A Derivations of statistical estimators

A.1 Conformation probability

The probability that, under Boltzmann equilibrium conditions, a DNA complex will when sampled adopt a sequence-level conformation corresponding to the $i$th domain-level conformation is estimated empirically by sampling sequence-level conformations from the Boltzmann equilibrium distribution and tabulating the number of such conformations corresponding to the domain-level conformation in question. This estimate is computed using two pieces of information: the total number of samples, $N$, and the number of samples corresponding to the domain-level conformation, $N_i$.

The Bayesian estimate for the conformation is appropriate because the maximum likelihood approach produces misleading results when $N_i = 0$ or $N$, which can occur when $N$ is small compared to $\frac{1}{p_i}$ or $\frac{1}{1-p_i}$, respectively.

We use a uniform prior distribution on $p_i$, so $P(p_i) = 1$ for all $p_i$. Thus, we have,

$$\hat{p}_i = E[p_i|N, N_i] = \int_0^1 p_i P(p_i|N, N_i)dp_i,$$

where $E[p_i|N, N_i]$ is the expectation of $p_i$ given the observed values of $N$ and $N_i$ and $P(p_i|N, N_i)$ is the probability mass function of $p_i$ given the same.

We can compute $P(p_i|N, N_i)$ exactly using Bayes’s law and combinatorial techniques. Bayes’s law states that

$$P(p_i|N, N_i) = \frac{P(N, N_i|p_i)P(p_i)}{P(N, N_i)}$$

where $P(p_i) = 1$ and $P(N, N_i) = \int_0^1 P(N, N_i|p_i)dp_i$. We know that

$$P(N, N_i|p_i) = \binom{N}{N_i} p_i^{N_i} (1-p_i)^{N-N_i}$$

and so

$$P(N, N_i) = \binom{N}{N_i} \int_0^1 p_i^{N_i} (1-p_i)^{N-N_i}dp_i = \binom{N}{N_i} B(N_i+1, N-N_i+1)$$

where $B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$ is the beta function. Thus, we have

$$P(p_i|N, N_i) = \frac{p_i^{N_i} (1-p_i)^{N-N_i}}{B(N_i+1, N-N_i+1)}.$$
\[ \hat{p}_i = \int_0^1 p_i \mathbb{P}(p_i | N_i, N) \, dp_i \]

\[ = \frac{1}{B(N_i + 1, N - N_i + 1)} \int_0^1 p_i^{N_i+1}(1 - p_i)^{N-N_i} \, dp_i \]

\[ = \frac{B(N_i + 2, N - N_i + 1)}{B(N_i + 1, N - N_i + 1)} \]

\[ = \frac{(N_i + 1)! (N - N_i)!}{(N + 2)!} \frac{(N + 1)!}{N!(N - N_i)!} \]

\[ = \frac{N_i + 1}{N + 2} \]

where we use the fact that \(B(\alpha, \beta) = \frac{(\alpha - 1)! (\beta - 1)!}{(\alpha + \beta - 1)!}\) when \(\alpha\) and \(\beta\) are positive integers. So we have the estimator for conformation probability \(p_i\):

\[ \hat{p}_i = \frac{N_i + 1}{N + 2}. \]

### A.2 Conformation probability error

As a measure of the confidence in the estimate for \(p_i\), we compute the standard deviation of \(p_i\) given the observed data \(N\) and \(N_i\). We first compute the \textit{a posteriori} variance of \(p_i\) to be:

\[ \text{Var}(p_i | N, N_i) = \mathbb{E}[p_i^2 | N, N_i] - \mathbb{E}[p_i | N, N_i]^2 \]

\[ = \int_0^1 p_i^2 \mathbb{P}(p_i | N, N_i) \, dp_i - \hat{p}_i^2 \]

\[ = \frac{1}{B(N_i + 1, N - N_i + 1)} \int_0^1 (p_i^{N_i+2}(1 - p_i)^{N-N_i}) \, dp_i - \left( \frac{N_i + 1}{N + 2} \right)^2 \]

\[ = \frac{B(N_i + 3, N - N_i + 1)}{B(N_i + 1, N - N_i + 1)} \left( \frac{N_i + 1}{N + 2} \right)^2 \]

\[ = \frac{(N_i + 2)! (N - N_i)!}{(N + 3)!} \frac{(N + 1)!}{N!(N - N_i)!} - \left( \frac{N_i + 1}{N + 2} \right)^2 \]

\[ = \frac{(N_i + 1)(N_i + 2)}{(N + 2)(N + 3)} \left( \frac{N_i + 1}{N + 2} \right)^2 \]

\[ = \frac{(N_i + 1)(N - N_i + 1)}{(N + 2)^2(N + 3)} \]

where we again use the fact that \(B(\alpha, \beta) = \frac{(\alpha - 1)! (\beta - 1)!}{(\alpha + \beta - 1)!}\) when \(\alpha\) and \(\beta\) are positive integers.

The standard deviation is the square root of the variance, which is given by

\[ \hat{\sigma}_{p_i} = \sqrt{\frac{(N_i + 1)(N - N_i + 1)}{(N + 2)^2(N + 3)}} = \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{N + 3}} \]

### A.3 \(k_1\) estimate for bimolecular reactions

For the first-step model reaction

\[ A + B \xrightarrow{k_1} AB \]

we estimate \(k_1\) using Bayesian inference on the observed simulated sequence-level reaction trajectories. Specifically, assume we observe \(N\) Multistrand trajectories, each of which begins with two conformations.
of $A$ and $B$ sampled from the Boltzmann distribution of secondary structures. Each trajectory is characterized by an indicator variable $S^n_k$, which is 1 if and only if the final product multiset $P_n$ equals $P_i$, and its collision rate constant $k^n_{\text{coll}}$, which is computed as the net rate of forming any initial base pair between $A$ and $B$ in the first step of the simulation. For simplicity in this and the following sections, in the context of estimating $k^n_1$ and $k^n_2$, we refer to trajectories in which $S^n_k = 1$ as successful trajectories, and all others as unsuccessful.

