for all k (*SI Appendix*, Fig. S3*C*). In the σ dimension, unlike in the t_S dimension, the class s_k itself has a peak. As σ increases, the k of the class that peaks at a given σ increases. Consequently, the Q_{multi} of each length class in isolation will show a "fake" prozone with increasing σ, due entirely to the polymerization wave passing through class k as it moves toward higher k while flattening (Fig. 4*C*, dotted lines). Since there is no site inflation, the overall Q_{poly} increases monotonically.

Effects of ligand imbalance and unequal ligand-binding affinities are discussed in the *SI Appendix*, section 11.

Interaction Horizon. The assumption that every A can interact with every B on the same scaffold can be relaxed by introducing an "interaction horizon," $q_{max}(l, h)$, defined as the number h of scaffold bonds within which a bound A can interact with

a bound B on a polymer of size l. Thus, an A can interact with at most $2h + 1$ substrate agents B: h to its "left," h to its "right," and the one bound to the same protomer. The interaction horizon only affects the Q_{max} of a polymer of length l, thus replacing the interaction factor l^2 with (*SI Appendix*, section 6):

$$q_{max}(l, h) = \begin{cases} l(2h+1) - h(h+1), & \text{for } 0 \le h \le l-1 \\ l^2, & \text{for } h \ge l \end{cases}.$$

In the most restrictive scenario, we assume a fixed horizon h, independent of l. With this assumption, Eq. **6** becomes (*SI Appendix*, section 6)

$$Q_{poly} = p(t_S, t_A, \alpha)\, p(t_S, t_B, \beta)\, \frac{s\left(1 + \sigma s - 2(\sigma s)^{h+1}\right)}{(1 - \sigma s)^3}. \quad [7]$$

In Eq. **7**, the numerator of the Q_{max} term of Eq. **6** is corrected by $-2s(\sigma s)^{h+1}$. Since $\sigma s < 1$ for all finite t_S and σ, even

moderate values of h yield only a small correction to the base case of a limitless horizon.

The Discrete Case in Equilibrium

In the discrete case, we replace concentrations with particle numbers $t_S, t_A, t_B \in \mathbb{N}$ in a specified reaction volume V. In this setting, we must convert deterministic equilibrium constants, such as σ, to corresponding "stochastic" equilibrium constants σ_s through $\sigma_s = \sigma/(\mathcal{A}V)$, where $\mathcal{A}$ is Avogadro constant. For simplicity, we overload notation and use σ for σ_s.

The basic quantity we need to calculate is the average catalytic potential $\langle Q_{\text{poly}} \rangle = \sum_{l,i,j} i\,j \langle n_{lij} \rangle$, where $\langle n_{lij} \rangle$ is the average number of occurrences of a polymer of length l with i and j ligands of type A and B, respectively. Conceptually, $\langle n_{lij} \rangle$ counts the occurrences of an assembly configuration $A_i S_l B_j$ in every possible state of the system weighted by that state's Boltzmann probability. In *SI Appendix*, section 7, we show that $\langle n_{lij} \rangle$ is given by the number of ways of building one copy of $A_i S_l B_j$ from given resources (t_S, t_A, t_B) times the ratio of two partition functions—one based on a set of resources reduced by the amounts needed to build configuration $A_i S_l B_j$, the other based on the original resources. The posited independence of all binding processes in our model implies that the partition function is the product of the partition functions of polymerization and dimerization, which are straightforward to calculate (*SI Appendix*, section 8). While exact, the expressions we derive for $\langle Q_{\text{poly}} \rangle$ (*SI Appendix*, section 8, Eq. 66) and $\langle Q_{\text{multi}} \rangle$ (*SI Appendix*, section 8, Eq. 69) are sums of combinatorial terms and therefore not particularly revealing. For numerical evaluation of these expressions, we change the size of the system by a factor ξ (typically $\xi = 0.01$), i.e., we multiply volume and particle numbers with ξ and affinities with $1/\xi$. Such resizing preserves the average behavior. Our numerical examples therefore typically deal with 10 to 1,000 particles and stochastic affinities on the order of 10^{-2} to 10^{-1} molecules.

The key aspect of the discrete case is the existence of a largest polymer consisting of all t_S protomers. We refer to it as the "maximer"; no maximer exists in the continuum case because of the infinite fungibility of concentrations (Fig. S9). Since there is only one maximer for a given t_S, its expectation is the probabil-

ity of observing it: $\langle s_{\text{max}} \rangle = t_S!\,\sigma^{t_S - 1}/Z_{t_S}^{(\text{poly})}$, where $Z_{t_S}^{(\text{poly})}$ is the partition function of polymerization (*SI Appendix*, sections 8 and 9). This probability is graphed as a function of t_S and σ in Fig. 5A. At any fixed t_S, the probability of observing the maximer will tend to 1 in the limit $\sigma \to \infty$. This puts a ceiling to Q_{max} that is absent from the continuum description. In the t_S dimension, the maximer probability decreases as t_S increases at constant σ.

Polymerization as considered here has a natural analogy to bond percolation on a one-dimensional lattice. In the case of polymerization, the probability p that any two protomers are linked by a bond is a function of t_S and σ: $p = 1 - 2/(1 + \sqrt{1 + 4\sigma t_S})$ (Fig. 5B and *SI Appendix*, section 9). The salient observation is that for small t_S, the maximer has a significant probability of already occurring at modest affinities: for 10 protomers and $\sigma = 1$, p is already 0.78 and the maximer probability a respectable 0.06. For larger t_S, the maximer loses significance unless the affinity is scaled up correspondingly (*SI Appendix*, section 10). This is also reflected in the mass distribution (Fig. 5C).

Fig. 6A compares the discrete polymerizing-scaffold system with discrete multivalent scaffolds, much like Fig. 4A for the continuum case. The behavior of the discrete case is similar to that of the continuum case—with a few nuances that are prominent at low particle numbers and high affinities, such as the topmost orange curve. Its $\langle Q_{\text{poly}} \rangle$ profile does not hug the monovalent profile (bottom green chevron curve) to then increase its slope into the prozone peak as in the continuum case (Fig. 4A). A behavior like in the continuum case is observed for the lower orange and red curves, for which σ is much weaker. In the continuum case, the affinity does not affect slope—the slope always changes from 1 to 2—but determines where that change occurs (Fig. 4A). The higher the affinity, the earlier the change. The topmost orange curve could be seen as realizing an extreme version of the continuum behavior in which an exceptionally high affinity causes a change to slope 2 at unphysically low protomer concentrations. That such a scenario can be easily realized in the discrete case is due to the significant probability with which the maximer occurs at low particle numbers, similar to finite-size percolation. It bears emphasis that, as the number t_S of protomers increases, the maximer probability decreases
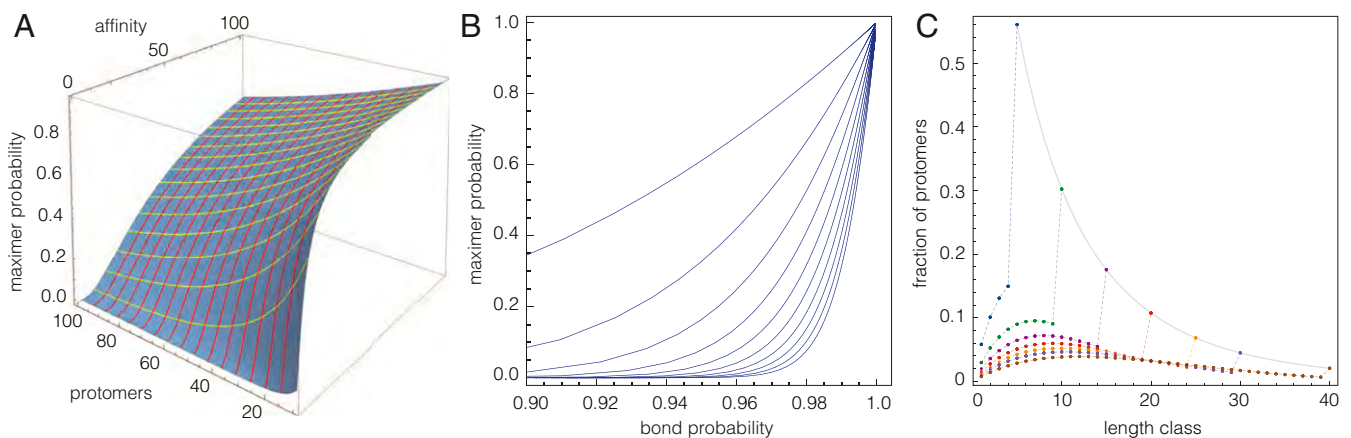
Fig. 5. Maximer. (*A*) The surface depicts the probability of observing the maximer as a function of t_S and σ. (*B*) Here, the maximer probability is graphed as function of the probability $p(t_S, \sigma)$ that a bond exists between two protomers. Each curve corresponds to a particular t_S with varying σ. t_S ranges from 10 (topmost curve) to 100 (bottom curve) in increments of 10, while σ ranges from 1 to 1,000. (*C*) Mass distributions in the polymerizing-scaffold model. Any curve depicts the fraction of protomers in all length classes n, computed as $n\,\sigma^{n-1} t_S!/(t_S - n)!\,Z_{t_S - n}^{(\text{poly})}/Z_{t_S}^{(\text{poly})}$, with $Z_{t_S}^{(\text{poly})}$ the partition function for polymerization with t_S protomers (*SI Appendix*, section 8). Each curve corresponds to a given number of protomers: $t_S = 5$ (blue), 10 (green), 15 (plum), 20 (red), 25 (orange), 30 (purple), 40 (brown); affinity $\sigma = 3$ in all cases. When t_S is small, the longest possible polymer—the "maximer"—is realized with appreciable frequency and dominates the mass distribution. As t_S increases, at fixed σ, the maximal length class increases too, but its dominance fades.
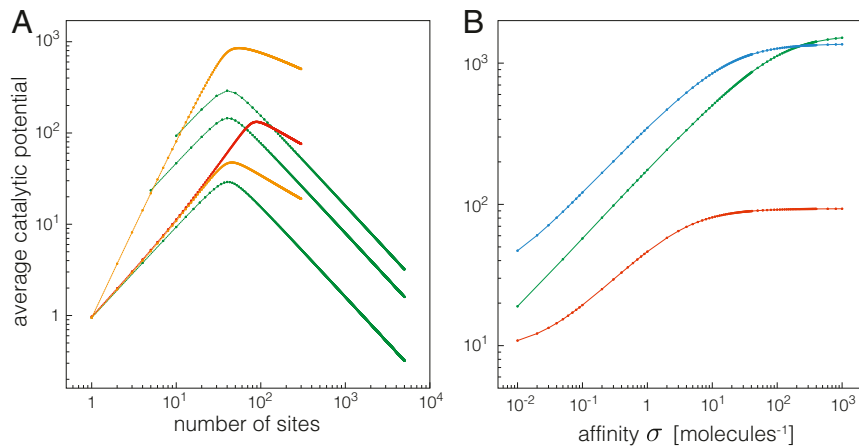
Fig. 6. Multivalent and polymerizing scaffolds in the discrete case. (*A*) Comparison of polymerizing scaffold (orange and red) with multivalent systems of various valencies (green). Orange: $t_A = t_B = 40$, $\alpha = \beta = 0.9$, $\sigma = 10$ (upper), and $\sigma = 0.01$ (lower). All affinities are in units of molecules^{-1}. Red: $t_A = t_B = 80$, $\alpha = \beta = 0.9$, and $\sigma = 0.01$. Green: $t_A = t_B = 40$, $\alpha = \beta = 0.9$, valency $n = 10$ (top), $n = 5$ (middle), and $n = 1$ (bottom). (*B*) $\langle Q_{\mathrm{poly}} \rangle$ as a function of affinity σ. $t_A = t_B = 40$, $\alpha = \beta = 0.9$, $t_S = 300$ (green), $t_S = 10$ (red), and $t_S = 50$ (blue).

(Fig. 5*C*), since the length of the maximer is t_S. However, once the maximer has receded in dominance, the increased number of length classes below it have gained occupancy and control the catalytic potential much like in the continuum case. Likewise, affinity does not appear to affect the slope of the downward leg as t_S increases.

The discrete multivalent scaffold system behaves much like its continuum counterpart. In the affinity dimension (Fig. 6*B*), the discrete system shows a behavior similar to the continuum case with the qualification that $\langle Q_{\mathrm{poly}} \rangle$ must level off to a constant, rather than increasing indefinitely. This is because, at constant t_S, an ever increasing affinity will eventually drive the system into its maximer ceiling. Because of the volume dependence of stochastic equilibrium constants, such an increase in affinity at constant protomer number can be achieved by any reduction of the effective reaction volume, for example, by confinement to a vesicle or localization to a membrane raft.

We determined SDs using stochastic simulations of the cases presented in Fig. 6*A* (*SI Appendix*, section 12). For a given $\langle Q \rangle$, the SD is larger after the prozone peak than before. Upon adding ligand-binding sites, the ratio of SD to mean (noise) increases much slower for the polymerizing system than for multivalent scaffolds.

Main Conclusions

Our theoretical analysis of a polymerizing-scaffold system shows that, at constant chemical potential, the system can be driven into criticality not only by increasing protomer concentration or affinity but by just increasing ligand concentrations. In equilibrium, polymerizing scaffolds exhibit a different type of prozone effect than multivalent ones: the polymerizing system provides a greater increase in catalytic potential on the upward leg above a certain protomer concentration t_S, delays the pro-

zone peak, and significantly mitigates the collapse at high t_S. This behavior is caused by the response of the polymer-length distribution to changes in t_S and could be exploited to identify the presence of a polymerizing scaffold in an experimental setting. When particles are present in discrete numbers, not continuous concentrations, system behavior is affected by the existence of a maximal polymer length. Behavior easily attainable at small protomer numbers requires extreme parameter values in the concentration-based description. Other than that, the discrete and the concentration-based systems behave similarly.

A polymerizing scaffold concentrates ligands locally and facilitates their interaction. In this it functions like a compartment, but through a mechanism that can be readily regulated by varying protomer concentration t_S and polymerization affinity σ. We surmise that the regulation of catalytic potential in such systems is best modulated through the affinity σ, as this would not incur a prozone effect (Fig. 3*B* and Eq. **5**) while being presumably faster and less costly than modulating t_S. Regulation of σ could occur through posttranslational modification of the scaffold protomer. Regulation through σ is most effective at a t_S close to the prozone peak, whose location with respect to t_S is robust to changes in σ (Fig. 3*B*). Our analysis suggests that conformational constraints may not qualitatively subvert these observations, since adding a bond-distance constraint to ligand interactions did not fundamentally alter the combinatorial picture. Still, taking into account polymer conformation would increase realism. A significant extension of this work would consider scaffolding units of distinct types that form multiply interconnected aggregates. Such aggregates could control a larger diversity of ligand interactions and affect the aggregation/disaggregation transition.

