

Electronic Supplementary information (ESI) Computational design of a minimal catalyst using colloidal particles with programmable interactions

Maitane Muñoz-Basagoiti, Olivier Rivoire, Zorana Zeravcic

I. INTERACTION MODEL AND NUMERICAL SIMULATIONS

We model the interactions between particles with an isotropic interaction potential introduced by Wang et al. in [1], which is given by

$$U(r, \epsilon) = \epsilon \alpha(r_{\text{cutoff}}, \sigma) \left[\left(\frac{\sigma}{r} \right)^2 - 1 \right] \times \left[\left(\frac{r_{\text{cutoff}}}{r} \right)^2 - 1 \right]^2 \quad (\text{ESI Eq. 1})$$

if $r < r_{\text{cutoff}}$ and zero elsewhere, and where $\alpha(r_{\text{cutoff}}, \sigma)$ fixes $U(r_{\text{min}}, \epsilon) = \epsilon$. The equilibrium position r_{min} is given by the minimum of $U(r, \epsilon)$, which for $r_{\text{cutoff}}/\sigma = 1.1$ results in $r_{\text{min}}/\sigma \approx 1.03$. With this potential, the association of substrate/product monomers is diffusion-limited (see Fig. 1B in the main text). To simulate the irreversible case, i.e., when substrate bond formation is not allowed, we forbid product particles from interacting with each other. Particle valence is realized by keeping track of the bonds between particles in the system: each catalyst particle can bind one catalyst particle and one substrate/product particle, while each substrate/product particle can bind one catalyst and one substrate/product particle at the time. When two particles come within the interaction range $r < r_{\text{cutoff}}$, they can form a bond if they are of the right type and if they have not already formed a bond of that type. When the distance between two particles that cannot interact is $r < r_{\text{min}}$, they repel via a soft-harmonic potential,

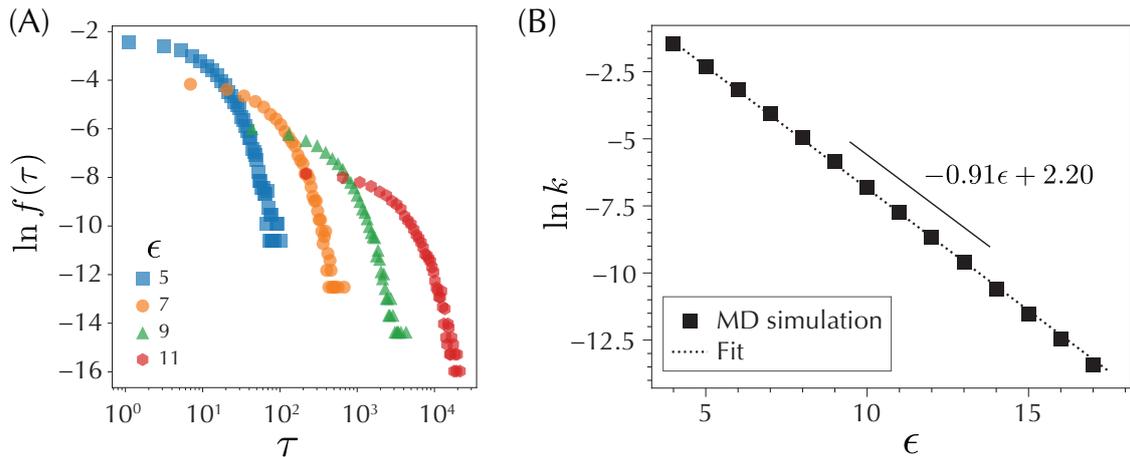
$$U_H(r) = \frac{1}{2}k(r - r_{\text{min}})^2 \quad (\text{ESI Eq. 2})$$

where we take $k = 1000$.

We perform Langevin Molecular Dynamics (MD) simulations using an in-house code that implements a modified velocity-Verlet algorithm to integrate the equations of motion [2]. All simulations are performed with temperature $T = 1.0$, friction coefficient $\gamma = 12.5$ and time-step $\Delta t = 10^{-4}$ time units, and we used periodic boundary conditions. To simplify the simulations, the rigid bond in the catalyst is realized by fixing the catalyst particles in space within the simulation box, i.e., the catalyst remains immobile during the simulation. To investigate the effect of product inhibition (Fig. 3C and Fig. 4 in the main text), the side of the simulation box is chosen as $L = 2R_{\text{Diff}} + \sigma$, where R_{Diff} is the radius of a disk centered around the catalyst (see ESI Fig. 8). The latter is placed in the center of the simulation box.

II. MEAN REACTION TIME FOR THE SPONTANEOUS DISSOCIATION

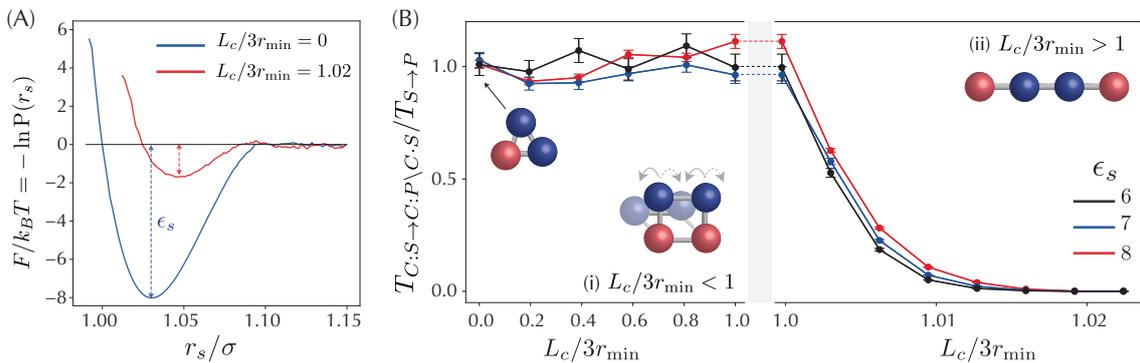
Using MD simulations, we verify that the time it takes the substrate dimer to spontaneously dissociate into two free product monomers for the first time, that is, the first-passage time to overcome the interaction potential (see ESI section I), is exponentially distributed (ESI Fig. 1A). For fixed interaction range $r_{\text{cutoff}}/\sigma = 1.1$, the rate constant associated with the spontaneous reaction, defined as the inverse of the mean first-dissociation time [3, 4], $k_{S \rightarrow P} = 1/T_{S \rightarrow P}$, decreases exponentially with the depth of the potential ϵ (ESI Fig. 1B). We fit the simulation data with an Arrhenius-like expression $k = e^{-A\epsilon+B}$, with $A = 0.91$ and $B = 2.20 \pm 0.07$ within the range $\epsilon \in [3, 17]$. Since $A < 1$, the $\epsilon_{cs} < A\epsilon_s < \epsilon_s$ necessary condition for catalysis holds.



ESI Figure 1. (A) Distribution of first-dissociation times τ obtained from MD simulations for different potential depths ϵ , with fixed interaction range $r_{\text{cutoff}}/\sigma = 1.1$. The dissociation occurs when the distance between particles in the substrate dimer, initially at equilibrium $r = r_{\text{min}}$, exceeds $r > r_{\text{cutoff}}$ for the first time. (B) Logarithm of the dissociation rate k , defined as the inverse of the mean first-dissociation time, $k = 1/\langle\tau\rangle$, as a function of ϵ . Simulation data (squares) can be fit with an Arrhenius-like expression $\ln k = -0.91\epsilon + 2.20$ (dashed line).