The rate constant $k^n_1$ represents the net rate with which $A$ and $B$ will collide in a CME trajectory leading to $P_i$, so that

$$k^n_1 = E[S^n_k | E[k^n_{\text{coll},t} | S^n_k = 1] = p_t k^n_{\text{coll},t}$$

where $p_t = E[S^n_k]$ is the probability of simulating a successful FSM trajectory and $k^n_{\text{coll},t}$ is the average value of $k^n_{\text{coll}}$ over successful FSM trajectories.

The estimator $\hat{k}_1^n$ is defined as

$$\hat{k}_1^n = E[k^n_1 | P, k_{\text{coll}}] = E[p_t k^n_{\text{coll},t} | P, k_{\text{coll}}]$$

where $P = (P_1, P_2, \ldots, P_N)$ and $k_{\text{coll}}$ are the vectors of product multisets $P_i$ and collision rates $k^n_{\text{coll}}$, respectively, for each of the $N$ observed trajectories. Note that in contrast to $E[S^n_k]$, which is an expectation taken over a particular distribution of trajectories, $p_t$ and $k^n_{\text{coll},t}$ are underlying parameters of the system and do not vary between trajectories. $E[p_t k^n_{\text{coll},t}]$ refers to an expectation taken over the space of possible $(p_t, k^n_{\text{coll},t})$, which is $[0, 1] \times [0, \infty)$. This expectation is only well-defined with priors for each random variable, described below.

To make the algorithm that follows more tractable, we make the simplifying assumption that $p_t$ and $k^n_{\text{coll},t}$ are independent random variables, so that

$$\hat{k}_1^n = E[p_t | P, k_{\text{coll}}] E[k^n_{\text{coll},t} | P, k_{\text{coll}}].$$

It remains to compute the conditional expectations $E[p_t | P, k_{\text{coll}}]$ and $E[k^n_{\text{coll},t} | P, k_{\text{coll}}]$.

Observe that $p_t$ is a random variable of the same form as a conformation probability (see Section A.1). Thus, making a similar assumption of a uniform prior on $p_t$ over $[0, 1]$, we can follow identical steps to estimate the expectation as:

$$E[p_t | P, k_{\text{coll}}] = \frac{N_i + 1}{N + 2}$$

where $N_i$ is the number of trajectories for which $S^n_k = 1$.

To compute the expectation on $k^n_{\text{coll},t}$, we first assume that the individual $k^n_{\text{coll},t}$ values for each FSM trajectory are sampled from an exponential distribution with mean $k^n_{\text{coll},t}$. This is justified by the fact that the exponential distribution maximizes informational entropy, so that choosing this distribution assumes the least amount of prior knowledge about the $k^n_{\text{coll},t}$. In addition, this assumption makes the following math tractable.

By definition, we have,

$$E[k^n_{\text{coll},t} | P, k_{\text{coll}}] = \int_0^\infty k^n_{\text{coll},t} P(k^n_{\text{coll},t} | P, k_{\text{coll}}) dk^n_{\text{coll},t}$$

$$= \int_0^\infty k^n_{\text{coll},t} P(P | k^n_{\text{coll},t}) \frac{P(k^n_{\text{coll},t})}{P(P, k_{\text{coll}})} dk^n_{\text{coll},t}$$

$$= \int_0^\infty k^n_{\text{coll},t} P(P, k_{\text{coll}} | k^n_{\text{coll},t}) P(k^n_{\text{coll},t}) dk^n_{\text{coll},t}$$

where the second equality is due to Bayes’s law and $P(k^n_{\text{coll},t})$ is the prior probability distribution on $k^n_{\text{coll},t}$.

To compute the value of $P(P, k_{\text{coll}} | k^n_{\text{coll},t})$, first observe that all trajectories are independent, so that

$$P(P, k_{\text{coll}} | k^n_{\text{coll},t}) = \prod_{1 \leq n \leq N} P(k^n_{\text{coll} | k^n_{\text{coll},t}}).$$

Because $k^n_{\text{coll},t}$ gives no information about unsuccessful trajectories, any terms due to these trajectories will be canceled out by the normalization term, and we may drop these terms. For successful trajectories,
our assumption regarding the distribution of \( k_{\text{coll}}^n \) (above) gives us:

\[
P(k_{\text{coll}}^n|k_{\text{coll},i}) = \frac{1}{k_{\text{coll},i}} \exp\left(-\frac{k_{\text{coll}}^n}{k_{\text{coll},i}}\right)
\]

so that

\[
P(P, k_{\text{coll}}|k_{\text{coll},i}) = \prod_{S_i^n = 1} \frac{1}{k_{\text{coll},i}} \exp\left(-\frac{\sum_{S_i^n = 1} k_{\text{coll}}^n}{k_{\text{coll},i}}\right)
\]

\[
= \left(\frac{1}{k_{\text{coll},i}}\right)^{N_i} \exp\left(-\frac{\gamma_i}{k_{\text{coll},i}}\right), \text{ where } \gamma_i = \sum_{S_i^n = 1} k_{\text{coll}}^n.
\]

Substituting into the equation for the expectation of \( k_{\text{coll},i} \) yields

\[
E[k_{\text{coll},i}|P, k_{\text{coll}}] = \int_0^\infty \left(\frac{1}{k_{\text{coll},i}}\right)^{N_i - 1} \exp\left(-\frac{\gamma_i}{k_{\text{coll},i}}\right) P(k_{\text{coll},i}) dk_{\text{coll},i}
\]