1. R. P. Bhattacharyya, A. Reményi, B. J. Yeh, W. A. Lim, Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680 (2006).
2. M. C. Good, J. G. Zalatan, W. A. Lim, Scaffold proteins: Hubs for controlling the flow of cellular information. *Science* **332**, 680–686 (2011).
3. C. Liu *et al.*, Control of β-catenin phosphorylation/degradation by a dual-kinase mechanism. *Cell* **108**, 837–847 (2002).
4. S. Ikeda *et al.*, Axin, a negative regulator of the wnt signaling pathway, forms a complex with GSK-3β and β-catenin and promotes GSK-3β-dependent phosphorylation of β-catenin. *EMBO J.* **17**, 1371–1384 (1998).
5. S. Kishida *et al.*, DIX domains of dvl and axin are necessary for protein interactions and their ability to regulate ?-catenin stability. *Mol. Cell. Biol.* **19**, 4414–4422 (1999).
6. M. Fiedler, C. Mendoza-Topaz, T. J. Rutherford, J. Mieszczanek, M. Bienz, Dishevelled interacts with the dix domain polymerization interface of axin to interfere with its function in down-regulating β-catenin. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1937–1942 (2011).
7. J. Behrens *et al.*, Functional interaction of an axin homolog, conductin, with β-catenin, apc, and gsk3β. *Science* **280**, 596–599 (1998).
8. E. J. Deeds, J. Krivine, J. Feret, V. Danos, W. Fontana, Combinatorial complexity and compositional drift in protein interaction networks. *PLoS One* **7**, e32032 (2012).

9. R. Suderman, E. Deeds, Machines vs. ensembles: Effective MAPK signaling through heterogeneous sets of protein complexes. *PLoS Comput. Biol.* **9**, e1003278 (2013).

10. B. J. Mayer, M. L. Blinov, L. M. Loew, Molecular machines or pleiomorphic ensembles: Signaling complexes revisited. *J. Biol.* **8**, 81 (2009).

11. J. Behrens *et al.*, Functional interaction of an axin homolog, conductin, with -catenin, APC, and GSK3. *Science* **280**, 596–599 (1998).

12. C. Brangwynne *et al.*, Germline p granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).

13. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).

14. L. Bergeron-Sandoval, N. Safaee, S. Michnick, Mechanisms and consequences of macromolecular phase separation. *Cell* **165**, 1067–1079 (2016).

15. S. Boeynaems *et al.*, Protein phase separation: A new phase in cell biology. *Trends Cell Biol.* **28**, 420–435 (2018).

16. D. Bray, S. Lay, Computer-based analysis of the binding steps in protein complex formation. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13493–13498 (1997).

17. J. E. Ferrell Jr, What do scaffold proteins really do. *Science STKE* **2000**, pe1 (2000).

18. A. Levchenko, J. Bruck, P. W. Sternberg, Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5818–5823 (2000).

Supplementary Information for

Combinatorial protein-protein interactions on a polymerizing scaffold

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

Walter Fontana.
E-mail: walter_fontana@hms.harvard.edu

This PDF file includes:
> Supplementary text
> Figs. S1 to S17
> Table S1
> References for SI reference citations

Contents

List of Figures

List of Tables

List of symbols used in the main text and this Supplementary Information

Table S1

Symbol	Meaning
A	ligand molecule of type A ("enzyme")
B	ligand molecule of type B ("substrate")
S	polymerizing scaffold protomer
$S_{(k)}$	multivalent scaffold with valency k for each A and B
$\mathcal{A}$	Avogadro's constant
V	system volume
$A_p S_n B_q$	a complex that is member of a class consisting of p A-molecules and q B-molecules bound to an S-polymer of length n
$\{A_p S_n B_q\}$	the class of complexes consisting of p A-molecules and q B-molecules bound to an S-polymer of length n
a	equilibrium concentration of free (unbound) A
b	equilibrium concentration of free (unbound) B
s	equilibrium concentration of free (unbound) S
s_k	equilibrium concentration of polymers of length k
n	for multivalent scaffold: number of binding sites per ligand type
t_A	total concentration of A
t_B	total concentration of B
t_S	total concentration of S
$t_{S_{(k)}}$	total concentration of k-valent scaffold (non-polymeric)
t_X	total concentration of molecular species X
t_{sit}	total number of binding sites (for A, or for B) in the system
α	equilibrium constant of A binding to S
β	equilibrium constant of B binding to S
σ	equilibrium constant of S binding to S (polymerization)
γ	generic affinity
$W(\cdot)$	energy-weighted generating function of molecular species
Q	catalytic potential: number of possible interactions between A-molecules and B-molecules on a scaffold species
Q_{direct}	catalytic potential of the direct interaction mechanism (Michaelis-Menten)
Q_{multi}	catalytic potential of a multivalent scaffold system
Q_{poly}	catalytic potential of a polymerizing scaffold system
Q_{max}	maximum catalytic potential of a system
$q_{max}(l, h)$	maximal catalytic potential for a polymeric scaffold of length l for ligands that can interact when separated by at most h polymer bonds
h	"catalytic horizon": max distance within which two ligands can interact
$p(t_{\text{sit}}, t_X, \gamma)$	probability that one among t_{sit} scaffold sites is occupied with a ligand of type X at concentration t_X and binding affinity γ
$\vec{n}$	a system state $(n_1, n_2, \ldots, n_C)$, where n_i is the number of particles of species Y_i
$\vec{t}$	the vector of atomic resources $(t_1, \ldots, t_T)$, where t_i is the number of atoms of type i
$\varepsilon(\vec{n})$	Boltzmann factor of state $\vec{n}$
$d(\vec{n})$	degeneracy of state $\vec{n}$
Y_i	generic molecular species
X_i	"atomic" building block in the assembly of Y_i
T	number of "atom" types available for assembly (here $T = 3$)
ω_i	number of symmetries (graph automorphisms) of the ith generic species Y_i
K_r	rate constant of the rth reaction
$\mathcal{P}$	assembly path
ε_i	free energy of formation of Y_i
$\varrho(\vec{t}, Y_i)$	the number of distinct ways of realizing a single instance of Y_i given resources $\vec{t}$
ξ	resizing factor
$\langle \cdot \rangle$	expectation value
Z	partition function

Supporting Information Text

1. W and Q in the polymerizing scaffold model

In this section we step through the treatment of the polymerizing scaffold model with more granularity.

A polymerizing scaffold protomer S has 1 binding site for each ligand A and B. Let $\{A_p S_n B_q\}$ be the set of complexes (configurations) consisting of a scaffold polymer with n protomers, p agents of type A and q agents of type B; let $[\{A_p S_n B_q\}]$ denote their aggregate equilibrium concentration. The equilibrium concentration of any particular representative $A_p S_n B_q$ of that class is given by

$$[A_p S_n B_q] = \sigma^{n-1} \alpha^p \beta^q s^n a^p b^q = \sigma^{n-1} s^n (\alpha a)^p (\beta b)^q, \qquad [1]$$

where a, b, s are the equilibrium concentrations of *free* A, B, and S, respectively; α denotes the equilibrium constant of A binding to S and, similarly, β and σ are the equilibrium constants for B binding to S and for S binding to S, respectively. All

binding interactions are posited to be mechanistically independent of one another.

In an equilibrium treatment, a system of reactions only serves to define a set of reachable complexes and could be replaced with any other mechanism, no matter how unrealistic, as long as it produces the same set of reachable configurations. Hence we could posit that a polymer of length n is generated by a reversible "reaction" in which all constituent protomers come together at once. The equilibrium constant of such an imaginary reaction must be the exponential of the energy content of a polymer of length n, which in our case is simply $(n-1)$ times the energy content of a single bond, i.e. $\ln \sigma$. Thus, the equilibrium constant of the fictitious one-step assembly reaction is σ^{n-1} and [1] follows.

To aggregate the equilibrium concentrations of all molecular configurations in the class $\{A_p S_n B_q\}$ we note that the set $\{A_p S_n B_q\}$ includes $\binom{n}{p}\binom{n}{q}$ configurations with the same energy content $\sigma^{n-1}\alpha^p\beta^q$. Summing over all p and q, yields the contribution of the polymer length class n, $\{A_* S_n B_*\}$

$$[\{A_* S_n B_*\}] = \sigma^{n-1} s^n \left[\sum_{p=1}^n \binom{n}{p}\alpha^p\, a^p\right] \left[\sum_{q=1}^n \binom{n}{q}\beta^q\, b^q\right] = \sigma^{n-1} s^n (1+\alpha\, a)^n (1+\beta\, b)^n = \frac{1}{\sigma} \left(\sigma\, s\, (1+\alpha\, a)(1+\beta\, b)\right)^n \quad [2]$$

Summing over all equilibrium concentrations defines a function W:

$$W = a + b + \frac{1}{\sigma} \sum_{n=1}^\infty \left(\sigma\, s\, (1+\alpha a)\, (1+\beta b)\right)^n = a + b + s(1+\alpha a)(1+\beta b) \sum_{n=0}^\infty \left(\sigma s(1+\alpha a)(1+\beta b)\right)^n \quad [3]$$

When viewing a, b and s as formal variables, W acts as a generating function of energy-weighted configurational counts. By differentiating W with respect to s, each s-containing term gets multiplied with the exponent of s, which is the S-content of the respective configuration. Multiplying by s then restores the exponent and recovers the equilibrium concentration of the respective configuration. Summing over all configurations so treated, yields the total amount of S protomers in the system and thus a conservation relation. This holds for all formal variables representing the "atoms", or building blocks, of the system:

$$t_A = a\frac{\partial W(a,b,s)}{\partial a}, \quad t_B = b\frac{\partial W(a,b,s)}{\partial b} \quad t_S = s\frac{\partial W(a,b,s)}{\partial s}. \quad [4]$$

By solving the equations [4], we obtain the equilibrium concentrations of free A, B, and S needed to compute the equilibrium concentration of any configuration:

$$a = \frac{\alpha t_A - \alpha t_S - 1 + \sqrt{(\alpha t_A + \alpha t_S + 1)^2 - 4\alpha t_A \alpha t_S}}{2\alpha} \quad [5]$$

$$b = \frac{\beta t_B - \beta t_S - 1 + \sqrt{(\beta t_B + \beta t_S + 1)^2 - 4\beta t_B \beta t_S}}{2\beta} \quad [6]$$

$$s = \frac{2}{\sigma^2 t_S} \frac{2\sigma t_S + 1 - \sqrt{4\sigma t_S + 1}}{\left(\alpha t_A - \alpha t_S + 1 + \sqrt{(\alpha t_A + \alpha t_S + 1)^2 - 4\alpha t_A \alpha t_S}\right)\left(\beta t_B - \beta t_S + 1 + \sqrt{(\beta t_B + \beta t_S + 1)^2 - 4\beta t_B \beta t_S}\right)} \quad [7]$$

Carrying out the geometric sum in [3] yields Eq. [3] in the main text:

$$W(a,b,s) = a + b + \frac{s(1+\alpha a)(1+\beta b)}{1 - \sigma s(1+\alpha a)(1+\beta b)}. \quad [8]$$

The same manipulation of W used to obtain [4] can be carried out twice, once for a and once for b, to yield the catalytic potential of the system:

$$Q = a\, b\, \frac{\partial^2}{\partial a \partial b} W(a,b,s), \quad [9]$$

given as Eq. [4] in the main text.

By setting $a = b = 0$, we recover the standalone polymerization system with

$$W(s) = \frac{s}{1 - \sigma s} \quad [10]$$

and s obtained from solving $t_S = dW(s)/ds$:

$$s = \frac{1}{4\sigma}\left(\sqrt{4 + \frac{1}{\sigma t_S}} - \sqrt{\frac{1}{\sigma t_S}}\right)^2, \quad [11]$$

as in Eq. [2] of the main text. We discuss the main properties of the standalone polymerization system in section 3 of this Appendix. In an equilibrium setting, the critical point of the model with ligands A and B should be the same as that of the polymerization system without ligands, namely $t_S \to \infty$ or $\sigma \to \infty$. This is not obvious from W (whose critical point Q inherits) as given in [8] with solutions [5]-[7]. However, it is made explicit in an alternative, more insightful derivation of the equilibrium catalytic potential Q given in section 2 of this Appendix.

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

2. Derivation of the general expression for the catalytic potential

In this section we derive Eq. [5] of the main text.

We consider a *multivalent* scaffold agent S with n_A binding sites for A and n_B binding sites for B. Our goal is to calculate the catalytic potential Q_{multi} of a system consisting of A-agents at concentration t_A, B-agents at concentration t_B, and S-agents at concentration t_S.

The function $W(a, b, s)$, introduced in the main text for the polymerizing scaffold system, sums up the equilibrium concentrations of all possible entities in the system. The same concept applies to a multivalent scaffold:

$$W_{\text{multi}}(a, b, s) = a + b + s(1 + \alpha a)^{n_A}(1 + \beta b)^{n_B} \tag{12}$$

with a, b, and s the equilibrium concentrations of the free A, B, and S, respectively. The catalytic potential Q_{multi} of the multivalent scaffold system is

$$Q_{\text{multi}} = a\, b\, \frac{\partial^2}{\partial a \partial b} W_{\text{multi}}(a, b, s) = s\, \alpha\, \beta\, a\, b\, n_A\, n_B\, (1 + \alpha a)^{n_A - 1}(1 + \beta b)^{n_B - 1}. \tag{13}$$

The equilibrium concentrations a, b, and s are determined by the system of conservation equations

$$a\frac{\partial}{\partial a}W = t_A, \quad b\frac{\partial}{\partial b}W = t_B, \quad s\frac{\partial}{\partial s}W = t_S. \tag{14}$$

However, we can bypass solving these equations by calculating the concentrations directly, which serendipitously gives us an intelligible expression for the catalytic potential Q in general.