III. A SPHERICAL PARTICLE CANNOT CATALYZE THE DISSOCIATION REACTION

A single spherical particle, i.e. $L_c = 0$, cannot catalyze a dissociation reaction within our model. This is shown in ESI Fig. 2A, where we plot the free energy of the system along the substrate bond r_s when the substrate is completely attached to the catalyst, i.e., $r_{cs} < r_{\text{cutoff}}$ for the two substrate-catalyst bonds. We compare two different catalyst designs: a single catalyst particle and a rigid dimer with $L_c > 3r_{\text{min}}$. The activation barrier to break the substrate bond in the presence of a single catalyst particle with $\epsilon_{cs}/\epsilon_s = 1$ equals the barrier in the absence of the catalyst. In other words, the interaction between the substrate dimer and the catalyst particle does not modify the average time to cleave the substrate bond. For the same binding energy, a rigid dimer satisfying $L_c > 3r_{\text{min}}$ reduces the activation barrier by straining the substrate bond, consequently accelerating the cleaving of the bond.



ESI Figure 2. (A) Free energy of the system along the substrate bond r_s when the substrate is fully attached to the catalyst ($r_{cs} < r_{\text{cutoff}}$ for both substrate-catalyst bonds) for two catalyst designs: a single spherical particle (blue line) and a rigid dimer with $L_c/3r_{\text{min}} = 1.02$ (red line). Simulation parameters correspond to $\epsilon_s = \epsilon_{cs} = 8$. (B) Comparison of the mean first-passage times to break the substrate bond with and without the catalyst, $T_{C:S \rightarrow C:P \setminus C:S} / T_{S \rightarrow P}$, as a function of L_c , where we set $\epsilon_{cs} = 30 \gg \epsilon_s$ to avoid unbinding events. Catalysis requires $T_{C:S \rightarrow C:P \setminus C:S} / T_{S \rightarrow P} < 1$. When $L_c \leq 3r_{\text{min}}$, all three bonds in the $C:S$ configuration can exist at equilibrium. This is no longer true when $L_c > 3r_{\text{min}}$, and as a result, the substrate bond is strained by the catalyst.

IV. COMPLEMENTARITY TO THE TRANSITION STATE AND THE $L_c > 3r_{\min}$ NECESSARY CONDITION FOR THE CATALYST GEOMETRY

Catalysis requires $T_{C:S \rightarrow C:P \setminus C:S} < T_{S \rightarrow P}$, where $T_{C:S \rightarrow C:P \setminus C:S}$ is a function of the catalyst geometry L_c , the ϵ_{cs} binding energy and the substrate bond ϵ_s . To investigate the effect of L_c on $T_{C:S \rightarrow C:P \setminus C:S}$, we initiate the system in $C:S$ and record the time it takes to reach $C:P$ using MD simulations. To prevent unbinding events, we set $\epsilon_{cs} \gg \epsilon_s$. Simulation results in ESI Fig. 2B show that catalysis requires $L_c/3r_{\min} > 1$. In this limit, the $C:S$ configuration is linear and the catalyst geometry does not allow the three bonds in $C:S$, namely, the substrate bond and the two catalyst-substrate bonds, to simultaneously exist at equilibrium. As a consequence, the substrate bond is strained.

We can approximate the effective barrier for $C:S \rightarrow C:P$ by

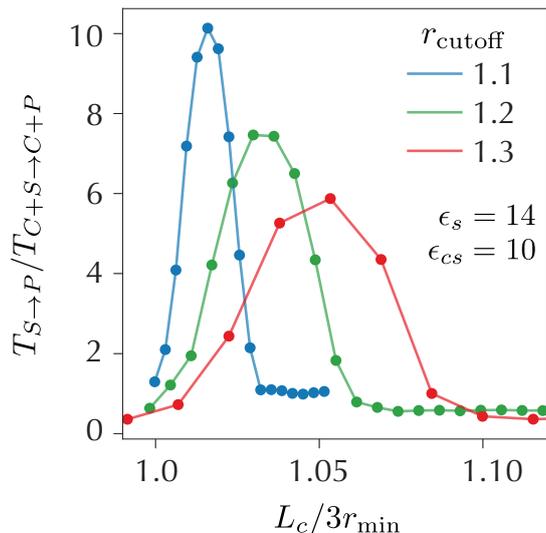
$$\Delta U_{C:S \rightarrow C:P} = 2U\left(\frac{L_c - r_{\text{cutoff}}}{2}, \epsilon_{cs}\right) - \min_{r_s^*, r_{cs}^*} \left[U(r_s, \epsilon_s) + U(r_{cs}, \epsilon_{cs}) + U(L_c - r_s - r_{cs}, \epsilon_{cs}) \right], \quad (\text{ESI Eq. 3})$$

where r_s represents the substrate bond length, r_{cs} the substrate-catalyst bond length, the first term in the equation is the potential energy for $r_s = r_{\text{cutoff}}$, i.e., when the substrate is at the transition state, and the second term is the minimum of the potential energy when the substrate is completely attached to the catalyst. The same approach for the effective barrier for a one-dimensional catalyst has been reported in [5]. The smaller $\Delta U_{C:S \rightarrow C:P}$, the faster the substrate bond will break, as shown in ESI Fig. 2. The first term in (ESI Eq. 3) reaches its minimum when $L_c = 2r_{\min} + r_{\text{cutoff}}$. This threshold marks when the catalyst is geometrically complementary to the transition state. As a consequence, the substrate bond can be strained and the cleaving of the bond accelerated. Note that despite the large interaction energy ϵ_{cs} available, the catalyst only accelerates the reaction for a small subset of geometries, in agreement with Pauling's principle of transition state stabilization [6]. For this subset of geometries, the isotropic interaction with the catalyst becomes directional.

V. CONSTRAINTS ON PARTICLE POLYDISPERSITY

Particle polydispersity δ_σ (particle diameter is $\sigma \pm \delta_\sigma$) affects the three bonds formed along the catalytic axis in configuration $C:S$, changing the distance at which the particles in the catalyst are fixed to $L_c \pm 3\delta_\sigma$. Since catalysis requires $L_c = L_0 \pm \Delta L_c/2$, with $\Delta L_c = 3 \cdot (r_{\text{cutoff}} - r_{\min})$ and L_0 the middle of the allowed range, catalysis is guaranteed if $3\delta_\sigma < \Delta L_c/2$. For the parameters in our model, that is, $r_{\text{cutoff}} = 1.1$ and $r_{\min} \approx 1.03$, $\Delta L_c = 3.3 - 3.09 = 0.21\sigma$, which leads to maximal allowed polydispersity $\delta_\sigma^{\max} = 3.5\%$. Results in Fig. 2A in the main text, which correspond to a specific slice of the parameter space for $\epsilon_s = 14$, show that catalysis occurs for $\Delta L_c^{\text{strict}} = 0.09\sigma$ and maximal allowed polydispersity $\delta_\sigma^{\max} = 1.5\%$. Therefore, for $r_{\text{cutoff}} = 1.1$, the allowed particle polydispersity is $\delta_\sigma^{\max} \in (1, 3.5)$ approximately. Such polydispersity control has been previously reported in the fabrication of colloidal spheres with diameters of a few hundreds of nanometers [7], where authors report polydispersity of 2%, or in the fabrication of emulsion droplets with microfluidic devices [8].