To allow this expression to be evaluated analytically for all nonnegative \( N_i \), we use the prior distribution

\[
P(k_{\text{coll},i}) = \left(\frac{1}{k_{\text{coll},i}}\right)^3.
\]

Although this is an improper prior over \([0, \infty)\), the form of the integrand allows both the numerator and denominator of \( E[k_{\text{coll},i}|P, k_{\text{coll}}] \) to converge. First, considering the numerator,

\[
\int_0^\infty \left(\frac{1}{k_{\text{coll},i}}\right)^{N_i - 1} \exp\left(-\frac{\gamma_i}{k_{\text{coll},i}}\right) P(k_{\text{coll},i}) dk_{\text{coll},i} = \int_0^\infty \left(\frac{1}{k_{\text{coll},i}}\right)^{N_i + 2} \exp\left(-\frac{\gamma_i}{k_{\text{coll},i}}\right) dk_{\text{coll},i}
\]

\[
= \int_0^\infty \frac{u^{N_i}}{\gamma_i^{N_i + 1}} e^{-u} du, \text{ where } u = \frac{\gamma_i}{k_{\text{coll},i}}
\]

\[
= \frac{N_i!}{\gamma_i^{N_i + 1}}
\]

where in the last equality we use the fact that \( m! = \int_0^\infty x^m e^{-x} dx \) for nonnegative integral \( m \).

Similarly, the denominator can be simplified to

\[
\int_0^\infty \left(\frac{1}{k_{\text{coll},i}}\right)^{N_i + 3} \exp\left(-\frac{\gamma_i}{k_{\text{coll},i}}\right) dk_{\text{coll},i} = \frac{(N_i + 1)!}{\gamma_i^{N_i + 2}}
\]

and we have

\[
E[k_{\text{coll},i}|P, k_{\text{coll}}] = \frac{\gamma_i}{N_i + 1} \frac{N_i + 1}{N + 2} = \frac{\gamma_i}{N + 2}
\]
A.4 $k_1$ error estimate for bimolecular reactions

As with conformation probabilities, the spread in $k_1$ given the observed trajectories can be computed as the standard deviation of the posterior distribution of $k_1$. That is,

\[
\text{Var}(k_1^i | P, k_{\text{coll}}) = E[(k_1^i)^2 | P, k_{\text{coll}}] - E[k_1^i | P, k_{\text{coll}}]^2
\]

\[
= E[(k_{\text{coll},i})^2 | P, k_{\text{coll}}] E[\gamma_i^2 | P, k_{\text{coll}}] - \left( \frac{\gamma_i}{N+2} \right)^2
\]

\[
= \left( \int_0^\infty (k_{\text{coll},i})^2 P(k_{\text{coll},i}) dk_{\text{coll},i} \right) \left( \frac{(N_i + 1)(N_i + 2)}{(N + 2)(N + 3)} \right) - \left( \frac{\gamma_i}{N+2} \right)^2
\]

\[
= \left( \frac{(N_i - 1)!}{\gamma_i^{N_i + 1}} \right) \left( \frac{(N_i + 1)(N_i + 2)}{(N + 2)(N + 3)} \right) - \left( \frac{\gamma_i}{N+2} \right)^2
\]

\[
= \gamma_i^2 \left( \frac{N_i(N_i + 1)}{N_i + 2} \right) \left( \frac{N_i + 2}{N + 2} \right) - \left( \frac{\gamma_i}{N+2} \right)^2
\]

\[
= \gamma_i^2 \left( \frac{2N - N_i + 1}{N_i(N + 2)(N + 3)} \right)
\]

using again the fact that $\int_0^\infty x^n e^{-x} dx = n!$ for integral nonnegative $n$.

The width of the posterior distribution is measured by the standard deviation, or square root of the variance. So we estimate the error of the estimate for $k_1$ to be:

\[
\hat{\sigma}_{k_1} = \sqrt{\gamma_i^2 \left( \frac{2N - N_i + 1}{N_i(N + 2)(N + 3)} \right)} = \hat{k}_1 \sqrt{\frac{2N - N_i + 1}{N_i(N + 3)}}
\]

Note that in the edge case with $N_i = 0$ and $N > 0$ we estimate an upper bound on $k_1$ as:

\[
k_1 \leq \max_n \frac{\gamma_i}{N_i + 2}
\]

and report error bounds as infinite.

A.5 $k_2$ error estimate for bimolecular reactions

Using Equation (6) of the main text, it is apparent that the unimolecular reaction rate $k_2$ is computed as the weighted average of the reaction times, where weights are derived as the $k^{n}_{\text{coll}}$ corresponding to each simulated reaction time $\tau^n_{2,i}$. We first derive the variance of the mean reaction time $\tau_{2,i}$ for the $i$th first-step model reaction, from which the variance of $k_2$ may be computed.

As previously noted, we estimate $\tau_{2,i}$ as the weighted average:

\[
\hat{\tau}_{2,i} = \sum_{n=1}^{S_{\text{coll}}} \frac{k^{n}_{\text{coll}} \tau^n_{2,i}}{S_{\text{coll}}} = \sum_{n=1}^{S_{\text{coll}}} w^n_{i} \tau^n_{2,i}
\]

where $w^n_{i}$ is the normalized weight $\frac{\gamma_i}{\sum_{n=1}^{S_{\text{coll}}} k^{n}_{\text{coll}}}$ for each simulated trajectory.