We first calculate the equilibrium concentration of the fully occupied scaffold configuration, $[A_{n_A}SB_{n_B}]$ by reasoning at the level of binding *sites*. The concentration of *sites available* for binding to S are denoted by a, which is also the concentration of free A-agents. Since each A-binding site on S is independent, the equilibrium fraction of S-agents that are fully occupied with A-agents is simply

$$\frac{[\{A_{n_A}S\}]}{t_S} = \left(\frac{\alpha a}{1 + \alpha a}\right)^{n_A} \tag{15}$$

The expression in parentheses is the single-site binding equilibrium. Likewise, let $[s]$ be the concentration of free A-binding sites on S-agents and $[as]$ the concentration of bonds between A- and S-agents. In equilibrium we have that

$$\alpha a\, [s] = [as], \quad n_A t_S = [s] + [as], \quad t_A = a + [as]. \tag{16}$$

Hence, $a = [as]/(\alpha[s])$ or $a = (t_A - a)/(\alpha[s]) = (t_A - a)/(\alpha(n_A t_S - t_A + a))$, which yields a quadratic in a whose solution is

$$a = \frac{1}{2\alpha}\left(\alpha t_A - n_A \alpha t_S - 1 + \sqrt{(\alpha t_A - n_A \alpha t_S - 1)^2 + 4\alpha t_A}\right). \tag{17}$$

We plug [17] into [15] to obtain

$$\frac{[\{A_{n_A}S\}]}{t_S} = \left(\frac{\alpha t_A - n_A \alpha t_S - 1 + \sqrt{(\alpha t_A - n_A \alpha t_S - 1)^2 + 4\alpha t_A}}{\alpha t_A - n_A \alpha t_S + 1 + \sqrt{(\alpha t_A - n_A \alpha t_S - 1)^2 + 4\alpha t_A}}\right)^{n_A}. \tag{18}$$

The same reasoning holds for the (independent) binding of B to S:

$$\frac{[\{SB_{n_B}\}]}{t_S} = \left(\frac{\beta t_B - n_B \beta t_S - 1 + \sqrt{(\beta t_B - n_B \beta t_S - 1)^2 + 4\beta t_B}}{\beta t_B - n_B \beta t_S + 1 + \sqrt{(\beta t_B - n_B \beta t_S - 1)^2 + 4\beta t_B}}\right)^{n_B}. \tag{19}$$

At this point it is useful to abbreviate

$$\begin{aligned}
a_\pm &\equiv a_\pm(t_A, t_S, \alpha, n_A) = \alpha t_A - n_A \alpha t_S \pm 1 + \sqrt{(\alpha t_A - n_A \alpha t_S - 1)^2 + 4\alpha t_A} \\
b_\pm &\equiv b_\pm(t_B, t_S, \beta, n_B) = \beta t_B - n_B \beta t_S \pm 1 + \sqrt{(\beta t_B - n_B \beta t_S - 1)^2 + 4\beta t_B}
\end{aligned} \tag{20}$$

Note that these abbreviations are dimensionless functions of the parameters t_A, t_S, α and $n_{A/B}$. Because A and B bind independently, we can combine [18] and [19] to obtain:

$$[A_{n_A}SB_{n_B}] = t_S\frac{a_-^{n_A}\, b_-^{n_B}}{a_+^{n_A}\, b_+^{n_B}} = (\alpha a)^{n_A}(\beta b)^{n_B} s, \tag{21}$$

where the last equation is the equilibrium concentration in terms of free A, free B, and free S, as mentioned in the Introduction of the main text (and section 1 of this Appendix). The expression a for free A is given by [17], or $a = a_-/(2\alpha)$. The expression b for free B is analogous, $b = b_-/(2\beta)$. Equation [21] now yields s:

$$s = t_S\frac{1}{(\alpha a)^{n_A}(\beta b)^{n_B}}\frac{a_-^{n_A}\, b_-^{n_B}}{a_+^{n_A}\, b_+^{n_B}} = t_S\frac{2^{n_A}2^{n_B}}{a_+^{n_A}b_+^{n_B}} \tag{22}$$

To summarize, using abbreviations [20]:

$$a = \frac{a_-}{2\alpha}, \quad b = \frac{b_-}{2\beta}, \quad s = t_S \left(\frac{2}{a_+}\right)^{n_A} \left(\frac{2}{b_+}\right)^{n_B}. \tag{23}$$

Keep in mind that $a_{+/-}$ and $b_{+/-}$ are not constants, but functions of the system parameters. We now insert [23] into [13] to obtain

$$
\begin{aligned}
Q_{\text{multi}} &= n_A\, n_B\, s\, \left(\frac{\alpha a}{1+\alpha a}\right) \left(\frac{\beta b}{1+\beta b}\right) (1+\alpha a)^{n_A}(1+\beta b)^{n_B} \\
&= n_A\, n_B\, t_S \left(\frac{2}{a_+}\right)^{n_A} \left(\frac{2}{b_+}\right)^{n_B} \left(\frac{\alpha a}{1+\alpha a}\right) \left(\frac{\beta b}{1+\beta b}\right) (1+\alpha a)^{n_A}(1+\beta b)^{n_B} \\
&= n_A\, n_B\, t_S \left(\frac{\alpha a}{1+\alpha a}\right) \left(\frac{\beta b}{1+\beta b}\right) \left(\frac{2+2\alpha a}{a_+}\right)^{n_A} \left(\frac{2+2\beta b}{b_+}\right)^{n_B} \\
&= n_A\, n_B\, t_S \left(\frac{\alpha a}{1+\alpha a}\right) \left(\frac{\beta b}{1+\beta b}\right) \\
&= n_A\, n_B\, t_S \frac{a_-}{a_+} \frac{b_-}{b_+}. \tag{24}
\end{aligned}
$$

The cancellations are due to $2\alpha a = a_-$ (from [23]) and $a_+ = a_- + 2$ (from [20]).

Return to equation [18] and set $n_A = 1$. This gives the fraction of A-binding sites (of monovalent scaffold agents) that are occupied, that is, the probability that an A is bound:

$$p(t_S, t_A, \alpha) = \frac{a_-(t_A, t_S, \alpha, 1)}{a_+(t_A, t_S, \alpha, 1)} = \frac{\alpha t_A - \alpha t_S - 1 + \sqrt{(\alpha t_A - \alpha t_S - 1)^2 + 4\alpha t_A}}{\alpha t_A - \alpha t_S + 1 + \sqrt{(\alpha t_A - \alpha t_S - 1)^2 + 4\alpha t_A}} \tag{25}$$

In the site-oriented view it does not matter whether an A-binding site belongs to a monovalent scaffold agent or to an n-valent scaffold agent. At the same agent concentration t_S, the n-valent agent simply provides n times more sites. Thus, the probability that an A is bound if the scaffolds are n-valent is

$$p(nt_S, t_A, \alpha) = \frac{a_-(t_A, t_S, \alpha, n)}{a_+(t_A, t_S, \alpha, n)} = \frac{a_-(t_A, nt_S, \alpha, 1)}{a_+(t_A, nt_S, \alpha, 1)}, \tag{26}$$

since the number of binding sites only scales t_S in [20]. With these observations, we can rephrase [24] as the product of two terms:

$$Q_{\text{multi}} = \underbrace{p(n_A t_S, t_A, \alpha) p(n_B t_S, t_B, \beta)}_{I} \underbrace{n_A\, n_B\, t_S}_{II}. \tag{27}$$

Term (I) is the probability that a site of *some* S is occupied by A and a site of *some* S is occupied by B. Term (II) counts the maximal number of possible interactions between A and B agents in the system.
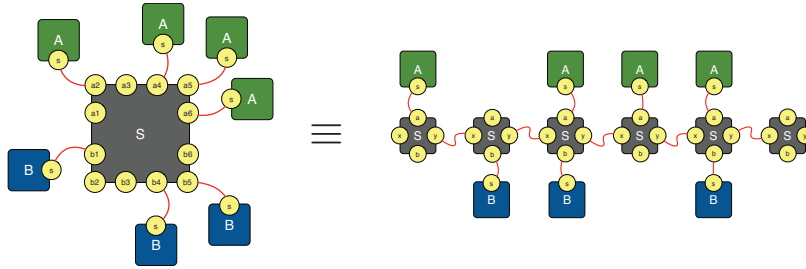
Fig. S1. A multivalent scaffold agent can be thought as representing a particular scaffold polymer configuration.

Let $S_{(i)}$ denote an agent of valency i for both ligands and let $t_{S_{(i)}}$ denote its concentration. In a mixture of multivalent scaffold types of distinct valencies $i = 1, \ldots, n$ present at concentrations $t_{S_{(i)}}$, the catalytic potentials of each type add up to that of the mixture, Q_{mix}:

$$Q_{\text{mix}} = p\left(\sum_{i=1}^{n} i\, t_{S_{(i)}}, t_A, \alpha\right) p\left(\sum_{i=1}^{n} i\, t_{S_{(i)}}, t_B, \beta\right) \sum_{i=1}^{n} i^2 t_{S_{(i)}}. \tag{28}$$

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

Generally, we can write Q_{mix} as

$$Q_{\text{mix}} = p(t_{\text{sit}}, t_A, \alpha)p(t_{\text{sit}}, t_B, \beta) \, Q_{\text{max}}(\vec{t}_S). \qquad [29]$$

In [29], t_{sit} is the total concentration of binding sites, regardless of how they are partitioned across scaffold agents, $\vec{t}_S = (t_{S(i)}, \ldots, t_{S(n)})$ is a partition of sites across scaffold molecules of different valencies, and Q_{max} is the maximal attainable number of enzyme-substrate interactions in the system, which depends on the concentration of scaffolds and their valency.

If the mixture results from a polymerization process between monovalent scaffolds $S \equiv S_{(1)}$, we identify a polymer of length l with an l-valent scaffold agent (Figure S1).

The concentrations $t_{S_{(l)}}$ are endogenously determined by polymerization at equilibrium:

$$t_{S_{(l)}} = \sigma^{l-1}s^l,$$

where the expression for s is given by the expression for the equilibrium concentration of free monomer in the polymerization system *absent ligands*, expression [11] in section 1 (Eq. [2] in the main text). Using these $t_{S_{(l)}}$ in the sum [28], which in the continuum case runs to $n = \infty$, yields the Eq. [6] for Q_{poly} in the main text:

$$Q_{\text{poly}} = p(t_S, t_A, \alpha)p(t_S, t_B, \beta) \sum_{n=1}^{\infty} n^2 \sigma^{n-1} s^n = p(t_S, t_A, \alpha)p(t_S, t_B, \beta)\frac{s(1 + \sigma s)}{(1 - \sigma s)^3}, \qquad [30]$$

with $p(\cdots)$ given by [25].

3. Overview of the polymerization system
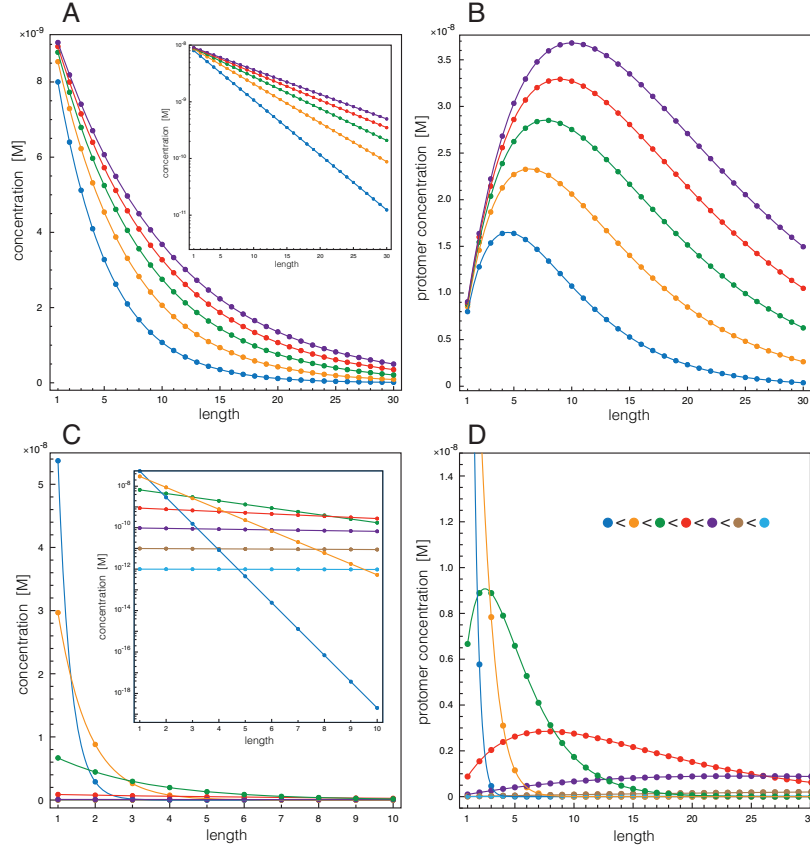
Fig. S2. The dependence of the length distribution on the protomer concentration t_S and the affinity σ. **A:** The curves depict the length distribution s_i of the linear polymerization subsystem with varying t_S at $\sigma = 10^8$ M^{-1}. Blue: $t_S = 2 \cdot 10^{-7}$ M, orange: $t_S = 4 \cdot 10^{-7}$ M, green: $t_S = 6 \cdot 10^{-7}$ M, red: $t_S = 8 \cdot 10^{-7}$ M, purple: $t_S = 1 \cdot 10^{-6}$ M. The inset plots the same curves in lin-log. **B:** The curves depict the concentrations of protomers in each length class, that is, the "mass" distribution $i \, s_i$ under the same conditions as in panel A. **C:** The curves depict the length distribution s_i with varying polymerization affinity σ at $t_S = 6 \cdot 10^{-8}$ M. Blue: $\sigma = 10^6$ M^{-1}, orange: $\sigma = 10^7$ M^{-1}, green: $\sigma = 10^8$ M^{-1}, red: $\sigma = 10^9$ M^{-1}, purple: $\sigma = 10^{10}$ M^{-1}, brown: $\sigma = 10^{11}$ M^{-1}, light blue: $\sigma = 10^{12}$ M^{-1}. **D:** As in panel B, but with varying affinity σ (as in panel C) at $t_S = 6 \cdot 10^{-8}$. For all panels $\alpha = \beta = 10^7$ M^{-1}, $t_A = 15 \cdot 10^{-9}$ M and $t_B = 5 \cdot 10^{-7}$ M.

In this section we summarize some combinatorial properties of the polymerization subsystem. Understanding the concentration profile of the polymer length distribution is useful for rationalizing the overall behavior with respect to catalytic potential, because we can view the polymerizing scaffold system as a mixture of multivalent scaffolds whose concentration is set by

polymerization. Since this is the simplest conceivable polymerization system, it would surprise us if anything being said here isn't already known in some form or another. Some of the features described can be found in Flory (1).

Let S_n be a polymer of length n and let s_n denote the equilibrium concentration of polymers in length class n. To conform with our previous notation, we shall refer to the equilibrium concentration of the monomer as $s \equiv s_1$ and to the monomer species as $S \equiv S_1$. As stated repeatedly,

$$s_n = \sigma^{n-1} s^n \quad \text{with} \quad s = \frac{1}{4\sigma} \left(\sqrt{4 + \frac{1}{\sigma t_S}} - \sqrt{\frac{1}{\sigma t_S}} \right)^2 \tag{31}$$

Figure S2 shows the dependency of s_n on the total protomer concentration t_S (panels A and B) and the affinity σ (panels C and D). Obviously, s_n is a geometric progression, thus linear in a lin-log plot for all parameter values (insets of panel A and C).