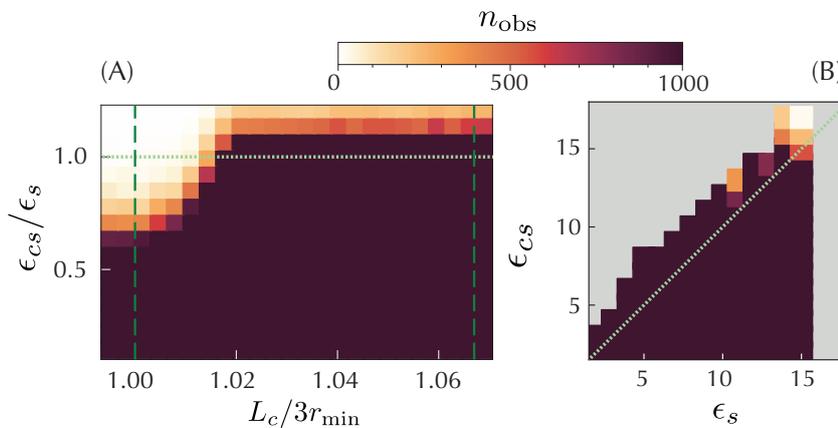
Increasing the interaction range allows us to increase the tolerated particle polydispersity, since $L_c \sim f(r_{\text{cutoff}} - r_{\min})$. While for $r_{\text{cutoff}} = 1.1$, $\delta r = r_{\text{cutoff}} - r_{\min} = 0.07$, by choosing the experimentally accessible values $r_{\min}^{\text{new}} = 1.07$ and $r_{\text{cut}}^{\text{new}} = 1.3$, as is the case for DNA-mediated interactions [9, 10], we can increase the range of catalyst geometries for which catalysis can occur to $\delta r_{\text{new}} = 0.23$. This would then yield allowed polydispersity $\delta_\sigma^{\max} > 10\%$ for the L_c within the necessary conditions. Likewise, by increasing r_{cutoff} , we expect the growth of the parameter range in the more efficient (strict) regime for catalysis (the red area in Fig. 2(a)) by some factor α , i.e., $\Delta L_c^{\text{strict, new}} = \alpha \cdot \Delta L_c^{\text{strict}}$. This would then relax the constraints on particle polydispersity. We verify this by running simulations for $r_{\text{cutoff}}^{\text{new}} = 1.2$ ($r_{\min}^{\text{new}} = 1.05$) and $r_{\text{cutoff}}^{\text{new}} = 1.3$ ($r_{\min}^{\text{new}} = 1.07$). Results in ESI Fig. 3 show that for $r_{\text{cutoff}}^{\text{new}} = 1.3$ the range of L_c values for which catalysis occurs increases by a factor $\alpha \sim 2.5$, relaxing the constraints on particle polydispersity.



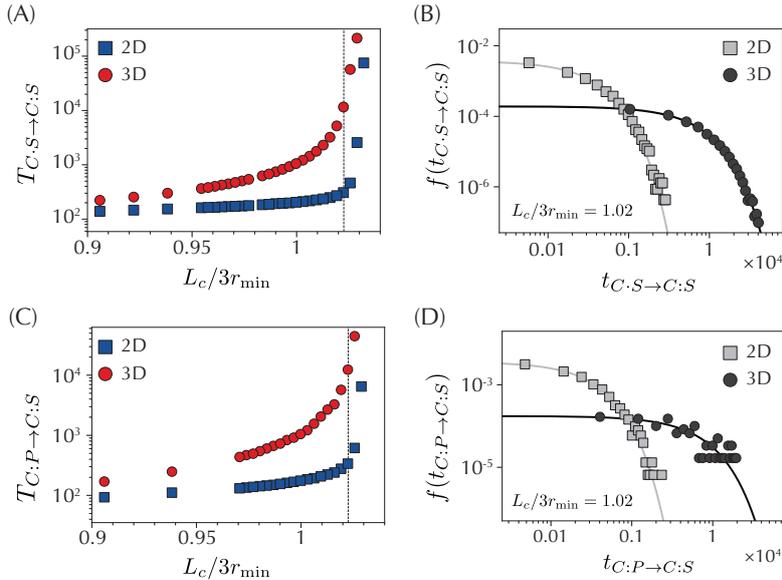
ESI Figure 3. Simulated catalytic efficiency $T_{S \rightarrow P} / T_{C+S \rightarrow C+P}$ as a function of the catalyst geometry L_c for three different interaction ranges r_{cutoff} . Results correspond to an average over > 100 simulations.

VI. SIMULATION DATA AND STATISTICS FOR FIG. 2 IN THE MAIN TEXT

To produce Fig. 2 in the main text, we initiate an ensemble of simulations in $C + S$ by placing the substrate at a random $r > r_{\text{cutoff}}$ distance from both particles in the catalyst in a simulation box with side $L/\sigma = 7.5$, and record the time it takes two monomers to be released in solution in the presence of the catalyst. This time is compared to the timescale of the spontaneous reaction (see ESI Information section II). Each time a monomer is released in solution, it is systematically blocked, that is, it is not allowed to interact with any other particle. We only perform simulations for the reversible case, where the substrate bond can reform in the presence of the catalyst through the $C:P \rightarrow C:S$ transition. To account for the irreversible case, which concerns Fig. 2 in the main text, we use the same simulation trajectories, but only keep the trajectory up to the first time the substrate bond breaks. If the bond breaks after a $C:S \rightarrow C:P$ or a $C:S \rightarrow C:P$ transition, we then simulate the release of the monomer(s) bound to the catalyst by drawing from an exponential distribution with the rate $k(\epsilon_{cs})$ (ESI Fig. 1B). The number of simulations run for each set of parameters is shown in ESI Fig. 4.



ESI Figure 4. Number of observations n_{obs} for each data point reported in Fig. 2 in the main text. The color bar is capped at $n_{\text{obs}} = 1000$ observations. The gray area in panel (B) indicates that there is no data in this region. The $\epsilon_{cs} < \epsilon_s$ (green dotted lines) and $3r_{\text{min}} < L_c < 3r_{\text{cutoff}}$ (green dashed lines) conditions for catalysis are indicated.



ESI Figure 5. (A, C) Simulated mean first-passage time for the $C:S \rightarrow C:S$ (A) and $C:P \rightarrow C:S$ (C) transitions as a function of L_c for a system in 2D (blue squares) and 3D (red circles). The vertical dashed line indicates the $L_c = 2r_{\min} + r_c$ threshold. (B, D) First-passage time distributions for $C:S \rightarrow C:S$ (B) and $C:P \rightarrow C:S$ (D) when $L_c/3r_{\min} = 1.02$, for simulations in two (gray squares) and three (black circles) dimensions. The solid lines are exponential distributions with rate $k = 1/T_{C:S \rightarrow C:S}$ and $k = 1/T_{C:P \rightarrow C:S}$ taken from (A) and (C), respectively, for the corresponding geometry.