We assume that all reaction times are drawn from a distribution with mean $\mu_{\tau^n_{2}}$ and variance $\sigma^2_{\tau^n_{2}}$. The variance of the individual $\tau^n_{2}$ differs from the variance of their weighted sum, $\sigma^2_{\tau_{2,i}}$. Assuming fixed weights $w^n_{i}$, this is given by

\[
\sigma^2_{\tau_{2,i}} = \sum_{n=1}^{S_{\text{coll}}} (w^n_{i})^2 \sigma^2_{\tau^n_{2}}.
\]
Thus, we can estimate the variance of the weighted sum from the variance from which the individual reaction times are drawn. Note that since \(w_i^n\) is actually a random variable, our estimates are neglecting this source of variability.

To estimate \(\sigma_{\tau_2}^2\), the variance of the distribution for individual reaction times, we first propose the estimator:

\[
\hat{\sigma}_{\tau_2}^2 = \sum_{i=1}^{S_i} w_i^n (\tau_{2,i}^n - \hat{\tau}_{2,i})^2.
\]

This estimator is biased, and the bias can be quantified by computing the expectation of \(\hat{\sigma}_{\tau_2}^2\), again treating \(w_i^n\) as fixed:

\[
E \left[ \hat{\sigma}_{\tau_2}^2 \right] = \sum_n w_n^2 E \left[ (\tau_{2,i}^n - \hat{\tau}_{2,i})^2 \right]
\]

where the summation is over only successful trajectories (i.e. for which \(S_i^a = 1\)). All summations in the remainder of this derivation are similarly performed over only successful trajectories.

Expanding the definition of \(\hat{\tau}_{2,i}\) yields

\[
E \left[ (\tau_{2,i}^n - \hat{\tau}_{2,i})^2 \right] = E \left[ \left( \sum_m w_i^m \tau_{2,i}^m \right)^2 \right] - 2 \left( \sum_m w_i^m \right)^2 E \left[ \tau_{2,i}^n \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right]^2
\]

\[
= E \left[ \left( \sum_m w_i^m \tau_{2,i}^m \right)^2 \right] - 2 \left( \sum_m w_i^m \right)^2 E \left[ \tau_{2,i}^n \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right]^2
\]

\[
= E \left[ \left( \sum_m w_i^m \right)^2 \right] - 2 \left( \sum_m w_i^m \right)^2 E \left[ \tau_{2,i}^n \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right]^2
\]

\[
= E \left[ \sum_m w_i^m \right] - 2 \left( \sum_m w_i^m \right)^2 E \left[ \tau_{2,i}^n \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right] + \sum_m w_i^m E \left[ \tau_{2,i}^m \right]^2
\]

\[
= E \left[ \sum_m w_i^m \right] - 2 \sum_m w_i^m \mu_{\tau_{2,i}^m} + \sum_m w_i^m \sigma_{\tau_{2,i}^m}^2 + \sum_m w_i^m \mu_{\tau_{2,i}^m}^2 + \sum_m w_i^m \sigma_{\tau_{2,i}^m}^2
\]

\[
= E \left[ \sum_m w_i^m \mu_{\tau_{2,i}^m}^2 + \sum_m w_i^m \sigma_{\tau_{2,i}^m}^2 \right] + \sum_m w_i^m \mu_{\tau_{2,i}^m}^2 + \sum_m w_i^m \sigma_{\tau_{2,i}^m}^2
\]

\[
= E \left[ \sum_m w_i^m \mu_{\tau_{2,i}^m}^2 + \sum_m w_i^m \sigma_{\tau_{2,i}^m}^2 \right] + \sum_m w_i^m \sigma_{\tau_{2,i}^m}^2
\]

\[
= E \left[ \mu_{\tau_{2,i}^m}^2 + \sigma_{\tau_{2,i}^m}^2 \right] - \mu_{\tau_{2,i}^m}^2 - 2 w_i^m \sigma_{\tau_{2,i}^m}^2 + \sum_m (w_i^m)^2 \sigma_{\tau_{2,i}^m}^2
\]

\[
= \sigma_{\tau_{2,i}^m}^2 - 2 w_i^m \sigma_{\tau_{2,i}^m}^2 + \sum_m (w_i^m)^2 \sigma_{\tau_{2,i}^m}^2
\]

\[
= \sigma_{\tau_{2,i}^m}^2 (1 - 2 w_i^m + \sum_m (w_i^m)^2).
\]
Substituting into the expression for \( E \left[ \sigma_{\tau_2}^2 \right] \),

\[
E \left[ \sigma_{\tau_2}^2 \right] = \sum_n w_n^2 \sigma_{\tau_2}^2 \left( 1 - 2w_n^2 + \sum_m (w_m^n)^2 \right)
\]

and, simplifying,

\[
E \left[ \sigma_{\tau_2}^2 \right] = \sigma_{\tau_2}^2 (1 - \sum_n (w_n^2)^2).
\]

Thus, an unbiased estimator for \( \sigma_{\tau_2}^2 \) is

\[
\hat{\sigma}_{\tau_2}^2 = \frac{\sigma_{\tau_2}^2}{1 - \sum_n (w_n^2)^2}
\]

and an unbiased estimator for \( \sigma_{\tau_2,i}^2 \) is

\[
\hat{\sigma}_{\tau_2,i}^2 = \frac{\sum_n (w_n^2)^2}{1 - \sum_n (w_n^2)^2} \hat{\sigma}_{\tau_2}^2.
\]