In the t_S dimension, s_n approaches $1/\sigma$ from below for each n and there is no value of t_S that maximizes s_n. In the σ dimension, s_n approaches 0 like $1/\sigma$ (in the lin-log plot, inset of panel C, the straight lines become less tilted and sink toward 0); see also expansions [36] and [37] below. However, for any given length class n, there is a σ that maximizes the concentration of that class:

$$\sigma = \frac{n^2 - 1}{4t_S}. \tag{32}$$

At that σ, the respective s_n is the most frequent, i.e. the most dominant, length class. It does not mean that s_n is at its most frequent, for s_n rises to $1/\sigma$ as $t_S \to \infty$. In the continuum description, the most frequent polymer class is always the monomer, for any t_S or σ. This is much more pronounced in the t_S dimension than the σ dimension.
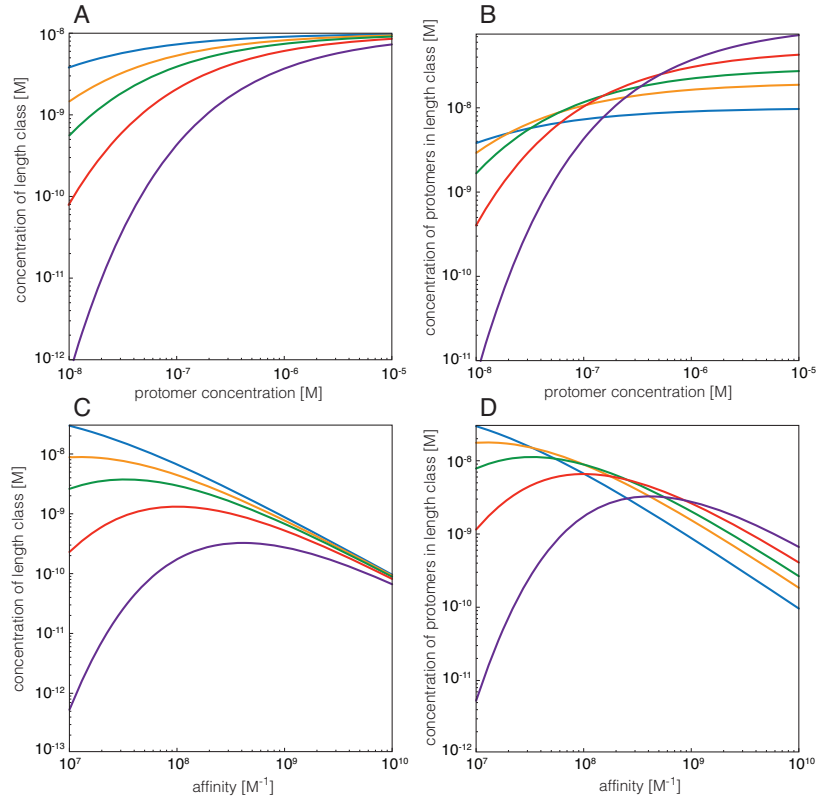
Fig. S3. Concentrations within length classes. These panels are complementary to those in Figure S2. Each curve tracks the concentration of a particular length class n as protomer concentration t_S and affinity σ are varied, effectively following the changes along a vertical cut across the curves in Figure S2. Blue: $n = 1$, orange: $n = 2$, green: $n = 3$, red: $n = 5$, purple: $n = 10$. All other parameters as in Figure S2. **A:** Concentration s_n of length class n with varying t_S. **B:** Concentration ns_n of the mass in length class n with varying t_S. Panel **C:** Concentration s_n of length class n with varying σ. Panel **D:** Concentration ns_n of the mass in length class n with varying σ.

Panels B and D of Figure S2 show the "mass" distribution, ns_n, i.e. the concentration of protomers in each length class. For all values of t_S and σ the mass exhibits a maximum at some class length. This maximum wanders towards ever larger n with increasing t_S and σ, while its value steadily increases with t_S, whereas it decreases with increasing σ. The length class n whose mass is maximized at a given t_S and σ is

$$n_{\max} = \left[\log \left(\frac{4t_S \sigma}{\left(\sqrt{1 + 4t_S \sigma} - 1 \right)^2} \right) \right]^{-1}, \tag{33}$$

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

and, for given σ and n, the t_S at which the class n becomes the most massive of all classes is given by

$$t_S = \frac{\exp(1/n)}{\sigma(1 - 2\exp(1/n) + \exp(2/n))}. \tag{34}$$

The pink squares on the blue multivalent scaffold curves in Figure 4B of the main text correspond to the catalytic potential Q that obtains at this concentration of sites. The same expression obtains for σ by swapping t_S and σ. At the t_S at which the mass in class n peaks, the concentration of the class is

$$s_{n_{\max}} = \frac{1}{e\sigma}, \tag{35}$$

independent of $n_{\max}$. Equation [33] assumes a continuous n; thus, to account for the discrete nature of polymer length, the actual $n_{\max}$ should be the nearest integer to the $n_{\max}$ given in [33]. Accordingly, the actual value of $s_{n_{\max}}$ in expression [35] will wobble slightly.

Switching perspective from the length distribution to the behavior within a length class yields Figure S3. The expansion of s_n shows how each length class approaches its limit as $t_S \to \infty$ or $\sigma \to \infty$ (multiply by n for the mass distribution):

$$\text{As } t_S \to \infty, \; s_n \to \frac{1}{\sigma} \text{ with } \frac{1}{\sigma} - \frac{n}{\sigma^{3/2}} \frac{1}{t_S^{1/2}} + O\left(\frac{1}{t_S}\right) \tag{36}$$

$$\text{As } \sigma \to \infty, \; s_n \to 0 \text{ with } \frac{1}{\sigma} - \frac{n}{t_S^{1/2}} \frac{1}{\sigma^{3/2}} + O\left(\frac{1}{\sigma^2}\right) \tag{37}$$

4. Mixtures of multivalent scaffolds

Figure S4A shows the Q_{mix}-surface [28] of a bivalent and trivalent scaffold mixture. The main observation is the asymmetry in the effect on Q upon adding $S_{(3)}$ to a fixed amount of $S_{(2)}$ compared to the other way around—blue versus red mesh lines in Figure S4. Upon adding $S_{(3)}$, the ligands A and B re-equilibrate over the available binding sites. Over a range of $[S_{(2)}]$, this equilibration is more likely to result in A and B agents ending up on the same $S_{(3)}$ scaffold than on the same $S_{(2)}$ scaffold. This is most pronounced at small $[S_{(2)}]$ and disappears gradually as the addition of binding sites drives the system past the prozone peak due to the p^2 term in [28]. The orange curve shows the Q-profile of a mixture in which $S_{(3)}$ and $S_{(2)}$ are increased in equal amounts. The dotted curves are the projections of the mixture curve on each component axis for the purpose of comparison with the Q-curves of each component in isolation. This behavior is more dramatic in binary mixtures of multivalent scaffolds with large valency differences (Figure S4B).
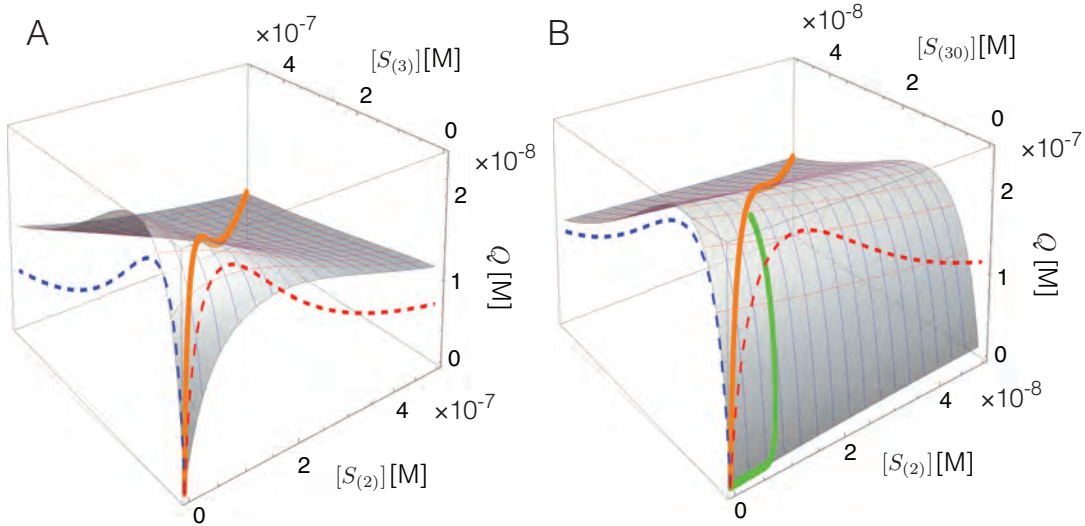
Fig. S4. Mixtures of multivalent scaffolds. **A:** The graphics renders the Q_{multi}-surface of a mixture of a bivalent and trivalent scaffold. The orange line is the Q-profile when both agents are added in equal amounts to the mix. The dotted lines are projections of the orange line for comparison with the homogeneous scaffold systems. **B:** Same as in panel **A** but for a mixture of $S_{(2)}$ and $S_{(30)}$; only the portion of the surface at low scaffold concentrations is shown. The green curve shows the Q-trajectory for the binary mixture that would obtain when $[S_{(2)}]$ and $[S_{(30)}]$ are set by the polymerizing scaffold system with increasing t_S. The green curve is the *whole* trajectory, because both $[S_{(2)}]$ and $[S_{(30)}]$ converge to $1/\sigma = 10^{-8}$ M (Figure S3). Other parameters: $\alpha = \beta = 10^7$ M^{-1}, $t_A = 15 \cdot 10^{-9}$ M, $t_B = 5 \cdot 10^{-7}$ M.

In a polymerizing scaffold system, the concentrations $s_i \equiv [S_{(i)}]$ and $s_j \equiv [S_{(j)}]$ do not increase in equal amounts when t_S is increased, but are related by a factor $(\sigma s)^{i-j}$. Since $\sigma s < 1$ for $t_S < \infty$, there is a lag between the rise of $S_{(i)}$ and $S_{(j)}$, where $S_{(i)}$ increases before $S_{(j)}$ for $i < j$; this lag is more dramatic the bigger the difference $|i - j|$ (Figure S4B, green curve). In the polymerizing system, as t_S increases, the ratio of $S_{(i)}$ and $S_{(j)}$ will tend to 1, but by then the between-class prozone is taking its toll. In sum, the "stealing" of ligands by higher length classes from lower ones is the reason for the turn towards a steeper

slope of Q_{poly} at t_S values at which polymerization becomes effective (Figure 4A in the main text). Incidentally, the shift of ligands from lower towards higher valency classes also tends to flatten the intrinsic slope of the downward leg of lower valency classes after the prozone peak, contributing further to prozone mitigation in the overall system.

5. Comparison between polymerizing and multivalent scaffold systems

In the main text, Figure 4A and 4B, we compare multivalent scaffolds with the polymerizing scaffold system. Figure S5 places that comparison in the context of the full Q_{poly} surface to show the effectiveness of regulating the affinity σ.

While even for $n_A = n_B = n$ and $\alpha = \beta$, Q_{multi} is a cumbersome expression, determining the concentration of scaffold agents t_S for which $dQ_{\text{multi}}/dt_S = 0$ yields a simple solution

$$t_S = \frac{1}{n}\left(\frac{1}{\alpha} + \frac{t_A + t_B}{2}\right). \tag{38}$$

Equation [38] shows that when plotting Q_{multi} against the concentration of sites $t_{\text{sit}} = nt_S$, as in Figure S5 and Figure 4A of the main text, the prozone peaks line up for all valencies n.
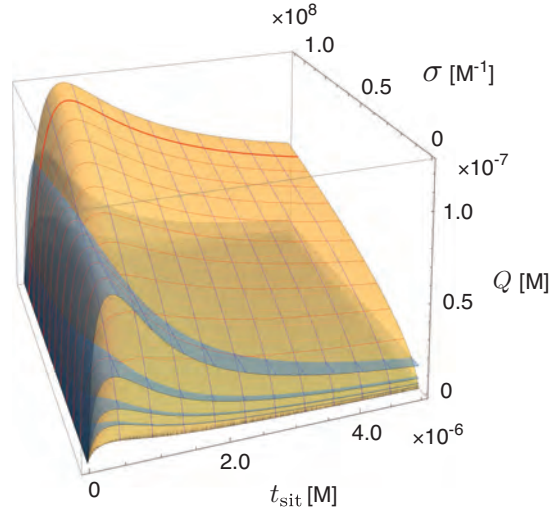
Fig. S5. Polymerizing scaffold and multivalent scaffolds. The surface shows Q_{poly} as function of t_S and σ, giving more context to Figure 4B in the main text. The emphasized mesh line (red) at $\sigma = 10^8$ corresponds to the Q-function of the polymerizing scaffold system shown in Figure 4B of the main text. $\alpha = \beta = 10^7$ M^{-1}, $t_A = 15 \cdot 10^{-9}$ M, $t_B = 5 \cdot 10^{-7}$ M.

Expanding Q_{multi} (assuming $n_A = n_B = n$) in t_S near zero, yields

$$Q_{\text{multi}} = \frac{\alpha t_A \beta t_B}{1 + \alpha t_A + \beta t_B + \alpha\beta t_A t_B} n^2 t_S + O(t_S^2). \tag{39}$$

Hence in a log-log plot, the up-leg of $Q_{\text{multi}}(n)$ has, to leading order, slope 1 and offset n when plotted against sites $t_{\text{sit}} = nt_S$ as in Figure 4A of the main text. Similarly, expanding Q_{multi} in t_S near infinity, yields

$$Q_{\text{multi}} = t_A t_B \frac{1}{t_S} + O(1/t_S^2), \tag{40}$$

and hence, to leading order, a slope of -1 in a log-log plot in the down-leg after the prozone peak and an offset of n when plotted against t_{sit} as in Figure 4A of the main text.

The expansion of Q_{poly} in t_S ($= t_{\text{sit}}$) around zero yields

$$Q_{\text{poly}} = \frac{\alpha t_A \beta t_B}{1 + \alpha t_A + \beta t_B + \alpha\beta t_A t_B} t_S + \left[f(\alpha, \beta, t_A, t_B) + g(\alpha, \beta, t_A, t_B)\sigma\right] t_S^2 + O(t_S^3) \tag{41}$$

with $f()$ and $g()$ functions of the indicated parameters. The leading-order term is the same as the Q_{multi} of the monovalent scaffold, *and is independent of* σ, which enters the second-order term. Accordingly, for small t_S, Q_{poly} hugs the Q of the monovalent scaffold as if there was no polymerization; as t_S increases, σ (i.e. polymerization) becomes effective and Q_{poly} doubles its slope upward. This is clearly seen in Figure 4A of the main text. Some microscopic consequences from building up a length distribution as t_S increases are discussed in section 4.

Expanding Q_{poly} in t_S at infinity yields

$$Q_{\text{poly}} = 2t_A t_B \sqrt{\sigma}\sqrt{\frac{1}{t_S}} + O(1/t_S^{3/2}), \tag{42}$$

where the $p(t_S, t_A, \alpha)p(t_S, t_B, \beta)$ component scales with $t_A t_B/t_S^2$ and the Q_{max} component with $2t_S^{3/2}\sqrt{\sigma}$ to leading order. As a result, the slope of the down-leg of Q_{poly} after the prozone peak in a log-log plot is $-1/2$.