VII. FIRST-PASSAGE TIMES TO BIND AND REFORM THE SUBSTRATE BOND

The $C:S \rightarrow C:S$ transition is diffusion-limited and only depends on the catalyst geometry. To measure the mean first-binding time associated to it, we initiate the system in a random configuration corresponding to state $C:S$ and record the time it takes to reach $C:S$ for the first time, i.e., the time for the free particle in the substrate to attach to the catalyst. The average first-binding time diverges as $L_c \rightarrow 3r_{\text{cutoff}}$, as shown for a system in 2D and 3D in ESI Fig. 5A. This gives rise to the $L_c/3r_{\min} < r_{\text{cutoff}}/r_{\min}$ necessary condition for catalysis and impacts the trade-off for optimal catalyst geometry. Indeed, the further apart the two particles in the catalyst, the longer it will take the substrate to fully bind. We note that when $L_c/3r_{\min} \gg 1$ (narrow escape limit), the first-binding time distribution is dominated by an exponential (ESI Fig. 5B).

The $C:P \rightarrow C:S$ transition is similar to the $C:S \rightarrow C:S$ process: it is also limited by diffusion and only depends on the catalyst geometry L_c . We provide the average first-time to reform the substrate bond, and first-passage time distribution in ESI Fig. 5C and D. These are obtained by initiating the simulations in $C:P$ and recording the time it takes to reform the substrate bond, i.e., reach $C:S$ for the first time.

VIII. MARKOV STATE MODELS

We use Markov State Models (MSMs) to extend our results beyond the parameter range that we can explore with MD simulations. We first construct a minimal model to account for the system when free monomers are systematically removed once they are released in solution. The structure of this model is shown in Fig. 3A in the main text, and it is used to produce Fig. 3B (black lines). We then extend the model to account for the system when free monomers are removed only if they diffuse sufficiently far from the catalyst, i.e., $r > R_{\text{Diff}}$. This extended model is used to produce the results in Fig. 3C and Fig. 3D in the main text.

A. Minimal MSM

The minimal MSM, depicted in Fig. 3A in the main text, consists of six states, $C+S$, $C\cdot S$, $C:S$, $C:P$, $C\cdot P$ and $C+P$, where each state is characterized by the number and the types of bonds (see configurations in Fig. 1C in the main text).

We first infer transition rates from numerical simulations by coarse-graining the MD trajectories into discrete states. The rates are given by $k_{ij} = p_{ij}/\tau_i$, where k_{ij} is the rate from state i to state j , τ_i is the average time the system stays at state i , and p_{ij} is the jump probability from i to j [11]. To extract τ_i and p_{ij} from simulations, we initiate the system in state $C+S$ and sample the system every $\tau_{\text{Lag}} = 50$ time units recording the formation and breaking of bonds, where τ_{Lag} is chosen sufficiently large to ignore barrier recrossings (e.g. immediate reformation of a bond after breaking) while still allowing us to resolve the states along the catalytic pathway. This procedure leads to the merging of states $C:S$ and $C:P$. Transition probabilities are then computed by measuring the average transition frequency between states [12]. Examples of waiting time distributions, transition probabilities and simulation-inferred rates are shown in ESI Fig. 6.

To explore the parameter space beyond simulations, we next construct an analytical model for the rates by classifying the transitions in the MSM into escape and diffusive processes.

1. Escape processes

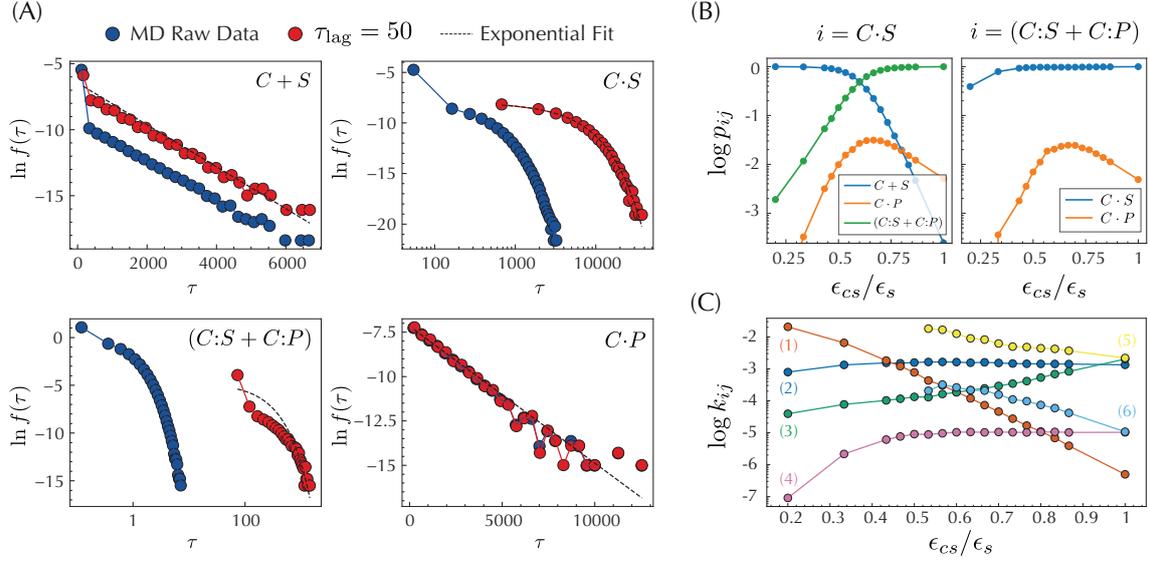
We describe bond breaking events as barrier escape problems with Arrhenius-like expression for the rate of the transition, $k(\epsilon) = e^{-A\epsilon+B}$ with $A = 0.91$ and $B = 2.20$ (ESI Fig. 1). The broken bond for transitions $C\cdot S \rightarrow C+S$ and $C\cdot P \rightarrow C+P$ corresponds to ϵ_{cs} , and hence, the rate is $k(\epsilon_{cs})$. The same bond is broken during the $C:P \rightarrow C\cdot P$ transition, but $k_{C:P \rightarrow C\cdot P} \approx 2k_{C\cdot P \rightarrow C+P}$ as any of the two monomers attached to the catalyst can be released independently. The bond broken during the $C\cdot S \rightarrow C\cdot P$ transition is the scissile bond in the substrate, and therefore, the rate is $k(\epsilon_s)$. The barrier for $C:S \rightarrow C:P$ is described by (ESI Eq. 3). The barrier for the $C:S \rightarrow C\cdot S$ transition is approximated by

$$\Delta U_{C:S \rightarrow C\cdot S} = U(r_{\min}, \epsilon_s) + U(r_{\min}, \epsilon_{cs}) - \min_{r_s^*, r_{cs}^*} \left[U(r_s, \epsilon_s) + U(r_{cs}, \epsilon_{cs}) + U(L_c - r_s - r_{cs}, \epsilon_{cs}) \right], \quad (\text{ESI Eq. 4})$$

where r_s represents the distance between the particles in the substrate and r_{cs} the substrate-catalyst particle distance. The calculations for $\Delta U_{C:S \rightarrow C\cdot P}$ and $\Delta U_{C:S \rightarrow C\cdot S}$ assume that $C:S$ is strictly one dimensional, which we have shown is true within the range of L_c values necessary for catalysis (ESI Fig. 2).