So we can estimate the standard deviation of \( \tau_{2,i} \) with

\[
\hat{\sigma}_{\tau_{2,i}} = \sqrt{\frac{\sum_n (w_n^2)^2}{1 - \sum_n (w_n^2)^2}} \hat{\sigma}_{\tau_2}
\]

and, using the fact that \( k_{\tau}^2 = \frac{1}{\tau_{2,i}} \), we make a linear approximation for the relationship between \( k_{\tau}^2 \) and \( \tau_{2,i} \), and propagate that to the variance:

\[
\hat{\sigma}_{k_{\tau}} = \left( \frac{1}{\tau_{2,i}} \right)^2 \hat{\sigma}_{\tau_{2,i}}.
\]

We can further simplify this expression, by letting

\[
N_{i,\text{eff}} = \frac{1}{\sum_n (w_n^2)^2} = \frac{\left( \sum_n k_n^{n_{\text{coll}}} \right)^2}{\sum_n (k_n^{n_{\text{coll}}})^2}
\]

so that

\[
\hat{\sigma}_{k_{\tau}} = \left( \frac{1}{\tau_{2,i}} \right)^2 \sqrt{\frac{1}{N_{i,\text{eff}} - 1}} \hat{\sigma}_{\tau_2}
\]

where

\[
\hat{\sigma}_{\tau_2} = \sqrt{\frac{\sum_n k_n^{n_{\text{coll}}} (\tau_{2,i}^n - \tau_{2,i})^2}{\sum_n k_n^{n_{\text{coll}}}}}
\]

This form of the expression shows the inverse square-root relationship with respect to the number of samples, which is characteristic of standard errors of the mean over multiple independent samples.

Unfortunately, a few of our reported plots and calculations used a previous, incorrect, version of this estimate for \( \hat{\sigma}_{k_{\tau}} \), but the deviations were less than 5% in cases that we tested.

Finally, we note that this calculation does not account for the additional uncertainty due to variation in \( k_n^{n_{\text{coll}}} \). Future work may aim to develop a more rigorous estimate that accounts for this source of variation.
B Case Study: Entropy-driven Catalyst

The full system is given in Figure 6A (main text). In addition to the productive reactions shown there, unproductive reactions were included between every pair of resting macrostates for a total of 31 non-spurious reactions (3 productive, 28 unproductive).

Sequences are taken from [2], Table 1.

| Domain | Length (nt) | Sequence   |
|--------|-------------|------------|
| 1      | 10          | CTTTCTTACA |
| 2      | 24          | CCAAGCTCGG |
| 3      | 4           | CCCT       |
| 4      | 16          | CATTCAACCTAC |
| 5      | 6           | TCTCCA     |
| 6      | 16          | CCACATACATCAT |

Table S1. Sequences for entropy-driven catalyst. Taken from Table 1 of [2].

All Multistrand simulations were performed at 25°C with sodium concentration of 1 M, magnesium concentration of 0 M, GT wobble pairing enabled, dangles energy parameter ‘Some’, and Metropolis rate method. The values of $k_{bi}$ and $k_{run}$ were 8.01171383 $\times 10^5$ M$^{-1}$s$^{-1}$ and 2.41686715 $\times 10^6$ s$^{-1}$, respectively. These values were determined as the mode of the posterior distribution after training a simplified Multistrand-like model on an extensive experimental dataset [2].

A single spurious reaction was observed between Substrate and Fuel under these parameters. We expect this spurious reaction to occur via 0-to-0ehold strand displacement. This spurious reaction becomes more prominent at higher temperatures.