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

6. Interaction horizon

Structural constraints might prevent every catalyst A on a polymeric scaffold from interacting with all substrates B bound to the same polymer. To obtain a rough sense of how such constraints could impact the catalytic potential Q, we define an "interaction horizon", h, Figure S6. The horizon h is the farthest distance in terms of scaffold bonds that a bound A can "reach". This means that a given bound enzyme A can interact with at most $2h + 1$ substrate agents B: h to its "left", h to its "right" and the one bound to the same protomer, Figure S6A. For example, in Figure S6B, the 2-horizon of the A at position 1 includes the Bs at positions 2 and 3, but not at position 5. Likewise, the B at position 2 is outside the 2-horizon of the A at position 5, whereas all Bs are within reach of the A at position 3. Clearly, the interaction horizon only modulates the Q_{max} in equation [29] of a polymer of length n; more precisely, it modulates the interaction factor—the n^2 in the first equation of [30]. We now write this factor as $q_{max}(n, h)$; it replaces the n^2 in [30].
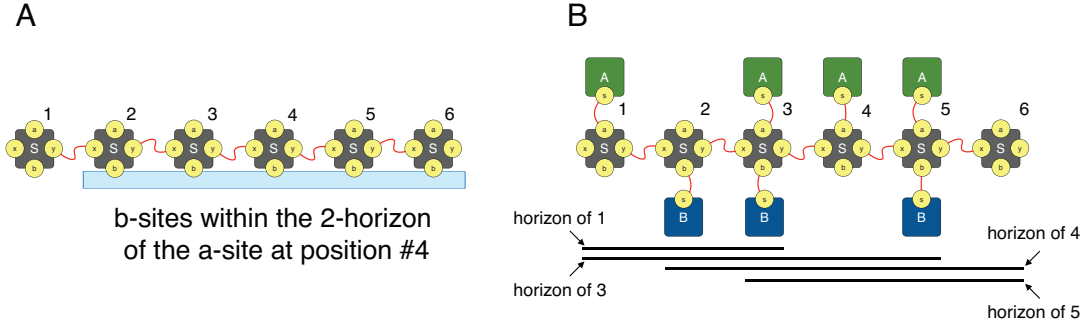
Fig. S6. Interaction horizon. The schematic illustrates the case in which the horizon h is less than the polymer length n. In this case, each A-binding position can interact with at most h B-binding positions on its "left" or "right" side. When $h \geq n$, every A-position can interact with every B-position.

To reason about the catalytic combinations, we first consider the case $0 \leq h \leq \lfloor n/2 \rfloor$:

$$q_{max}(n, h) = \underbrace{(n - 2h)(2h + 1)}_{\text{I}} + \underbrace{2h(h + 1)}_{\text{II}} + \underbrace{2 \sum_{k=1}^{h-1} (h - k)}_{\text{III}} = \quad n(2h + 1) - h(h + 1) \qquad [43]$$

Term I refers to the $n - 2h$ positions in the middle region of the chain that can interact with the full complement of $2h + 1$ sites within its horizon. Term II refers to the h positions at each end of the chain and accounts for all $h + 1$ sites reachable towards the interior of the chain. Term III accounts for the remaining $h - k$ locations towards the end of the chain that can be reached from a position considered in term II; these locations depend on that position's distance k from the end of the chain. For $\lfloor n/2 \rfloor < h \leq n - 1$ we obtain

$$q_{max}(n, h) = \underbrace{(2h - n)n}_{\text{I'}} + \underbrace{2(n - h)(h + 1)}_{\text{II'}} + \underbrace{2 \sum_{k=1}^{n-h} (k - 1)}_{\text{III'}} = \quad n(2h + 1) - h(h + 1) \qquad [44]$$

In analogy to [43], Term I' refers to the $2h - n$ positions that can access the whole chain; term II' accounts for the $h + 1$ locations spanned by the inward-facing side of the remaining $n - h$ positions at each end of the chain. Finally, term III' accounts for the locations covered by the outward facing side of these $n - h$ positions.

If the horizon h is larger than the polymer length n, then every A-position can interact with every B-position on the polymeric scaffold and $q_{max}(n, h) = n^2$. Merging this with [43] and [44] yields

$$q_{max}(n, h) = \begin{cases} n(2h + 1) - h(h + 1), & \text{for } 0 \leq h \leq n - 1 \\ n^2, & \text{for } h \geq n \end{cases} \qquad [45]$$

which appears in the main text. The corner cases are covered correctly: $q_{max}(n, 0) = n$ and $q_{max}(n, n - 1) = n^2$. (Note that $h = n$ yields the same result as $h = n - 1$, which is useful below.)

We use [45] to calculate two scenarios. In scenario 1, h is a simple linear function of the length n: $h = \xi n$ with $0 \leq \xi \leq 1$. In other words, every A can monitor the same fraction ξ of B-binding sites on a polymer of any size. This seems rather unrealistic (and makes h a continuous variable, although that appears to work just fine). However, scenario 1 may serve as a comparison with the subsequent, more realistic scenario 2.

When $h = \xi n$, h is always less or equal than n and the first case of [45] applies. Using $q_{max}(n, h)$ with $h = \xi n$ instead of n^2 in the first equation of [30] yields

$$Q_{max}(\xi) = \sum_{n=1}^{\infty} [n(2h+1) - h(h+1)] \sigma^{n-1} s^n = \sum_{n=1}^{\infty} [n(2\xi n + 1) - \xi n(\xi n + 1)] \sigma^{n-1} s^n$$

$$= \frac{1}{\sigma} \left[\xi(2-\xi) \sum_{n=1}^{\infty} n^2 \sigma^n s^n + (1-\xi) \sum_{n=1}^{\infty} n\sigma^n s^n \right] = \xi(2-\xi) \frac{s(1+\sigma s)}{(1-\sigma s)^3} + (1-\xi) \frac{s}{(1-\sigma s)^2}, \quad [46]$$

which leads to

$$Q = p(t_S, t_A, \alpha) p(t_S, t_B, \beta) \left(\xi(2-\xi) \frac{s(1+\sigma s)}{(1-\sigma s)^3} + (1-\xi) \frac{s}{(1-\sigma s)^2} \right) \quad [47]$$

For $\xi = 1$, the expression [47] becomes [30], as a horizon that equals the length of any polymer does not affect Q_{max}. For $\xi = 0$ we get

$$Q = p(t_S, t_A, \alpha) p(t_S, t_B, \beta) \frac{s}{(1-\sigma s)^2} = p(t_S, t_A, \alpha) p(t_S, t_B, \beta) t_S, \quad [48]$$

because of $t_S = s\, dW/ds$ for the polymer-only system. Thus, for $\xi = 0$, we recover the Q of the simple monovalent scaffold, since in this case the organization of protomers into polymers doesn't affect catalytic potential. Scenario 1 is shown in Figure S7, panels A and B.
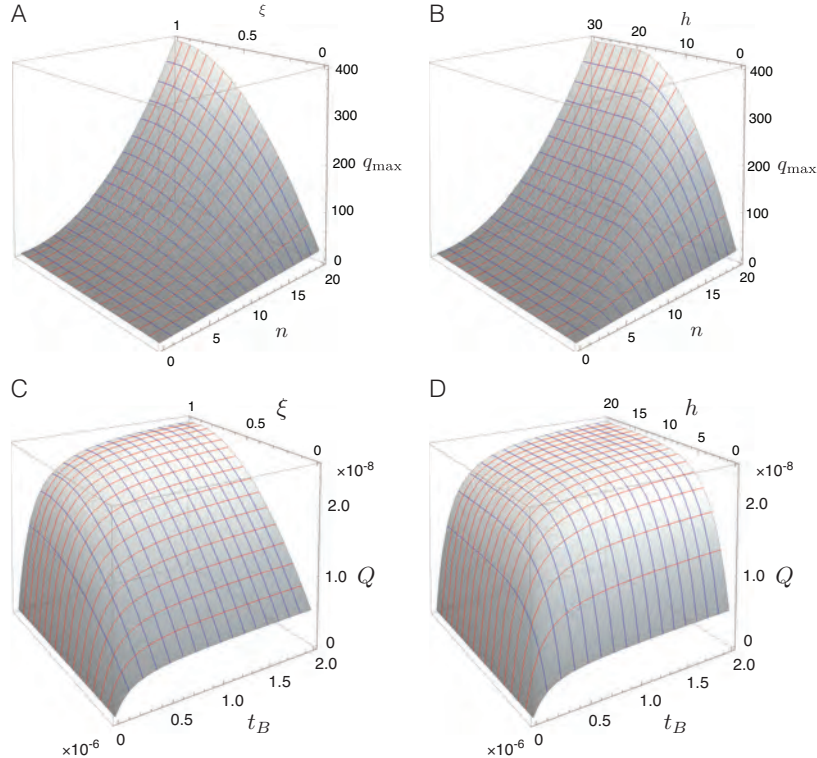
Fig. S7. Interaction horizon scenarios. **A:** $q_{max}(n, h)$, equation [45], for scenario 1 when $h = \xi n$ ($0 \le \xi \le 1$). **B:** $q_{max}(n, h)$, equation [45], for scenario 2 when h is a constant independent of n. The difference to panel **A** is that the surface of scenario 2, once h exceeds n, is a quadratic extension of the surface of scenario 1 in panel **A** at $\xi = 1$. **C:** The Q-surface [47] for scenario 1 as a function of substrate concentration t_B. **D:** The Q-surface [50] for scenario 2 as a function of substrate concentration t_B. In Figure S8, this surface is compared against the Michaelis-Menten case. The parameter values in **C** and **D** are: $\alpha = \beta = 10^7$ M and $\sigma = 10^8$ M, $t_A = 15 \cdot 10^{-9}$ M, and $t_S = 60 \cdot 10^{-9}$ M.

In scenario 2, $h = \text{const}$ for all lengths n, which means a "hard" horizon independent of polymer size. This scenario is more

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

realistic. $Q_{max}(h)$ becomes

$$Q_{max}(h) = \sum_{n=1}^{\infty} q_{max}(n,h)\sigma^{n-1}s^n = \sum_{n=1}^{h} n^2\sigma^{n-1}s^n + \sum_{n=h+1}^{\infty} [n(2h+1) - h(h+1)]\sigma^{n-1}s^n$$

$$= \frac{1}{\sigma}\left\{\sum_{n=1}^{h} n^2(\sigma s)^n + (2h+1)\sum_{n=h+1}^{\infty} n(\sigma s)^n - h(h+1)\sum_{n=h+1}^{\infty}(\sigma s)^n\right\}$$

$$= \frac{1}{\sigma}\left\{\frac{\sigma s(1+\sigma s) - (\sigma s)^{h+1}[(h+1)^2 - (2h^2+2h-1)\sigma s + h^2(\sigma s)^2]}{(1-\sigma s)^3}\right.$$

$$\left. + (2h+1)\frac{(\sigma s)^{h+1}(h+1-h\sigma s)}{(1-\sigma s)^2} - h(h+1)\frac{(\sigma s)^{h+1}}{1-\sigma s}\right\}$$

$$= \frac{s\left(1+\sigma s - 2(\sigma s)^{h+1}\right)}{(1-\sigma s)^3}, \tag{49}$$

yielding

$$Q = p(t_S, t_A, \alpha)p(t_S, t_B, \beta)\frac{s\left(1+\sigma s - 2(\sigma s)^{h+1}\right)}{(1-\sigma s)^3}, \tag{50}$$

which is Eq. [7] of the main text. Expression [50] becomes [48] for $h = 0$, as we would expect. As h increases, [50] quickly converges to the infinite horizon case [30], since $\sigma s < 1$ raised to the power of h becomes negligible. Scenario 2 is shown in Figure S7, panels B and D. As suggested in Figure S8, even restrictive structural constraints (small h) make only a relatively modest dent in the catalytic potential of the polymerizing scaffold when compared to that of the plain Michaelis-Menten scenario.
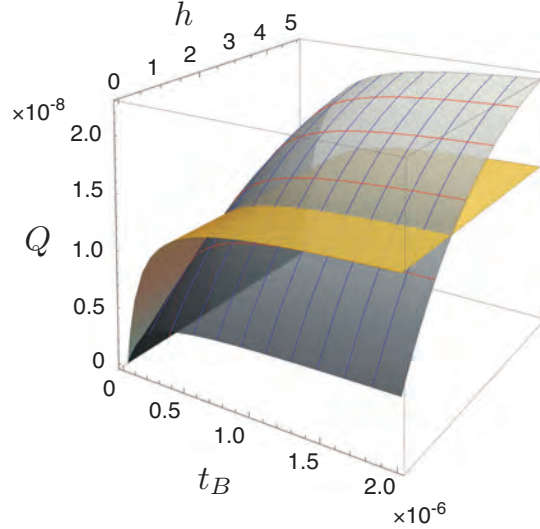
Fig. S8. The impact of the interaction horizon. The Q-surface [50] with hard horizon h, gray, and the plain dimerization (Michaelis-Menten) surface, orange, for the parameter settings corresponding to Figure 3A in the main text ($\alpha = \beta = 10^7$ M^{-1} and $\sigma = 10^8$ M^{-1}, $t_A = 15 \cdot 10^{-9}$ M, $t_S = 60 \cdot 10^{-9}$ M). At $t_S = 60$ nM (the curve with the red dot in Figure 3A of the main text) a horizon $h = 2$ is already sufficient to achieve a higher catalytic potential than the direct binding of enzyme to substrate. This suggests that structural constraints forcing a small interaction horizon might not undermine the efficacy of a polymerizing scaffold.

7. The discrete case

While we strive for a reasonably self-contained exposition, some details are only asserted for brevity and are developed in a forthcoming manuscript providing a more general treatment of equilibrium assembly.

In the following, we use the same symbols for the binding affinities α, β, and σ as in the continuum case, but they must now be understood as "stochastic affinities". Specifically, if γ' is a binding affinity in the continuum case, the stochastic affinity γ (in units of molecules^{-1}) is related as $\gamma = \gamma'/(\mathcal{A}V)$, where V is the effective volume hosting the system and $\mathcal{A}$ is Avogadro's constant. Thus a polymerization affinity of 3 molecules^{-1} in the discrete case corresponds to about $1.8 \cdot 10^{12}$ M^{-1} in a cell volume of 10^{-12} L in the continuum setting.

A. Average catalytic potential. Our objective is to calculate the average catalytic potential $\langle Q \rangle$ of a scaffold mixture, defined as

$$\langle Q \rangle = \sum_{i=0}^{\min(t_A, n)} \sum_{j=0}^{\min(t_B, n)} i\, j\, \langle S_{ij} \rangle, \qquad [51]$$

where S_{ij} is any scaffold (polymer or multivalent) with n A-binding sites, of which i are occupied, and n B-binding sites, of which j are occupied. More precisely, S_{ij} is the set of all configurations, or molecular species, with i and j agents of type A and B bound, respectively. $\langle S_{ij} \rangle$ is the average or expected total number of such configurations in an equilibrium system with resource vector $\vec{t} = (t_A, t_B, t_S)' \in \mathbb{N}_0^3$. The $'$ means a transpose. (t_S is typically the number of scaffolds of a given valency n or the number of protomers in a polymerizing system. When considering mixtures of scaffolds of different valencies i, t_S is generalized accordingly.)

This raises the need to compute $\langle S_{ij} \rangle$, which requires a little detour. We start by defining a few well-known quantities.

Assume a system of molecular interactions with a set of atomic building blocks, or atoms for short, $\{X_1, \ldots, X_T\}$ (in the main text typically $T = 3$, namely A, B, and S) that give rise to a set of configurations $\{Y_1, \ldots, Y_C\}$. Since we are interested in equilibrium, the precise nature of the interactions is irrelevant as long as the resulting systems have the same set of reachable molecular species. The assembly scenarios considered in the main text only require binding and unbinding interactions.