2. Diffusive processes

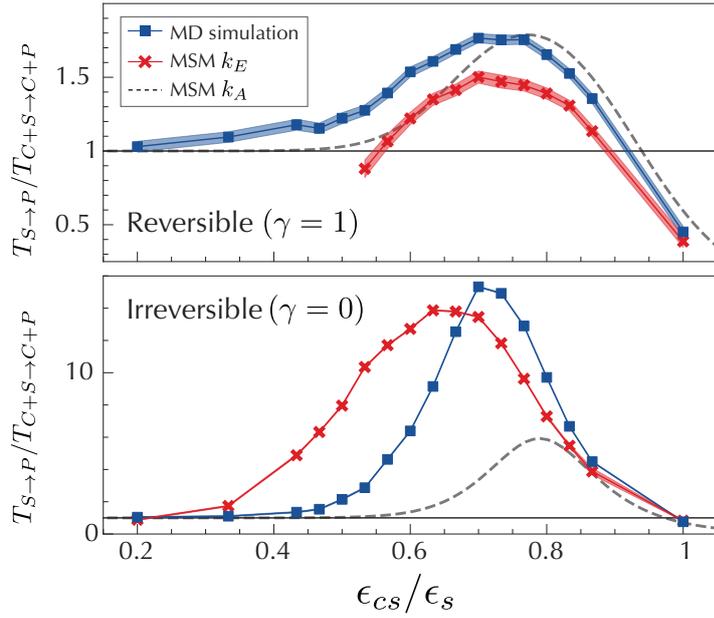
Transitions $C+S \rightarrow C\cdot S$, $C\cdot S \rightarrow C:S$ and $C:P \rightarrow C:S$ are limited by diffusion. While the first transition depends on the volume of the system, the two latter are intrinsic to the catalyst design and only depend on L_c . We verify that in the narrow escape limit, i.e., $L_c/3r_{\min} > 1$, the first-passage time distributions for $C\cdot S \rightarrow C:S$ and $C:P \rightarrow C:S$ are described by a single timescale (ESI Fig. 4). To interpolate between the reversible and irreversible cases, we introduce a parameter γ , such that $\tilde{k}_{C:P \rightarrow C:S} = \gamma k_{C:P \rightarrow C:S}$, where $\gamma = 1$ accounts for the reversible (diffusion-limited) case and $\gamma = 0$ accounts for the irreversible case. The first-passage time distribution associated with the $C+S \rightarrow C\cdot S$ transition is not exponential and depends on the initial distance of the substrate with respect to the catalyst [13]. To estimate a rate, we map the transition to a search process in a disk of radius R_{Diff} with reflecting boundary and an absorbing trap with radius $r = r_{\text{cutoff}}$ in the center. To compute the mean first-passage time to reach the absorbing trap starting from the reflecting boundary, we solve $D\Delta t(r) = -1$, where D is the diffusion constant, with boundary conditions $\nabla t(r)|_{r=R_{\text{Diff}}} = 0$ and $t(r = r_{\text{cutoff}}) = 0$. We use the inverse of the mean first-passage time $k_{C+S \rightarrow C\cdot S} = 1/t(R_{\text{Diff}})$ as the rate. We note that the necessary and sufficient conditions for catalysis do not depend on this particular transition, and hence, the value we set for this rate only impacts the catalyst's efficiency and not the regions where catalysis emerges. In particular, the maximal catalytic efficiency, depicted in Fig 3B in the main text, is obtained in the limit when $k_{C+S \rightarrow C\cdot S} \gg 1$.



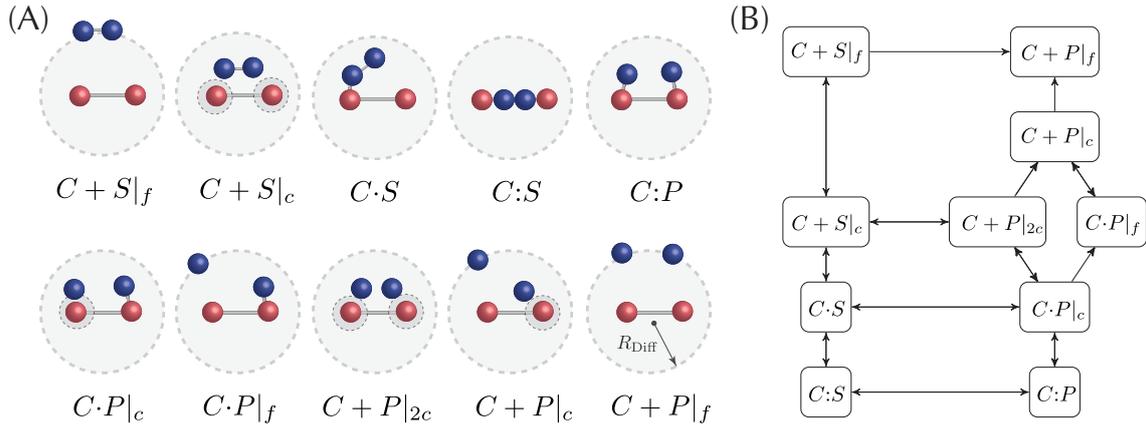
ESI Figure 6. (A) Waiting time distributions for states $C+S$, $C\cdot S$, $C\cdot P$ and the merged state $(C:S+C:P)$, obtained by recording the time it takes the system to form and break bonds (blue) or by coarse-graining the MD trajectories with a lag time $\tau_{\text{lag}} = 50$ time units, which allows us to fit an exponential distribution to the data (dashed line) and to extract the average waiting time $\langle \tau_i \rangle$. Simulation parameters are $\epsilon_s = 15$ and $\epsilon_{cs} = 10.5$. Results correspond to the reversible limit. (B) Transition probabilities p_{ij} from states $i = C\cdot S$ and $i = (C:S + C:P)$ as a function of the catalyst binding strength ϵ_{cs} . Legend indicates states j where the system transitions next. (C) Transition rates $k_{ij} = p_{ij} / \langle \tau_i \rangle$ as a function of the catalyst binding energy ϵ_{cs} : (1) $k_{C\cdot S \rightarrow C+S}$, (2) $k_{C+S \rightarrow C\cdot S}$, (3) $k_{C\cdot S \rightarrow (C:S+C:P)}$, (4) $k_{C\cdot S \rightarrow C\cdot P}$, (5) $k_{(C:S+C:P) \rightarrow C\cdot S}$ and (6) $k_{(C:S+C:P) \rightarrow C\cdot P}$.

B. Validation

To validate the coarse-graining and discretization of the catalytic path into states, we compare the efficiency of the catalyst design, i.e., $T_{S \rightarrow P} / T_{C+S \rightarrow C+P}$, in simulations, where the catalytic pathway is *not* divided into elementary transitions, to the efficiency using the minimal MSM with rates inferred from simulations and rates from the analytical model. The comparison is shown in ESI Fig. 7 for the reversible ($\gamma = 1$) and irreversible ($\gamma = 0$) cases. The agreement in the reversible case supports the coarse-graining of the system into states. The way in which data is generated for the irreversible case (see ESI section VI) might be one of the reasons why the results do not quantitatively agree for small ϵ_{cs} for the MSM with the simulation inferred rates. Although the analytical model for the rates underestimates the efficiency of the catalyst in the irreversible case, the overall scaling agreement between the MSM and simulation data in Fig. 3B in the main text for both reversible and irreversible cases further support the coarse-graining.