| Reactants | Products | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|-----------|----------|--------------|-------------------------|-----------------|
| Substrate + Catalyst | Intermediate + Signal | 869/20881 | 2.94±0.14 $\times 10^9$ | 111±4 |
| Intermediate + Signal | Substrate + Catalyst | 151/63000 | 1.78±0.20 $\times 10^9$ | 112±8 |
| Fuel + Intermediate | Output + Catalyst + Waste | 278/20258 | 0.20±0.79 $\times 10^7$ | 2.15±0.14 |
| Catalyst + Waste | Catalyst + Waste | 19992/20000 | 2.83±0.02 $\times 10^7$ | 20.5±0.6 |
| Substrate + Signal | Substrate + Signal | 19997/20000 | 1.14±0.01 $\times 10^8$ | 7.33±0.24 $\times 10^5$ |
| Fuel + Waste | Fuel + Waste | 19992/20000 | 3.84±0.03 $\times 10^7$ | 433±53 |
| Catalyst + Signal | Catalyst + Signal | 19992/20000 | 5.69±0.04 $\times 10^7$ | 7.24±0.70 $\times 10^6$ |
| Catalyst + Catalyst | Catalyst + Catalyst | 20000/20000 | 1.97±0.03 $\times 10^7$ | 1.14±0.10 $\times 10^6$ |
| Fuel + Catalyst | Fuel + Catalyst | 20000/20000 | 5.31±0.04 $\times 10^7$ | 3.34±0.41 $\times 10^6$ |
| Fuel + Fuel | Fuel + Fuel | 19999/20000 | 7.41±0.05 $\times 10^7$ | 6.50±0.77 $\times 10^6$ |
| Waste + Waste | Waste + Waste | 19977/20000 | 4.34±0.03 $\times 10^7$ | 3.39±0.16 $\times 10^6$ |
| Substrate + Substrate | Substrate + Substrate | 19998/20000 | 2.00±0.01 $\times 10^7$ | 6.87±0.26 $\times 10^6$ |
| Signal + Waste | Signal + Waste | 19989/20000 | 4.23±0.03 $\times 10^7$ | 2.81±0.05 $\times 10^6$ |
| Output + Signal | Output + Signal | 19999/20000 | 5.86±0.04 $\times 10^7$ | 4.06±0.70 $\times 10^6$ |
| Catalyst + Intermediate | Catalyst + Intermediate | 19999/20000 | 4.86±0.03 $\times 10^7$ | 9.54±0.55 $\times 10^6$ |
| Fuel + Signal | Fuel + Signal | 19998/20000 | 7.76±0.05 $\times 10^7$ | 1.93±0.20 $\times 10^7$ |
| Intermediate + Signal | Intermediate + Signal | 629803/63000 | 7.20±0.01 $\times 10^7$ | 8.82±0.14 $\times 10^6$ |
| Signal + Signal | Signal + Signal | 19994/20000 | 8.08±0.06 $\times 10^7$ | 4.22±0.26 $\times 10^6$ |
| Substrate + Intermediate | Substrate + Intermediate | 20000/20000 | 8.20±0.06 $\times 10^7$ | 9.98±0.50 $\times 10^6$ |
| Substrate + Fuel | Substrate + Fuel | 1819975/182000 | 1.05±0.00 $\times 10^7$ | 1.02±0.02 $\times 10^7$ |
| Output + Substrate | Output + Substrate | 20000/20000 | 7.61±0.05 $\times 10^7$ | 1.45±0.25 $\times 10^7$ |
| Intermediate + Intermediate | Intermediate + Intermediate | 20000/20000 | 4.68±0.03 $\times 10^7$ | 3.13±0.10 $\times 10^7$ |
| Output + Catalyst | Output + Catalyst | 19998/20000 | 3.99±0.03 $\times 10^7$ | 5.51±0.00 $\times 10^7$ |
| Output + Waste | Output + Waste | 19988/20000 | 2.73±0.02 $\times 10^7$ | 597±62 |
| Output + Output | Output + Output | 19996/20000 | 4.81±0.03 $\times 10^7$ | 2.03±0.73 $\times 10^7$ |
| Substrate + Catalyst | Substrate + Catalyst | 20012/20881 | 7.57±0.05 $\times 10^7$ | 8.26±0.31 $\times 10^7$ |
| Fuel + Intermediate | Fuel + Intermediate | 19980/20258 | 6.47±0.05 $\times 10^7$ | 1.86±0.25 $\times 10^7$ |
| Output + Fuel | Output + Fuel | 19985/20000 | 5.42±0.04 $\times 10^7$ | 2.15±0.56 $\times 10^7$ |
| Substrate + Waste | Substrate + Waste | 19983/20000 | 4.05±0.03 $\times 10^7$ | 4.07±0.13 $\times 10^7$ |
| Output + Intermediate | Output + Intermediate | 19997/20000 | 4.61±0.03 $\times 10^7$ | 2.47±0.07 $\times 10^7$ |
| Intermediate + Waste | Intermediate + Waste | 19985/20000 | 2.13±0.02 $\times 10^7$ | 3.27±0.09 $\times 10^7$ |

Table S2. Reaction rate estimates for all productive, unproductive, and (observed) spurious reactions in the entropy-driven catalyst. The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.
| $[C]_0$ (nM) | KinDA (min) | Zhang et al. [?] (min) | Ratio (Zhang et al./KinDA) |
|-------------|-------------|------------------------|---------------------------|
| 10          | 2.2         | 8.0                    | 3.6                       |
| 5           | 3.4         | 14.5                   | 4.3                       |
| 2           | 7.2         | 36.2                   | 5.0                       |
| 1           | 13.6        | 73.1                   | 5.4                       |
| 0.5         | 26.4        | 147.0                  | 5.6                       |
| 0.2         | 65.0        | 367.8                  | 5.7                       |
| 0.1         | 129.4       | 731.5                  | 5.7                       |
| 0.05        | 258.9       | 1443.0                 | 5.6                       |
| 0.02        | 652.7       | 3459.4                 | 5.3                       |

**Table S3.** System half-completion times for varying initial catalyst $C$ concentrations. Half-completion times estimated by simulating the mass-action ODEs for the rates in Table S2 and finding the time at which $[OB]=5$ nM. Initial concentrations of $S$ and $F$ are 10 nM and 13 nM for all simulations. Half-completion times were also estimated using the published reaction rates in Zhang et al. [?]. KinDA overestimates the reaction rates so all half times are much lower than those calculated with the published rates, by a factor of about 4-6.

| Resting Macrostate | Complex | Samples | $p$ (%) |
|--------------------|---------|---------|---------|
| Input              | Input   | 483792/500500 | 96.7±0.0 |
|                    | spurious| 16708/500500  | 3.3±0.03 |
| Output             | Output  | 431370/500500 | 86.2±0.0 |
|                    | spurious| 69130/500500  | 13.8±0.0 |
| Waste              | Waste   | 500020/500500 | 99.9±0.0 |
|                    | spurious| 480/500500    | 0.1±0.44 |
|                    |         | (9.6±0.44)×10^{-2} |       |
| Signal             | Signal  | 382146/500500 | 76.4±0.1 |
|                    | spurious| 118354/500500 | 23.6±0.1 |
| Intermediate       | Intermediate | 196772/500500 | 99.3±0.0 |
|                    | spurious| 3728/500500   | (7.45±0.12)×10^{-1} |
| Substrate          | Substrate | 364393/500500 | 72.8±0.1 |
|                    | spurious| 136107/500500 | 27.2±0.1 |
| Fuel               | Fuel    | 433740/500500 | 86.7±0.0 |
|                    | spurious| 66760/500500  | 13.3±0.0 |

**Table S4.** Probabilities of $p$-approximations for each resting complex, with $p = 0.7$. The Samples column shows number of sampled conformations that $p$-approximate the resting complex (or are $p$-spurious) over the total number of sampled conformations.
C Case Study: Multiple Desired Pathways

The full system is given in Figure 8 (main text). The sequences were randomly generated from a four-letter alphabet with an equal probability of each base, with the exception of toeholds $s_w$ and $s_s$, which were chosen to be weak and strong toeholds, respectively.

| Domain | Length (nt) | Sequence          |
|--------|-------------|-------------------|
| t      | 6           | GGAGCC            |
| s = $s_w$ | 6         | ATATAT            |
| s = $s_s$ | 6         | GCGCGC            |
| 2      | 10          | GGCAAACAAG        |
| 3      | 10          | CGGCAGAATT        |
| a      | 10          | CGCATTTGCCC       |
| b      | 10          | TACCTTTTCCC       |
| c      | 10          | CAAAGCCCTT        |

Table S5. Sequences for the multiple fates case study.