B. Boltzmann factor of a molecular species. Each molecular species Y_i has a Boltzmann factor given by

$$\varepsilon_i = \prod_r \gamma_r, \qquad [52]$$

where $\gamma_r = \exp(-\frac{\Delta G_r^0}{kT})$ is the binding constant of the r-th reaction and the product runs over a series of reactions r that constitute an assembly path from atomic components (A, B, and S). Note that, in the discrete case, ε_i is not divided by the number of symmetries ω_i as in the continuum case (main text leading up to Eq. [1]). The effect of symmetries is accounted for in the state degeneracy, Eq. [54] below, which considers all instances of Y_i in a given state. As a consequence, $-kT \log \varepsilon_i$ is not the free energy of formation, but just the internal energy due to bond formation.

C. Boltzmann factor of a state. By extension, the Boltzmann factor of a system *state* $\vec{n} = (n_1, n_2, \ldots, n_C)'$, where n_i is the number of particles of species Y_i, is given by

$$\varepsilon(\vec{n}) = \prod_{i=1}^{C} (\varepsilon_i)^{n_i}. \qquad [53]$$

More precisely, [53] is the Boltzmann factor associated with a particular *realization* of the state $\vec{n}$, as all atoms are labelled (distinguishable).

D. Degeneracy of a state. A state $\vec{n}$ is the specification of a multiset of species in which atom labels are ignored. The degeneracy $d(\vec{t}, \vec{n})$ of a state $\vec{n}$ with resource vector $\vec{t} = (t_1, \ldots, t_T)$ is the number of distinct ways of realizing it by taking into account atom labels. Let $\mu_{i,j}$ denote the number of atoms of type X_j contained in one instance of Y_i. For a given resource vector $\vec{t}$ the set $\Sigma(\vec{t})$ of states $\vec{n}$ that are compatible with it satisfy $t_j = \sum_{i=1}^{C} \mu_{i,j} n_i$ for every atom type X_j. Hence, the degeneracy of a state $\vec{n} \in \Sigma(\vec{t})$ is given by

$$d(\vec{t}, \vec{n}) = \frac{\displaystyle\prod_{i=1}^{T} t_i!}{\displaystyle\prod_{i=1}^{C} n_i! \prod_{i=1}^{C} (\omega_i)^{n_i}}. \qquad [54]$$

The numerator counts all permutations of the atoms that constitute the system, the first product in the denominator corrects for all orderings among the n_i copies of species Y_i and the second product corrects for all symmetries associated with Y_i.

E. The partition function for a given resource vector. As usual,

$$Z(\vec{t}) = \sum_{\vec{n} \in \Sigma(\vec{t})} d(\vec{t}, \vec{n}) \varepsilon(\vec{n}), \qquad [55]$$

where the sum runs over all admissible states given resource vector $\vec{t}$. The equilibrium probability of a state $\vec{n}$ is given by

$$p(\vec{t}, \vec{n}) = \frac{d(\vec{t}, \vec{n}) \varepsilon(\vec{n})}{Z(\vec{t})}. \qquad [56]$$

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

F. The average number of instances of a specific configuration in equilibrium. For a given resource vector $\vec{t}$ a species Y_i occurs in various numbers n_i across the states $\vec{n}$ in the admissible set $\Sigma(\vec{t})$. The average abundance of Y_i, $\langle n_i \rangle$ then is

$$\langle n_i \rangle = \sum_{\vec{n} \in \Sigma(\vec{t})} n_i p(\vec{t}, \vec{n}) = \frac{1}{Z(\vec{t})} \sum_{\vec{n} \in \Sigma(\vec{t})} n_i d(\vec{t}, \vec{n}) \varepsilon(\vec{n}). \tag{57}$$

The workhorse for the discrete treatment of the scaffolding systems discussed in the main text is the following Theorem.

Theorem:
The average equilibrium abundance $\langle n_i \rangle$ of species Y_i in an assembly system with resource vector $\vec{t}$ is given by

$$\langle n_i \rangle = \varrho(\vec{t}, Y_i) \varepsilon_i \frac{Z(\vec{t} - \vec{\mu_i})}{Z(\vec{t})}, \tag{58}$$

where $\vec{\mu_i} = (\mu_{i,1}, \ldots, \mu_{i,T})'$ is the atomic content vector of species Y_i; $\varrho(\vec{t}, Y_i)$ is the number of distinct realizations of a single instance of Y_i given resources $\vec{t}$; and $Z(\vec{t} - \vec{\mu_i})$ is the partition function of a system in which the atomic resources have been decreased by the amount needed to build one instance of Y_i.

It is immediate from [54] that

$$\varrho(\vec{t}, Y_i) = d(\vec{t}, \vec{Y_i}) = \frac{\prod_{j=1}^{T} t_j!}{\prod_{j=1}^{T} (t_j - \mu_{i,j})! \, \omega_i}, \tag{59}$$

where $\vec{Y_i}$ denotes a unit vector in the Y_i direction. We provide a proof of the theorem using generating functions elsewhere. However, to see why the claim holds, we reason as follows. The subset of $\Sigma(\vec{t})$ in which we restrict ourselves to states $\vec{n}$ that contain at least one copy of Y_i stands in a 1-1 correspondence to the unrestricted state space $\Sigma(\vec{t} - \vec{\mu_i})$, because any realization of Y_i in $\Sigma(\vec{t})$ occurs in all possible contexts and these contexts are precisely the states of $\Sigma(\vec{t} - \vec{\mu_i})$. The question then is how the degeneracy and the energy content of a state $\vec{n} \in \Sigma(\vec{t} - \vec{\mu_i})$ change by adding $\vec{\mu_i}$ atoms to realize one instance of Y_i. The degeneracy of state $\vec{n} \in \Sigma(\vec{t} - \vec{\mu_i})$ is amplified (multiplied) by $\varrho(\vec{t}, Y_i)$ realizations of Y_i, but one instance of Y_i is added to those the state already had and so we also need to divide by $n_i + 1$ to compensate for indistinguishable permutations within the instances of Y_i, see [54]. Thus, $d(\vec{t}, \vec{n} + \vec{Y_i}) = (\varrho(\vec{t}, Y_i)/(n_i + 1))d(\vec{t} - \vec{\mu_i}, \vec{n})$ and the Theorem follows as summarized symbolically:

$$\frac{1}{Z(\vec{t})} \sum_{\substack{\vec{n} \in \Sigma(\vec{t}) \\ n_i \geq 1}} n_i d(\vec{t}, \vec{n}) \varepsilon(\vec{n}) = \frac{1}{Z(\vec{t})} \sum_{\vec{n} \in \Sigma(\vec{t} - \vec{\mu_i})} (n_i + 1) \frac{\varrho(\vec{t}, Y_i)}{n_i + 1} d(\vec{t} - \vec{\mu_i}, \vec{n}) \varepsilon_i \varepsilon(\vec{n}) = \varrho(\vec{t}, Y_i) \varepsilon_i \frac{Z(\vec{t} - \vec{\mu_i})}{Z(\vec{t})}. \tag{60}$$

It remains to compute the partition function of the assembly systems discussed in the main text, which is not too difficult and provided in the subsequent section 8.

## 8. Partition functions and average catalytic potential

A. Polymerizing scaffold without ligands. Let a state contain i bonds (not necessarily in the same polymer). Any such state has a Boltzmann factor σ^i, where σ is the binding affinity between two scaffold protomers. We count the number of ways to realize i bonds as follows. Line up the t_S (labelled) protomers and observe that there are $t_S - 1$ slots between protomers where a bond could be inserted. Thus there are $\binom{t_S - 1}{i}$ ways of inserting i bonds and the insertion of i bonds always creates $t_S - i$ molecules. For each choice of i slots there are $t_S!$ permutations of the protomers. Since the order in which a choice of bond locations creates the $t_S - i$ molecules is irrelevant, we must reduce the label permutations by $(t_S - i)!$ object permutations to obtain the degeneracy d_i of a state with i bonds. The partition function is therefore

$$Z_{t_S}^{\text{poly}} = \sum_{i=0}^{t_S - 1} \sigma^i \binom{t_S - 1}{i} \frac{t_S!}{(t_S - i)!} \tag{61}$$

The number of possible realizations of a single polymer s_n of length n is $t_S!/(t_S - n)!$, which yields with [58] for the average number of polymers of length n, $\langle s_n \rangle$:

$$\langle s_n \rangle = \frac{t_S!}{(t_S - n)!} \sigma^{n-1} \frac{Z_{t_S - n}^{\text{poly}}}{Z_{t_S}^{\text{poly}}}. \tag{62}$$

Figure S9 compares the length distributions of equivalent continuum and discrete polymerization systems
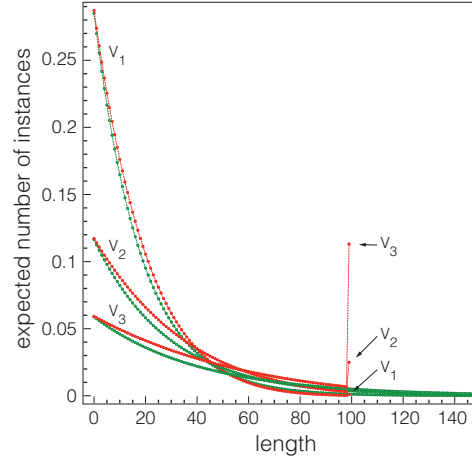
Fig. S9. Length distribution in continuum and discrete polymerization. A continuum and discrete polymerization system are set up with equivalent parameters assuming a base volume $V = 10^{-15}$ L (the order of magnitude of a bacterial cell). Their length distributions are compared for three volumes: $V_1 = 0.05V$, $V_2 = 0.02V$, $V_3 = 0.01V$. A change in volume means a change in affinity for the discrete system and a change in protomer concentration for the continuum system, i.e. $t_S = 100$ protomers or $t_S = 100/(\mathcal{A}V_i)$ M; discrete affinity $\sigma_s = 10^8/(\mathcal{A}V_i)$ molecules^{-1} or continuum affinity $\sigma_d = 10^8$ M^{-1}. The green curves are associated with the continuum system (equation 31 and the red ones with the discrete case (equation 62. Associated volumes are as indicated in the graph. Since the curves cross, the maximer is also marked with the corresponding volume. The continuum distribution is cut off at 150.

B. Average catalytic potential of the polymerizing scaffold with ligands. Because of binding independence, the partition function of this system is the product of three partition functions: $Z_{t_S}^{\mathrm{poly}} Z_{t_S,t_A}^{\mathrm{dimer}} Z_{t_S,t_B}^{\mathrm{dimer}}$, with $Z_{t_S,t_X}^{\mathrm{dimer}}$ the partition function of a system in which S-agents and X-agents can dimerize with affinity γ. $Z_{t_S,t_X}^{\mathrm{dimer}}$ is simple to obtain: choose i agents of type A, i agents of type S, and pair them:

$$Z_{t_S,t_X}^{\mathrm{dimer}} = \sum_{i=0}^{\min(t_S,t_X)} \gamma^i \binom{t_S}{i}\binom{t_X}{i} i!. \tag{63}$$

Putting this together yields the partition function for resource vector $\vec{t} = (t_A, t_B, t_S)$

$$Z(\vec{t}) = \left[\sum_{k=0}^{t_S-1} \sigma^k \binom{t_S-1}{k} \frac{t_S!}{(t_S-k)!} \right] \left[\sum_{i=0}^{\min(t_S,t_A)} \alpha^i \binom{t_A}{i}\binom{t_S}{i} i! \right] \left[\sum_{j=0}^{\min(t_S,t_B)} \beta^j \binom{t_B}{j}\binom{t_S}{j} j! \right]$$

$$= Z_{t_S}^{\mathrm{poly}} Z_{t_S,t_A}^{\mathrm{dimer}} Z_{t_S,t_B}^{\mathrm{dimer}} \tag{64}$$

The total number of realizations, $\varrho(\vec{t}, \{A_i S_l B_j\})$ of polymers of length l with i A-agents and j B-agents attached, and thus each with Boltzmann factor $\sigma^{l-1}\alpha^i\beta^j$, is given by

$$\varrho(\vec{t}, \{A_i S_l B_j\}) = \frac{t_S!}{(t_S-l)!}\binom{l}{i}\binom{t_A}{i} i! \binom{l}{j}\binom{t_B}{j} j!$$

$$= \binom{l}{i}\binom{l}{j}\frac{t_S!}{(t_S-l)!}\frac{t_A!}{(t_A-i)!}\frac{t_B!}{(t_B-i)!}$$

$$= \binom{l}{i}\binom{l}{j}\frac{\vec{t}!}{(\vec{t}-\vec{v})!} \tag{65}$$

where $\vec{v} = (i, j, l)$ is the composition vector of the configuration and we define for brevity the factorial of a vector as the product of the factorials of its components. Putting all this together yields the average catalytic potential $\langle Q \rangle$

$$\langle Q_{\mathrm{poly}} \rangle = \sum_{l=1}^{t_S} \sum_{i=0}^{\min\{l,t_A\}} \sum_{j=0}^{\min\{l,t_B\}} \underbrace{i\,j}_{\substack{\text{\# of} \\ \text{interactions}}} \underbrace{\binom{l}{i}\binom{l}{j}\frac{\vec{t}!}{(\vec{t}-\vec{v})!}\sigma^{l-1}\alpha^i\beta^j \frac{Z(\vec{t}-\vec{v})}{Z(\vec{t})}}_{\substack{\text{total realizations of} \\ \text{configurations with } \vec{v}}} \tag{66}$$

$$\underbrace{}_{\text{average total counts}}$$

C. Average catalytic potential of the multivalent scaffold with ligands. The case of a multivalent scaffold with m binding sites for A and n binding sites for B follows the lines of section B. For each type of binding sites one can formulate a partition

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana

function in full analogy to $Z_{t_S,t_X}^{\text{dimer}}$, but with $m\,t_S$ (or $n\,t_S$) sites available to bind i agents of type A (or j agents of type B) to yield a state with Boltzmann factor $\alpha^i\beta^j$. Thus, the partition function for a multivalent scaffold system is

$$Z_{t_A,t_B,t_S}^{\text{multi}} = \sum_{i=0}^{\min(m\,t_S,t_A)} \sum_{j=0}^{\min(n\,t_S,t_B)} \alpha^i\,\beta^j \binom{t_A}{i}\binom{m\,t_S}{i} i! \binom{t_B}{j}\binom{n\,t_S}{j} j! \qquad [67]$$

The average number of scaffolds loaded with i ligands of type A and j ligands of type B *in a particular configuration* then becomes

$$\langle n_{ij}\rangle = \frac{t_A!}{(t_A - i)!}\frac{t_B!}{(t_B - j)!} t_S\,\alpha^i\,\beta^j \frac{Z_{t_A-i,t_B-j,t_S-1}^{\text{multi}}}{Z_{t_A,t_B,t_S}^{\text{multi}}}. \qquad [68]$$

Finally, for the average catalytic potential we have

$$\langle Q_{\text{multi}}\rangle = \sum_{i=0}^{\min(t_A,m)} \sum_{j=0}^{\min(t_B,n)} i\,j \binom{m}{i}\binom{n}{j}\langle n_{ij}\rangle. \qquad [69]$$

D. Remarks on numerical evaluation. While expressions [66] and [69] are explicit, their use with large particle numbers—t_S, t_A and t_B—is limited by numerical instabilities (even after efficiency rearrangements). In a separate paper we connect assembly systems with the theory of analytic combinatorics (2), which provides direct approximations based on viewing generating functions as analytic functions over the complex numbers. In our hands, these approximations are not accurate enough over the entire parameter range for the present context. Our figures were therefore generated using the exact expressions [66] and [69], using arbitrary-precision calculations (to 100 significant digits) in `Mathematica` (3), and employing relatively modest particle numbers to keep computation times reasonable.