ESI Figure 7. Comparison of the catalytic efficiency, i.e., $T_{S \rightarrow P} / T_{C+S \rightarrow C+P}$, for a fixed substrate bond ($\epsilon_s = 15$) and catalyst geometry ($L_c / 3r_{\min} = 1.02$) and different ϵ_{cs} for numerical simulations (blue points), the MSM with the rates inferred from simulation (red crosses, k_E) and the MSM using the analytical rates (dashed black line, k_A). Top panel corresponds to the reversible case ($\gamma = 1$), while the bottom panel corresponds to the irreversible case ($\gamma = 0$). In all cases, monomers are removed as soon as they are released in the solution. Simulation results correspond to a box with side $L/\sigma = 15$.



ESI Figure 8. (A) States considered in the extended MSM, when product monomers (blue) are not immediately removed after they are released in solution and have to diffuse a distance $r > R_{\text{Diff}}$ from the center of the system, where the catalyst is placed (red), to leave it. We now consider that the substrate and product particles can be far (f) or close (c) to the catalyst. (B) Extension of the MSM in Fig. 3A in the main text to account for the diffusion of the monomers away from the catalyst.

C. Extended MSM to account for volume

To explore the impact of product inhibition and other volume effects on catalysis, we require monomers to diffuse sufficiently far away from the catalyst, i.e., $r > R_{\text{Diff}}$, in order to leave the system, where R_{Diff} is the radius of a disk (sphere) centered around the catalyst (see ESI Fig. 8A). We extend the MSM in Fig 3A in the main text to 10 states and introduce transitions that describe the substrate and product particles diffusing towards or away from the catalyst. To estimate a rate for these transitions, we take the inverse of the mean first-passage time of a search process in a bounded domain. We compute the mean first-passage time by solving $D\Delta t(r) = -1$ with the appropriate boundary conditions, where D is the diffusion constant [14]. For example, we estimate the mean first-passage time for the $C + S|_f \rightarrow C + S|_c$ transition, where the subscripts f and c indicate that the substrate is 'far' ($r = R_{\text{Diff}}$) and 'close' ($r = r_{\text{cutoff}} + \delta$, with $\delta \ll 1$) from the catalyst, by mapping the transition to a diffusion process in a disk (sphere) with boundary conditions $t(r = r_{\text{cutoff}} + \delta) = 0$ and $\nabla t(r)|_{r=R_{\text{Diff}}} = 0$. The first condition represents an absorbing target of radius $r = r_{\text{cutoff}} + \delta$ in the center of the domain and the second one accounts for the reflecting boundary at $r = R_{\text{Diff}}$. The mean first-passage time will depend on the starting point of the process $t = t(r_0)$, which we set at $r_0 = R_{\text{Diff}}$, corresponding to state $C + S|_f$. We estimate the first return time for $C + S|_c \rightarrow C \cdot S$ by assuming a similar setup, but setting the starting point of the process at $r_0 = r_{\text{cutoff}} + \delta$ and absorbing boundary at $r = r_{\text{cutoff}}$. Transitions $C \cdot P|_c \rightarrow C \cdot P$, $C + P|_{2c} \rightarrow C \cdot P|_c$ and $C + P|_c \rightarrow C \cdot P|_f$ also fall within this category. We note that results depend on the choice of δ , and we use $\delta = 0.01$ to produce Fig. 3C and D in the main text.

For transitions requiring the monomers to diffuse away from the catalyst, such as $C \cdot P|_c \rightarrow C \cdot P|_f$ and $C + P|_{2c} \rightarrow C + P|_c$, we consider $t(r = R_{\text{Diff}}) = 0$ and $\nabla t(r)|_{r=r_{\text{cutoff}}} = 0$ as boundary conditions, and set the starting point of the process at $r_0 = r_{\text{cutoff}}$. We note that despite our procedure to estimate the rates, the first-passage time distributions for these processes are not exponential [13, 15], and therefore, a rate cannot be properly defined. The comparison between MD simulation data in Fig. 3C in the main text, and ESI Fig. 10 below show qualitative agreement. Further improvements of the model regarding diffusive processes are left as future work.

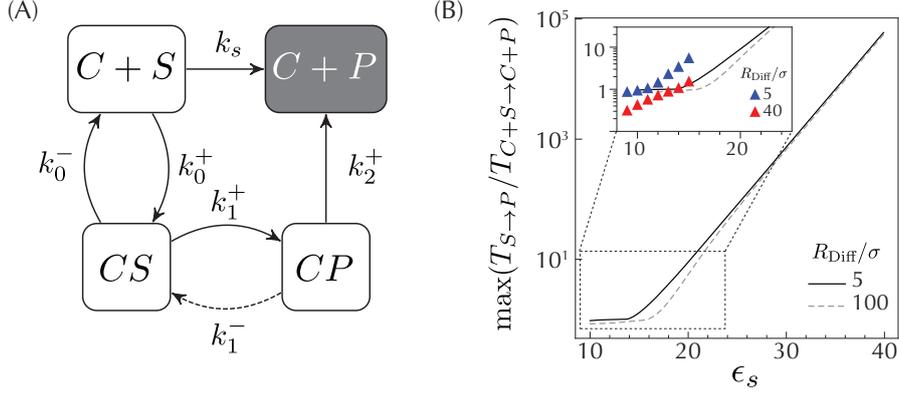
IX. NECESSARY CONDITIONS FOR CATALYSIS.

We derive the necessary conditions for catalysis when monomers are systematically removed from the system (MSM in Fig. 3A in the main text) by comparing the mean first-passage time from $C + S$ to $C + P$, i.e., $T_{C+S \rightarrow C+P}$, to the mean reaction time in the absence of catalyst, $T_{S \rightarrow P} = 1/k_{S \rightarrow P}$. $T_{C+S \rightarrow C+P}$ can be analytically computed [16] and the criterion for catalysis, $T_{S \rightarrow P}/T_{C+S \rightarrow C+P} > 1$, takes the form

$$\frac{1}{k_{S \rightarrow P}} > \frac{1}{k_{C:S \rightarrow C:P}} + \frac{1}{k_{C:P \rightarrow C \cdot P}} + \frac{1}{k_{C \cdot P \rightarrow C+P}} + \frac{k_{C:P \rightarrow C:S}}{k_{C:S \rightarrow C:P} k_{C:P \rightarrow C \cdot P}} + \frac{k_{S \rightarrow P}}{k_{C:S \rightarrow C:S} k_{C \cdot P \rightarrow C+P}} \times \left(1 + \frac{k_{C:S \rightarrow C:S}}{k_{C:S \rightarrow C:P}} \frac{k_{C:P \rightarrow C:S}}{k_{C:P \rightarrow C \cdot P}} + \frac{k_{C:S \rightarrow C:S}}{k_{C:S \rightarrow C:P}} \right), \quad (\text{ESI Eq. 5})$$

where $k_{i \rightarrow j}$ is the rate to transition from state i to state j . The above equation is a sufficient condition for catalysis of the form $T_{S \rightarrow P} > \sum_n T_n$. Each individual term on the right hand side of the equation leads to a necessary condition for catalysis that can be subsequently translated into physical and geometrical constraints in our model.