For both the unmodified (with $s = s_w$) and modified (with $s = s_s$) systems, the reactions $A + B \rightarrow C + D$ and $A + B \rightarrow E + F$ were analyzed. Multistrand simulations were run with the same parameters as the entropy-driven catalyst (Appendix B).

| Reactants | Products | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|-----------|----------|--------------|--------------------------|-----------------|
| Gate + Interloper | Fate1_Cpx1 + Fate1_Cpx2 | 13532/5000000 | $(7.71\pm0.09) \times 10^3$ | $(1.54\pm0.02) \times 10^5$ |
| Gate + Interloper | Fate2_Cpx1 + Fate2_Cpx2 | 7/5000000 | $19.2\pm10.3$ | $862\pm107$ |
| Gate + Interloper | Gate + Interloper | 4976331/5000000 | $(3.37\pm0.00) \times 10^7$ | $(6.04\pm0.02) \times 10^5$ |

Table S6. Reaction rate estimates for reactions in the multiple fates case study (unmodified system, $s = s_w$). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.

| Reactants | Products | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|-----------|----------|--------------|--------------------------|-----------------|
| Gate + Interloper | Fate1_Cpx1 + Fate1_Cpx2 | 1611704200 | $(5.55\pm0.20) \times 10^5$ | $279\pm16$ |
| Gate + Interloper | Fate2_Cpx1 + Fate2_Cpx1 | 200/5042000 | $672\pm67$ | $268\pm18$ |
| Gate + Interloper | Gate + Interloper | 5001511/5042000 | $(1.95\pm0.00) \times 10^7$ | $(3.93\pm0.02) \times 10^5$ |

Table S7. Reaction rate estimates for reactions in the multiple fates case study (modified system, $s = s_s$). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.
| Resting Macrostate | Complex      | Samples       | $p$ (%) |
|-------------------|--------------|---------------|---------|
| $B$ (unmodified)  | $B_1$        | 14477/201612  | 7.18±0.06 |
|                   | $B_2$        | 83652/201612  | 41.5±0.1  |
|                   | $B_3$        | 11072/201612  | 5.49±0.05 |
|                   | $B_4$        | 50/201612     | $(2.53±0.35) \times 10^{-2}$ |
|                   | spurious     | 92361/201612  | 45.8±0.1  |
| $B$ (modified)    | $B_1$        | 39381/200465  | 19.6±0.1  |
|                   | $B_2$        | 0/200465      | $(4.99±4.99) \times 10^{-4}$ |
|                   | $B_3$        | 46345/200465  | 23.1±0.1  |
|                   | $B_4$        | 0/200465      | $(4.99±4.99) \times 10^{-4}$ |
|                   | spurious     | 114739/200465 | 57.2±0.1  |

**Table S8.** Resting macrostate conformation probabilities for resting macrostate $B$ (see main text) for both $s = s_w$ and $s = s_s$. The Samples column shows number of sampled conformations that $p$-approximate the resting complex (or are $p$-spurious) over the total number of sampled conformations.
Case Study: Mechanisms Combining 3-way and 4-way Branch Migration

| Domain | Length (nt) | Sequence               |
|--------|-------------|------------------------|
| a      | 22          | CAGTCCCAAGTCACCACCTAGC |
| b      | 22          | GCACCTGCGATACGAGGCCTGG |
| c      | 22          | CCGATCGCAGCCATCGTTCC   |
| t1     | 6           | CGTTTT                 |
| t2     | 6           | ACATCC                 |
| t3     | 10          | CCTCTACTCA             |
| T2     | 2           | TT                     |
| d1s    | 16          | CCAAACCTTCATCTTC       |
| d2     | 6           | TACTCG                 |

Table S9. Sequences for case study, taken from Kotani & Hughes [7]. Note that our simulations used a modified reporter complex from [7] that uses domain d1s rather than d1. d1s is produced by removing 2 nt from the 5’ end of d1.

| Reactants | Products | Trajectories | k1 (M−1s−1) | k2 (s−1) |
|-----------|----------|--------------|-------------|----------|
| S1 + C1   | P1 + H   | 19/50000     | (5.23±1.70)×10^5 | 174±28   |
| P1 + I1   | S1 + C1  | 0/50000000   | < 78.7      | —        |
| I1 + S2   | C1 + P3  + P2 | 21/50000 | (9.49±2.92)×10^5 | 0.66±0.13 |
| P2 + R    | D + RW   | 72/50000     | (1.34±0.22)×10^6 | 2.66±0.32 |
| S2 + R    | S2-R     | 154/247      | (3.18±0.30)×10^7 | 1.57±0.13×10^7 |
| I1 + S2-R | C1 + P3  + D + RW | 14/25000 | (1.14±0.43)×10^6 | 1.14±0.24 |

Table S10. Reaction rate estimates for the Kotani & Hughes (2017) [7] case study. All simulations were run in ordered-complex mode, except for the unexpected reaction, S2 + R → S2-R, which has only one product and was run in count-by-domain mode. The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants. Simulation parameters were the same as for the entropy-driven catalyst (Appendix B).
E Case Study: Binding Reactions and Flavors of Macrostates