9. The maximer probability and 1D percolation

The probability of observing the longest possible polymer, given protomer resources, is obtained from [62] by setting $n = t_S$:

$$\langle s_{\text{max}}\rangle = \frac{t_S!\,\sigma^{t_S-1}}{Z_{t_S}^{\text{poly}}}. \qquad [70]$$

This probability is graphed as a function of t_S and σ in Figure 5A of the main text.

There is an analogy between 1D bond percolation and polymerization at our level of abstraction. The analogy is an exact correspondence in the case of continuum polymerization and bond percolation on an infinite 1D lattice.

A basic quantity in 1D percolation is the mean number of chains (clusters) of size n normalized per lattice site, which is given by $p^{n-1}(1-p)^2$, where p is the probability of a bond between adjacent lattice sites and functions as a parameter. The same expression obtains in terms of the concentration of polymers of length n normalized per protomer (1, 4):

$$\frac{s_n}{t_S} = p^{n-1}(1-p)^2. \qquad [71]$$

In the context of polymers, the bond probability is not the primary parameter, but a function of the basic parameters t_S and σ. Following Flory (1), we can express p as

$$p = \frac{t_S - W}{t_S} = 1 - \frac{1}{t_S}\frac{s}{1 - \sigma s}, \qquad [72]$$

with W the concentration of all polymers as defined in [3] for $a = b = 0$ and given more compactly by [10]. The first equality defines p in terms of the difference between the maximal possible concentration of objects in the system (t_S) and the actual concentration of objects; this difference is the concentration of bonds. Using [31] for s yields

$$p = 1 - \frac{2}{1 + \sqrt{1 + 4\sigma t_S}}. \qquad [73]$$

Together, expressions [71] and [73] are equivalent to [31] and connect simple polymerization to percolation. As well-known, in the infinite/continuum case, percolation can only occur at $p = 1$, which is to say in the limit of $t_S \to \infty$ or $\sigma \to \infty$.

The analogy persists but the exact correspondence breaks down in the finite, i.e. discrete, case. The percolation probability in the polymerization case is $\langle s_{\text{max}}\rangle$ as given by [70]. The bond probability, p_{bond}, is the expected fraction of bonds and can be computed following the arguments that led to [61]. We obtain

$$p_{\text{bond}} = \frac{1}{t_S - 1} \frac{\displaystyle\sum_{i=1}^{t_S-1} i\sigma^i \binom{t_S - 1}{i}\frac{t_S!}{(t_S - i)!}}{Z_{t_S}^{\text{poly}}}. \qquad [74]$$
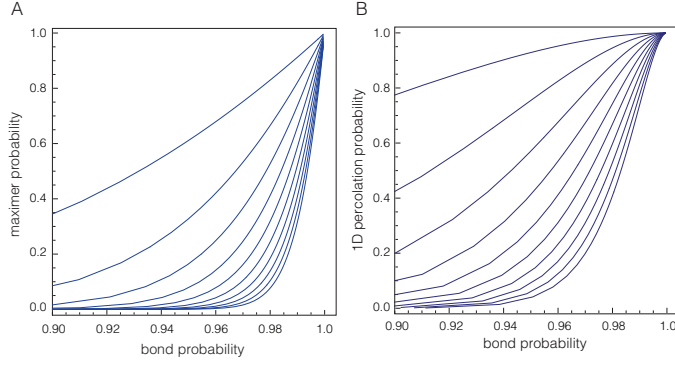
Fig. S10. Finite size 1D bond percolation and polymerization. **A:** This panel is panel B of Figure 5 in the main text. It depicts the probability of the maximer [70] as a function of p_bond as given by [74]. Each curve represents a particular t_S-value for which σ sweeps from 1 to 1000 molecules $^{-1}$. t_S ranges from 10 (topmost curve) to 100 (lowest curve) in increments of 10. **B:** The plot depicts the 1D bond percolation probability [75] as a function of the same bond probabilities used in panel A. The comparison serves to illustrate the difference between 1D bond percolation and polymerization while also emphasizing the analogy. On the other hand, bond percolation on an infinite 1D lattice is equivalent to polymerization described in terms of continuous concentrations.

In 1D bond percolation, the percolation probability is

$$p_\text{perc} = 1 - (1-p)^2 \sum_{i=0}^{t_S-2} ip^{i-1} = p^{t_S-2}(t_S - p(t_S - 2) - 1), \tag{75}$$

with t_S the size of the lattice and p the bond probability.

In Figure 5B of the main text we sweep across a range for t_S and σ. For each (t_S, σ) pair we calculate the corresponding p_bond via [74] as the abscissa and $\langle s_\text{max} \rangle$ via [70] as the ordinate. This graph is reproduced as Figure S10B for comparison with finite-size bond percolation, Figure S10A. Clearly in [75] p is just a parameter, but in Figure S10A we compute it via [74] using the same sweep over t_S and σ as for Figure S10B to make comparison meaningful. The view from percolation is useful because it packages the dependency on t_S and σ into the single quantity p (or p_bond).

10. Scaling behavior
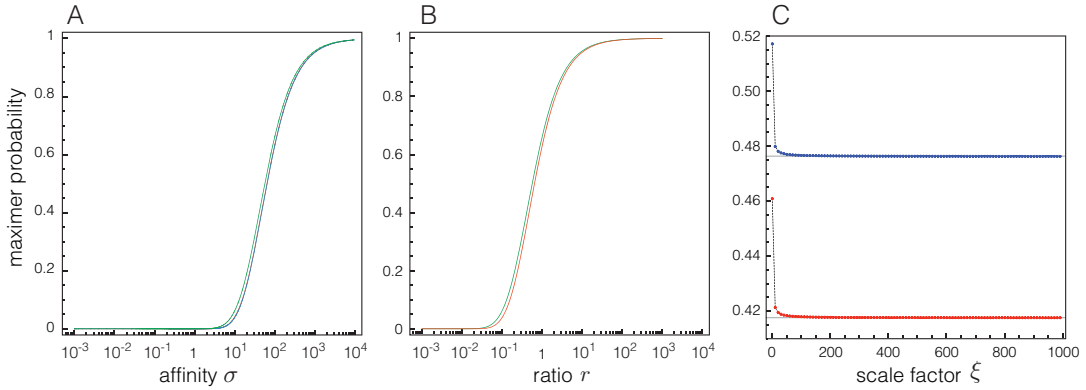
Fig. S11. Scaling behavior of the maximer distribution. The panels illustrate the approximate scaling behavior of $\langle s_\text{max} \rangle$ from different perspectives implied by [77]. In all three panels, the ordinate is the maximer probability as given by [70]. **A:** The graph exemplifies the relation [77] by plotting three curves, blue: $\langle s_\text{max} \rangle[10, 0.1\sigma]$, red: $\langle s_\text{max} \rangle[100, \sigma]$, and green: $\langle s_\text{max} \rangle[1000, 10\sigma]$ as a function of the affinity σ. The blue and green graphs are related to the (arbitrary) red baseline graph by scale factors $\xi = 0.1$ and $\xi = 10$, respectively. The red and blue graphs sit on top of each other, while green has a slight (and slightly σ-dependent) shift to the left. **B:** This panel illustrates the scaling version [78], comparing red: $\langle s_\text{max} \rangle[1000, r\,1000]$ with green: $\langle s_\text{max} \rangle[10, r\,10]$, sweeping along r. **C:** The graph in this panel shows an integer sweep of the scale factor ξ, as per [77], for two pairs, $[t_S, \sigma] = [10, 5]$ (red), $[t_S, \sigma] = [10, 6]$ (blue). The scaling relation is well fulfilled except for very small particle numbers.

We refine the notation for the maximer probability [70] to emphasize the dependence on the parameters t_S and σ,

$$\langle s_\text{max} \rangle[t_S, \sigma] \equiv \langle s_\text{max} \rangle, \tag{76}$$

in order to note an approximate scaling relation that we observe numerically:

$$\langle s_\text{max} \rangle[t_S, \sigma] \approx \langle s_\text{max} \rangle[\xi t_S, \xi \sigma], \tag{77}$$

with $\xi > 0$ a dimensionless scale factor. Two systems are approximately equivalent if their protomer numbers and affinities are related by the same scale factor: $t_S^{(1)} = \xi t_S^{(2)}$ and $\sigma^{(1)} = \xi \sigma^{(2)}$. This implies that $t_S^{(1)}/t_S^{(2)} = \sigma^{(1)}/\sigma^{(2)}$ or $r = \sigma^{(1)}/t_S^{(1)} = \sigma^{(2)}/t_S^{(2)}$.

**Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana**

The latter says that two systems behave approximately the same if the ratio r of their respective affinity to protomer number is the same, which yields another way of expressing the scaling observation as

$$\langle s_{\max}\rangle[t_S^{(1)}, r\, t_S^{(1)}] \approx \langle s_{\max}\rangle[t_S^{(2)}, r\, t_S^{(2)}].$$ [78]

These relations are depicted in Figure S11.

11. Unequal ligand concentrations and ligand binding affinities

A. Polymerizing scaffold system. As in Figure 6 of the main text, Figure S12A evidences the σ-dependence of the initial slope in the discrete system and illustrates the effect of ligand imbalance: Once the scarcer ligand, here A, is mostly bound up and the number of scaffold protomers increases further, A-ligands must spread across an increasingly wider range of length classes, thereby reducing the likelihood of multiple occupancy on the same polymer. As a result, although the binding opportunities for the more abundant ligand, here B, increase (up to the overall prozone peak), B-particles bound to a particular polymer are less likely to encounter any As bound to it. The result is a slope reduction compared to a situation in which both ligands are present in equal numbers. A substantive difference between ligand binding constants causes not only a slope reduction prior to the prozone but has, in particular, the effect of delaying the prozone peak considerably beyond what one would expect based on particle numbers alone. It is worth noting that in the Wnt signaling cascade, ligand affinities——enzyme-scaffold, i.e. GSK3β–Axin, and substrate-scaffold, i.e. β-catenin–Axin—are regulated by the signaling process (5, 6).
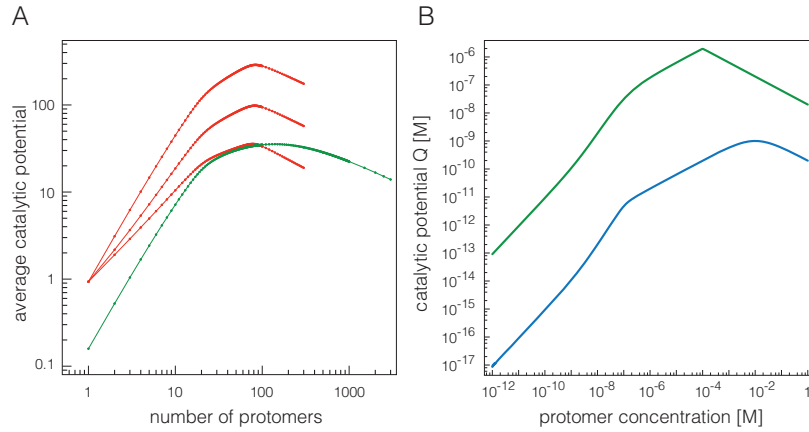

Fig. S12. Effects in discrete and continuum polymerizing scaffold systems. **A:** The panel illustrates the effects of the polymerization constant σ, of ligand imbalance, and of unequal ligand affinities on discrete polymerization. Red, ligand imbalance: $t_A = 20$, $t_B = 80$, $\alpha = \beta = 0.9$ molecules^{-1}, $\sigma = 0.01$ (lower), $\sigma = 0.1$ (middle), $\sigma = 1$ (upper). Green, unequal ligand affinities: $t_A = t_B = 20$, $\alpha = 0.01$, $\beta = 1$ molecules^{-1}, $\sigma = 1$ molecules $^{-1}$. t_S on the abscissa. **B:** This panel illustrates the effects of ligand imbalance and of unequal ligand binding constants on continuum polymerization. Blue, unequal binding constants: $\alpha = 10^2$ M^{-1}, $\beta = 10^9$ M^{-1}, $t_A = t_B = 10^{-7}$ M, $\sigma = 10^8$ M^{-1}. Green, ligand imbalance: $t_A = 10^{-8}$ M, $t_B = 10^{-4}$ M, $\alpha = \beta = 10^7$ M^{-1}, $\sigma = 10^8$ M^{-1}.

In the continuum case, unlike the discrete case, the initial slope is independent of the polymerization constant σ until a level of protomer abundance is reached sufficient for making polymerization effective, as discussed in section 5 (equation 41. The inflection point at which the slope changes from 1 to 2 (in a log-log plot) will shift accordingly. After that slope change, the responses to ligand imbalance and to differences between ligand binding constants are analogous to the discrete case, as seen in Figure S12B.

Neither ligand imbalance or differences in binding constants appear to affect the downward slope at large t_S in the continuum or the discrete case.