The first necessary condition in (ESI Eq. 5), $k_{S \rightarrow P} < k_{C:S \rightarrow C:P}$, pertains to the catalytic mechanism and leads to $L_c > 3r_{\text{min}}$ and $\epsilon_{cs} > 0$. The second and third necessary conditions in (ESI Eq. 5), $k_{S \rightarrow P} < k_{C:P \rightarrow C \cdot P}$ and $k_{S \rightarrow P} < k_{C \cdot P \rightarrow C+P}$, are associated to product release and lead to $\epsilon_{cs} < \epsilon_s$. The fourth necessary condition in (ESI Eq. 5) is non trivial only in the reversible case, when $C:P \rightarrow C:S$ is possible. In the irreversible case, this transition is not allowed and hence, $k_{C:P \rightarrow C:S} = 0$. The last necessary condition on the right hand side in equation (ESI Eq. 5) stems from the alternative pathway that the substrate may take in the presence of the catalyst to produce products, i.e. $C \cdot S \rightarrow C + P$, without visiting $C:S$. This shortcut to the final state of the reaction relaxes the constraint on $k_{C:S \rightarrow C:S}$, which does not have to be as fast as the spontaneous reaction for catalysis to emerge.



ESI Figure 9. (A) Minimal reaction scheme that we use to explain the scaling of the maximal catalytic efficiency in the reversible and irreversible limits. (B) MSM model results for the maximal catalytic efficiency when $\gamma = 0$ and monomers are removed from the system if $r > R_{\text{Diff}}$, for $R_{\text{Diff}}/\sigma = 5$ (black solid line) and $R_{\text{Diff}}/\sigma = 100$. (Inset) Zoom in the small ϵ_s region showing MD simulation data for $R_{\text{Diff}}/\sigma = 5$ (blue triangles) and $R_{\text{Diff}}/\sigma = 40$ (red triangles).

Note that (ESI Eq. 5) is independent of $T_{C+S \rightarrow C.S}$, the average time it takes the substrate to diffuse to the catalyst. This is because the substrate can always dissociate spontaneously into two monomers in a single step in the presence of the catalyst. As here we consider that the reaction ends when two monomers have been released into solution (systematic product removal), any catalyst design that successfully accelerates the production of monomers will only contribute to reducing the mean reaction time. In other words, as long as the product is systematically removed, catalysis is an intrinsic property of our catalyst design, and therefore independent of the volume of the system. The efficiency of the catalyst, however, depends on volume.

X. SCALING OF THE MAXIMAL EFFICIENCY

To understand why the maximal catalytic efficiency scales exponentially with ϵ_s in the irreversible case and why it saturates in the reversible case (Fig 3B in the main text), we propose a minimal reaction scheme that captures the essential features of our model. This scheme is shown in Fig. 9 and it contains four states, where the $CS \rightarrow CP$ transition accounts for the chemical transformation step and $CP \rightarrow C + P$ represents product release. We model the rates as

$$k_s = e^{-\epsilon_s} \quad k_1^+ = k_s e^{\alpha \epsilon_{cs}} \quad k_2^+ = e^{-\epsilon_{cs}} \quad k_1^- = \gamma k_{CP \rightarrow CS}, \quad (\text{ESI Eq. 6})$$

and we leave k_0^+ and k_0^- unspecified as they do not take part in the sufficient condition for catalysis (see ESI section IX). Here α represents the catalyst's ability to reduce the rate of the spontaneous reaction, $k_{CP \rightarrow CS}$ is the diffusion-limited rate of the reverse reaction in the catalyst and γ is the parameter that regulates such rate. Note that for fixed ϵ_s , the larger ϵ_{cs} , the larger k_1^+ will be, but the smaller k_2^+ , recovering the Sabatier principle of optimal intermediate binding strength. To make the notation compact, we use $\gamma k_{CP \rightarrow CS} \equiv \tilde{\gamma}$. Catalysis requires $\alpha > 0$ and $\epsilon_{cs} < \epsilon_s$. When $\alpha = 1$, the catalyst cancels the barrier of the spontaneous reaction.

We determine the efficiency of the catalyst by computing the mean first-passage time to reach the product state in the presence and absence of the catalyst, which yields

$$\frac{T_{S \rightarrow P}}{T_{C+S \rightarrow C+P}} = \frac{k_0^+ k_1^+ k_2^+ + k_1^+ k_2^+ k_s + k_0^- (k_1^- + k_2^+) k_s}{[k_1^+ k_2^+ + k_0^- (k_1^- + k_2^+) + k_0^+ (k_1^- + k_1^+ + k_2^+)] k_s}. \quad (\text{ESI Eq. 7})$$

A. Irreversible case: exponential scaling

In the limit when $k_0^+ \rightarrow \infty$ and $\tilde{\gamma} = 0$, the mean-first passage time to produce two product monomers in the presence of the catalyst is

$$T_{C+S \rightarrow C+P} = \frac{1}{k_1^+} + \frac{1}{k_2^+} = e^{\epsilon_{cs}} + e^{\epsilon_s - \alpha \epsilon_{cs}}, \quad (\text{ESI Eq. 8})$$

which is minimized by

$$\epsilon_{cs}^* = \frac{\epsilon_s + \log \alpha}{1 + \alpha}. \quad (\text{ESI Eq. 9})$$

As a result, the maximal catalytic efficiency for $\epsilon_s \rightarrow \infty$ scales as

$$\lim_{\epsilon_s \rightarrow \infty} \left(\frac{T_{S \rightarrow P}}{T_{C+S \rightarrow C+P}} \right) \Big|_{\epsilon_{cs} = \epsilon_{cs}^*} = \frac{1}{1 + \alpha} e^{\frac{1}{1+\alpha}(\epsilon_s + \log \alpha)} \sim e^{\tilde{\alpha} \epsilon_s}, \quad (\text{ESI Eq. 10})$$

recovering the scaling in Fig. 3B in the main text for the irreversible case, where $\tilde{\alpha} = (1 + \alpha)^{-1}$. Note that when $\alpha = 1$, the optimal binding strength $\epsilon_{cs}^* = \epsilon_s/2$ and $\tilde{\alpha} = 0.5$. For our MD results in Fig 3B, $\alpha_{MD} = 0.50 \pm 0.04$.