DNA sequences were taken from Groves et al. [?]. All Multistrand simulations used the same parameters as with the entropy-driven catalyst (Appendix B).

| Domain | Length (nt) | Sequence       |
|--------|-------------|----------------|
| a      | 16          | GTAGGAGTGGAGGTGA|
| 1      | 6           | GGGAAAT         |
| 2      | 6           | TCTTAC          |
| b      | 16          | CAACACACACACACC|
| 3      | 6           | TGATGA          |
| 4      | 6           | AACTAC          |

Table S11. Sequences for case study, taken from Groves et al. [?].

| Reactants | Products | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|-----------|----------|--------------|------------------------|-----------------|
| InA + OR  | OR, InA_Waste + OR, InA_Sig | 3160/104931   | (1.66±0.04) x 10^6 | 50.7±0.9 |
| InB + OR  | OR, InB_Waste + OR, InB_Sig | 3136/59496    | (2.99±0.07) x 10^6 | 189±3 |
| InA + OR  | InA + OR     | 101490/104931| (4.71±0.02) x 10^7 | (4.85±0.23) x 10^3 |
| InB + OR  | InB + OR     | 56360/59496   | (5.01±0.02) x 10^7 | (4.39±0.35) x 10^4 |

Table S12. Reaction rate estimates for the OR gate (mode: ordered-complex). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.

| Reactants | Products | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|-----------|----------|--------------|------------------------|-----------------|
| InA + OR  | OR, InA_Waste + OR, InA_Sig | 3163/103313   | (1.67±0.04) x 10^6 | 50.3±0.9 |
| InB + OR  | OR, InB_Waste + OR, InB_Sig | 3136/61394    | (2.88±0.07) x 10^6 | 189±3 |
| InA + OR  | InA + OR     | 99856/103313  | (4.71±0.02) x 10^7 | (5.28±0.25) x 10^3 |
| InB + OR  | InB + OR     | 58258/61394   | (5.02±0.02) x 10^7 | (5.04±0.42) x 10^4 |

Table S13. Reaction rate estimates for the OR gate (mode: count-by-complex). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.

| Reactants | Products | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|-----------|----------|--------------|------------------------|-----------------|
| InA + OR  | OR, InA_Waste + OR, InA_Sig | 3159/106333   | (1.63±0.04) x 10^6 | 50.9±0.9 |
| InB + OR  | OR, InB_Waste + OR, InB_Sig | 3121/62556    | (2.80±0.07) x 10^6 | 183±3 |
| InA + OR  | InA + OR     | 102874/106333 | (4.71±0.01) x 10^7 | (5.02±0.23) x 10^3 |
| InB + OR  | InB + OR     | 59435/62556   | (5.02±0.02) x 10^7 | (4.76±0.37) x 10^4 |

Table S14. Reaction rate estimates for the OR gate (mode: count-by-domain). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.
| Reactants  | Products          | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|------------|-------------------|--------------|--------------------------|-----------------|
| InB + AND  | AND$_{InB}$       | 2583/6708    | (1.26±0.03) × 10$^7$    | (1.13±0.03) × 10$^7$ |
| InA + AND$_{InB}$ | AND$_{Sig}$ + AND$_{Waste}$ | 3149/94505 | (1.12±0.03) × 10$^7$ | 60.3±1.1 |
| InB + AND  | InB + AND         | 4125/6708    | (1.88±0.03) × 10$^7$    | (2.01±0.05) × 10$^7$ |
| InA + AND$_{InB}$ | InA + AND$_{InB}$ | 31065/94505 | (2.86±0.01) × 10$^7$   | (3.16±0.11) × 10$^3$ |

Table S15. Reaction rate estimates for the AND gate (mode: ordered-complex). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.

| Reactants  | Products          | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|------------|-------------------|--------------|--------------------------|-----------------|
| InB + AND  | AND$_{InB}$       | 3066/36102   | (3.10±0.08) × 10$^6$    | (5.12±0.08) × 10$^6$ |
| InA + AND$_{InB}$ | AND$_{Sig}$ + AND$_{Waste}$ | 3149/97322  | (1.09±0.03) × 10$^5$   | 57.5±1.0 |
| InB + AND  | InB + AND         | 30788/33859  | (2.85±0.02) × 10$^5$    | (2.85±0.14) × 10$^5$ |
| InA + AND$_{InB}$ | InA + AND$_{InB}$ | 93002/97322 | (2.87±0.01) × 10$^7$   | (3.20±0.11) × 10$^3$ |

Table S16. Reaction rate estimates for the AND gate (mode: count-by-complex). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.

| Reactants  | Products          | Trajectories | $k_1$ (M$^{-1}$s$^{-1}$) | $k_2$ (s$^{-1}$) |
|------------|-------------------|--------------|--------------------------|-----------------|
| InB + AND  | AND$_{InB}$       | 3066/36102   | (2.89±0.07) × 10$^4$    | 30±3            |
| InA + AND$_{InB}$ | AND$_{Sig}$ + AND$_{Waste}$ | 3151/96431  | (1.10±0.03) × 10$^6$   | 57.2±1.1 |
| InB + AND  | InB + AND         | 33036/36102  | (2.86±0.02) × 10$^7$    | (2.28±0.08) × 10$^4$ |
| InA + AND$_{InB}$ | InA + AND$_{InB}$ | 93002/96431 | (2.87±0.01) × 10$^7$   | (2.94±0.10) × 10$^3$ |

Table S17. Reaction rate estimates for the AND gate (mode: count-by-domain). The Trajectories column shows the number of trajectories corresponding to each reaction over the total number of trajectories simulated for the reactants.