B. Multivalent scaffold system. The responses to ligand and affinity imbalances in a multivalent scaffold system follow similar lines as in the polymerizing case. When both ligand types are present with the same number of particles, the ligand with higher affinity experiences the prozone later, since the amount of scaffold-bound ligand is higher compared to the other type. This is seen in Figure S13B with the steepening of the downward slope associated with the stronger binding ligand. The situation with ligand imbalance is analogous. The ligand with higher abundance keeps binding while the scarcer ligand is undergoing its prozone; thus the subdued effect on catalytic potential, which, in the example of Figure S13C is mainly holding a constant level until the prozone for the more abundant ligand sets in. Although affinity and number imbalance mimic each other, the affinity imbalance exhibits a much less pronounced plateau around the prozone peak and consequently the drop-off is less sharp than in the case of number imbalance. Extremely high affinity differences would be required to generate a plateau similar to number imbalance. This is seen in the continuum case, shown in Figure S14A, where affinities differ by 7 orders of magnitude. The concentration imbalance in the continuum case yields a similar picture as in the discrete case (Figure S14B).
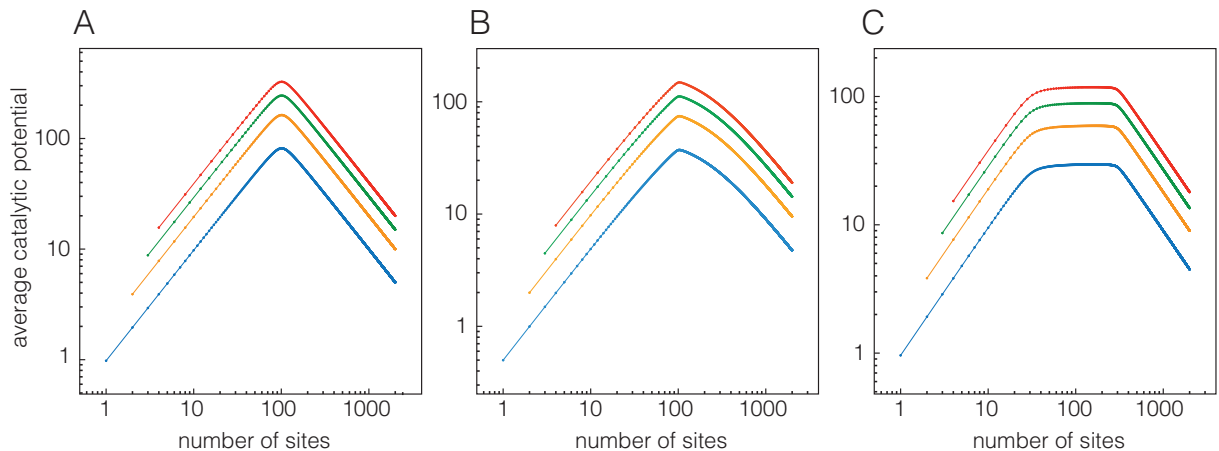
Fig. S13. Catalytic potential of multivalent scaffolds (discrete case). **A:** $\langle Q_{\text{multi}} \rangle$, equation [69], when particle numbers and binding affinities are the same for both ligand types: A and B are 100 particles each, binding affinities are 0.9 molecules^{-1}. Valencies: 1 (blue), 2 (orange), 3 (green), 4 (red). The abscissa shows the total number of sites, but $\langle Q_{\text{multi}} \rangle$ is calculated for site increments that reflect the valency of each scaffold. **B:** Like panel A, but unequal ligand binding affinities: $\alpha = 0.01$ and $\beta = 9$ molecules^{-1}. **C:** Like panel A, but unequal numbers of ligand particles: $A = 30$ and $B = 300$, binding affinities for both are 0.9 molecules^{-1}. Colors indicate valencies as in panel A.
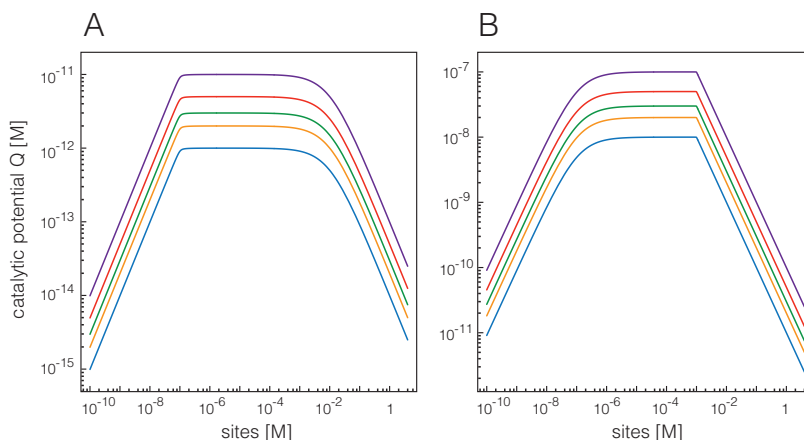

Fig. S14. Catalytic potential of multivalent scaffolds (continuum case). **A:** The panel provides an example for the effect of unequal ligand binding affinity. $t_A = t_B = 10^{-7}$ M, $\alpha = 10^2$ M^{-1}, $\beta = 10^9$ M^{-1}, valencies: $1, 2, 3, 4$. **B:** The panel illustrates the effect of ligand concentration imbalance. $t_A = 10^{-8}$ M, $t_B = 10^{-3}$ M, $\alpha = \beta = 10^7$ M^{-1}, valencies: $1, 2, 3, 4$.

12. Stochastic simulations

Our analysis of the discrete case focuses on average behavior. Analytic techniques for higher moments are beyond the scope of this contribution and will be presented elsewhere. In lieu of an analytic treatment, we performed several stochastic simulations using the Kappa platform (7, 8) and GNU Parallel (9). Figure S15 displays the essential observations in the context of Figures 3A and 6A of the main text and S11B of this Supplement.

Fluctuations in the binding of ligands translate into Q-fluctuations on the basis of how sites are partitioned into agents. There are three regimes, which we describe in the case of a monovalent scaffold system for simplicity (lowest green curve in Figure S15; green curve in Figure S16; and Figure S17): (i) At low scaffold numbers, prior to the prozone peak, most scaffolds are fully occupied by both ligands. Fluctuations cause transitions between system states with similar Q and variance is therefore low (see red distributions in Figure S17). (ii) Just past the prozone peak, many scaffolds are still occupied by both ligands, but there is an increasing number of singly bound and some empty scaffolds. Unbinding from a fully occupied scaffold is statistically offset by re-binding to the pool of singly-bound scaffolds, which yields a net effect similar to situation (i). However, in addition, singly-bound scaffolds may also lose their ligand. This event is neutral in Q, but free ligands may re-bind an already singly-bound scaffold, thereby increasing Q. Likewise, dissociation from a fully occupied scaffold an re-association with an empty one will decrease Q. As a result of this expanded Q-range, the variance has increased compared to a situation with similar average Q prior to the prozone peak (see green distributions in Figure S17). (iii) Well past the prozone peak, a number of scaffolds are bound by one ligand and many have no ligands at all. Ligand binding fluctuations will mainly shift ligands from singly-bound scaffolds to empty scaffolds with no effect on Q. As a result, Q-variance is now decreasing again (see blue distributions in Figure S17).

Andrés Ortiz-Muñoz, Héctor F. Medina-Abarca and Walter Fontana
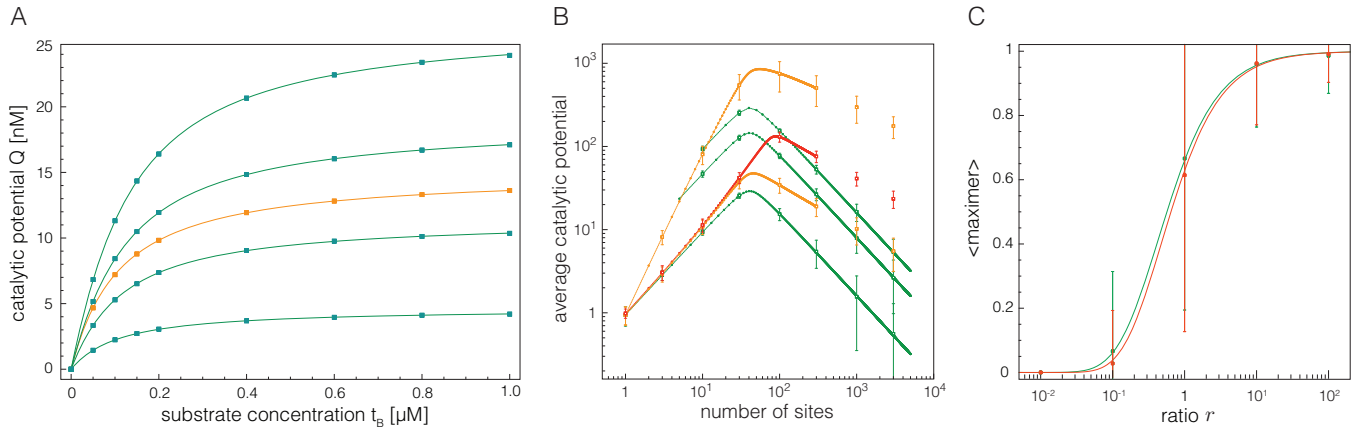
Fig. S15. Stochastic simulations. For all stochastic simulations, we used a volume on the order of an human erythrocyte, $V = 10^{-12}$ L. All summary statistics were computed with 500 samples, each an independent and equilibrated state. **A:** The solid curves in this panel are identical to those in Figure 3A of the main text. Stochastic simulations were performed by converting deterministic affinities into stochastic affinities as described in the main text (section "The discrete case in equilibrium") and by converting concentrations into particle numbers at the given volume V. Averages of catalytic potential are indicated by filled squares. Green: polymerizing system at various protomer numbers, descending from top: 36120 molecules (60 nM), 27090 molecules (45 nM), 18060 molecules (30 nM), 9030 molecules (15 nM). Orange: reference Michaelian system with 60200 (100nM) enzymes. Because of the large numbers of particles, the standard deviation is smaller than the squares at the chosen scale. This panel is meant as a sanity check that simulations at large particle numbers indeed reproduce the continuum picture as we derived it analytically. **B:** The curves in this panel are identical to those in Figure 6A of the main text and refer to discrete scaffolding systems. Stochastic simulations were performed using the same parameters listed in that Figure. The squares mark the average catalytic potential, which coincides with the theoretical calculations; the error bars mark one standard deviation. In the polymerizing scaffold case, the simulation allowed us to extend the range of the rather time-consuming calculations using the analytical expression 66. Note the log-log scale of the axes distorting the error bars; for a linear-log scale see Figure S16. Green: multivalent scaffolds of valencies $n = 10$ (upper), $n = 5$ (middle), and $n = 1$ (lower). Orange: polymerizing scaffold system with polymerization affinities $\sigma = 10$ (upper) and $\sigma = 0.01$ (lower). Red: polymerizing scaffold system at the same affinity as the lower orange curve, but with twice the number of ligand particles. **C:** The curves are identical to those in Figure S11B. As in that Figure, r is the ratio of affinity to the number of protomers. Squares mark the average number of maximers and error bars mark one standard deviation. Green: system with 10 protomers. Red: system with 1000 protomers.
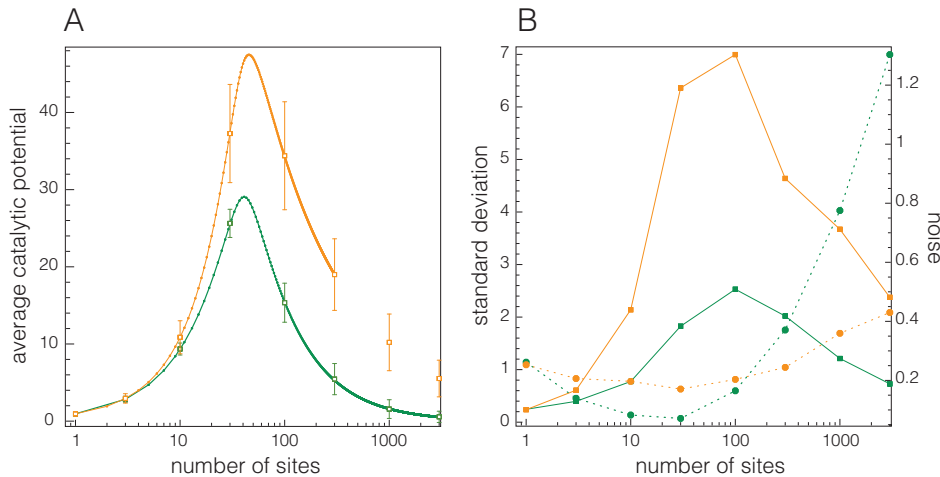
Fig. S16. Variance and noise. **A:** This panel reproduces a subset of data from Figure S15B on a linear-log scale to enable a more direct visual interpretation of fluctuations. The green curve in this panel corresponds to the lowest green curve in Figure S15B. It belongs to a system of multivalent scaffolds with valency 1. The orange curve belongs to the polymerizing scaffold system and corresponds to the lowest orange curve in Figure S15B. Because the valency of individual scaffolds in both systems is 1, the number of sites on the abscissa corresponds to the number of scaffold agents, polymerizing or not. The main observation is that for the same average catalytic potential $\langle Q \rangle$ the standard deviation is larger after the prozone peak than prior to it. **B:** This panel recasts the information in panel A by directly displaying the standard deviation (solid curves). The dashed curves (right ordinate) depict the noise, i.e. the ratio of standard deviation to the mean. The main observation here is that the polymerizing system (orange) is significantly less noisy than the monovalent scaffold system (green).
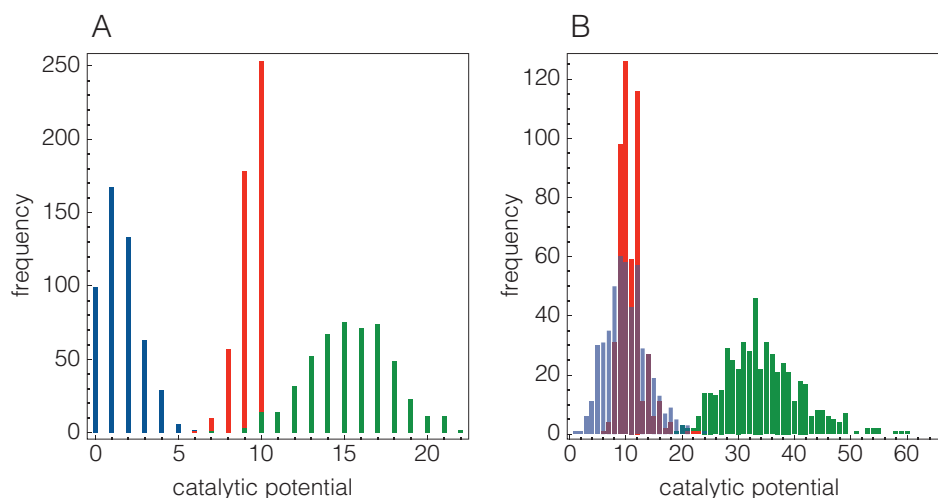
Fig. S17. Distributions of catalytic potential. Panels **A** (monovalent scaffold system) and **B** (polymerizing scaffold system) depict the distribution of catalytic potential for a state sampled prior to the prozone peak (10 scaffold particles, red), just past the peak (100 particles, green) and well past the peak (1000 particles, blue). Other parameters as in Figure 6A of the main text.

References

1. Flory PJ (1936) Molecular size distribution in linear condensation polymers I. *Journal of the American Chemical Society* 58(10):1877–1885.
2. Flajolet P, Sedgewick R (2009) *Analytic combinatorics.* (Cambridge University Press, Cambridge; New York), pp. xiii, 810.
3. Wolfram Research Inc. (2019) Mathematica, Version 12.0. Champaign, IL.
4. Reynolds P, Stanley H, Klein W (1977) Ghost fields, pair connectedness, and scaling: exact results in one-dimensional percolation. *Journal of Physics A: Math. Gen.* 10(11):L203–L209.
5. Luo W, et al. (2007) Protein phosphatase 1 regulates assembly and function of the β-catenin degradation complex. *The EMBO Journal* 26(6):1511–1521.
6. Willert K, Shibamoto S, Nusse R (1999) Wnt-induced dephosphorylation of axin releases β-catenin from the axin complex. *Genes & Development* 13(14):1768–1773.
7. Boutillier P, et al. (2018) The Kappa platform for rule-based modeling. *Bioinformatics* 34(13):i583–i592.
8. Boutillier P, Feret J, Krivine J, Fontana W (2018) The Kappa Language and Kappa Tools: A User Manual and Guide [v4] (https://kappalanguage.org).
9. Tange O (2018) *GNU Parallel 2018.* (Ole Tange).