B. Reversible case: Saturation

When $\tilde{\gamma} \neq 0$, $T_{C+S \rightarrow C+P}$ in the limit when $k_0^+ \rightarrow \infty$ is given by

$$T_{C+S \rightarrow C+P} = \frac{1}{k_1^+} + \frac{1}{k_2^+} + \frac{k_1^-}{k_1^+ k_2^+} = e^{\epsilon_{cs}} + e^{\epsilon_s - \alpha \epsilon_{cs}} + \tilde{\gamma} e^{(1-\alpha)\epsilon_{cs} + \epsilon_s}, \quad (\text{ESI Eq. 11})$$

and the catalytic efficiency is

$$\frac{T_{S \rightarrow P}}{T_{C+S \rightarrow C+P}} = \frac{e^{\epsilon_s}}{e^{\epsilon_{cs}} + e^{\epsilon_s - \alpha \epsilon_{cs}} + \tilde{\gamma} e^{(1-\alpha)\epsilon_{cs} + \epsilon_s}} = \frac{1}{e^{\epsilon_{cs} - \epsilon_s} + e^{-\alpha \epsilon_{cs}} + \tilde{\gamma} e^{(1-\alpha)\epsilon_{cs}}} \quad (\text{ESI Eq. 12})$$

Since catalysis requires $\epsilon_s > \epsilon_{cs}$, in the limit when $\epsilon_s \rightarrow \infty$,

$$\frac{T_{S \rightarrow P}}{T_{C+S \rightarrow C+P}} = \frac{1}{e^{-\alpha \epsilon_{cs}} + \tilde{\gamma} e^{(1-\alpha)\epsilon_{cs}}}. \quad (\text{ESI Eq. 13})$$

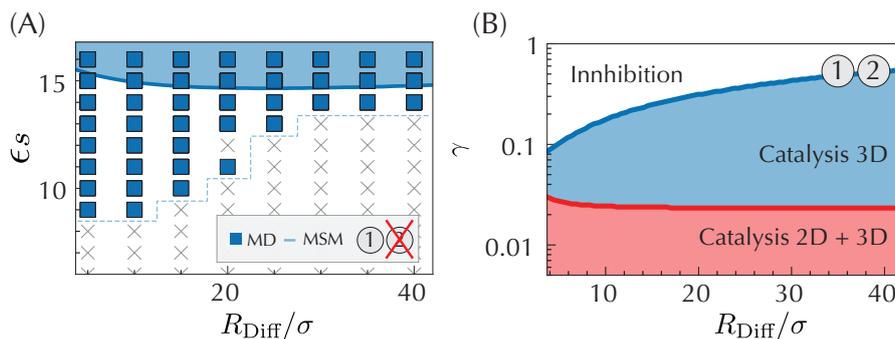
The catalytic efficiency is maximal when the denominator is the smallest, which is minimized when $\epsilon_{cs} = 0$. As a result, the maximal efficiency saturates,

$$\frac{T_{S \rightarrow P}}{T_{C+S \rightarrow C+P}} = \frac{1}{e^{-\alpha \epsilon_{cs}} + \tilde{\gamma} e^{(1-\alpha)\epsilon_{cs}}} \Big|_{\epsilon_{cs}=0} = \frac{1}{1 + \tilde{\gamma}}, \quad (\text{ESI Eq. 14})$$

in agreement with the cases for $\gamma \neq 0$ in Fig. 3B in the main text.

XI. EMERGENCE OF CATALYSIS IN 3D

In ESI Fig. 10, we produce Figs. 3C and D in the main text for a system in 3D. To construct ESI Fig. 10A, we simulate the system at varying ϵ_s and R_{Diff} , with fixed catalyst geometry $L_c/3r_{\text{min}} = 1.02$ and we explore a range of ϵ_{cs} that satisfy $\epsilon_{cs} < \epsilon_s$. We consider that catalysis emerges for a given ϵ_s and R_{Diff} if there is at least one ϵ_{cs} for which $T_{S \rightarrow P}/T_{C+S \rightarrow C+P} > 1$. We note that since our resolution in binding energy is $\Delta \epsilon_{cs}/k_B T = 1$, we may miss designs for which the criterion for catalysis is satisfied. The results qualitatively agree with the MSM (blue region). MD results in 3D indicate that catalysis emerges at smaller values of ϵ_s than in the 2D case, i.e., $\epsilon_{s,\text{min}}^{(3D)} < \epsilon_{s,\text{min}}^{(2D)}$. In ESI Fig. 10B, we show MSM results where catalysis was observed for a range of $\epsilon_s \in [2, 30]$.



ESI Figure 10. Simulation and MSM data for the system in 3D. (A) Substrate bonds ϵ_s for which there is catalysis when monomers are removed from the system if they diffuse sufficiently far from the catalyst, i.e., $r > R_{\text{Diff}}$. Simulation data is shown as squares and crosses and the MSM results are shown as a shaded blue region. (B) MSM results showing values of γ for which catalysis can be observed in 2D (red region) and 3D (red and blue regions) when monomers are removed from the system if they have diffused a distance R_{Diff}/σ from the catalyst.

-
- [1] Xipeng Wang, Simón Ramírez-Hinestrosa, Jure Dobnikar, and Daan Frenkel. The lennard-jones potential: when (not) to use it. *Phys. Chem. Chem. Phys.*, 22:10624–10633, 2020.
 - [2] Robert D. Groot, Timothy J. Madden, and Dominic J. Tildesley. On the role of hydrodynamic interactions in block copolymer microphase separation. *The Journal of Chemical Physics*, 110(19):9739–9749, 1999.
 - [3] Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.
 - [4] Reinhard Müller, Peter Talkner, and Peter Reimann. Rates and mean first passage times. *Physica A: Statistical Mechanics and its Applications*, 247(1-4):338–356, 1997.
 - [5] Olivier Rivoire. Geometry and flexibility of optimal catalysts in a minimal elastic model. *The J. of Phy. Chem. B*, 124(5):807–813, 2020.
 - [6] Linus Pauling. Molecular architecture and biological reactions. *Chemical and Engineering News*, 24(10):1375–1377, 1946.
 - [7] Jason D. Forster, Heeso Noh, Seng Fatt Liew, Vinodkumar Saranathan, Carl F. Schreck, Lin Yang, Jin-Gyu Park, Richard O. Prum, Simon G. J. Mochrie, Corey S. O’Hern, Hui Cao, and Eric R. Dufresne. Biomimetic Isotropic Nanostructures for Structural Coloration. *Advanced Materials*, 22(26-27):2939–2944, 2010.
 - [8] Lingling Shui, E. Stefan Kooij, Daniël Wijnperlé, Albert van den Berg, and Jan C. T. Eijkel. Liquid crystallography: 3d microdroplet arrangements using microfluidics. *Soft Matter*, 5:2708–2712, 2009.
 - [9] Sung Yong Park, Abigail K. R. Lytton-Jean, Byeongdu Lee, Steven Weigand, George C. Schatz, and Chad A. Mirkin. Dna-programmable nanoparticle crystallization. *Nature*, 451(7178):553–556, 2008.
 - [10] Tatiana Schmatko, Behnaz Bozorgui, Nienke Geerts, Daan Frenkel, Erika Eiser, and Wilson C. K. Poon. A finite-cluster phase in -DNA-coated colloids. *Soft Matter*, 3(6):703–706, 2007.
 - [11] Christof Schütte, Frank Noé, Jianfeng Lu, Marco Sarich, and Eric Vanden-Eijnden. Markov state models based on milestoning. *The Journal of Chemical Physics*, 134(20):204105, 2011.
 - [12] B. R. Bhat. Some Properties of Regular Markov Chains. *The Annals of Mathematical Statistics*, 32(1):59 – 71, 1961.
 - [13] Denis S Grebenkov, Ralf Metzler, and Gleb Oshanin. Full distribution of first exit times in the narrow escape problem. *New Journal of Physics*, 21(12):122001, 2019.
 - [14] Sidney Redner. *A Guide to First Passage Processes*. Cambridge University Press, 2001.
 - [15] Carlos Mejía-Monasterio, Gleb Oshanin, and Grégory Schehr. First passages for a search by a swarm of independent random searchers. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(06):P06022, jun 2011.
 - [16] Iyer-Biswas and Anton Zilman. First-Passage Time Events in Biology. *Advances in Chemical Physics*, 2016.